adyen Payments made easy

Adyen Mobile Payments Index

Adyen Mobile Payment Index Reveals Emerging as M-Commerce Powerhous

UK maintains global lead, and mobile authorization rates be desktop

Delft University of Technology

Amsterdam, 23 October, 2014 – Adyen, the global payment technology company, today announced findings from its latest global Mobile Payments Index, which covers the period of July through September 2014. The Index, which captures data from web-based transactions across Adyen's customer base of more than 3,500 businesses, found that globally payments over mobile devices contributed 23.3% of total online payments for the quarter, up from 21.5% last quarter.

Asia experiences greatest growth, Europe maintains global lead

Asia now has the second-highest proportion of mobile transactions among global regions, at 17% of total online payments. Comparing August 2013 to August 2014, Asia experienced the strongest mobile payments growth of all regions, increasing by 58%.

Europe maintained its lead among global regions, with mobile payments averaging at 24% for the quarter. It also experienced strong growth of 34% from August 2013 to August 2014. The UK leads the pack in Europe and globally, with mobile payments averaging 41% for Q3, followed by the Netherlands and Spain at 26%, France 18% and Germany 16%. North America remains steady, at 16.7% and Latin America remains below other regions, at 6% for the quarter.

Roelant Prins, Chief Commercial Officer, Adyen, said, "The future belongs to businesses that will continue to adapt to local mobile payment preferences and streamline the checkout flow. Evernote, for example, saw an uplift of 10% after implementing a mobile-optimized checkout for Alipay, the most popular online payment method in China."

Mobile authorization rates beat desktop

Interestingly, the Index reveals that authorization rates for online payments are higher on mobile devices than on desktops, averaging 88.1% versus 86.7% for the months of August and September.

TU Delft University of Technology

Furthermore, the proportion of transactions refused by bank mobile devices is 1.5 percentage points lower than it is on deskto

This reflects the increased emphasis businesses are placin streamlining mobile payment processing now that it has beco primary sales channel.

Digital goods versus physical goods across desktop tablet, and smartphone

The index examines vertical industries split by digital goods (su games, services like club memberships, hotel reservations, and tic versus physical goods/retail (such as clothing, furniture, applia groceries).

The data shows that throughout Q3, people used smartphones than tablets when purchasing digital goods, but the opposite is with physical goods:

- Digital goods: desktop 72%; smartphone 20%; tablet 8%
- Physical goods: desktop 71%; smartphone 11%; tablet 18%

The Average Transaction Value (ATV) of physical goods is highe €86.1 versus €26.2 for digital goods according to Index data. correlation suggests people are more comfortable making h value purchases over tablet devices. Shoppers' tendency to purc physical goods on a tablet rather than smartphone also suggest larger screens offer a better way to view and browse physical prod

Smartphone share is up, tablet share is steady

The Index shows a clear preference for smartphones over tablets, 57% of total mobile payments between July and September made smartphone, compared with 43% on a tablet. As shown in the M Payment Index Infographic, smartphone transaction share conti on an upward gradient, while tablet transaction share appears t levelling out. This correlates with a consumer trend toward smartph with larger screen sizes, and a levelling out of tablet sales.

Challenge the future

Why My Payment Got Rejected

A Method to Mine Payment Refusal Clues

by

R.J.A. (Roy) van der Valk

in partial fulfillment of the requirements for the degree of

Master of Science
in Systems Engineering, Policy Analysis and Management,
with a specialisation in Information Architecture,
and an Entrepreneurship Annotation

at the Delft University of Technology,
to be defended publicly on Monday January 12, 2015 at 4:30 PM.

Student number:	1357212	
E-mail address:	hello@royvandervalk.nl	
Project duration:	May 15, 2014 – January 12, 2015	
Daily supervisors		
	Dr. D. (Dina) Hadžiosmanović,	TU Delft
	S. (Samaneh) Tajalizadehkhoob MSc.,	TU Delft
	M.P.C. (Mark) Oostdam MSc.,	Adyen
	L.W.M. (Maikel) Lobbezoo MSc.,	Adyen
Thesis committee		
Chair:	Prof. dr. ir. J. (Jan) van den Berg,	TU Delft
First supervisor:	Prof. dr. ir. M.F.W.H.A. (Marijn) Janssen,	TU Delft
Second supervisor:	Dr. H.G. (Haiko) van der Voort,	TU Delft

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft Delft University of Technology

"Make everything as simple as possible, but not simpler."
Albert Einstein

Executive Summary

Introduction

In October 2014 President Obama's credit card was rejected when he was trying to pay after having dinner with his wife in New York. The president explained in a humorous way: "I guess I don't use it enough. I was trying to explain to the waitress, 'No – I really think that I've been paying my bills'." Something triggered the system of the president's own bank to refuse the payment. In this case the president seems to expect the bank's system refused the payment by qualifying his card as high risk because of infrequent usage.

A significant amount of payments fail for similar 'mysterious' reasons when refused by the cardholder's bank (*issuer*). On average about 20% of the payments are refused by an issuer. An issuer can refuse a payment because of insufficient funds or incorrect cardholders details. However, as we show in this thesis, a large amount is refused because of other reasons which are not contained in the electronic message send by the issuer. Hence it is very difficult for cardholders and merchants to find clues for the issuer's refusal reason (see Figure 1). We refer to this as *information asymmetry*.

This information asymmetry also complicates the monitoring of the issuer by the card network. In this thesis we argue the interests between the issuer and the card network are not always aligned. In some scenarios an issuer might prioritise its business strategy or its exposure to liability over the interests of the cardholder and the merchant, the two main stakeholders in payments. This can lead to issuers refusing more payments than desired from the perspective of the cardholder and the merchant. We refer to this as *moral hazard effects*. In this thesis we argue that partially dissolving the information asymmetry reduces the moral hazard effects. Hence the network becomes more economically efficient if this information asymmetry is dissolved.

Nowadays, there are millions of parties (and thus systems) communicating with each other in the complete payment network. All these parties, divided over different geographical regions (and jurisdictions) with different norms, values and interests, introduce a *social complexity*. All the systems, and their mutual interfaces and dependencies, introduce *technical complexity* and potential points of failure. These complexities hinder the development of policies, which impose an issuer to share its 'honest' reason, or an information architecture which can cope with this diversity of reasons. This complicates the *coordination* of the payment process by policy makers or card networks. Hence we argue for other methods to dissolve the information asymmetry.

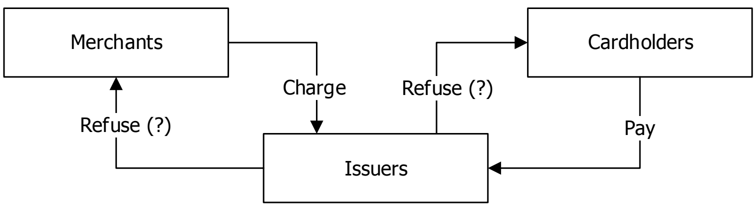

Figure 1: A diagram in which the payment ecosystem is reduced to merchants wanting to charge cardholders via an issuer which leaves both parties in the dark about the underlying refusal reason.

Aim and Research Approach

In this thesis we aim at empowering other parties in the payment network to break the issuer's 'black box system' by revealing the decision rules an issuer uses to authorise some payments, while refusing others. In other words, we look for a system (i.e. a set of principles according to which something is done) behind this logic. Hence we aim at inducing the decision rules which describe the distinctive features of payments which are systematically refused by an issuer. The leads to our main research question:

Main RQ. How to induce the authorisation decision rules an issuer uses to refuse certain payments from the data available to Payment Service Providers (PSPs)?

This empowerment enables parties to act on an issuer's decision. This can provide issuers with incentives to make decisions which better reflect the interests of other parties in the payment network. We argue this makes the payment network economically more efficient.

PSPs process high volumes of payments on behalf of merchants. To realise this PSPs have connections to the other parties in the network. This umbrella position allows a PSP to reverse engineer the refusal logic of an issuer and dissolve the information asymmetry. However PSPs need tools to extract this knowledge about issuer refusals from their payment data. Obtaining this knowledge is challenging, because a) the volumes of payments for large PSPs is in the order of millions of transactions per day and b) there are hundred thousands of different combinations of characteristics (e.g. different currencies, amounts, locations, etc.) that describe a payment. Hence our objective is **to provide insights about refusals to experts from PSPs to act on the groups of refusals on the merchant's behalf**.

Data Mining (DM) is the field of research dealing with discovering knowledge from data. In this thesis we explore the potential for techniques from this field to reach our objective. We use the state-of-the-art in DM research to tailor a specific technique to the requirements necessary to perform this task. We test the implementation and validate the practical relevance of the results in the form of a case study at a large PSP (with connections to thousands of merchants).

Association Rule Mining using a New Interest Measure

We apply a customised implementation of a DM technique called *association rule mining*. Association rule mining is a DM technique to find associations between specific features in a large data set. We select this technique after an extensive exploration.

This exploration starts with designing a tool to gather incidents with prior systematically refused payments. Later we explore the payment data related to these incidents and interview an issuer about its refusal logic. From this knowledge, we first derive the interesting dimensions necessary to find groups of systematically refused payments. Second, we determine what criteria a DM technique should meet in order to find these groups and present them in an usable way.

We also build on this knowledge to customise association rule mining. In order to find the most interesting groups of systematic refusals we introduce a new *pattern evaluation measure* (or *interest measure*), which we call Unique Confidence (UC), to the field of association rule mining. In essence UC makes it possible to find the distinctive features to which we can ascribe a certain consequence (in this case refusing payments). Besides this measure being essential in the discovery of the required insights, it also improves on one of the major challenges in association rule mining: to cope with the sheer amount of patterns found.

Findings on Authorisation Decision Rules

The title of this thesis is *"Why my payment got rejected?"* We broadly answer this question using the intelligence mined by our method from 11,5 million payments processed via one PSP. [redacted]

[redacted]

[redacted]

[redacted]

(a) Authorisation Decision Rules

[redacted]

(b) Issuing Countries

[redacted]

(c) Issuers

Figure 2: Different perspectives on the groups of systematic refusals as a consequence of specific authorisation decision rules.

Evaluation

We evaluate the results on three aspects: *correctness*, *completeness*, and *practical relevance*. First, we use expert validation to confirm the correctness of the results. On the basis of a survey among experts, we conclude the logic of our DM application is similar to how an expert is able to find the distinctive groups of systematically refused payments. Hence we conclude that on the basis of this validation the authorisation decision rules we find are indeed the ones of interest.

Second, we validate the completeness of the results on the basis of gathered information about incidents with prior systematically refused payments. We check if our application is able to find all relevant groups of payments. We conclude that our implementation is able to find all relevant groups above a threshold for minimum impact.

Third, we validate the practical relevance of the results. We validate the found groups with the relevant parties in the payment network. We make the following observations:

- An issuer confirms it systematically refuses one of the found decision rules
- Merchants confirm two of the found decision rules
- A PSP experiments with the results and is able to increase its authorisation rate significantly

[redacted]

Table 1: Results of PSP experiment. The experiment was conducted during one week from November 6 2014 until November 12 2014.

[redacted]

Conclusion

In the context of the main research question "How to induce the authorisation decision rules an issuer uses to refuse certain payments from the data available to PSPs?", we can conclude that the method we describe in this thesis is very effective in inducing the decision rules of an issuer. We additionally find significant proof that this method achieves our objective to provide insights about refusals to experts from PSPs to act on the groups of refusals on the merchant's behalf.

Hereby we effectively empower other parties in the payment network to break the issuer's black box system and dissolve the information asymmetry related to refusal reasons. This empowerment enables parties to act on an issuer's decision. Hereby issuers are provided with incentives to make decisions which better reflects the interests of other parties in the payment network. We argue this makes the payment network economically more efficient. Additionally the results of an experiment indicate there is a significant business case for PSPs to act on the findings.

Future Research

We suggest future research to mainly focus on improving the (computational) performance of the method. For instance by applying techniques from the field of constraint-based mining or by using an efficient framework to manipulate large datasets in order to post-process the rules.

[redacted]

Another suggestion is regarding the moral hazard assumption. We underpin this assumption with the observation that issuers systematically refuse payments which are risky in terms of liability. However to formulate the assumption as an hypothesis and make claims, requires another type of research.

Preface

I never expected I would end up doing my thesis internship in the payment industry. It sounded a bit dull to me at first. I asked myself: "How complex can it be to process a payment? Surely there are no exciting things happening here? I mean, we are already paying with cards for decades and nothing really changed, right?" After a few conversations with payment experts I was completely convinced of the contrary.

I hope this thesis reads like a novel, like if we knew everything upfront. I can promise you this was not the case. We found ourselves on completely new territory. As the title describes, it was a complete mystery what logic consumer banks were using to authorise payments and how to (data) mine this knowledge. Expert knowledge was highly dispersed and scientific literature about this topic is very scarce. My first task was to gather this knowledge. As someone with an IT background in an IT company this resulted in building tools for experts to log and verify all incidents related to bad authorisation rates of certain payments. Based on this knowledge we could ultimately design and implement the analysis presented in this thesis.

I'm very thankful for all the great, fun, open-minded, supportive and inspiring people which helped in putting this work together. In the first place, I would like to thank my daily supervisors. Dina for making sure we never lost focus, guarding the overall structure while keeping eye for detail, and for inspiring me to apply association rule mining. Samaneh for her pointers to fundamental literature, making sure we were using the common terminology, and her suggestion to reflect on the methodology. Mark, besides all the feedback, for his skills in guiding me in managing the multi-actor problem of all the stakeholders of this thesis. Maikel for his fundamental role in recognising the scientific and business potential of this research and trusting me to it, sharing his extensive industry knowledge, and making sure everything was in alignment with the goals and other efforts of the company. Additionally I would like to thank Maikel for his openness, honesty and devotion, even when things got more hectic surrounding the birth of his daughter.

Secondly I would like to thank my thesis committee for their support, inspiration and feedback during the process, especially in guarding the generalisability of this thesis. Besides I want to thank my committee for our earlier encounters on the faculty. Jan, Marijn – whom I have both witnessed being appointed as full professor – and Haiko, for their openness and for inspiring me for their fields of research during all the lectures and discussions we had.

Furthermore I'd like to thank all the other people who have supported me. My loving parents Leo and Karin, who guided me to where I am today. My brothers, family and friends who provided me with support and a welcome distraction when necessary. All my colleagues from Adyen for their openness and support, especially Andrada (who edited the cover photo), Bert, Chris, Felix, Faris, Huub, Jaap, Jan, Mark, Milena, Steffen, Tettri, Thibaut, and Tim. The founders and my former colleagues of Mendix, who are responsible for a spark of big ambition in my life which still drives me today.

Finally, I hope the generalisability of our work is recognised by researchers and practitioners in other domains which are confronted with comparable challenges.

R.J.A. (Roy) van der Valk
Delft, December 2014

Contents

1

Introduction

In the current global economy, with the wide adoption of the Internet, companies are serving customers throughout the world. This poses a challenge to develop and maintain an infrastructure which can cope with the global electronic payment traffic. Since the card network Visa introduced electronic payments in 1979, the payment network grew into today's complex network of numerous computer systems linked together [1].

In a typical card payment, where a *shopper* buys something in a physical or online shop of a *merchant*, several parties are involved to support the payment process (see Figure 1.1). The merchant connects with its *Payment Service Provider (PSP)* who further processes the payment and connects with all the other parties in the value chain, the risk management service supplier, the bank of the merchant (*acquirer*), the *card network* (or *scheme*), the bank of the shopper (*issuer*).

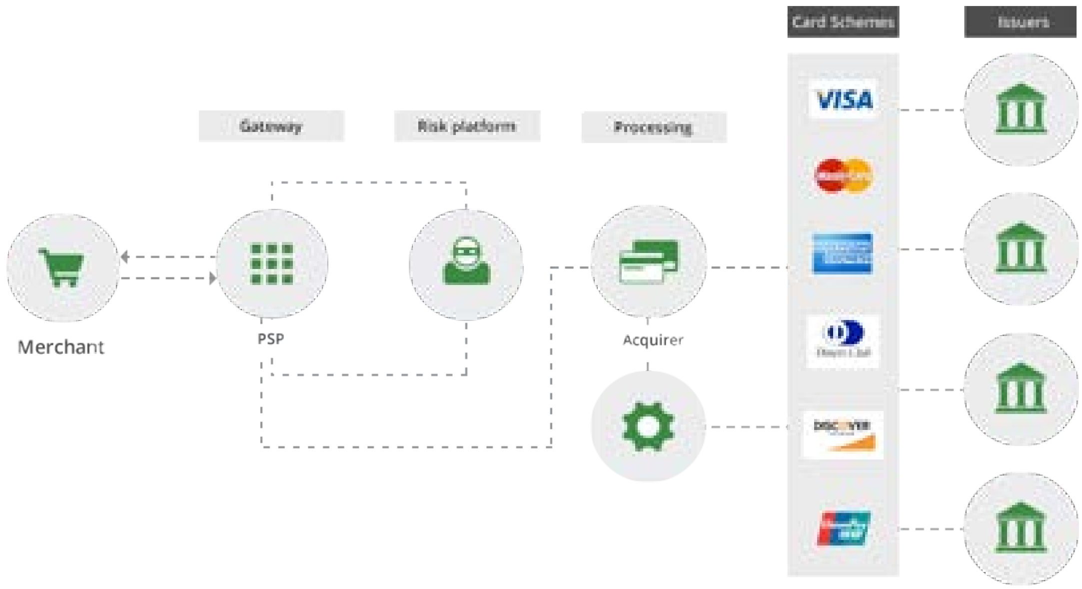

Figure 1.1: Parties involved in payment processing from the perspective of a merchant [2].

In principle, merchants could also renounce from using a PSP and connect to the other parties themselves. However because of the complexity of the payment network this has proven to be a negative business case for many merchants, especially the ones doing business globally [3]. As noted by Evans and Schmalensee [3] payment systems circumventing the current payment network have significant problems in getting accepted. For this reason innovators in the payment industry are probably more effective to find ways to improve the current payment network, rather than to replace it entirely. Because of their umbrella position in the value chain PSPs and acquirers who find unique ways to improve the payment network have a competitive advantage.

PSPs and acquirers process payments on behalf of the merchant. The merchant considers a number of Key Performance Indicators (KPIs) when evaluating the performance of the payment process. At the highest level, the merchant aims to improve the *conversion rate* of payment attempts by shoppers into a bank transfer which it can add to its assets without any liabilities. The conversion rate can be affected at numerous stages in the payment process. This can range from a shopper leaving the checkout page of the website (or a malfunctioning payment device at a physical store) to a shopper disputing a payment and claiming a refund of the payment at the end of the process.

In this thesis we focus on the stage where a card payment is sent for authorisation to the issuer. The related KPI is the *authorisation rate*. We define the authorisation rate as the fraction of authorised payments in the total number of payments which are sent to the issuer for authorisation.

On average about 20% of the payments are refused by an issuer. An issuer can refuse a payment because of insufficient funds or incorrect details of the *cardholder* (i.e. shopper). However there can also be more systematic reasons for an issuer to refuse a payment, for instance when the issuer qualifies a payment as high risk, does not recognise the acquirer or does in general not allow a certain payment to be made with the issued card. Every issuer can have a different logic for authorising payments. This can lead to strange situations.

For instance The New York Times reported that President Obama's credit card was rejected when he was trying to pay at a dinner with his wife in New York [4]. The president explained in a humorous way: "I guess I don't use it enough. I was trying to explain to the waitress, 'No – I really think that I've been paying my bills'." In this case it is questionable if any party in the payment industry is in favour of the outcome of the payment process.

The standards which describe the electronic payment messages being passed back and forth in the payment network have a limited variety in refusal reasons. Hence the actual reason behind a refusal is often not transparent and this makes it hard for merchants, PSPs, and acquirers to structurally solve these issues and re-enable the affected shoppers to pay for their services. Therefore providing merchants with the service to dissolve this information asymmetry and to effectively deal with systematic refusal reasons to boost the authorisation rate has a substantial business case. Additionally it has the potential to improve the entire payment network by preventing undesired refusals (from the perspective of cardholders and merchants).

PSPs processing high volumes of payments from multiple merchants to multiple acquirers are in an umbrella position. For this reason PSPs can hypothetically extract the required knowledge from their payment data to distinguish groups of refused payments which are likely to be issuer-specific instead of specific to certain merchants, acquirers or even shoppers. Data Mining (DM) is the field of research dealing with discovering knowledge from data.

This thesis explores techniques from the field of DM for **proving insights about groups of systematic refusals to experts from PSPs to act on the merchant's behalf**. We use the state-of-the-art in DM research to tailor a specific technique to the requirements necessary to perform this task. We test the implementation and validate the practical relevance of the results in the form of a case study at Adyen, a large PSP, with own acquiring capabilities, operating globally. Adyen processes millions of payments on a daily basis in over 250 payment methods, over 180 currencies, and via more than 85 acquirers [5].

The structure of this thesis is as follows. In Chapter 2 we outline the research. As a point of departure, Chapter 3 contains the background knowledge for this thesis from literature related to the payment industry and the field of DM. We expand on this knowledge to analyse the stakeholders and form expectations about the refusals behaviour of issuers in Chapter 4. We base the design of our method on the results from exploring the data related to the refusals, which we describe in Chapter 5. Additionally we base our design on a comparison between the main features of the relevant DM techniques, which we describe in Chapter 6. We present the method itself in Chapter 7 and the implementation in Chapter 8. In Chapter 9 we show the results, which we evaluate in Chapter 10. Finally, we conclude and make recommendations in Chapter 11 and end with an epilogue about the DM approach in Chapter 12.

2

Research Description

In this chapter we describe the research. First we provide a clear statement of the central problem in this research. Second we explain our objective regarding this problem. In the following section we describe the knowledge gaps we address in this thesis by formulating several research questions. Finally we outline our approach to answer these questions and to close the knowledge gaps.

2.1. Problem Statement

The number of merchants (millions), acquirers (hundreds) and issuers (thousands) communicating via a number of card networks (around ten major networks), gives an indication of the amount of parties (and thus systems) communicating with each other in the complete payment network [6]. All the parties, divided over different geographical regions (and jurisdictions) with different norms, values and interests, introduce a *social complexity*. All the systems, and their mutual interfaces and dependencies, introduce *technical complexity* and potential points of failure.

The effect of this social and technical complexity also manifests itself in the authorisation phase of the payment process. Issuers can refuse a payment in their own interest and leave other parties in the dark about the reason. We refer to this as *information asymmetry*. Because of this information asymmetry other parties in the network are powerless to argue or act on the issuer's authorisation decision.

The cardholder and the merchant (i.e. the two sides in this two-sided market) are unlikely to always agree with the issuer's decision to authorise or refuse a certain payment because they have different interests. Issuers want to encourage or discourage certain forms of payment based on their business strategy. Cardholders on the other hand, want to pay for any desired good or service, while merchants want to charge any cardholder which is interested in their goods or services. All parties want to minimise their liability. We depict the payments ecosystem reduced to a two-sided market with merchants, issuers and cardholders in Figure 2.1.

As Anderson and Moore [7] note, when issuers are aware that they cannot be held accountable by shoppers this can give rise to *moral hazard effects*. Klick and Mitchell [8] define a moral hazard as 'inefficiencies that occur when risks are displaced or cannot be fully evaluated'. Moral hazard typically occurs when the following preconditions are met: *separation of ownership* (i.e. the party taking actions is different from the party experiencing the consequences), *information asymmetry* (i.e. the parties do not have similar information available), *hidden action* (i.e. privately taken actions affect the probability

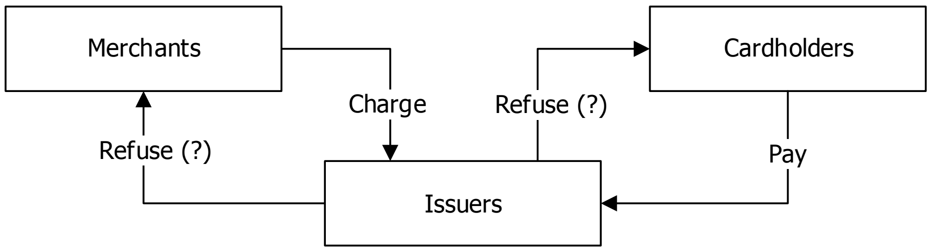

Figure 2.1: A diagram in which the payment ecosystem is reduced to a two-sided market between merchants wanting to charge cardholders via an issuer which leaves the two sides of the market in the dark about the underlying reason in the case of a refusal.

distribution of the outcome) [9–11].

Moral hazard effects for instance occurred in the United Kingdom (UK). In the UK shoppers had to prove that the issuer was at fault when they were a victim of fraud, while in most other countries it was vice versa [7]. Issuers knew that in the majority of cases shoppers could not access the information required to prove the issuer was at fault. Hence, there was no incentive to improve the security vulnerabilities in the system. This situation led to a steep increase in (successful) fraud and high costs for the issuer as a consequence [7]. This leads us to believe that it is also in the (long term) interest of issuers to be more open about their reasoning for authorising payments.

Because the payment network is a two-sided market, all other parties in the network derive their existence from the cardholders and merchants. We argue that the identified information asymmetry (together with the other preconditions for moral hazards) causes some of the issuer's authorisation decisions to be neither a) in the shopper's interest, b) in the merchant's interest and c) in the long term also not in the issuer's interest. Hence the network becomes more *economically efficient* (i.e. *Pareto optimal*) if this information asymmetry is dissolved.

Dissolving the information asymmetry is technically complex as well. An issuer can have an abundance of different reasons to make a certain decision. The standards which describe the electronic messages being passed back and forth in the payment network have a limited variety in refusal reasons which causes the systematic refusal reasons to be be described as 'other reason'. This description provides no information about the actual reason.

The described social and technical complexity make it difficult to design a) an information architecture which can cope with this diversity of reasons or b) an organisational structure which imposes an issuer to share its 'honest' reason. This complicates the *coordination* of the payment process by policy makers or the card networks.

PSPs process high volumes of payments on behalf of merchants. To realise this PSPs have connections to multiple merchants to multiple issuers via multiple acquirers and card networks. Conceptually, this umbrella position allows a PSP to reverse engineer the logic an issuer uses to refuse payments and dissolve the information asymmetry. PSPs also have the motivation to act, because their business depends on allowing shoppers to pay for the products and services of the merchant.

PSPs need means to extract the knowledge about issuer refusals from their payment data to be able to dissolve the information asymmetry with the issuer. Obtaining this knowledge is challenging, because the volumes of payments for large PSPs is in the order of millions of transactions per day and there are hundred thousands of different combinations of characteristics (e.g. different Bank Identification Numbers (BINs), currencies, amounts, locations, etc.) that describe a payment. A BIN is related to the first 6 digits on a card, and identify the issuer (i.e. the bank that issued the card).

2.2. Objective

We aim at abstracting knowledge about issuer refusals from a PSP's payment data. More specifically we aim at reverse engineering the authorisation decision made by an issuer on a specific BIN. We argue PSPs can use these insights to (partially) dissolve the identified information asymmetry and improve their service.

To reverse engineer the authorisation decisions of issuers, we aim at exposing the characteristics of payments which are systematically refused on certain BINs. Because the number of BINs are in the order of ten thousands, we also aim at providing insights in the similarities and impact of payments which are refused over multiple BINs. These insight can help experts from PSPs to act on similar groups of refusals in an efficient way. In essence, we aim at **proving insights about groups of systematic refusals to experts from PSPs to act on the merchant's behalf**.

2.3. Research Questions

In order to precisely describe the knowledge gaps we need to fill to achieve our objective and solve the stated problem, we formulate Research Questions (RQs). The main knowledge gap directly derives from the information asymmetry between issuers and other parties in the network, in this thesis specifically PSPs. PSPs have no clue about why a payment is refused, because the issuer makes the decision in private and provides other parties with little explanation. A method to uncover the reasoning of an issuer to refuse certain payments can in principle dissolve the information asymmetry. The concept of reasoning we narrow down to 'a system of rules on which the decision is based'. In other words we are looking for a method (thus 'how?') to induce[1] the authorisation decision rules[2] an issuer uses to refuse certain payments from the data available to PSPs. Hence we formulate the main RQ as follows:

Main RQ. How to induce the authorisation decision rules an issuer uses to refuse certain payments from the data available to PSPs?

As described earlier, information asymmetry may lead to moral hazards. In our context, a moral hazard may occur when an issuer decides to refuse certain payments to safeguard own interests. Knowingly or unknowingly this might conflict with the interests of other parties in the payment network, causing moral hazard effects. In practice, moral hazard effects will materialise as unjustified refusal payments, rooted in authorisation decision rules of the card issuer. To understand the indicators of moral hazard effects we formulate the following RQ:

RQ 1. What kind of authorisation decision rules would signal moral hazard effects?

In the first RQ we explore the stakeholder landscape and issuer incentives with respect to moral hazards. Next, we explore the empirical clues how particular refusals manifest. In essence, a good understanding of systematic issuer refusals is necessary to design a method to mine for decision rules on large scale. Decision rules imply that issuers authorise and refuse payments in a systematic way. Hence we formulate the following RQ:

RQ 2. What are the characteristics of systematic issuer refusals?

Besides understanding issuer refusals, we also need to understand the available payment data. This is necessary to precisely characterise systematic issuer refusals and understand how we can recognise these characteristics in the data. Hence we formulate the following RQ:

[1] Meaning to derive a general principle (in this case decision rules) from specific observations (in this case processed payments) [12].

[2] Authorisation decision rules are rules describing data patterns leading to authorisations or refusals which most likely reflect the distinctive characteristics of a payment on which the issuer bases its decision.

RQ 3. What kind of payment data can be used to analyse systematic issuer refusals?

After answering the last two RQs we understand the characteristics of systematic issuer refusals and understand how to recognise these characteristics in the data. Based on this knowledge we can determine which DM technique is most suitable for finding systematic issuer refusals. Hence we formulate the following RQ:

RQ 4. Which DM technique has the functionality required to find systematic issuer refusals?

After answering this RQ we are able to select a DM technique. We need to customise the technique to be able to apply it to find systematic issuer refusals and output the results in a usable format. Hence we formulate the following RQ:

RQ 5. How to apply the DM technique to find systematic issuer refusals and present the results in an usable format to experts?

After answering the five RQs we know a) what kind of authorisation decision rules to expect (RQ 1); b) how these rules portray itself in the payment data of a PSP (RQ 2 & 3); and c) how to induce these rules using DM (RQ 4 & 5). Based on this knowledge we are ultimately able to answer the main RQ and induce the authorisation decision rules an issuer uses to refuse certain payments from the data available to PSPs.

2.4. Approach

The work in the thesis follows the general design science methodology [13]. To answer particular RQs, we use empirical research, exploring and processing heavy loads of transactional data in payment data available to PSPs. We use both quantitative and qualitative aspects of empirical research. The research is quantitative in the sense that we mainly experiment on (quantitative) payment data available to PSPs. The qualitative aspects of this thesis are in the *exploration* and *evaluation*. For instance we use (qualitative) data regarding incidents with issuer refusals. Besides we use interviews to explore the behaviour of one issuer (essentially a small *case study* [14]), as well as surveys and interviews to evaluate the quality of the method.

Figure 2.2 describes the structure of this thesis and the steps in our approach. Our approach has methodological similarities with DM approaches which are used in practice and research [16], like for instance the Cross Industry Standard Process for Data Mining (CRISP-DM) (see Figure 2.3).

First phase in our approach is the literature research. We describe the literature on the domain of international card payment processing and data mining in Chapter 3. The converging phase of the research which follows is based on this knowledge.

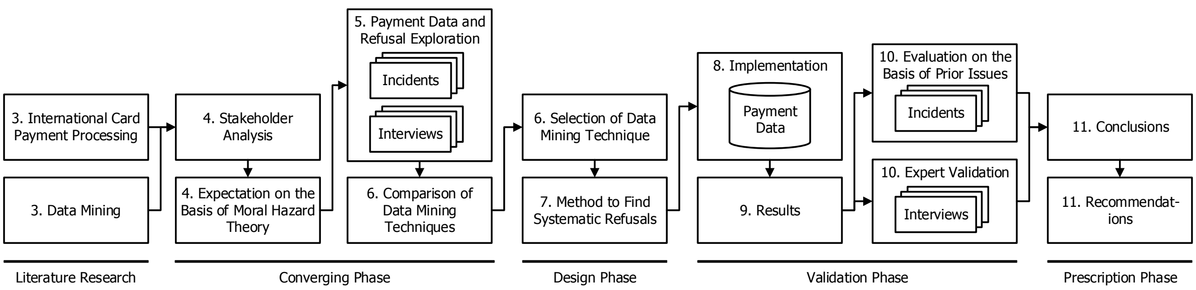

Figure 2.2: An overview of the structure of this thesis, the steps in our approach and the data sources we use (if applicable). The numbers refer to the respective chapter in which the step is discussed. The outline of the figure is based on Oei [15].

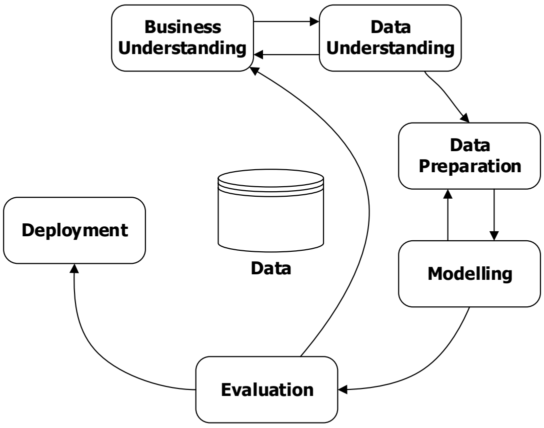

Figure 2.3: Diagram showing the different phases of CRISP-DM [17]

In this phase we first outline our expectation of the refusal behaviour of issuers in Chapter 4. We do this by analysing the driving factors of the main stakeholders in payment processing and relating this to the theory about moral hazard. Second we explore the available payment data and incidents related to issuer refusals in Chapter 5. We interview experts and gather incidents about issuer refusals using a special tool in which experts log all the incident information. We end this phase with a comparison between the DM techniques in Chapter 6. We base this comparison on the gathered knowledge about DM and issuer refusals.

In the same chapter we select the most suitable DM technique. This is the start of design phase. Next we describe the method based on the selected DM technique in Chapter 7.

In the validation phase we implement the application of the DM technique on a large set of payment data which we describe in Chapter 8. We show the results in Chapter 9, which we evaluate in Chapter 10. The evaluation consists of two parts. First we evaluate if the prior reported groups of refused payments are found. These are reported in the same tool which we use for the exploration of refusals. Next we validate if the results of our implementation are correct and usable using expert validation.

During the prescription phase we conclude and recommend on the basis of the validation phase. Chapter 11 outlines these two elements.

3

Background

In this chapter we present the background knowledge for this thesis. First we discuss the domain of international card payment processing, the domain in which we perform the research. Second we discuss the relevant literature from the field of DM. This chapter provides the theoretical foundation for the remainder of this thesis.

3.1. International Card Payment Processing

Card payments are the most widely used form of payment [18]. In this section we discuss the typical model of card payments and a number of alternative models. Hereafter we discuss when parties are at risk due to liability for the payments they process. Next we discuss the economics and policies in card payment processing. Finally we discuss the research efforts in payment processing.

3.1.1. Four-Party Model

Figure 3.1 depicts the process of the typical four-party model of card payment processing.[1] The main parties are the merchant, acquirer, card network and issuer, hence this model is called a *four-party model*. There are two main phases.

The first phase is called the *authorisation phase*. In this phase the funds of the cardholder (i.e. shopper) are committed, but no money is exchanged. After the cardholder presents the card to the merchant and accepts the amount of the payment charged by the merchant, the payment request is transmitted to the merchant's PSP. Then the PSP routes the payment request to an acquirer with a connection the appropriate card network which corresponds with the card presented by the cardholder. The *issuer*, which issued the card, receives the payment from the network and decides whether to authorise the request or not. Then in reversed order, the response of the issuer is sent to the merchant again.

In order to recognise the card network and issuer which should receive the payment the Bank Identification Number (BIN) is used. The BIN is indicated by the first six digits of a card number. Each card scheme has its own ranges of BINs which are further assigned to issuers. Issuers can issue cards on multiple BINs, and use each BIN for a different purpose - e.g. for different issuing countries, or card types.

[1]When mentioning the payment network in the remainder of this thesis we mainly refer to the network according to the four-party model.

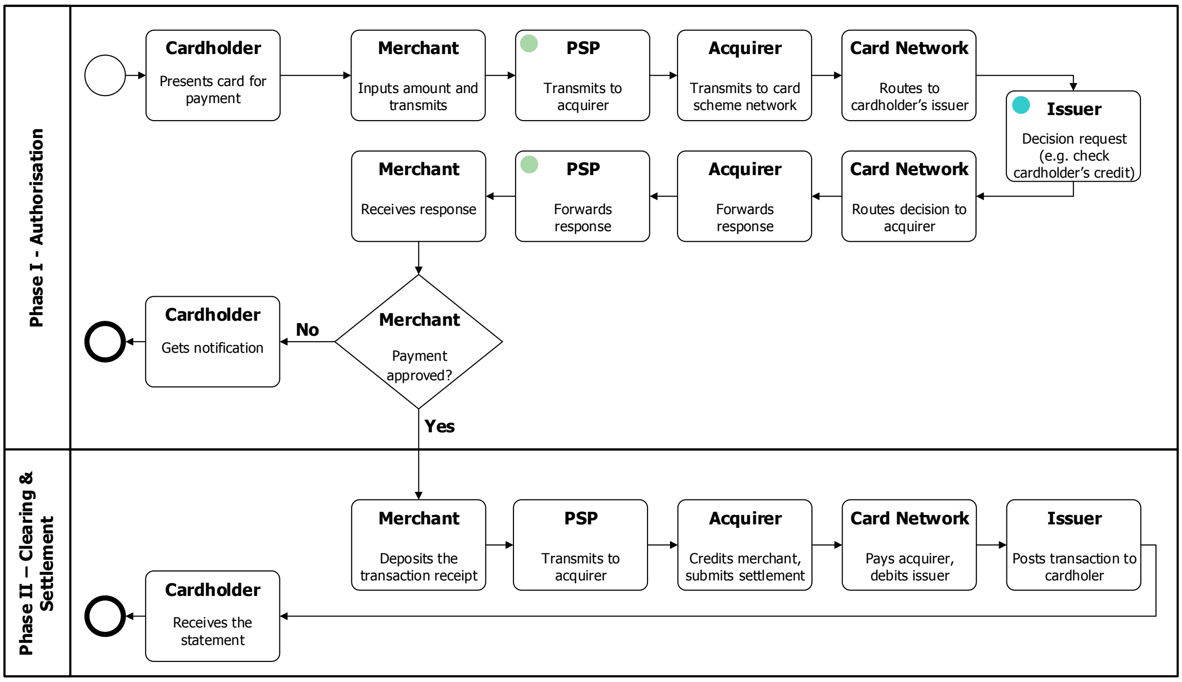

Figure 3.1: Typical process of the stakeholder actions involved in a card payment on a four-party network [6, 19]. The elements relate to the specification of the Business Process Model and Notation (BPMN). The green dots indicate the start en end location which the electronic payment message travels with respect to this thesis. The message is send from the PSP via other parties to the issuer (blue dot), which then decides to either authorise or refuse the payment. Then ultimately the response is retrieved by the PSP again.

The second phase is called *clearing and settlement* and is initiated only when the payment request is approved by the issuer. This is where the merchant ultimately receives the money. First the merchant deposits the transaction receipt at the acquirer via the PSP. The acquirer credits the merchant account and together with a bulk of other daily transactions it is sent to the card network's clearing system, which debits the issuer and credits the merchant. Hereafter the *settlement* takes place, where the issuer sends the actual payment via the card scheme to the acquirer. After *reconciliation* all balances of accounts are in agreement, and the cardholder has been billed and the merchant has been paid. Typically the actual money doesn't travel via the PSP's account, but is directly paid by the acquirer to the merchant.

3.1.2. Alternative Models

The payment model from the previous section is a simplified version of reality. In fact the payment industry comprises many different types of parties, because many parties fulfil multiple roles, form alliances, or outsource certain parts (like risk, or processing services) [3, 6]. Variations on the described model are for instance private-label cards, and three-party (instead of four) networks (like American Express and Discover) [6, 20]. Private-label cards, are cards which are only accepted at one merchant and issued only by a specific issuer. Only one merchant and one processing entity is involved in this case [6]. In a three-party network the acquirer and the issuer are the same entity [21]. This set-up typically evolves when acquirers form partnerships with issuers [6].

Also PSPs can have acquiring capabilities as well and depending on the contract between the merchant and PSP, a PSP can offer different *service levels*. A service level can range from only acting as a gateway, up to full managed service, where the actual money flows through the PSP and the PSP holds all the risk. When a PSP manages the full service, then the merchant only needs to sign a contract with the PSP and the PSP has contracts with all the acquirers. For lower service levels the merchant needs contracts with the acquirers as well [6]. A disadvantage of a lower service level is that the merchant needs to do additional reconciliation for all acquirers.

3.1.3. Research in Payment Processing

A large amount of research in the payment industry has been from network economics, for example topics such as: the effect of price incentives for payment methods; externalities related to interchange fees; liquidity management of payment networks; contracting issues; and supply chain coordination [22–28]. Social studies mainly research the acceptance of new payment methods [29, 30]. Research attention from the field of engineering - mainly computer science - revolves around the design and evaluation of new payment methods [31, 32], the security of payment methods [33, 34], and the detection of fraudulent payments [35–37].

About a decade ago DeGennaro [6] noted that academic research on the payments industry is almost non-existent. We observe that the amount of payment industry related research performed has increased over recent years. Despite this observation we believe there are still a lot of unexplored areas. One of these areas is the authorisation stage of the payment process where we focus on in this thesis. We see a lot of potential to improve the transparency and hence the economical efficiency of the payment process for studies in this direction.

We aim at contributing to research studies on payment processing via the use of DM. DM is the field of research dealing with discovering knowledge from data. The background knowledge for DM is the subject of the next section.

3.2. Data Mining

The first part of this section introduces the most basic concepts and techniques related to DM. There are different definitions for DM and related concepts and techniques. In this thesis we use the terminology of Han *et al.* [38].

In most cases the data needs to be preprocessed independent of which technique is used. We the discuss the literature related to data preprocessing in the second part of this section. In the following part we discuss the different DM functionalities. Finally we discuss the current applications of DM in payment processing.

3.2.1. Introduction into the Concepts and Techniques

There are different definitions for DM. In a narrow view DM is defined as applying intelligent methods to extract data patterns, and in a broad view it entails the whole process of Knowledge Discovery from Data (KDD). Figure 3.2 depict the KDD process. This typically involves *data cleaning*, *data integration*, *data selection*, *data transformation*, *pattern discovery*, *pattern evaluation*, and *knowledge presentation* [39]. Similarly to Han *et al.* [38], we adopt the broad view in this research, because this is definition is most commonly used in media, as well as industry and science. DM is a highly application-driven domain with successful applications in Business Intelligence (BI), Web search, finance, health informatics, and digital libraries [40]. It relies on many technologies from other domains such as statistics, information retrieval, database and data warehouse systems, and machine learning.

A patterns is interesting when it contains knowledge which is hidden in the data. There are different criteria to determine if a pattern is interesting: a) it has to be valid on test data to some degree; b) it has to be novel; c) it has to be potentially useful (e.g. actionable or verifies a hypothesis); and d) it should be easily understandable by humans [38]. Objective or subjective measures can be used to guide the pattern discovery.

The selection of a suitable DM technique is mainly affected by the by the type of data to be mined (e.g. *database data*, *data ware house data*, or *transactional data*) and the sought functionalities specifying the kind of pattern or knowledge to be found [41]. Next we discuss the data preprocessing steps that can be helpful irrespective of the DM functionality and hereafter we discuss the different DM functionalities.

3.2.2. Data Preprocessing

It is evident that low-quality data will lead to low-quality results. Data pre-processing can be used to (1) improve the quality of the data, and (2) improve the efficiency and ease the mining process [41]. Han *et al.* [38] define a number of preprocessing techniques which can be used alongside each other:

- *Data cleaning* is applied to correct inconsistencies and fill in missing values. It can also be used to remove the noise from a dataset. In this case outliers are identified and transformed to behave according to the data pattern.
- *Data integration* takes data from several sources and merges it into a single, coherent data store.
- *Data reduction* is aimed at reducing the size of the data. This can for instance be achieved by *dimensionality reduction*, which reduces the number of random variables or attributes under consideration. Another technique is called *numerosity reduction*, where the actual data is approximated with the use of a model distribution. *Data compression* methods apply transformations to create a reduced or "compressed" version of the original data.
- *Data transformation* can be used to improve the accuracy and efficiency of mining algorithms

6 Chapter 1 *Introduction*

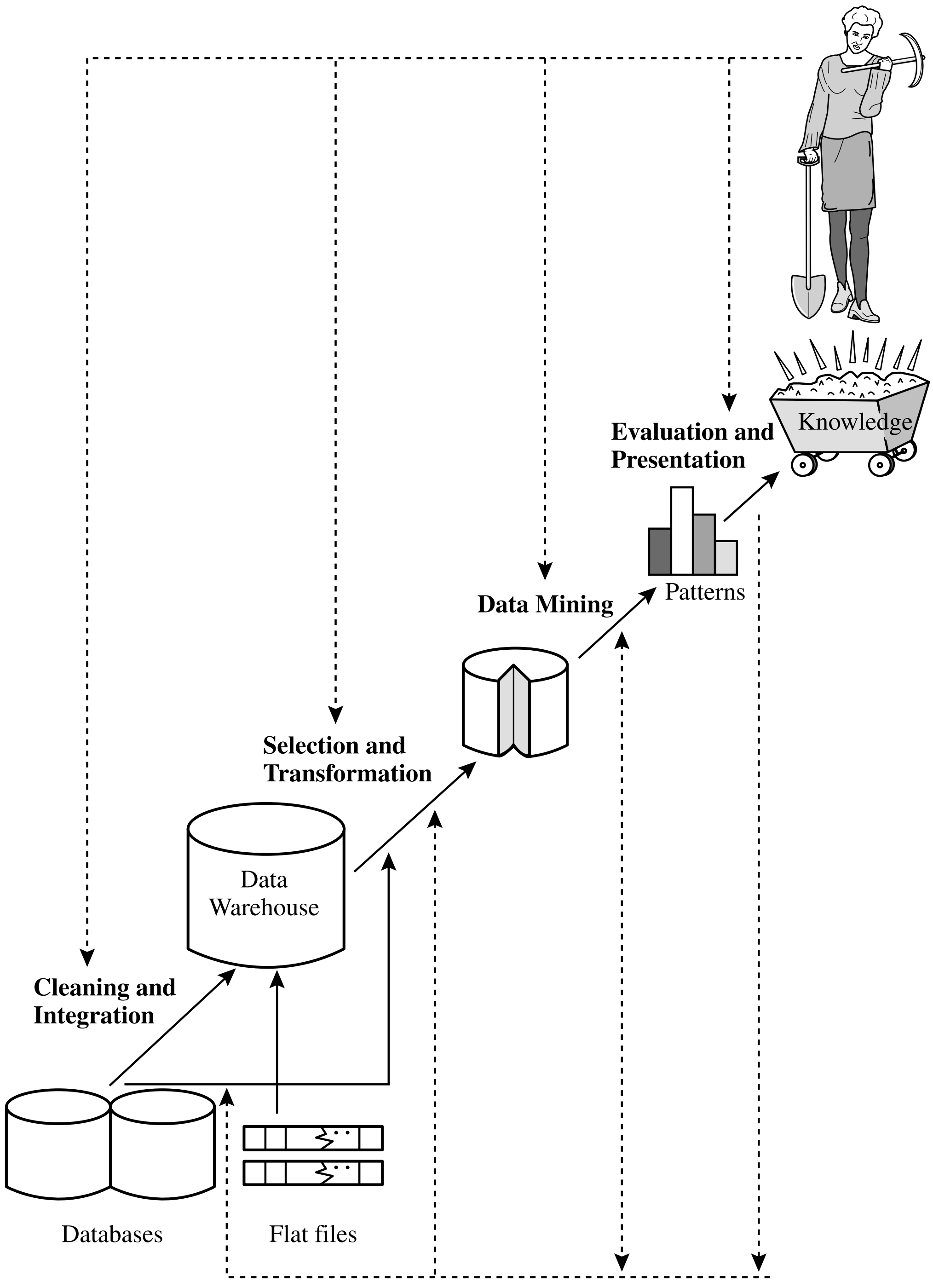

Figure 1.4 Data mining as a step in the process of knowledge discovery.

Figure 3.2: Data mining as a step in the knowledge discovery process. In our definition data mining refers to the complete knowledge discovery process instead of one specific step. Diagram from Han *et al.* [38].

involving distance measures. A common example is *normalisation*, where data is scaled to fall within a smaller range, like 0,0 to 1,0. Other examples are data *discretisation*, and *concept hierarchy generation*.

3.2.3. Functionalities

The DM functionalities include *characterisation and discrimination*; *mining frequent patterns, associations and correlations*; *classification and regression*; *cluster analysis* and *outlier detection*. Each functionality has its own strengths and limitations. In the next parts we discuss the relevant functionalities. We start with a relatively simple functionality and expand on this with more sophisticated functionalities.

Characterisation and Discrimination

Data characterisation entails summarising the general characteristics or features of a specific target class in a data set [38]. Typically these results are collected by a query and the output can be presented in various forms, like charts or tables. Data discrimination is a comparison of the general features of the object of the target class and the general features from contrasting classes [38]. In many cases, techniques from this functionality are able to meet the demands of experts to extract certain information from the data.

For some case this relatively simple functionality is not able to deliver the required insights. Nonetheless, as Witten and Frank [41] note it is important to always try the simple things first (as we do in the next chapter). Repeatedly in DM researchers and practitioners eventually, after an extended struggle, obtain good results on a sophisticated technique only to find out later that simpler methods perform just as well [41].[2] Also more sophisticated functionalities often expand on the insights from a more simpler functionality. We discuss these more sophisticated functionalities in the next parts.

Mining Frequent Patterns, Associations and Correlations

Finding frequent patterns is essential for finding associations, correlations and other interesting relationships between items in a large transactional or relational data set [42]. Han *et al.* [38] divide frequent patterns into three types. A set of items, like a transaction flagged e-commerce with a Card Verification Code (CVC), that appear frequently together in one transaction, is called a *frequent itemset*. If items appear in sequence frequently together, it is called a *frequent sequential pattern*. Like when a merchant always first performs a transaction to validate the shopper's card which is followed by a regular e-commerce (payment) transaction. When a substructure, which is are for instance subgraphs or subtrees, appear frequently together this is called a *frequent structured pattern*. The number of transactions that contain a certain itemset is defined as the *support count*.

Items that are frequently associated together can be represented in the form of *association rules* [43]. For instance an association rule may look like,

$$\begin{aligned} issuer(X, \text{"Banco Dinero"}) \wedge amount(X, \text{"0"}) \Rightarrow authorised(X, \text{"FALSE"}) \\ [support = 0,2\%, confidence = 85\%], \end{aligned} \tag{3.1}$$

where X is representing a payment. The rule indicates that there is a 85% chance, or *confidence*, transactions of Banco Dinero with an amount of zero lead to a refusal. A 0,2% *support* means that in the total set of transactions under analysis, there are 0,2% payments from Banco Dinero with a zero amount which are refused. Confidence and support have the following formulas [38]:

$$support(A \Rightarrow B) = P(A \cup B) \tag{3.2}$$

[2]This is the inspiration for the quote on the first page.

$$confidence(A \Rightarrow B) = P(B|A) \tag{3.3}$$

Typically a threshold is defined for the *minimum confidence* and *minimum support* that rules should pass in order to be interesting, these rules are called strong [38]. If an itemset meets the support threshold, then it is regarded as a frequent itemset.

The process of association rule mining generally exists out of two steps [38]:

1. **Find all frequent itemsets:** The itemsets that meet the minimum support count.
2. **Generate strong association rules from the found frequent itemsets:** The rules that meet the minimum confidence.

Performance is mainly dependent on the first step, because the second step is much less costly [41]. A major challenge is that usually for large data sets a high number of patterns are found which meet the minimum support count [42]. One of the reasons for this is that if an itemset is frequent, all the sub-itemsets are frequent as well. For example if there are 100 items in the set, than a total of $1{,}27 \times 10^{30}$ sub-itemsets are frequent as well. An itemset with k items is often named a k-itemset.

To overcome this the concepts *closed frequent itemset* and *maximal frequent itemset* are introduced. An itemset is closed when there are no super-itemsets with the same support count [44]. If there exists no super-itemset that has the itemset as subset and is still frequent, then the itemset is a maximal frequent itemset [45]. Additional *pattern evaluation measures* can be applied in order to reduce the set of rules even more and find the patterns of interest [38].

Work of Bauer *et al.* [46] shows resemblance with the our problem. The study aims to detect policy misconfiguration in access-control systems. Association rule mining was used to predict changes in users' behaviour that are likely to prevent legitimate access to a resource for which currently no access is given. System administrators can be noticed in advance of misconfiguration and resolve the issue, before it becomes a problem. Another interesting aspect of this study is that a feedback mechanism was utilized, to filter the results if repeatedly incorrect predictions are given.

Classification

Classification extracts models, called *classifiers*, from the data which describe important data classes, and are able to predict categorical *class labels* [41]. This can for instance be used to build a classification model to categorise which issuers are either healthy or unhealthy (according to expected behaviour or not). Many of the older classification techniques are very memory consuming, but more scalable techniques have been developed in recent DM research [38].

Classification is a two-step process, consisting of a *learning step*, where the classifier is build, and a *classification step*, where the classifier is applied to predict class labels for a given data set. The data used for the learning step is represented by *tuples* and the associated class labels. The class labels are discrete and unordered [47]. A tuple, X, is an attribute (or feature) vector, $X = (x_1, x_2, \ldots, x_n)$, depicting n tuple measurements from n attributes, respectively, $A_1, A_2, \ldots, A_n$. Each tuple X belongs to a class as described by the class label attribute. The tuples in the *training set* are randomly selected from the overall dataset [38]. The learning step described is also known as *supervised learning* (i.e. the classifier is told to which class label each training tuple belongs), as opposed to *unsupervised learning* (or *clustering*, where the class label is not known which we discuss in a later part of this section). We use the definition of Mohri *et al.* [48] for supervised learning. Mohri *et al.* [48] define supervised learning as "the machine learning task of inferring a function from labelled training data."

In the second step of the process, the model is used for classification. First a *test set* is used to determine the predictive *accuracy* of the classifier. The test set contains data that was not used when constructing the classifier. This is done to avoid an optimistic accuracy, because the classifier tends to *overfit* the data [47]. Overfitting means that during learning some specific anomalies from the training set are incorporated, which are not part of the overall data set. *Holdout*, *random sampling*, *cross-validation*, and *bootstrapping* are typical methods to partition the data in a training and test set [38].

In order to evaluate the classifier's quality a *confusion matrix* can be used. For a two class prediction problem, the *true positives*, *true negatives*, *false positives*, and *false negatives* are shown in a matrix [41]. Measures to evaluate a classifier's predictive ability are *accuracy*, *sensitivity* (also knows as *recall*, *specificity*, *precision*, F, and F_β. Relying on the accuracy only can be deceiving when the class label to be predicted is in the minority [47]. Dealing with such classification problems, called *class imbalance problems*, is also addressed in the sub-field *outlier detection*, which is discussed in one of the next parts. There are several strategies to address this problem like *oversampling*, *threshold moving*, and *ensemble techniques* [38].

Popular algorithms for classification are *decision tree induction*, a top-down recursive tree induction algorithm, *Naïve Bayesian classification*, based on Bayes' theorem of posterior probability, *rule-based classifiers*, using a set of IF-THEN rules for classification. *Significance tests* and *ROC curves* can be useful in selecting a specific classifier [38]. Significance tests give a measure to what extend the difference in accuracy between two classifiers is due to chance. ROC curves plot the true positive rate (or sensitivity) against the false positive rate (or 1 - specificity) of one or multiple classifiers [38].

Cluster Analysis

Unlike classification, *cluster analysis* or *unsupervised learning*, focuses on problems where the class label is unknown [41]. When the number of objects and the number of attributes is very large, it can be very costly or even infeasible to determine these by hand. In a clustering process, a set of data objects are grouped into multiple groups (or *clusters*), so that the objects within a cluster have high similarity, while being very dissimilar to objects from other clusters. The similarity and dissimilarity is based on the attribute values and is often assessed using distance measures.

Many clustering techniques have been developed. Several basic clustering techniques are: *partitioning methods*, *hierarchical methods*, *density-based methods*, and *grid-based methods*. In a partitioning method, a predefined number of clusters are created, and then object are iteratively relocated in order to improve the clusters [38]. A hierarchical decomposition is created, either *bottom-up* or *top-down*, on the given data set for hierarchical methods [47]. For density-based methods clusters are grown either to the density of neighbouring objects, or according to a density function [38]. In grid-based methods, the object space is first quantified into a finite number of cells to form a grid structure, before the clustering is performed [38].

In order to assess the feasibility of a cluster analysis and assess the quality of the formed clusters, a *clustering evaluation* is performed. Evaluation includes assessing the clustering tendency, determining the number of clusters, and assessing the clustering quality [38].

Outlier Detection

Outlier - or anomaly - detection, is a sub-field of DM which deals with finding patterns in data that do not conform to expected behaviour [49]. Outlier detection techniques are applied in a number of domains, for example for the detection of network intrusion [50], tumours [51], and faulty components of a space craft [52]. In the payment industry these techniques are predominantly used to detect fraudulent transactions [36]. It can be expected that most transactions processed are handled correctly, and problematic transactions are rare. For this reason outlier detection is relevant to this research.

Typically data doesn't follow a strict pattern, there is some randomness present. A shopper can decide to go for a dinner one day, while the shopper would normally get dinner from the supermarket.

These transactions should not be marked as outliers and handled as fraud. It is key to set up the detection in such a way that only 'real' pattern deviation is detected [38], and the balance between incorrectly marked fraud (false positives) and incorrectly market legitimate (false negatives) is optimal. Often this is achieved by making assumptions about what is normal data and determining when these assumptions are violated significantly [38].

Outlier detection methods can be categorised in two ways [38, 53]. First way to categorise the methods is according to whether data is provided, which is labelled by a domain expert as "normal" or "outlier". If labels are provided, the problem can be regarded as a classification problem and a supervised method for classifying can be used which is trained and tested on the labelled data. When labelled data is not available, unsupervised methods can be used. These methods are based on an implicit assumption that the normal data follow a pattern far more frequently than outliers. However this doesn't have to be the case, the normal data can be uniformly distributed and the outliers form a small cluster. Semi-supervised methods also exist and are useful when only a small set of labelled data is available. The small set is extended with data that is similar. In the case of disputable issuer refusals probably some labelled outliers are available from past issues, however it is unlikely these represent all possible outliers. In this case Han *et al.* [38] state that the detection can be improved by information about normal data learned from unsupervised methods.

Second way to categorise is by dividing the methods into groups, according to the underlying assumptions about outliers compared to the other data [38]. When a labelled data set is provided, a classification-based method is used to train a classification model on the data. Methods based on the assumption that data follows a statistical (stochastic) model are named statistical methods. Proximity-based methods assume that the set of feature values of outliers are significantly different from the ones of the normal data. Clustering-based methods assume that normal data forms large and dense clusters while outliers form smaller and sparser clusters or are not present in any cluster at all.

Recent studies often use a hybrid procedure. By combining multiple methods, deficiencies of a certain method can be overcome [54]. A recent study of Kuna *et al.* [54] studies a problem that is somewhat similar to the DM problem in this thesis. Kuna *et al.* [54] study several outlier detection methods, for assisting an auditor in the process of detecting anomalous data within audit logs. Ultimately a the authors design a procedure to find outliers using a proximity-based method in combination with a clustering-based method. Classification-based methods decrease the false positives and negatives in the earlier found outliers.

3.2.4. Data Mining in Payment Processing

The payments processing industry uses DM techniques at large scale to detect fraudulent payments [37]. Consequently this application has been widely researched. Fraud detection utilises the DM functionality of *outlier detection* [36]. Other applications for DM techniques in the payment processing industry are not commonly researched. DM applications in related industries, like in e-commerce (i.e. customer profiling, recommendation systems, and buying behaviour) or, accounting and banking (e.g. fraud, loan payment prediction and accounts payable) have been widely researched [40, 55, 56]. Hence we aim at contributing to the field of DM by presenting a new method for the payment industry. We focus on techniques for *descriptive analytics* (as opposed to *predictive analytics*), because we are mining for insights (instead of predictions) on refusals.

As we discuss later, this method in potential can also be relevant to other domains with similar DM problems. Additionally studies in the field of DM are criticised to not contribute to business in a large scale by focussing on refining algorithms, instead of describing how the algorithms are used [57–59]. We follow up on this critique by refining a technique based on a use case instead of solely on the basis of theoretical assumptions.

This chapter provides the background knowledge for the remainder of this thesis. For instance to provide the industry context for our expectation on the refusal behaviour of issuers in the next chapter. Additionally we explicitly build on this chapter when the relevant DM techniques are compared on their applicability for this thesis.

4

Expectation of Refusal Behaviour of Issuers

In the previous chapter we outline the research approach. An important step in the approach is to form an initial idea on what the refusal behaviour of issuer will be. This gives an idea on what kind of *authorisation decision rules* to search for. We base our expectations on the assumption that there is a *moral hazard* in authorising payments. Hence this chapter answers **RQ 1.**: "What kind of authorisation decision rules would signal moral hazard effects?" If we confirm the presence of these rules with our method, this provides a basis for our moral hazard assumption, which provides a basis for the social relevance of this thesis.

We discuss our expectation of the refusals behaviour of issuers as follows. First we describe the driving forces in the payment industry. Second we expand on our argumentation that this environment can give rise to moral hazard and what the implications are for the (expected) refusal behaviour of issuers.

4.1. Driving Forces in the Payment Industry

We describe the driving forces in the payment industry from three perspectives. First we outline the financial risks due to liability. Second we present the breakdown of the revenues and expenses, and how this affects the interests of the different parties in the payment network. Third we describe the international policies which can affect the behaviour of the different parties.

4.1.1. Financial Risks Due to Liability

Depending on the role, the actual transaction value can flow trough the parties accounts or not, and the amount of risk the party is exposed to due to liability can differ. Typically the card issuer is liable for the shopper's payment obligations (in case of credit) and the acquirer for the merchant's payment obligations [6]. A merchant can be become liable once credit transactions are disputed by shoppers, called *chargebacks*. Shoppers can initiate a chargeback up to three months after the purchase if they are unsatisfied with the products of services received [6]. Once a merchant is unable to pay the chargebacks, for instance due to bankruptcy, the acquirer is obliged to compensate the issuer and its cardholder. Thus it is important for acquirers to keep the chargeback rates at minimum. [redacted]

[redacted]. [redacted].

Besides chargebacks, another source of risk is fraud. Fraud is defined by Kahn and Roberds [60] as the risk that a claim cannot be collected because the person who is in debt can not be identified. The types they distinguish are a) *existing account fraud*, usually by using stolen account information, b) *new account fraud*, when a new account is opened by a thief in a third parties name, and c) *friendly fraud*, when someone later denies a transaction which the person has legitimately made [60].

The liability in case of fraud depends on the details of transaction and the applicable rules and regulations, which are mostly set by the card network [61, 62]. The cardholder liability is mostly capped (around $50) by regulation [61]. The card network mostly covers the cost of fraud in the card-present (i.e. physical) world [61]. The merchant is mostly held liable for fraud in the card-not-present (e.g. online) world [61]. However some card networks offer authentication methods where the liability of 'friendly-fraud' shifts to the issuer [62].

An example is 3-D Secure (3DS). In this case the issuer's system and the merchant's website set up a direct connection to make sure it is (or to *authenticate*) the cardholder who initiates the card-not-present payment. MasterCard and Visa brand this service as Verified by Visa and MasterCard SecureCode [61]. More ancient is the CVC, which has also been developed by card networks to authenticate the cardholder (or in this case the physical card).

4.1.2. Revenues and Expenses

Another angle on the different stakeholder interests can be provided by looking at the breakdown of the revenue and costs of the parties involved. As shown in Figure 4.1, issuers mainly derive their revenue from interest paid by cardholders over the offered credit (61,8%), and *interchange fees* paid by the parties in the payment network (23,3%). Interchange fee is a share of the transaction value paid to the issuer to cover expenses. Their costs mainly consist of charge-offs (57,4%), operations- and marketing (30,4%), and cost of funds (11,0%) [63, 64]. No major changes have occurred in the market which influence these trends, except for price cap regulations regarding interchange in for instance the European Union and the United States of America (USA). For this reason the share of interchange in the issuers' revenue has dropped significantly for European issuers [65].

On the other side of the network, acquirers and PSPs derive virtually all their revenue from an additional *markup* added to the interchange fees charged to the merchant [20]. Unlike issuers, which compete over individual cardholders, acquirers compete over merchants and have fewer ways to differentiate their apart from price and conversion [20]. Conversion is defined as the percentage of transactions which get authorised. Networks also receive a fee of approximately 0,10 percent of the transaction volume, called the *dues and assessments fee* [20]. Although this is much smaller than the interchange fee which is roughly 1 to 3 percent of the transaction volume [66], and the additional markup for acquirers and PSP, which if combined is roughly 0,1 to 1,8 percent mainly depending on the overall transaction volume of the merchant [20]. So in regular four-party networks, merchants pay a share of the total transaction volume to the issuer, network, PSP, and acquirer. The total fee is called the Merchant Service Charge (MSC), or *merchant discount* [20]. Some constructions are devised to reduce the on interchange fees, like private-cards [21], and no interchange fees are due for three-party networks because issuer, acquirer and network are the same entity [22].

Most scholars and practitioners in the payment card markets follow the pioneering work of Baxter [67], and emphasize *two-sided market externalities* [24]. Both sides of the market, receive positive externalities from being part. On one side the cardholder benefits from using its card at a merchant,

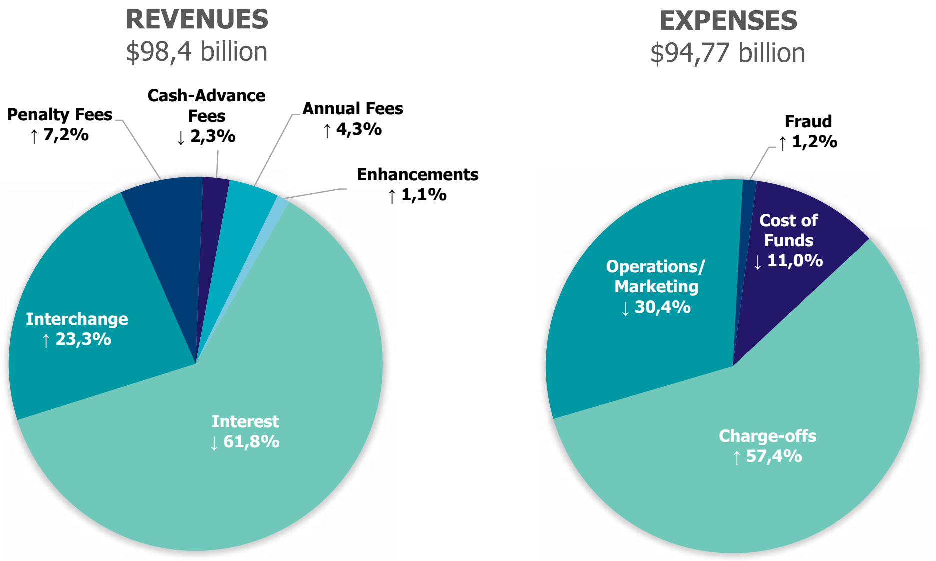

Figure 4.1: A pie chart showing the overall cost and revenue breakdown for Visa and MasterCard issuers in 2010 [63, 64]. The arrows indicate the difference with 2009.

on the other side the merchant benefits from making the sale. However, clearly, acquirers, serving the merchants, and issuers, serving the cardholders, mainly have opposing interests. Although the networks suppose to have an objective role, there is a general consensus that the networks operate more with the issuer interests in mind [20]. There are two likely reasons for this. One historical reason is that Visa and MasterCard used to be bank-owned organisations [20]. Additionally issuers provide the main revenue stream by paying for their card programs and services (38% for Visa) [68]. Other revenue streams are processing fees for authorisation, clearing and settlement (25% for Visa); and fees for international transactions (24% for Visa) [20, 68]. However the share of the service revenues is declining [68], and Visa and MasterCard converted into publicly traded companies in 2006 [20], and it is expected that this will move the network's interest more to the acquirers on the long term [20].

4.1.3. Policies

Policies in the payment industry highly differ per country [65]. Most market interventions are aimed at correcting the negative externalities which resulted from the power position of the networks [22, 65]. In several jurisdictions merchants are allowed to surcharge customers for certain types of payments. Most scholars argue that the interchange fees, the fees collected to cover the costs of the issuer and the network [6], would be neutral if merchants would be allowed to surcharge card transactions [65]. Depending on the jurisdiction interchange fees can also be regulated, to force networks to reduce the fee. Some payment networks have a so-called "honour-all-cards rule", this implies that if a merchant accepts one of the network's cards then it must accept all the network's cards. Around 5 million merchants sued MasterCard and Visa over this obligation and the case got settled out of court in 2003 [65, 69].

4.2. Moral Hazard in Authorising Payments?

In this section we expand on our argumentation from the problem statement in Chapter 2. In the problem statement we argue that the different driving forces of the parties in the payment industry (outlined in the previous section) can give rise to *moral hazard* in authorising payments. First we define our interpretation of moral hazard. Second we argue why this applies to the authorisation phase in the payment process. Third we describe what this might implicate for the refusal behaviour of issuers.

4.2.1. Interpretation of Moral Hazard

We interpret moral hazard according to the definition from Klick and Mitchell [8] which is: *"inefficiencies that occur when risks are displaced or cannot be fully evaluated"*. Moral hazard typically occurs when certain preconditions are met. These preconditions include the following: *separation of ownership* (i.e. the party taking actions is different from the party experiencing the consequences), *information asymmetry* (i.e. the parties do not have similar information available), *hidden action* (i.e. privately taken actions affect the probability distribution of the outcome) [9–11]. In other words, moral hazard occurs when a party with more information about its actions (or intentions) has a tendency to behave inappropriately from the perspective of a party with less information.

The nature of the three preconditions make moral hazard hard to observe [9]. Hence it is hard, if not impossible, to proof that moral hazard effects actually exist. However by forming an idea of the 'hidden action' an issuer might make in combination with our method to dissolve (part) of the information asymmetry we can in potential find evidence of moral hazard effects.

4.2.2. Authorising Payments: A Principal – Agent Problem

Moral hazard often arises when there is a *principal – agent problem* [70]. In a principal – agent problem, one party (the *agent*) acts in the interest of another party (the *principal*). The agent has more information about their actions (or intentions) than the principal. Because of this information asymmetry the principal cannot fully monitor the agent. A principal – agent problem arises when the interests of the principal and the agent are not (completely) aligned and the agent has incentives to act inappropriately (from the perspective of the principal).

We argue this also applies to authorising payments. Figure 4.2 shows how we frame the principal – agent problem between the card network (the principal) and the issuer (the agent). In this case the issuer authorises payments which is also in the interest of the card network. The issuer has the information on which criteria (or 'decision rules') it uses to authorise or refuse payments. For this reason the card network cannot fully monitor the behaviour of the issuer. Hence the issuer can decide to refuse more payments than desired by the card network when it is in their best interest to do so.

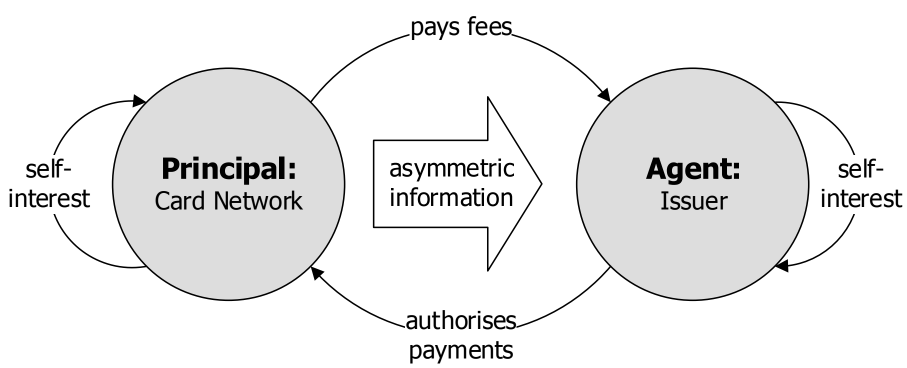

Figure 4.2: A diagram of the principal – agent problem related to the card network and the issuer.

While issuers serve their own interest, we argue the card network from its intertwined position has two-sided interests. On the one hand it serves the *merchant-side* of the platform (including acquirers and PSPs) and on the other the *cardholder-side* (including issuers), because their business model depends on both sides. However there is a general consensus that the networks operate more with the issuer interests in mind, because card networks a) used to be bank-owned organisations, and b) derive most their revenue from issuers (see Section 4.1.2). This adds more complexity to this principal – agent problem from the perspective of the merchant. Ultimately the merchant and the cardholder willing to pay at the merchant experience the effects of undesired (from their perspectives) refused payments.

The cardholder and the merchant are unlikely to always agree with the issuer's decision to authorise or refuse a certain payment. Issuers want to encourage or discourage certain forms of payment based on their business strategy. Cardholders on the other hand, want to pay for any desired good or service, while merchants want to charge any cardholder which is interested in their goods or services. All parties want to minimise their liability.

Other authors have also identified moral hazard and principal – agent problems in the payment industry. It has been used in relation to the fee structure of card networks (i.e. interchange fees) [71–74]. Others have associated it with the distribution of costs of fraud and what the implications of this distribution are for the prevention of fraud [7, 61]. We do not find work which relates it specifically to the authorisation of payments as we argue.

4.2.3. Implications for Refusal Behaviour of Issuers

On the basis of the moral hazard assumption we can infer the implications for the refusal behaviour of issuers. Issuers want to encourage or discourage certain forms of payment based on their business strategy and minimise their risks due to liability. We expect issuers to safeguard these interests more often than desirable by the other parties in the payment network.

[redacted]

4.3. Summary

In this chapter we discuss the driving forces in the payment industry from three perspectives. The three perspectives are: *financial risks due to liability*, *revenues and expenses*, *policies*. Table 4.1 outlines the main findings on each perspective.

Cardholder liability is mostly capped by regulation. Merchant are mostly liable in cases of chargebacks and card-not-present fraud not via 3DS. When merchants go bankrupt and PSPs provide a full-managed service then the PSP becomes liable in cases of fraud. If a PSP doesn't provide this service or itself goes bankrupt the acquirer becomes liable. Card networks mostly escape from any liability, while issuers are mostly liable in cases of card-present fraud or card-not-present fraud via 3DS.

In terms of revenues merchants pay a small amount per transaction to each party in the network. This is the main revenue source for PSPs and acquirers. On the other hand card networks gain slightly more from the revenue from service fees paid by the issuers, and issuers itself gain most more than twice as much from interest paid by cardholders than from the fee per transaction (i.e. interchange fee).

Policies depend on the jurisdiction. Possibly merchants can be subject to an honour-all-cards rule from the card network or can surcharge certain payment methods to incentivise shoppers too choose for one with relatively low transaction fees. In some jurisdictions regulation is in place which puts a cap on interchange fees.

Table 4.1: Simplified overview of driving factors for each main actor in the payment network from different perspectives.

Actor	**Perspective**		
	Liability Risks	**Revenue**	**Possible Policies**
Cardholder	• Mostly capped (around 50$)		
Merchant	• Chargebacks • Card-not-present fraud not via 3DS	• Sales	• Honour-all-cards rule • Surcharge payment methods
PSP	• Merchant bankruptcy with full-managed service	• Markup per transaction	
Acquirer	• Bankruptcy prior risk-bearing party (merchant or PSP)	• Markup per transaction	
Card Network	• Insignificant	• Service fees • Dues and assessments fees	• Caps on interchange
Issuer	• Card-present fraud • Card-not-present fraud via 3DS	• Interest • Interchange fees	

We argue that the different driving forces create an environment to give rise to *moral hazard effects* in authorising payments. We substantiate this claim by framing the card network – issuer relation as a *principal – agent problem*. Because of the unaligned interests and the existence of an *information asymmetry*, the issuer has incentives to refuse more payments than appropriately (from the perspective of the other actors).

On the basis of the moral hazard assumption we can infer the implications for the refusal behaviour of issuers. Issuers want to encourage or discourage certain forms of payment based on their business strategy and minimise their risks due to liability. We expect issuers to safeguard these interests more often than desirable by the other parties in the payment network. We argue issuers are cautious with payments for which the risk of liability conflicts is higher.

On the basis of these conclusions we can answer **RQ 1.**: "What kind of authorisation decision rules would signal moral hazard effects?" [redacted]. We identify two of these types of payments:

- [redacted]
  - [redacted]
  - [redacted]
- [redacted]
  - [redacted]

In the next chapter we explore the problem further. First we research what information the data of a PSP contains. Then we perform a small case study on the refusal behaviour of one issuer. Finally we gather and research the incidents of a specific PSP on systematic issuer refusals. This is the first possibility to confirm the conclusions drawn in this chapter.

5

Exploration of Payment Data and Issuer Refusals

In Chapter 2 we outline the research approach. An important step in the approach is the exploration of payment data and issuer refusals. Based on this knowledge we are able to answer **RQ 2.** "What are the characteristics of systematic issuer refusals?" and **RQ 3.** "What kind of payment data can be used to analyse systematic issuer refusals?".

In this chapter we present the highlights of this exploration. First, we discuss the payment payment information available in the data of PSPs. Second, we discuss the refusal behaviour of a specific issuer. Third, we discuss the characteristics of systematic issuer refusals. We obtain these characteristics by gathering authorisation incidents in a special tool and labelling these incidents. We summarise this chapter in the last section. The data related to this chapter is obtained from Adyen, the PSP where we perform this study.

5.1. Payment Data Available to PSPs

The available information differs between the organisations in the payment network. In principle PSPs get the data about whether a payments is approved or refused from an acquirer, this acquirer is connected to the issuer via the card scheme or has a direct connection. A direct connection to an issuer system is very rare though, only some acquirers have realised this and only for some large issuers. In some rare cases, typically when the issuer system is down, the card network processes the payment on behalf of the issuer without consultation, this is called STand-In Processing (STIP).

The issuer has the most detailed information on which it bases its decision to authorise or refuse a payment request. Parties at later stages in the authorisation process (i.e. network, acquirer, and PSPs) have less information. For instance the issuer has more detailed information about the account, like the account balance. The network nor the acquirer have access to this information.

An issuer can have an abundance of different reasons to make a certain decision. The standards which describe the electronic messages (e.g. ISO 8583) being passed back and forth in the payment network have a limited variety in refusal reasons. Hence the refusal reason in the electronic message send to the card network does not always represent the 'true' reason behind the issuer's decision.

When the network receives the information from the issuer, some information is not contained in the message and the internal reasoning of the issuer is not fully represented in the electronic message send to the network. Once the network receives the payment request, it may have information available about the issuer and shoppers enrolled for the a certain BIN. When the network forwards the payment to the acquirer it does not share all this information. Consecutively an acquirer may also not share all information (e.g. if STIP was used or not) when it sends the payment to the PSP. Thus PSPs have to deal with limited information.

Table 5.1 depicts the entities and attributes of the core (card) payment data available to PSPs. This information can be used to find groups of payments which are structurally refused on certain BINs. The complexity of knowing exactly why transactions are refused, is partly due to the fact that acquirer responses are often vague, wrongly assigned, and there are multiple standards which are implemented in different ways. We elaborate on this statement in the next section.

Table 5.1: Information available at PSP level.

Entity	**Attribute**	**Type**	**Description**
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]

N.B. This table is continued on the next page

Table 5.1: (Cont.) Information known to Adyen on a PSP level.

Entity	Attribute	Type	Description
[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]

5.2. Case Study on the Refusal Behaviour of a Specific Issuer

In order to gain insight in the refusal behaviour of issuers we perform a small case study on the behaviour of a specific European issuer. First we study the data related to the BINs related to the issuer. We use information the card network (in this case Visa and MasterCard) provide to PSPs about which BIN ranges are used for which purposes (e.g. which issuer, which card type which card type, etc.). Second we present the highlights from an interview with an expert of the issuer.

5.2.1. Characteristics of BINs

[redacted]

There are also a number of BINs which are assigned to the same range (and thus these range of BINs are used for similar purposes) and some BINs which cover multiple ranges (and thus such a BIN can be used for different purposes). Hence we conclude that a 6 digit BIN might not be the perfect characteristic to differentiate payments and locate payments. However, only the first 6 and the last 4 digits of a card number may be used for analysis [75]. The PCI-DSS imposes this to companies processing payments. [redacted]

Table 5.2 describes the volume and authorisation rate on the BINs [redacted]

5.2.2. Acquirer Responses of Refusals

[redacted]

Table 5.3 shows the unique responses of refusals send by the different acquirers active on BIN B. [redacted]

5.2.3. Interview with an Expert from the Issuer

[redacted]

Table 5.2: Volume and authorisation rates for the BINs related to the studied issuer in the period between March 1 and May 31 2014.

[redacted]

Table 5.3: Overview of unique responses of the refusals on BIN B during the period between March 1 and May 31 2014.

[redacted]

[redacted]

5.3. Characteristics of Systematic Issuer Refusals

In order to establish the characteristics of systematic issuer refusals we gather authorisation incidents in a special tool and label these incidents. First, we will describe the highlights of this tool. Second, we describe the characteristics induced from these incidents.

5.3.1. Tool to Investigate Issuer Refusals and Gather Incidents

[redacted]

5.3.2. Labelling Incidents to Induce Refusal Characteristics

Figure 5.1 outlines the approach to induce the refusal characteristics. In order to realise this we gather 40 incidents on the authorisation of specific payments on specific BINs via the tool. Per incident we describe the characteristics of the refused payments. Next we label the incidents by first checking if the incident can be assigned to an existing label. If so, we assign the label, otherwise we create and assign a new label to describe the incident.

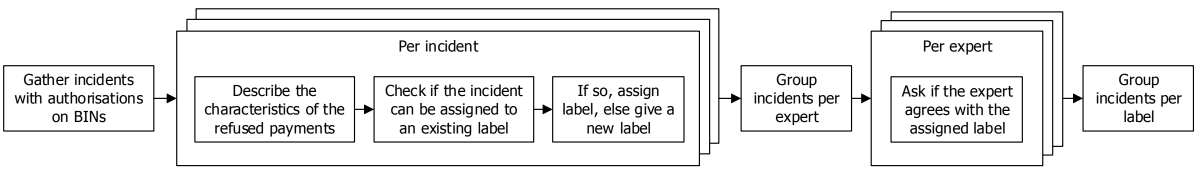

Figure 5.1: Outline of the approach to explore the characteristics of systematic issuer refusals by labelling the reported incidents.

[redacted]

Figure 5.2: Screenshots of the overview page and the detail page of the tool to administer incidents.

[redacted]

Figure 5.3: Screenshot of the page to manually investigate the authorisation rates for different types of payments per merchant, and per acquirer.

We validate the labels with experts to make sure that all incidents are assigned to the correct label. We realise this by grouping the incidents per expert and then we ask each expert if all the incidents which they reported have been assigned the correct label. Finally we group the incidents per label to create an overview. We show the highlights in Table 5.4. Appendix D describes the detailed information about the labelling of the incidents.

Table 5.4: Observed characteristics of systematic issuer refusals. Signalling attribute values are specific to the acquirer and issuer related to the incident.

Characteristic	Description	Signalling Attribute Values
[redacted]	[redacted].	[redacted]
[redacted]	[redacted].	[redacted])
[redacted]	[redacted].	[redacted]
[redacted]	[redacted].	[redacted]'
[redacted]	[redacted].	[redacted]'
[redacted]	[redacted].	[redacted]
[redacted]	[redacted].	[redacted]
[redacted]	[redacted].	[redacted]

N.B. This table is continued on the next page

Table 5.4: (Cont.) Observed characteristics of systematic issuer refusals.

Characteristic	Description	Signalling Attribute Values
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

N.B. This table is continued on the next page

Table 5.4: (Cont.) Observed characteristics of systematic issuer refusals.

Characteristic	Description	Signalling Attribute Values
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

5.4. Summary

In this chapter we explore the payment data at our disposal and the patterns describing systematic issuer refusals. First we describe the characteristics of the payment data. Second we study the refusal behaviour of a specific issuer and interview an expert from this issuer to ask for explanations for our observations. Third we discuss the characteristics of systematic issuer refusals. We obtain these characteristics by gathering and labelling incidents related to systematic issuer refusals.

Table 5.4 provides an overview of the characteristics of systematic issuer refusals which we observed in the incidents related to systematic refusals. Hence we answer **RQ 2.** "What are the characteristics of systematic issuer refusals?" In order to answer **RQ 3.** "What kind of payment data can be used to analyse systematic issuer refusals?" it is important to understand how the systematic issuer refusals manifest itself in the data. In this chapter we identify a number of challenges related to this:

1. Not in all cases is the BIN (equal to the first 6 digits of a card number) the lowest level on which an issuer makes a different decision. In some scenarios the issuer differentiates its decision rules on the basis of more digits. However we can not analyse these due to (information) security regulations.
2. In some scenarios the information related to systematic issuer refusals is delicate, due to potential liability of the parties involved.
3. The acquirer response in the electronic payment message is of limited use because:
   (a) Responses are highly acquirer- and issuer-dependent.
   (b) The majority of responses are vague (e.g. "not authorised"), so no clear refusal reason can be derived.
4. In some scenarios expert knowledge is required to derive the precise refusal reason, because the pattern contained in the data is not detailed enough.

Despite these challenges we observe that a lot of patterns can be identified on the basis of the value of specific payment attributes and in most cases the BIN is a good differentiator. Hence there is a lot of potential for DM techniques to mine for these patterns. To realise this we identify which payment attributes contain the necessary information to find these patterns.

There are also other contextual factors which can affect the authorisation rate. For instance the time, date and timezone of the payment, and load on the network. More extremely, the authorisation rate could even be affected by the weather. However we argue that correlations between the authorisation rate and these factors are of limited value. First of all, because we question if a PSP can act on these results (e.g. to date only Gyro Gearloose can influence the weather). Secondly because correlations can be misleading. If we find a correlation this does not mean there is a causation. Hence we only take the payment attributes into consideration which relate to the systematic issuer refusals found in incidents.

[redacted]

The findings from this chapter allow us to answer **RQ 3.** "What kind of payment data can be used to analyse systematic issuer refusals?". We conclude that the following attributes are useful to find the patterns of interest:

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

On the basis of the knowledge from this chapter, in the next chapter we determine which DM technique meets the requirements to be able to find systematic issuer refusals. Additionally we use the knowledge from this chapter to determine the focus when customising the selected DM technique.

6

Selection of Data Mining Technique

In the previous chapter we outline the characteristics of systematic issuer refusals, which we use to determine what payment data is useful to find these refusals. On the basis of this knowledge we determine in this chapter what criteria a DM technique should meet in order to find these refusals in a way that is usable for payment experts. These criteria allow us to compare the different DM techniques and select the most suitable technique for this research. Based on this knowledge we answer **RQ 4.** "Which DM technique has the functionality required to find systematic issuer refusals?". We expand on the background knowledge about the different DM techniques from Chapter 3 to achieve this.

We structure this chapter as follows. In the first section we present and justify the criteria on which we compare the DM techniques. We present the comparison itself in the next section. In the last section we summarise and conclude this chapter.

6.1. Criteria

Han *et al.* [38] note that a DM technique can only be useful if the results are actionable. This is specifically relevant for this application. Ideally experts act upon a specific group of refusals with the same (standard) strategy. Hence, above all, it is important that a DM technique can **find** groups of payments systematically refused on a BIN.

Second, the DM technique must **isolate** the groups of payment with a 'true' low authorisation rate. Meaning the technique should be able to isolate the most generic payment features that distinguish the groups of refusals. For instance when a payment is (only) described by feature X and Y. Isolating groups of payment with a 'true' low authorisation rate requires a technique to exclude a group with feature X, which on its own has a significant low authorisation rate, but has an acceptable authorisation rate if the group which has feature X and feature Y is not taken into account. In this case the group with feature X and feature Y is the reason for the low authorisation rate. Hence we want to find the group with feature X and Y and exclude the group with feature X.

Third, the DM technique must present the results on a high aggregation level in a clear fashion. Otherwise an expert could easily get lost in hundreds (if not thousands) of action items, depending on the size and dimensions in the data set. Hence the results must be easy to **aggregate**. The aggregation must be mutually exclusive to avoid double counting of refusals. Double counting is misleading because it can give a wrong impression about the impact of all group of refusals on the overall authorisation rate. On the basis of the impact an experts can determine the priority to follow up on these refusals. Correct isolation is a necessary ingredient to aggregate in a meaningful mutually exclusive way. For instance aggregating on payments with feature X from the example is misleading, because these payments are not the underlying reason of the low authorisation rate, the reason resides on a more specific level.

Fourth, the experts must be able to easily **interpret** the results to draw conclusions about the nature and priority of this group of refusals. Hence the results must be such that an expert can interpret them without deep knowledge of other domains. For instance the results must not require an expert to have knowledge about DM or the specific payment data format. Only the relevant knowledge about the payment industry is enough.

Fifth, the mining must have a reasonable **performance** in terms of processing time or required hardware. Depending on the application of DM, performance has a higher or lower importance. Refusals have a direct influence on the merchant's revenue, thus there is significant value in generating insights in real-time. Hence techniques which are efficient with processing power are preferred. Besides we evaluate if techniques can apply a 'divide and conquer' strategy. If we can break down a DM task into two or more sub-tasks we can perform the tasks recursively or distribute tasks over multiple processors and systems. Another option is to separate analysis from presentation. For instance when the aggregated results of the analysis are periodically updated and stored in a database, the visualisation (i.e. presentation) can be generated based on this data in real-time. Trends regarding the decision logic of issuers are not expected to change every second. Hence periodical updates are acceptable.

In summary, a DM functionality should meet the following requirements and criteria:

- **Find** groups of systematically refused payments on a BIN
- **Isolate** the most generic groups where the low authorisation rate is (nearly) unaffected by more specific groups
- **Aggregate** groups of systematic refusals per BIN to a mutually exclusive aggregation on PSP level
- **Interpretable** results by experts without knowledge about DM or the specific payment data format
- **Performance** should be acceptable to allow for periodical updates

6.2. Comparison of Data Mining Techniques

In this section we compare the DM techniques on the basis of the previously defined criteria. We outline the possible applications for the relevant techniques of the different functionalities, and discuss to which extend the techniques are able to meet the criteria and requirements. We do this by means of a qualitative comparison as outlined in Table 6.1. We mark functionalities with '+' if it outperforms the other functionalities on the specific criteria, '0' if it performs average, and '-' if it underperforms compared to the other functionalities. We start with the functionality with the least suitable techniques first and end with functionality we select.

Table 6.1: Criteria on which the DM techniques will be compared

Functionality	**Criteria**				
	Find	**Isolate**	**Aggregate**	**Interpretable**	**Performance**
Functionality			*Score*		

6.2.1. Outlier Detection

We can frame the search for groups of systematic refusals, as the search for payments which behave rather different from typical transactions. *Outlier detection* deals with such challenges [38]. In the payment industry outlier detection has contributed greatly in dealing with fraud [36].

However, in this case it is arguable if we can frame the search as a search for outliers. In the case of systematic refusals almost all related payments fail, thus the affected payments are not outliers with respect to each other. Still we can frame the problem as searching for *collective outliers*. However this is arguable as well. As Figure 6.1 shows, a significant group of BINs is having an authorisation rate which is significantly lower than the average for card transactions (+/- 80%). This makes it also questionable if it is correct to frame the search as a search for collective outliers.

If the patterns of interest can not be regarded as outliers the techniques from this functionality are not applicable. Hence we conclude that techniques from this functionality are not suitable to find groups of systematic refusals. Because this is a fundamental requirement we do not select a technique from this functionality. Table 6.2 depicts this conclusion.

Table 6.2: Rating of the outlier detection functionality on the relevant criteria.

Functionality	**Criteria**				
	Find	**Isolate**	**Aggregate**	**Interpretable**	**Performance**
Outlier Detection	- not outliers	N/A	N/A	N/A	N/A

Figure 6.1: Scatterplot showing the transaction of Adyen of an arbitrary month with all BINs shown as a point in the scatterplot on the basis of the location of the BIN in the range 000000-999999 on the horizontal axis. On the vertical axis the number of transactions directed to this BIN is shown using a logarithmic scale. The actual numbers are omitted for confidentiality reasons. The data points are coloured on the basis of the authorisation rate and the data points with the lowest authorisation rate are shown on top.

6.2.2. Characterisation and Discrimination

We can also query the database to get a summarised, concise, and yet precise description. *Characterisation and discrimination* deals with such tasks [38]. In many cases, techniques from this functionality are able to meet the demands of experts to extract certain information from the data.

For instance, experts within a PSP use a responsive visualisation we created. Experts use this visualisation to investigate the authorisation rate of certain groups of payments on certain BINs. Problem areas are pinpointed by comparing the authorisation rate of the different groups and limiting to certain merchants or acquirers to check for differences. Figure 6.2 shows this visualisation.

However, a query is highly inefficient and impractical to automate the task of pinpointing problem areas. In theory, we can create a data class per BIN containing the authorisation rate on all dimensions we want to check on. This has two major disadvantages.

First we need to search the complete combinatorial space which is highly inefficient. As we describe in the previous chapter there are nine important dimensions for a group of systematic refusals per BIN. On some dimensions, such as merchant identifiers, the amount of categories can be in the order of ten thousands. This leads to an explosion of combinations on which we need to search.

Second, if we have the authorisation rate on all dimensions it is hard to isolate the most generic groups where the low authorisation rate is (nearly) unaffected by more specific groups.

Hence we conclude that this functionality is limited in finding and isolating groups of structurally refused payments. Options to achieve this have such a low performance that these options are rendered infeasible. Table 6.3 depicts this conclusion.

Table 6.3: Rating of the characterisation and discrimination functionality on the relevant criteria.

Functionality	**Criteria**				
	Find	**Isolate**	**Aggregate**	**Interpretable**	**Performance**
Characterisation and Discrimination	- limited in dimensionality	- limited in dimensionality	+ complete control	+ highly customisable	- complete combinatorial space

Figure 6.2: Responsive visualisation for a specific BIN showing the authorisation rate (with daily precision) for certain groups of payments. The visualisation can be limited to certain merchants (i.e. companies) or acquirers only and the date range can be specified. The figure depicts the successful mediation of a Continuous Payment Authority (ContAuth) problem on this BIN.

6.2.3. Cluster Analysis

We can also organise the payments or refusals into similar groups and then induce the traits that define a group. *Clustering* deals with this type of tasks. In more generic terms, clustering tries to find a hidden structure from unlabelled data (i.e. *unsupervised learning* definition).

Figure 6.3 shows an example of how experts use clustering to detect fraud attacks for a merchant on certain time intervals. Based on a probability threshold clusters can be formed on the basis of the ratio between the fraud rate compared to the volume and the fraud rate compared to the refusals. If this ratio is in a certain area there has likely occurred a fraud attack on this time interval. In this case experts use clustering for outlier detection purposes.

Clustering can be useful to find groups of systematically refused payments. It can cluster payments on the relevant dimensions to find hotspots of certain groups which are not authorised (i.e. located mainly on the FALSE side of the authorised dimension axis). Some clustering techniques group the data in mutually exclusive clusters, however it is also possible to set it up in such a way that one data point may belong to multiple groups (e.g. probabilistic clustering techniques). This can be useful because groups of systematically refused payments could be partly overlapping.

An advantage of clustering is that these techniques are often highly performing. Clustering typically searches for local optimums, and hereby prune the search space significantly. This makes them highly performing [76].

However, most clustering algorithms work well on data with two or three (mostly numeric) dimensions [38]. In this case the data is high-dimensional (at least 10 relevant dimensions) and the dimension are mainly categorical. Getting the desirable clusters on such a dataset is challenging (curse of dimensionality), especially considering the fact that such data is likely to be very sparse and highly skewed [38].

Another challenge with clustering is that the clustering result needs to be interpretable and comprehensible, because especially with a lot of dimensions it is hard to semantically describe a cluster [38]. Additionally clustering algorithms require the number of clusters to create as input [38]. This is challenging, because we do not know how many problems there are.

In conclusion, it is impractical to find groups of systematically refused payments using clustering, because clustering techniques find fuzzy clusters which need to be interpreted visually and are hard to

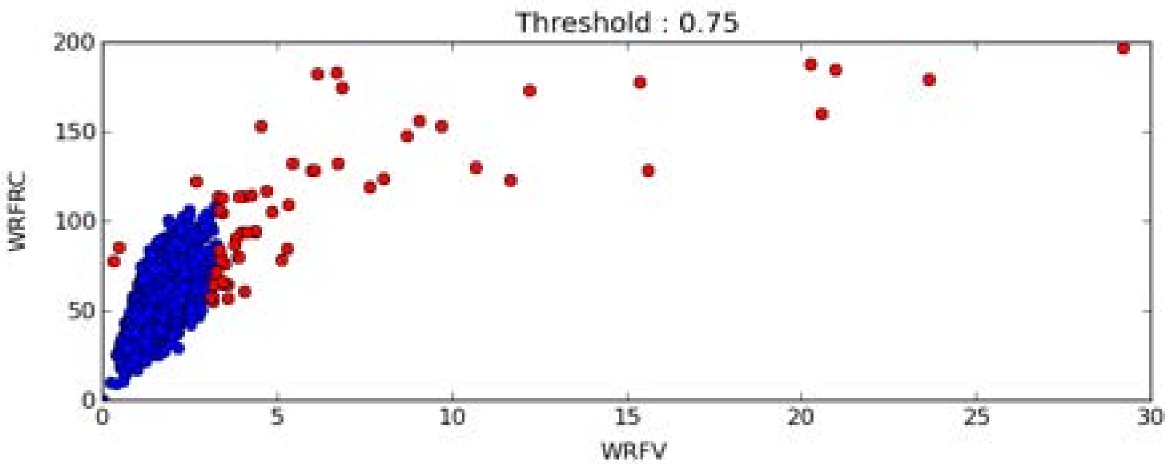

Figure 6.3: Visualisation of possible fraud attacks on one merchant. Each data point represents one hour of payments of a specific merchant. The horizontal axis shows the Weighted Rate Fraud versus Volume (WRFV) and the vertical axis the Weighted Rate Fraud versus Refusal Count (WRFRC). The blue dots represent the time slots which are most likely healthy. The red dots represent the time slots on which most likely fraud attacks occurred. The rates are weighted to make sure small volumes do not have a large impact on the visualisation.

describe semantically. The lack of a precise descriptions makes it hard to aggregate and isolate groups. Table 6.3 depicts this conclusion.

Table 6.4: Rating of the cluster analysis functionality on the relevant criteria.

Functionality	**Criteria**				
	Find	**Isolate**	**Aggregate**	**Interpretable**	**Performance**
Cluster Analysis	0 fuzzy refusal areas	- groups not identifiable	- groups not identifiable	- visual, hard to describe semantically	+ search for local optimum

6.2.4. Classification and Regression

[redacted]

[redacted] Figure 6.4 [redacted]

[redacted]

[redacted]

Figure 6.4: An example of how a decision tree ...

[redacted].

[redacted] [38, 77]. [redacted] [77]. [redacted] [77]. [redacted] Figure 6.4 [redacted].

[redacted] [38]. [redacted] [38]. [redacted].

[redacted].

In conclusion, we can not ensure a that techniques from this functionality find all ('real') groups of systematically refused payments and it is very complicated to isolate groups. We do not select techniques from this functionality because these criteria are fundamental for finding groups of systematically refused payments. Table 6.5 depicts the conclusion regarding this functionality.

Table 6.5: Rating of the classification and regression functionality on the relevant criteria.

Functionality	**Criteria**				
	Find	**Isolate**	**Aggregate**	**Interpretable**	**Performance**
Classification and Regression	0 limited due to 'butterfly' effect	0 limited due to overfitting	+ using meta-rules	+ concise and useful information	+ discards subset at each split

6.2.5. Mining Frequent Patterns, Associations and Correlations

[redacted].

[redacted]

[redacted]

[redacted] ⇒ [redacted]
[redacted] (6.1)

[redacted] [38, 78]. [redacted]

Isolation of Most Interesting Group of Refusals

[redacted]

Mutually Exclusive Aggregation

[redacted]

[redacted]

Performance

[redacted] [77]. [redacted]

[redacted] [38]. [redacted] [77]. [redacted]

[redacted] [38, 78]. [redacted]

[redacted]

In conclusion, association rule mining is able to find all groups of systematic refusals above a certain impact, provide highly interpretable information to the user, with an acceptable performance. Although we are able to create mutually exclusive (or even complete Mutually Exclusive and Collectively Exhaustive (MECE)) aggregations, these aggregations do not precisely describe the situation. This is because the 'true' rules are not isolated from 'false' rules for which the low authorisation rate is entirely caused by another rule (the 'true' rule). Table 6.6 depicts the conclusion regarding this functionality.

Table 6.6: Rating of the mining frequent patterns, associations and correlations functionality on the relevant criteria.

Functionality	**Criteria**				
	Find	**Isolate**	**Aggregate**	**Interpretable**	**Performance**
Mining Frequent Patterns, Associations and Correlations	+ all above certain impact	0 possibly with extra measures	+ using meta-rules	+ concise and useful information	0 possibly by pruning search space during mining

6.3. Summary

In this chapter we discuss the different DM functionalities in light of the relevant criteria. First we outline the possible applications for the techniques of the different functionalities. Table 6.7 summarises these possibilities.

Table 6.7: Possibilities per DM functionality in the context of this research.

Functionality	**Possibility**
Outlier Detection	N/A
Characterisation and Discrimination	Describe the (aggregated) performance of a BIN on several features (characterisation), and compare the performance of the features between different groups (discrimination)
Cluster Analysis	Organise payments or refusals into similar groups and then induce the features that define a group
Classification and Regression	Create a model which describes (i.e. predicts) which features of a payment are likely to lead to an authorisation and which features are likely to lead to a refusal
Mining Frequent Patterns, Associations and Correlations	Search for all combinations of payment features which often lead to refusals

Second we discuss how the relevant possibilities per functionality perform on the important criteria. Table 6.8 provides an overview of the scores of all functionalities.

We conclude that association rule mining has the highest overall score on the criteria, and thus we select this technique. Hence we answer **RQ 4.** "Which DM technique has the functionality required to find systematic issuer refusals?" on the basis of theory only. However to prove if our theory holds in practice we implement this technique and evaluate the results.

Table 6.8: Qualitative comparison between the relevant DM techniques per functionality in the context of this research. The order of the functionalities is related to its overall score, from the lowest score, to the highest score. The row marked in green relates to the selected functionality.

Functionality	**Criteria**				
	Find	**Isolate**	**Aggregate**	**Interpretable**	**Performance**
Outlier Detection	- not outliers	N/A	N/A	N/A	N/A
Characterisation and Discrimination	- limited in dimensionality	- limited in dimensionality	+ complete control	+ highly customisable	- complete combinatorial space
Cluster Analysis	0 fuzzy refusal areas	- groups not identifiable	- groups not identifiable	- visual, hard to describe semantically	+ search for local optimum
Classification and Regression	0 limited due to ‘butterfly’ effect	0 limited due to overfitting	+ using meta-rules	+ concise and useful information	+ discards subset at each split
Mining Frequent Patterns, Associations and Correlations	+ all above certain impact	0 possibly with extra measures	+ using meta-rules	+ concise and useful information	0 possibly by pruning search space during mining

Before we discuss the implementation of the technique we first define how we will apply this technique. This is the topic of the next chapter. In the application we mainly address the challenge related to the challenge with isolating the interesting rules from rules on a higher-level which are not interesting because their low authorisation rate is entirely caused by the main rule. We opt for this because this challenge is fundamental to the other challenges.

7

Method to Find Systematic Issuer Refusals

In the previous chapter we identify association rule mining as most promising technique for this research. We also note that isolating the rules of interest is the core challenge. We need to solve this challenge to find the interesting groups of systematic issuer refusals.

In this chapter we discuss how we apply association rule mining. Hereby we theoretically answer **RQ 5.** "How to apply the DM technique to find systematic issuer refusals and present the results in an usable format to experts?". In answering this question we emphasise how we plan to solve the core challenge of isolating the main rules of interest.

We structure this chapter according to the steps in the association rule mining process. First we discuss how the data is preprocessed. Second we discuss the frequent pattern mining. From these frequent patterns rules we induce BIN-specific rules about systematic refusals. In the third section we discuss pattern evaluation measures, which can help in the discovery of the 'true' interesting groups of systematic refusals. We also discuss the Unique Confidence (UC) measure in this section. This measure is fundamental to this research and our main contribution to the field of association rule mining.

7.1. Preprocessing

Data integrity is highly import in the payment industry. For this reason the quality of the data can be regarded as very high, with few inconsistencies. Missing values can exist on some non-critical dimensions, such as additional payment information or information related to the issuer. Because this information is unknown and following a certain pattern it is impossible to fill in these missing values. For this reason we do not perform data cleaning on the dataset with card payments. We treat missing values as a separate feature item in mining for frequent patterns.

[redacted]. Subsequently we reduce the data to only the relevant dimensions which we describe in Chapter 5. More specifically we perform the following operations:

- [redacted]. [redacted] [redacted].[redacted]. [redacted]

Table 7.1: Attributes and (nominal) values of transactions which will be used for association rule mining.

Attribute	Levels
[redacted]	[redacted]
[redacted]	[redacted], etc.
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted], etc.
[redacted]	[redacted], etc.
[redacted]	e.g. [redacted]
[redacted]	e.g. [redacted]
Authorised	TRUE, FALSE

[redacted].

- [redacted].
- [redacted].
- [redacted].

Table 7.1 describes the attributes and possible levels in the data after preprocessing.

7.2. Frequent Pattern Mining

The first step after preprocessing is mining for frequent patterns. We use a relative low threshold for minimum support and minimum confidence of respectively 5% and 70% to make sure most relevant patterns are found. In the remainder of this section we give our arguments on which algorithm we use and whether we mine all frequent itemsets or limit to specific frequent itemsets only.

7.2.1. Algorithm

Algorithms for finding frequent itemsets mainly differ in efficiency and scalability. The methods for mining frequent itemsets can be divided into three groups: a) *Apriori-like algorithms*, b) *frequent pattern growth-based algorithms*, and c) *algorithms using the vertical data format* [38]. Apriori is the basic algorithm for finding frequent itemsets [38].

The Apriori algorithm has strongly influenced the development of the mining for frequent itemsets [38]. It has been proposed by R. Agrawal and R. Srikant in 1994 for mining frequent itemsets for Boolean association rules [43]. The name Apriori stems from prior knowledge, as this algorithm uses an iterative approach, a *level-wise* search, where it uses the k-itemsets to explore the $(k+1)$-itemsets.

A level-wise search is possible because of the Apriori property, that all nonempty subsets of a frequent items must also be frequent [38].

In frequent pattern growth (FP-growth) algorithms, frequent itemsets are mined without the costly candidate generation and thus substantially improving performance. The transactional database is compressed into a highly compact data structure (an *FP-tree*). Then per frequent item or "pattern fragment", a separate database is created which is mined separately [38]. This approach substantially reduces the size of the data sets to be searched, as well as the "growth" of the patterns found in each iteration.

For algorithms using the vertical data format (Eclat), the horizontal data format of Transaction ID (TID)-itemset is transformed into a vertical item-TID_set. Then the transformed data set is mined, to find TID_set interactions based on the Apriori property and other optimisation techniques [38]. For this reason the Eclat algorithm generally outperforms Apriori on datasets with a relative low number of dimensions.

Our preprocessed data has 12 attributes of which 4 are expected to have a lot of levels (acquirer, merchant, Merchant Identifier (MID), Merchant Category Code (MCC)). On such a dataset we don't expect that Eclat will substantially outperform Apriori. FP-growth algorithms are generally known to be faster than Apriori [79], but they are also more complex and for this reason less commonly implemented in existing software packages. For this reason we will base our decision of Apriori versus FP-growth on the choice of software.

7.2.2. Closed and Maximal Frequent Itemsets

Instead of mining for all frequent itemsets, we can also mine for *closed frequent itemsets* or *maximal frequent itemsets*. An itemset is closed, if for an itemset no superset with the same support exists [44]. This can substantially reduce the number of patterns found while preserving all information regarding the complete set of frequent itemsets. Itemsets which are not closed frequent itemset can be safely omitted, because a subset with the same support can be regarded as a similar problem. Rules generated from the subset of itemsets are not interesting, because don't completely describe the related BIN problem.

An itemset is maximal when no proper superset exists [45]. As a result the support of the frequent sub-itemsets can not be derived from maximal frequent itemset, and thus information is lost. For this reason we can not omit itemsets which are not maximal frequent itemsets, because the root-cause of the related BIN problem can lie in a one of the rules generated from these itemsets.

We can conclude that closed frequent pattern mining can reduce the number of rules generated in a way where no valuable information is discarded. Maximal frequent pattern mining overlook certain patterns which contain valuable information. For this reason we will use closed frequent pattern mining in our approach.

7.3. Pattern Evaluation Measures

Good pattern evaluation measures can be of great help in coping with the sheer amount of patterns found. We use thresholds on appropriate evaluation measures to limit to patterns of interest only. We describe the pattern evaluation measures which we use in this section and suggest constraints based on these measures to omit uninteresting rules. Most measures are based on *correlation* [38]. First, we discuss these measures. Second, we discuss pattern evaluation measures which provide more details about the relation between the found rules.

7.3.1. Correlation Analysis

The value of correlations can be explained with an example. When mining for association rules on an arbitrary BIN A we find the following association rule:

$$[redacted] \Rightarrow \text{authorised} = \text{FALSE}) \quad [support = 12\%, confidence = 71\%] \tag{7.1}$$

Without any additional information this rule would probably be interpreted as [redacted] are problematic because 71% of these transactions are refused. However, given that 80% of all transactions on this BIN are refused, this rule self-evidently is not problematic any more. The authorisation rate of these transactions is even higher than the average. Based on this it is evident that unwise business decisions could be taken for rule $A \Rightarrow B$, by looking at the support and confidence alone. These measures do not measure the 'true' strength of the correlation between A and B.

There are many different correlation measures. A measure that is often used in statistics to specify the correlation is χ^2, known as *chi-square*. This measure shows how strongly one attribute implies the other according to the following formula:

$$\chi^2 = \Sigma \frac{(observed - expected)^2}{expected} \tag{7.2}$$

A simple correlation measure that is frequently used in association rule mining is *lift* [38].[1] If the lift is less than 1, this means A and B are negatively correlated, and vice versa. Lift is defined by the following formula:

$$lift(A, B) = \frac{P(A \cup B)}{P(A)P(B)} \tag{7.3}$$

The equation equals $confidence(A \Rightarrow B)/support(B)$. In simpler words, lift describes the degree to which the occurrence of one "lifts" the occurrence of the other. For example if A corresponds to a [redacted] and B corresponds to a payment being refused, then a [redacted] is said to increase or "lift" the likelihood of a payment being refused by a factor as large as the value returned by Equation 7.3.

A downside of these common measures is that they are highly influenced by the total numbers of transactions [38]. For instance, if in an arbitrary dataset of transactions would contain 10.000 transactions of A and B, 1.000 transaction of only A, and another 1.000 transactions of only B. The lift would be 9,26 and χ^2 would be 90.557. If there would be 100.000 transaction not containing either A or B. If this would be 100 transactions the lift would be 1 and χ^2 would be 0. Because the actual number of different transaction types could fluctuate highly this dependence on the total number of remaining transactions (known as *null-transactions*) is not desirable [38]. A good measure should be *null-invariant* [38].

A number of *null-invariant* measures which have been discussed in the literature are: *all_confidence*,

[1] Han *et al.* [38] describe lift as $\frac{P(A \cup B)}{P(A)P(B)}$, while the original authors Brin *et al.* [80] describe lift as $\frac{P(A \cap B)}{P(A)P(B)}$. Initially we assumed the formula of Han *et al.* [38] contained a typo because only the intersection (∩) between two sets contains information about the correlation. The union (∪) contains all elements in the set, thus says little about the dependency between A and B.

We remain with the notation of Han *et al.* [38] for *lift* as this is the community's standard notation. During e-mail contact Jiawei Han explained: "A and B are all item sets. For example, A = {diaper}, B = {beer}, A ∪ B = {diaper, beer}, A ∩ B = empty set. We explained this notation at the beginning of the two chapters and we just follow the same notation used in frequent item set mining. You can check Agrawal *et al.* [81] and Agrawal and Srikant [43] notations. This is the standard notation used in the community of frequent item set mining."

max_confidence, *Kulczynski (Kulc)*, and *cosine*. Han *et al.* [38] recommend to use Kulc, because this measure is unaffected by unbalanced conditional probabilities, a "balanced" skewness of the data. Kulc measures that if one object has a certain feature what the arithmetic mean probability is of another object having the same feature. The skewness of the data can be expressed numerically using the Imbalance Ratio (IR). Han *et al.* [38] state that if IR and Kulc are used together they provide a clear picture of the actual situation. The measures are defined according to the following formulas:

$$Kulc(A,B) = \frac{1}{2}(P(A|B) + P(B|A)) \tag{7.4}$$

$$IR(A,B) = \frac{|sup(A) - sup(B)|}{sup(A) + sup(B) - sup(A \cup B)} \tag{7.5}$$

A skewed case would for instance be if from one point of view there is a negative association between Zero Dollar Value Authorisation (Zero Auth) transactions and authorised, because only a low percentage of the authorised transactions are Zero Auth. However from another point of view, there appears to be a positive association, because a high percentage of Zero Auth transactions are authorised. For "balanced" skewed cases it would be fair to treat the correlation as neutral, as the Kulc measure does, and provide information about the skewness via the IR measure.

Although the Kulc and IR measure give valuable insights regarding correlation between the *antecedent*, or left-hand-side (*lhs*) and the *consequent*, or right-hand-side (*rhs*), it is hard to filter rules on the basis of a threshold on these measures. The only threshold that won't accidentally omit interesting rules is that there should at least be a positive correlation. There are other measures necessary to quantitatively assess which rule is interesting over the other.

7.3.2. Relative Measures

In the previous section we concluded that although the Kulc and IR measure give valuable insights regarding correlation, it is hard to filter rules on the basis of a threshold on these measures. In this section we explore if relative pattern evaluation measures can be of help in solving this challenge. First, we discuss a evaluation measure which looks top-down how much the confidence improves for sub-rules. Second, we discuss a measure which looks bottom-up how high the confidence of a rule would be without any specific sub-rule.

Improvement

Bayardo *et al.* [78] introduce a measure called *improvement*, which measures how much a sub-rule improves its super-rule in terms of confidence. More formally improvement is defined as the minimum difference between its confidence and the confidence of any proper super-rule with the same consequent [78]. Given a rule $A \Rightarrow B$ the equation looks as follows:

$$imp(A \Rightarrow B) = min(\forall A' \subset A, conf(A \Rightarrow B) - conf(A' \Rightarrow B)) \tag{7.6}$$

A positive improvement means that removing an item will at least drop the confidence by the improvement. This means that all items in the antecedent are important contributors to the predictive ability of the rule. Thus, rules with negative improvement are typically undesirable, because the rule can be simplified into a super-rule which is more predictive. A constraint greater than 0 for improvement is desirable in almost any application of association rule mining [78]. A larger minimum can often also be justified, because in dense datasets rules are not interesting where additional items only add a marginal increase in confidence.

Unique Confidence

In a way, the improvement measure suggests which sub-rules are not interesting, because they do not explain more patterns. However it can also be the case that a super-rule is not interesting anymore, because they are already explained by a sub-rule. We do not find any measure in literature which addresses this issue.

We introduce a new *uniqueness* measure which we term Unique Confidence (UC). UC measures the confidence of a rule without the maximum part accounted for by any sub-rule. More formally it is the minimum of the division, of support of a rule minus the support of a sub rule, divided by the support of the antecedent of the rule minus the support of the antecedent of the sub rule. Given a rule $A \Rightarrow B$ the equation looks as follows:

$$UC(A \Rightarrow B) = min(\forall A' \subset A, \frac{sup(A \Rightarrow B) - sup(A' \Rightarrow B)}{sup(A) - sup(A')}) \tag{7.7}$$

Similarly to the improvement measure, a threshold for this measure seems logical in almost any application of association rule mining. A value below 50% means that this rule decreases the predictive power with respect to a sub-rule, because the unexplained part of a rule actually performs worse than random in predicting the consequent of a rule. Actually it seems logical that the same confidence threshold applies for the unique part of the rule as applies for complete rules.

Also similarly to the improvement measure, the effect can be easily understood by an end-user and the measure works with the (limited) information which is already gathered during the mining of frequent itemsets and does not require the intractable task of enumerating and computing the support of all possible subsets. Because of this we expect the UC measure to have a minimum effect on performance.

In light of BIN problems, UC tells what the highest level is on which the problem occurs. If we have identified the highest level, than we can assess how severe the problem on lower levels is. In this case we are only interested if a problem on a lower level is at least 5% more severe, because a fluctuation of 5% in the authorisation between different subsets (e.g. different merchants or payment methods) is not considered problematic. Thus the improvement should at least be 5%.

7.4. Aggregation of Rules

To make the output of association rule mining useful for experts it is important to present the output in a concise format. This format should allow an expert to oversee the complete situation. There are two post-processing steps we perform to able to create an aggregation which provides these insights.

In the first step we aggregate similar rules which we find on multiple BINs. First we strip the 'BIN-part' from a rule. This allows us to group rules which are occurring on multiple BIN and calculate aggregated statistics for these rules.

In the second step we group rules which experts regard as very similar. We achieve this using meta-rules which represent a group of rules. Once this mapping is defined we can group on meta-rules and calculate aggregated statistics for these rules. For instance the meta-rule [redacted] could describe the following rules:

- [redacted] $\Rightarrow$ authorised = FALSE
- [redacted] $\Rightarrow$ authorised = FALSE

- [redacted] ⇒ authorised = FALSE
- [redacted] ⇒ authorised = FALSE

Once these post-processing steps are performed the output lends itself for visualisation which allow an expert to oversee the complete situation. For instance we can create a bar chart of the total support count (equal to the affected transaction) for each meta-rule. This graph shows the impact of each aggregated group of systematic refusals.

7.5. Summary

Figure 7.1 broadly outlines the application we describe in this chapter. First we preprocess the card payment data, where we mainly merge payment data from multiple sources, reduce the dimensions and group per BIN and per month. Second we mine per BIN for closed frequent items, induct the association rules and calculate the pattern evaluation (i.e. interest) measures. Third, we use thresholds on these measures to select only the interesting rules. These rules we aggregate over all BINs using meta-rules. Fourth we visualise these aggregated meta-rule rules to provide insight on the impact.

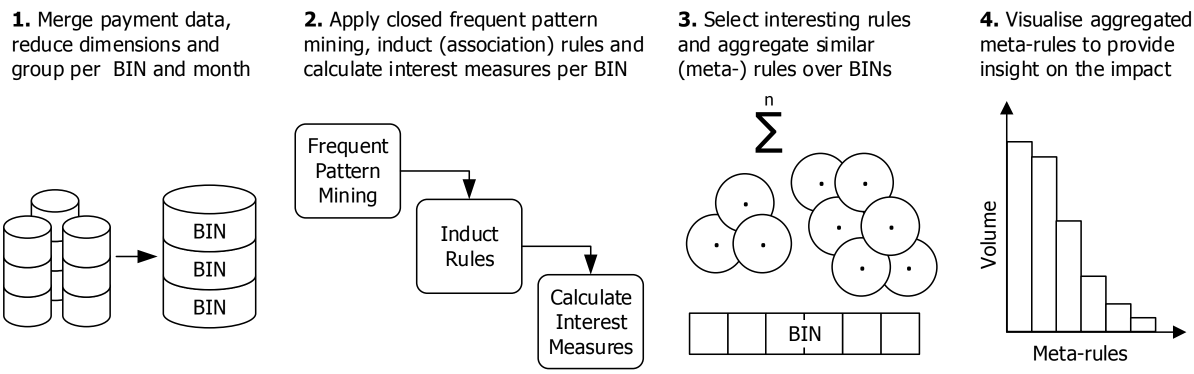

Figure 7.1: Broad outline of the application of association rule mining for this research.

We mainly address the core challenge of dealing with the sheer amount of patterns in our application. First we explain which approaches from prior work in the field of association rule mining could help solving this challenge. We find a number of approaches:

- Always look for the pattern which provides the most lengthy description of an identical problem (in terms of support/confidence), through *closed frequent pattern mining*
- Use the *Kulc* and *IR* measures to find patterns with a sufficient 'true' correlation
- Use *improvement* to find sub-rules which sufficiently improve the confidence of its super-rule

However the measures we find in literature do not address cases where the confidence of a super-rule might be completely due to a specific sub-rule. To address this issue we introduce the UC measure. UC measures the confidence of a rule without the maximum part accounted for by any sub-rule.

As merchants and PSPs are mainly interested in the groups of systematic refusals with the highest impact we look for rules with the highest support. However if we do not constrain using the UC measure we find rules which do not give a correct picture of the situation. This is essential information for experts to act on group of systematic refusals. Hence this measure is fundamental to this thesis.

Table 7.2: Evaluation measures with appropriate thresholds.

Measure	**Threshold**	**Description**
support	5%	A rule should at least comprise 5% of the transactions on a BIN
confidence	70%	A rule should at least be true in 70% of the cases
Kulc	0,5	A rule should explain a positive correlation
IR	None	N/A
UC	70%	A super-rule should be able to at least be true in 70% of the cases without the transactions comprised by an arbitrary sub-rule
improvement	5%	A sub-rule should be 5% more true than its super-rule

We argue this is essential in almost any application of association rule mining, especially when there needs to be searched for rules with one specific attribute in the consequent.

In conclusion, with respect to pattern evaluation measures, we select the *support*, *confidence*, *Kulc*, *IR*, *UC* and *improvement* measures to make it possible to filter out the interesting rules. Table 7.2 summarises and describes the constraints on the pattern evaluation measures.

In this chapter we theoretically answer **RQ 5.** "How to apply the DM technique to find systematic issuer refusals and present the results in an usable format to experts?". In next chapter we explain how this application is implemented. Based on this implementation we eventually evaluate how well this application is able to answer this research question in practice.

8

Implementation

In the previous chapter we describe the theoretical fundamentals of our application. In this chapter we discuss the implementation how we implement this. First we describe the software we use. Second, we describe how we translate the theoretical fundamentals to executable scripts. Third, we describe the dataset on which we experiment. In the last section we summarise this chapter.

8.1. Software

We implement the application in *R*, a programming language for statistical computing and graphics [82]. As an Integrated Development Environment (IDE) we use RStudio [83]. To experiment with different set-ups for association rule mining we mainly use the functionality of the package *arules* [84]. Additionally we use the packages [redacted] and *data.table* for respectively retrieving data from the database system, and fast aggregation and manipulation of large data [85, 86]. In order to ultimately create the visualisation we use *ggplot2* [87].

8.2. R Scripts

In the section we describe the (re-usable) building blocks that are essential for this research. First we describe the steps in the retrieving and preprocessing of the data. Second we describe the mining for association rules. Third we describe how we implement the interest measures which are not implemented in the *arules* package. Finally we describe how we aggregate and visualise the interesting rules.

8.2.1. Retrieving and Preprocessing of Data

Before we can use the data as input to mine for transactions it needs to be pre-processed according to the steps we describe in the previous section. First we retrieve the raw data from Adyen's database system. Second we process this data to cohere with the desired format.

We use two Structured Query Language (SQL) queries to retrieve the raw data as stored in the Adyen's database system. The first query selects the important attribute from the core payment table and joins this with tables containing other important related information (such as account information). The second query selects the information about BIN ranges provided by MasterCard. [redacted]. Because we run out of the 2GB memory available when retrieving the large amounts of transactions required, we divided the job to query the transactions in daily chunks.

Once the data is retrieved in the *data.table* format, we implement a smart join which locates the BIN information (for a range of BINs) which belongs to the BIN of the payment. We the basis of this information we determine the distance between the merchant account and the issuer. Additionally we determine if the currency is the issuer's local currency or not.

When all the desired feature values are determined we convert the *data.table* format to a format required by *arules*. First we convert all the attributes to factors. Second we convert the *data.table* records to objects of the class *transactions*. Once all the steps from this section are performed we can use the data to mine for association rules.

8.2.2. Mining for Association Rules

We separate the mining for rules into the following steps: a) run the Apriori algorithm to compute itemsets; b) induct rules; and [redacted]. The code block related to these steps is shown in Block 8.1. All the code blocks we discuss next work with the output from this block.

```
# create closed itemsets, induct rules and filter on interesting rules
itemsets <- apriori(txs, parameter = list(target="closed frequent itemsets", support = 0.05))
rules <- ruleInduction(itemsets, txs, confidence = 0.7, control = list(method = "ptree",
    reduce = FALSE, verbose = FALSE))
# confidential
```

Block 8.1: Code block to mine for frequent itemsets, and induce and select relevant rules.

8.2.3. Calculating Pattern Evaluation Measures

In this section we discuss the code blocks to implement the Kulc, IR, UC measures for pattern evaluation. To calculate these measures we need the support of the consequent (ride-hand-side of the rule, abbreviated as *rhs*) as input. Block 8.2 shows the code block to lookup this support.

```
rhsSupportLookup <- function(rhs) {
  # lookup itemsets containing the consequent (substr to remove "{" and "}")
  rhsItemsets <- subset(itemsets, items %ain% c(substr(rhs, 2, nchar(rhs)-1)))
  # max. support of these items (with closed pattern mining the single rhs itemset could be
      omitted if the support equals the one of a super-itemset)
  return(max(quality(rhsItemsets)$support))
}
```

Block 8.2: Code block to lookup the support of the consequent.

Kulc

In order to implement the Kulc measure for rule $A \Rightarrow B$, we rewrite the Kulc equation to work with the available support and confidence measures to allow for implementation. The following steps describe how the equation is translated to work with the support and confidence measures:

$$\begin{aligned} Kulc(A,B) &= \frac{1}{2}(P(A|B) + P(B|A)) = \frac{1}{2}(\frac{P(A \cap B)}{P(B)} + \frac{P(B \cap A)}{P(A)}) \\ &= \frac{1}{2}(\frac{sup(\text{rule})}{sup(\text{rhs})} + \frac{sup(\text{rule})}{sup(\text{lhs})}) = \frac{1}{2}(\frac{sup(\text{rule})}{sup(\text{rhs})} + conf(\text{rule})) \end{aligned} \tag{8.1}$$

Block 8.3 explains how we implement the Kulc measure using this equation.

```
# Kulc(A,B) = 1/2 ( P(A|B) + P(B|A) )
kulc <- 0.5 * (
  quality(rules)$support / sapply(labels(rhs(rules))$elements, function(x) rhsSupportLookup(x
      )) # P(A|B) = sup(rule) / sup(rhs)
  + quality(rules)$confidence # P(B|A) = conf(rule)
)
```

Block 8.3: Code block to calculate the Kulc measure for all the rules.

IR

In order to implement the IR measure for rule $A \Rightarrow B$, we rewrite the IR equation to work with the available support and confidence measures to allow for implementation. The following steps describe how the equation is translated to work with the support and confidence measures:

$$\begin{aligned} IR(A,B) &= \frac{|sup(A) - sup(B)|}{sup(A) + sup(B) - sup(A \cup B)} \\ &= \frac{|sup(lhs) - sup(rhs)|}{sup(lhs) + sup(rhs) - (sup(lhs) + sup(rhs) - sup(rule))} = \frac{|\frac{sup(rule)}{conf(rule)} - sup(rhs)|}{sup(rule)} \end{aligned} \tag{8.2}$$

Block 8.4 explains how we implement the IR measure using this equation.

```
# IR(A,B) =  | sup(A) - sup(B) | / ( sup(A) + sup(B) - sup(A u B) )
ir <- abs(
  quality(rules)$support / quality(rules)$confidence # sup(A) = sup(rule) / conf(rule)
  - sapply(labels(rhs(rules))$elements, function(x) rhsSupportLookup(x)) # sup(B) = sup(rhs)
) / (
  quality(rules)$support # sup(A) + sup(B) - sup(A u B) = sup(rule)
)
```

Block 8.4: Code block to calculate the IR measure for all the rules.

UC

In order to implement the UC measure for rule $A \Rightarrow B$, we rewrite the UC equation to work with the available support and confidence measures to allow for implementation. The following steps describe how the equation is translated to work with the support and confidence measures:

$$UC(A \Rightarrow B) = min(\forall A' \subset A, \frac{sup(A \Rightarrow B) - sup(A' \Rightarrow B)}{sup(A) - sup(A')}) \tag{8.3}$$

[redacted]

8.2.4. Aggregating and Visualising Rules

In order to make the output of association rule mining useful for experts we perform a number of steps. In the first step we aggregate similar rules which we find on multiple BINs. First we strip the 'BIN-part' from a rule. Second we group rules which experts regard as very similar. We achieve this using meta-rules which represent a group of rules. Block 8.5 explains how we implement this. Once this mapping is defined we can group on meta-rules and calculate aggregated statistics for these rules.

```r
# write rules to csv
write(rules, file = "~/rules.csv", quote=TRUE, sep = ",", col.names = NA)

# import as data.table
rules <- fread("~/rules.csv", sep="auto", sep2="auto", nrows=-1L, header="auto", na.strings="NA", stringsAsFactors=FALSE, verbose=FALSE, autostart=30L, skip=-1L, select=NULL, colClasses=NULL, integer64=getOption("datatable.integer64"))

# separate bin attribute from rule
rules[,bin:=as.integer(substr(ruleDescription, 6, 11))]
rules[,ruleDescription:=substr(ruleDescription, 13, nchar(ruleDescription)-23)]

# import meta-rules (manual mapping)
metaRules <- fread("~/meta-rules.csv", sep="auto", sep2="auto", nrows=-1L, header="auto", na.strings="NA", stringsAsFactors=FALSE, verbose=FALSE, autostart=30L, skip=-1L, select=NULL, colClasses=NULL, integer64=getOption("datatable.integer64"))

# merge meta-rules with rules (to append meta-rule attribute)
setkey(metaRules, ruleDescription)
setkey(rules,ruleDescription)
rules <- metaRules[rules,nomatch=NA]
```

Block 8.5: Code block to preprocess the data to create a chart from.

Based on the preprocessed data we create a stacked bar chart of the total support count (equal to the affected transaction) for each meta-rule with a separate bar for each card type. This graph shows the impact of each aggregated group of systematic refusals. Block 8.6 explains how we implement this using *ggplot2*.

```r
# plot stacked bar chart
chart <- ggplot(data=rules, aes(x=metaRuleDescription, y=supportCount, fill=txvariant)) +
  geom_bar(stat = "identity") +
  scale_y_continuous(limits=c(0, 80000), expand = c(0, 0)) +
  scale_fill_manual(values=c("#A9D6A6","#378888","#37638D","#152F47")) +
  guides(fill = guide_legend(reverse = TRUE)) +
  geom_text(aes(x=metaRuleDescription,y=groupSupportCount,label=paste(groupBinCount,"BINs", sep=" ")),vjust=0.5,hjust=-0.1, size = 4, stat="sum") +
  labs(list(x="",y="\nNumber of Transactions",fill="Card Type: ")) +
  coord_flip() +
  theme(text = element_text(size=14),
        text = element_text(family="Tahoma"),
        axis.title.y = element_text(size=16),
        axis.text.x = element_text(colour='black'),
        axis.text.y = element_text(colour='black'),
        axis.line = element_line(size=.7, color = "grey"),
        axis.ticks.y = element_blank(),
        legend.position="top",
        legend.title=element_text(size=16),
        legend.key.width=unit(0.4,"cm"),
        legend.key.height=unit(0.4,"cm"),
        panel.grid.major = element_line(size = .5, color = "grey"),
        panel.grid.major.y = element_blank())
```

Block 8.6: Code block to create the stacked bar chart.

8.3. Dataset

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]. Our implementation of association rule mining does not count missing values when creating frequent itemsets [88].

Table 8.1: Description of the (nominal) attributes and features contained in the dataset.

Attribute	Features
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

8.4. Summary

In summary, we describe how the steps discussed in the implementation relate back to the application discussed in the previous chapter using Figure 8.1. First, we retrieve the 11,5 million payments from 5564 BINs using SQL and preprocess the transactions. Second, we mine for frequent patterns, induct the rules, and calculate the improvement measure. Additionally we use a custom script to calculate the Kulc, IR and UC measures. Finally, we apply the constraints on the interest measures, we parse the data and merge the meta-rule description to all the mined rules. Using this data we plot the results in a stacked bar chart.

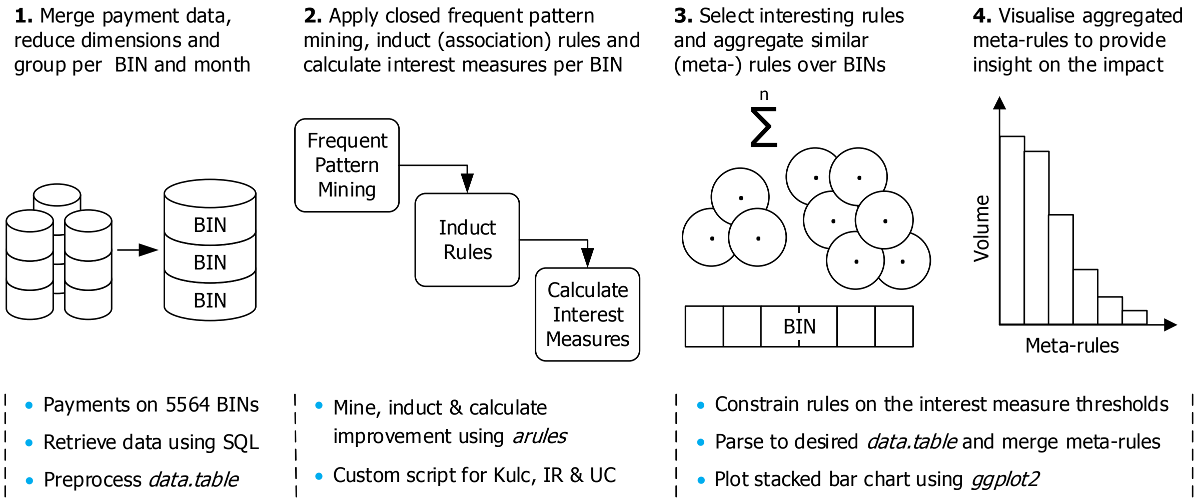

Figure 8.1: Mapping between the application from the previous chapter and the implementation of association rule mining from this chapter.

The next chapter presents the results of the implementation we describe in this chapter. We discuss the results the stages of *finding*, *isolating*, and *aggregating* separately to show the working of the method. Additionally we show the results in terms of *performance*.

9

Results

In this chapter we present the results of the implementation. We structure this chapter according to the steps of the method. First, we describe the found systematic authorisations and refusals. Second, we show how the method isolates the distinctive characteristics in the found systematic refusals. Third, we present a number of aggregations, each from a different perspective, to gain insight in the characteristics of systematic refusals. Fourth, we describe the results regarding the performance of the method. Finally, we summarise this chapter.

9.1. Found Systematic Authorisations and Refusals

The method a total of 600.908 different combinations of payment characteristics which are systematically refused or authorised on a BIN (in at least 70% of the cases). A combination of payment characteristics relates to an association rule found by the algorithm. A rule covers no less than 5% of the payments on a BIN and comprehends at least 50 payments (i.e. *minimum support threshold*).

On average there are 108 rules per BIN. It is highly intractable to manually inspect all these rules. Additionally all the rules can describe highly overlapping groups of payments which likely leads to confusion. To deal with this overlap, the method isolates the distinctive payment characteristics of systematic refusals.

9.2. Isolated Distinctive Systematic Refusal Characteristics

The method mainly uses two relative interest measures to isolate the distinctive payment characteristics of systematic refusals. One is an existing measure named *improvement* and a newly introduced measure named *UC*. Figure 9.1 shows the three arbitrary rules for one BIN with both measures translated to work with the authorisation rate. The authorisation rate is equal to the *confidence* of a rule (or inverted confidence for rules regarding refusals). Hence the unique confidence described by the UC measure and the confidence improvement can be translated to the authorisation rate. The portion is another word for *support* of a rule.

To isolate the rule pointing at the distinctive payment characteristics the method filters out (or 'prunes') rules. It uses the UC measure to filter out rules where the confidence is lower than the initial *minimum confidence threshold* if the sub-rules are deducted. [redacted]

[redacted]

Figure 9.1: Hierarchical representation of arbitrary rules regarding one BIN. Each rule shows the related characteristics, followed by information about the authorisation rate and the portion of BIN payments accounted for. The rules marked with a red bullet are rules where the unique authorisation rate is higher than the threshold of 30% or where the authorisation rate is higher than the super-rule.

[redacted]. In other words, payments with these distinctive characteristics are systematically refused on this BIN.

The method uses these measures in combination with a correlation measure to filter out the uninteresting rules. Table 9.1 shows the exclusion statistics of the filtering. First it limits to only rules which have a positive correlation (excludes 16%). Second it limits to only sub-rules which improve confidence (excludes 49%) and it limits to only super-rules which meet the confidence threshold without any arbitrary sub-rule (excludes 15%). About 58% of the rules are excluded once these constraints are applied simultaneously.

Table 9.1: Exclusion statistics about the constraints applied to limit the number of association rules.

Constraint	Individual Exclusion (%)	Cumulative Exclusion (%)
Rules should have a positive correlation (Kulc > 0.5)	16,0	16,0
Sub-rules should improve confidence (improvement > 0%)	48,8	53,9 (+37,9)
Super-rules should meet the confidence threshold without any sub-rule (UC > 70%)	14,5	58,0 (+4,1)

This leaves 252.150 association rules on 5196 BINs. Of these rules 27.980 rules predict transactions not getting authorised on 940 BINs. These are still impractically many to allow for manual inspection and still rules explain overlapping data which can lead to confusion.

[redacted]

Although the newly introduce UC measure filters out the smallest portion of rules, this does not necessarily mean it is the least useful. Because the filter works top-down it filters out rules with relative high support (i.e. portion of data accounted for) on the consequent of interest (in this case '⇒ refused'). In many situations experts are first considering the rules with the highest support, because these rules have the highest impact. Hence filtering out the deceiving high support rules can be of big value in preventing disastrous business decisions. The aggregation of our method also relies on this.

9.3. Aggregated Systematic Refusal Characteristics

In order to create a mutually exclusive aggregation the method selects the rule with the biggest impact (i.e. support) per BIN. Because a payment can only be routed to one BIN this ensures that no payments is double counted (and thus mutual exclusivity). We can regard these top-level rules (in terms of impact), because the filter on UC makes sure that all remaining rules uniquely explain a phenomena (in this case a refusal).

[redacted]

Table 9.2: Example of rules mapped to one descriptor in combination with their overall support and confidence. Overall in this case, for support is the support with respect to all the payments in the dataset and for confidence is the confidence of all the BIN-specific rules combined.

[redacted]

In a similar way we map the 369 unique rules to 103 unique meta-rules. The meta-rules refer to which payment characteristics are most likely to reflect the payment characteristics on which the issuers bases its decision. Hence in essence these are the decision rule which we set out to induce from the data. In the context of this thesis we regard the meta-rules and decision rules synonymous, and from now on we use the term decision rules.

Because the method enforces mutual exclusivity, we can aggregate the groups of systematic refusals related to a decision rule. From the information contained in the outputted rules we can derive the number of BIN on which a rule is found, and the overall measures for support and confidence. Table 9.3 shows all decision rules with a higher support than 0,01%.

Table 9.3: All grouped decision rules with a total support higher than 0,01%. The table shows for each decision rule how many of the rules relate to credit, debit or other BINs as well as aggregated support and confidence measures. It shows the average support (per BINs) of a decision rule and the total support of the decision-rule in the complete dataset of payments. Additionally it shows the overall confidence of all the decision rules combined.

[redacted]

[redacted]

Figure 9.2: Top decision rules. Next to each bar the number of BINs affected is shown and the colour of the bars represents the card type.

Figure 9.2 shows the decision rules from Table 9.3 graphically in terms of the total number of systematic refusals. [redacted]. [redacted].[1] [redacted].

[redacted].[2] [redacted]. [redacted].

[redacted]. [redacted]. [redacted]. [redacted]. [redacted].

[redacted]. [redacted]. [redacted]. [redacted] [6]. [redacted].

[1] These numbers can be obtained by the summation of the total support of the related rules in Table 9.3.

[2] [redacted]. [redacted].

[redacted] [3] [redacted]

[redacted]

9.3.1. Issuer-Centric Aggregation

The found decision rules allow for an issuer-centric aggregation. Each decision rule refers to a BIN and a BIN refers to an issuer, thus we can map issuer information to a decision rule. On the basis of the dataset concerning issuers provided by MasterCard (see Section 8.3), we map an issuer to a decision rule. [redacted]

Figure 9.3 shows the number of systematic refusals aggregated over all decision rules per issuer. [redacted]

Figure 9.4 shows a breakdown of the decision rules for the top-4 issuers. For each issuer there is one decision rule with a significant higher number of systematic refusals compared to other decision rules. The main decision rule for each issuer is:

- [redacted]
- [redacted]
- [redacted]
- [redacted]

We set out this thesis with the assumption that issuers use decision rules which describe which payments the issuer systematically refuses. The cohesion among the decision rules for each of the top-4 issuers provides a basis for this assumption. However it also shows that there is likely some noise contained in the results, especially in the decision rules describing small sets of systematic refusals (i.e. rules with low support).

[3] This is a disadvantage of the closed frequent pattern mining approach we use.

[redacted]

Figure 9.3: Top 52 issuers in terms of systematic refusal volume.

[redacted]

(a) Issuer A

[redacted]

(b) Issuer B

[redacted]

(c) Issuer C

[redacted]

(d) Issuer D

Figure 9.4: Breakdown of decision rules per issuer for the top 4 issuers in terms of systematic refusal volume.

9.3.2. Issuing Country-Centric Aggregation

Another perspective on the results is issuing country-centric.[4] By considering the issuing country we gain insight in the regional differences in authorisation decision rules. In other words, we gain insight in the authorisation behaviour between issuers operating in different countries.

Figure 9.5 shows the share of systematic refusals with respect to all payments in an issuing country. [redacted].[5] [redacted].

[redacted].

Figure 9.4 shows a breakdown of the decision rules for the top-4 issuing countries. Each country has a unique mix of decision rules which could stem from many things. For instance different qualities of the payment network in these countries, but also political climate, economy, demographics, and many other variables. The specific merchants in the dataset can similarly affect our observations. Nonetheless we formulate some possible explanations for these countries:

- [redacted].
- [redacted].
- [redacted].
- [redacted].

We set out this thesis with the assumption that issuers mainly decide to systematically refuse payments to ensure their own interest. However the decision rules related to the [redacted] issuers show that the actual situation might be more complex. In these cases we expect the issuers also have to take local economic policies into consideration when authorising payments.

[4] The issuing country is the country in which the cards are issued. In other words, the issuing countries of an issuer are the countries in which an issuer operates.

[5] [redacted].

Figure 9.5: Issuing countries (with at least 10 decision rules) in terms of share of systematic refusals with respect to all payments in that country.

(a) Venezuela

(b) Indonesia

(c) India

(d) Mexico

Figure 9.6: Breakdown of decision rules per country for the top 4 issuing countries in terms of share of systematic refusals with respect to all payments in that country.

9.4. Performance

An important requirement to use association rule mining in practice is that it has a reasonable (computational) performance. It takes about 11 seconds on average to find all relevant association rules and calculate all the patterns evaluation measures per BIN. However because we are performing the analysis in a loop over all 5564 BINs it takes about 18 hours (2,40GHz and 2GB RAM utilised) [redacted]. The implementation would be a lot more usable in practice if the performance would be improved.

It is important to understand where the bottlenecks are in order to improve performance. To facilitate this we logged the processing times of the important steps in the process. We also logged the number of transactions, frequent itemsets and relevant rules found. Table 9.4 and Figure 9.7 show that calculating the *UC* measure takes most of the time (12h), followed by calculating the *improvement* measure (3h) and subsetting the transactions (2h). The processing time for calculating UC and improvement show the highest standard deviation. The calculation of these measures are mainly dependent on the rules found, so for the BIN where the maximum number of rules (528) are found it also takes most time to calculate the improvement and UC measures.

We believe that the reason underlying the big processing times to calculate the UC and improvement measures is that for these measures for each rule all sub- or super-rules need to be found in order to calculate the measure. The current R package *arules* which we use for mining association rules is not optimised for this type of operations.

Table 9.4: Processing times and the transactions, frequent itemsets and rules found per iteration over a BIN. The mean, minimum (min), maximum (max), sum and standard deviation (SD) are shown.

Description	Mean	Min	Max	Sum	SD
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Frequent itemsets	246	6	1.095	1.366.163	202
Relevent rules	108	0	528	600.908	100
Subsetting transactions	1,19s	0s	2,21s	1,83h	0,06s
Coerce transactions to correct format	0,13s	0s	2,51s	0,20h	0,07s
Find all frequent itemsets	0,03s	0s	2,15s	0,04h	0,06s
Generate strong association rules	0,36s	0s	45,58s	0,56h	1,39s
Calculate Kulc	0,13s	0s	1,02s	0,21h	0,12s
Calculate IR	0,12s	0s	0,79s	0,19h	0,11s
Calculate improvement	1,81s	0s	10,81s	2,79h	1,75s
Calculate UC	7,89s	0s	38,05s	12,20h	7,22s

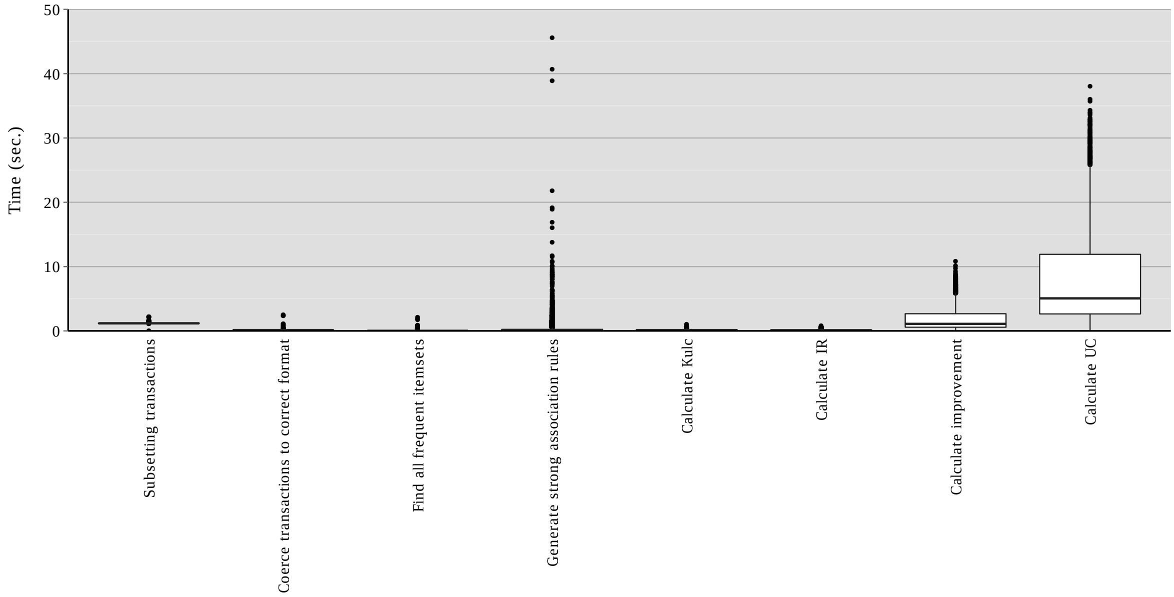

Figure 9.7: Boxplot of processing times per BIN of the most important steps in the process of our approach.

9.5. Summary

In summary, the results show how the distinctive characteristics of systematic refusals can be isolated using the suggested constraints on the pattern evaluation measures. We obtain a mutually exclusive aggregation by considering the top (authorisation) decision rule per BIN. We use meta-rules to group some of the similar rules in the results.

- **Breakdown:**
  - 600.908 found combinations of systematic authorisation or refusal characteristics (overlap)
  - 252.150 (-58%) isolated distinctive combinations of systematic authorisation or refusal characteristics (overlap)
  - 940 authorisation decision rules (no overlap, highest impact distinctive systematic refusal characteristic per BIN)
  - 369 unique rules assigned to 103 meta-rules
  - 3 different overviews created
    - *rule-centric:* [redacted]
    - *issuer-centric:* [redacted]
    - *country-centric:* [redacted]
- **Execution time:**
  - 18 hours in total (2,40GHz and 2GB RAM utilised)
  - 12 hours related to calculating the UC pattern evaluation measure
  - 3 hours related to calculating the improvement pattern evaluation measure
  - R library not optimised for our method

[redacted]

Table 9.5: Total support for different selections of the rules or issuers with the most systematic refusals. First from the perspective of meta-rules pointing at similar underlying decision rules. Second from the perspective of issuers.

[redacted]

The results outlined in this chapter forms the basis for the next chapter. In this chapter we evaluate the *correctness* and the *completeness* of the found decision rules. Additionally we discuss the *practical relevance* of the findings.

10

Evaluation

In the chapter we evaluate the results from previous chapter. On the basis of this evaluation we can determine to which extend our method answers the **Main RQ.**: "How to induce the authorisation decision rules an issuer uses to refuse certain payments from the data available to PSPs?" First we evaluate the *correctness* and the *completeness* of the found decision rules. Additionally we evaluate the *practical relevance* of the findings. Finally we summarise the findings from the evaluation.

10.1. Correctness of Found Decision Rules

In order to evaluate the correctness of the decision rules found by the method we surveyed two experts independently. The complete survey is included in Appendix F. Each expert was presented all unconstrained rules on five arbitrary BINs. We selected only BINs with no more than twenty found rules, to keep the amount of groups comprehensible, and no less then five to give experts an option of choice.

There are 39 rules contained in the survey. The method filters out 16 rules, because these do not meet one of the constraints. There are 13 which do not meet the UC constraint, 7 which do not improve the confidence of a parent rule, and 4 do not meet both constraints.

On the other hand the expert filters out 33 to 34 rules. Figure 10.1 provides an overview of the judgements of the experts. Although the experts confirm the 16 rules filtered out by the UC and improvement constraints (judgement A and B), the experts also indicate 17 to 18 rules are not interesting for other reasons. Another 3 to 4 rules are not interesting because they only marginally improve the confidence of a super-rule (judgement C). The thresholds the experts use is lower than 0,05% or lower than 0,1% depending on the expert. Additionally the experts note that 14 rules are highly similar in terms of the items in the rule, but also in terms of support and confidence (judgement D). Hence the experts note that probably these rules explain datasets which almost completely overlap.

Afterwards the experts were asked why they chose for specific rules. Experts stated: "When a child [rule] does not have a lower authorisation rate [equal to higher confidence] than its parent [rule] it is per definition not interesting", and "Probably the 'real' problem is at the highest level where the childs do not cause the parent to have a low authorisation rate [equal to a high confidence]". Also an expert noted: "Sometimes patterns can be almost completely overlapping, when patterns are very similar this is very likely to be the case. We need this information to judge what is important and what is not. Most probably the most generic pattern without overlap is the most interesting of the ones remaining".

Table 10.1: Results of two expert surveys on 39 decision rules.

Judgement		Rules
A:	Low UC	13
B:	No confidence improvement	7 (4 in A)
C:	Marginal confidence improvement	3 to 4
D:	Highly similar to another decision rule which probably almost completely overlaps	14
	Total:	33 to 34

In conclusion both experts independently confirmed that they were using a similar logic than the constraints on the *improvement* and *UC* measures. Thus these rules can be safely omitted. Additionally experts noted that in the remaining rules there are still a lot of less interesting rules, because they describe a highly similar dataset. In most case the most generic rule of the ones remaining will be the one of main interest.

10.2. Completeness of Found Decision Rules

Another measure on which we evaluate is the completeness of the found decision rules. This quite challenging because this is unexplored territory and there is no 'master-list' of decision rules. However we can rely on the incidents we initially collected to explore systematic issuer refusals. We can check if the method is able to find the decision rules related to the gathered incidents.

We collected data on all relevant incidents during the period between April 28 2014 to November 9 2014. The collection of the incidents was facilitated by a web-based tool which we specially built for this reason (see Appendix C). We use the incidents collected to check if all relevant rules are found by our implementation.

In total 62 incidents were collected on the concerning BINs, and 30 are actual issues which can be related to relevant (and significant) decision rules. The other 32 incidents are not relevant, because the issue was either 1) already solved at the time of analysis (September), or 2) experts didn't confirm the issue, or 3) the issue was not related to a group of refused transactions. In Appendix E contains the details concerning the evaluation on these incidents.

Our implementation is able to find 90% of the relevant issues. The 3 issues which the implementation does not find are on BINs with less than 100 transactions in the month September. For this reason these BINs do not meet the support threshold (i.e. minimum impact threshold) and no rules are searched for on these BINs. It is also important to note that in all cases the relevant rule was also the one which after the constraints were applied had the biggest impact (i.e. support). This provides a basis for the aggregations which we made on the rules with the highest impact only (e.g. in Figure 9.2). Thus we conclude that the implementation is able to find all relevant decision rules with a high enough severity and all this issues also show up in the aggregation.

10.3. Practical Relevance of Findings

[redacted].[1] It is important to know what actually are the biggest issues across the whole spectrum of BIN in order to achieve this. We aim at providing these insights through our aggregations (see Section 9.3. [redacted]. To validate whether the results can be used in practice we further investigated a number of issues with experts.

10.3.1. BIN Intelligence

[redacted]

[1] [redacted]

Table 10.2: Results of PSP experiment. The experiment was conducted during one week from November 6 2014 until November 12 2014.

[redacted]

[redacted]

10.3.2. Merchant Intelligence

[redacted]

[redacted]

[redacted]

10.4. Summary

In summary, this chapter shows the evaluation findings regarding the *correctness* and the *completeness* of the results obtained by the method presented in this thesis. Additionally this chapter shows the evaluation findings regarding the *practical relevance*. The evaluation findings are as follows:

- **Correctness constraints on evaluation measures:** Constraints identical to the logic of 2 experts, no interesting rules wrongfully omitted
- **Completeness of analysis:** 80% of the 32 gathered relevant BIN incidents found (100% of the severe ones)
- **Practical relevance:**
  - [redacted]
  - [redacted]
  - [redacted]
  - [redacted]

In conclusion, the implementation of association rule mining is capable to isolate the primary authorisation decision rules on BINs. The presented UC pattern evaluation measure is fundamental in this. Although the constraint on UC excludes the least percentage of rules, because of its top-down nature it excludes high-impact rules which do not uniquely explain a certain phenomena (in these case systematic refusals). Hence this is fundamental in the isolation and aggregation of the highest-impact rules per BIN. A point of attention concerning the method is its performance, however we argue that when using an implementation similar to Bayardo *et al.* [78] this will significantly improve. Table 10.3 contrasts these conclusions with the argumentation from Chapter 6, which we solely based on theory.

Table 10.3: Rating of the mining frequent patterns, associations and correlations functionality on the relevant criteria.

Basis of Scores	**Criteria**				
	Find	**Isolate**	**Aggregate**	**Interpretable**	**Performance**
Theoretical Argumentation (see Chapter 6)	+ all above certain impact	0 possibly with extra measures	+ using meta-rules	+ concise and useful information	0 possibly by pruning search space during mining
Empirical Evidence (see Chapter 9 & 10)	+ all above certain impact	+ using the UC measure	+ using most impactful meta-rule per BIN	+ mutually exclusive overview of top BIN-rules	- computationally inefficient approach

On the basis of this evaluation we conclude that the method is very effective in *inducing authorisation decision rules*. Hence we answer the main **Main RQ.**: "How to induce the authorisation decision rules an issuer uses to refuse certain payments from the data available to PSPs?" Additionally as the experiment from Table 10.2 shows there is a major potential in improving the overall authorisation rate using the insight delivered by the method.

11

Conclusions and Recommendations

In this chapter we conclude this thesis. First, we conclude on the findings in the context of areas in which we aim at contributing. Second, we discuss the limitations of the findings. Third, we reflect on our work. Finally we recommend a number of directions for future work on the basis of the limitations and the reflection.

11.1. Conclusions

In the following sections we conclude on each of the areas in which we aim to contribute. First, we conclude on the knowledge gap addressed by the main research question. Second, we answer the question from the title by summarising our result on why some payments are (systematically) rejected. Third, we conclude on our additional contribution to association rule mining, the DM technique used to mine the fill this knowledge gap. Fourth, we address the (social) relevance to 1) the payment network and and 2) PSPs. In summary our conclusions are as follows:

- **Main research question:** The method we present is very effective in *inducing decision rules* (and thus finding payment refusal clues).
- **Title question:** The method extracts clues for explaining why specific payments got rejected.
- **Additional contribution:** The method improves on one of the main challenges of association rule mining by filtering out a significant amount of irrelevant rules.
- **Social relevance (1):** The method has the potential to make the payment network economically more efficient by essentially dissolving 20% of the *information asymmetry*, and hereby better aligning the incentives of the parties in the payment network to reduce *moral hazard effects*.
- **Social relevance (2):** The method in combination with strategies to act on specific groups is a significant business case for PSPs, for example a small experiment generates a significant amount of additional revenue for a PSP (in the order hundreds of thousands of Euros per month).

11.1.1. A Method to Induce Authorisation Decision Rules

The main research question is *"How to induce the authorisation decision rules an issuer uses to refuse certain payments from the data available to PSPs?"* In this thesis we present a method making use of association rule mining to effectively achieve this. We support this claim by evaluating on a number of measures.

First, we use expert validation to confirm the *correctness* of the results. On the basis of a survey among experts we conclude that the logic of our method is in accordance to how an expert can pinpoint the group of systematically refused payments which is likely to reflect the relevant a decision rule. Hence we conclude that on the basis of this validation, the decision rules induced reflect reality to a large extend.

Second, we validate the *completeness* of the results. On the basis of information about incidents with refusals we check if our application is able to find all relevant groups of payments. We conclude that our implementation is able to find all relevant groups above a threshold for minimum impact.

In the context of the main research question *"How to induce the authorisation decision rules an issuer uses to refuse certain payments from the data available to PSPs?"*, we can thus conclude that the method we describe in this thesis is effective in inducing the decision rules.

11.1.2. Intelligence Extracted from Payment Data

The answer to the main research question provides significant proof that our method is effective in inducing decision rules. Hence we can further explore the intelligence extracted from the payment data to provide an answer to the title of this thesis *"Why my payment got rejected?"* We broadly answer this question using the intelligence mined by our method from 11,5 million payments processed via one PSP. [redacted]

[redacted]

(a) Authorisation Decision Rules

[redacted]

(b) Issuing Countries

[redacted]

(c) Issuers

Figure 11.1: Different perspectives on the groups of systematic refusals as a consequence of specific authorisation decision rules.

[redacted]

[redacted]

11.1.3. Contribution to Association Rule Mining

In the previous paragraph we conclude that the application we describe is able to deliver the required insights. We achieve this by introducing a new interest measure, we call UC, to the field of association rule mining. Besides this measure being essential in the discovery of the required insights we also improve on one of the main challenges in association rule mining.

In essence UC makes it possible to find the distinctive features to which we can ascribe a certain consequence (in this case systematically refusing payments). It does so by providing a measure which describes to which extend a rule (in this case the set of features leading a systematic refusal) is explained by a more specific rule (equal to the set of features of the generic rule with additional features). In other words, we are able to determine the unique 'explanation power' of a set features on a certain outcome (in this case the refusal of a payment). Hence we are able to pinpoint the distinctive features to explain an observation.

One of the biggest challenges in association rule mining on highly dimensional data is that when applied unconstrained a sheer amount of patterns are found (in our case 601k rules according to the pattern of interest). The information provided by UC allows us to significantly reduce the number of potentially interesting patterns found by association rule mining. Although the newly introduce UC measure filters out the smallest portion of rules (an additional 4% to the 54% filtered out by existing measures), this does not necessarily mean it is the least useful.

In contrast to existing measures UC works top-down. It can filter out uninteresting rules with relative high support on the consequent of interest (in this case systematically refusing payments). In many situations experts are first considering the rules with the highest support, because these rules have the highest impact. Hence filtering out the deceiving high support rules can be of big value in preventing disastrous business decisions.

In essence the interest measure works for DM challenges to find, precisely describe and cluster specific areas of significant over- or under-performance in a large, dense, categorical transactional datasets. This covers a very wide range ranges of challenges in the field of DM. Hence we conclude we provide an extension to association rule mining which makes the technique usable for a much wider range of DM problems. Appendix A contains the article related to this thesis which specifically focuses on this theoretical contribution and explains the method in another context than the payment industry.

11.1.4. Relevance to the Payment Network and PSPs

We set out this thesis with the notion of *information asymmetry* and how this makes the payment network prone to *moral hazard effects*. There is an information asymmetry, because other parties do not have access to the information to judge whether an authorisation decision of an issuer is an acceptable one. We argue, this creates an environment prone to moral hazard effects, because issuers can refuse payments to particularly serve their own interests, which we show are not always aligned with the interests of other parties. We argue this leads to an economically suboptimal system.

[redacted], our method provides information about likely refusals clues. We verify a small number of these clues with an issuer, two merchants and on the basis of clues which became apparent in earlier incidents at a PSP. These clues effectively empower other parties, (especially PSPs) in the payment network to break the issuer's black box system and dissolve the information asymmetry related to refusal clues. This empowerment enables parties to act on an issuer's decision (e.g. dispute or circumvent). Hereby issuers are provided with incentives to make decisions which better reflect the interests of other parties in the payment network. We argue this makes the payment network economically more efficient (i.e. Pareto optimal).

This is especially relevant to PSPs because there is a significant business case as well. Boosting the authorisation rate also boosts the overall payment conversion rate. Hereby the method directly contributes to one of the key value propositions of PSPs. PSPs can use the method as a unique selling point or as an additional service to their customers (i.e. merchants). Additionally PSPs charge merchants with a higher markup for authorisations than refusals. Thus PSPs have direct financial benefit from the boost in conversion rate as well.

[redacted]

[redacted]. Hereby we also meet our objective in *proving insights about groups of systematic refusals to experts from PSPs to act on the merchant's behalf*.

11.2. Limitations

In this section we discuss the limitations of this thesis that might hamper the generalisation of the results. For this we use the terms *validity* and *reliability*. Both validity and reliability are not 'sufficient' or 'insufficient', but moreover are degrees which can always be improved. Hence we will not argue our study is valid or reliable, but discuss the strong and weak point for each in the next paragraphs.

11.2.1. Validity

We use the term *validity* to indicate to which extend our conclusions scientifically answer the research question (i.e. if we measure what we suppose to measure) [89]. Yin [14] defines three types of validity: *construct validity*, *internal validity* and *external validity*. Each is important for our research.

Construct Validity

Construct validity refers to which extend the operationalisations of a construct (i.e. the experiments developed from theory) measure what they intend to [14]. There are few prior research studies related to our problem field. This limits the theory from which to develop the operationalisation. Hence our work is of exploratory nature and does not build upon a well established research typology from the payment domain.

We adopt our own definition of 'authorisation decision rules' which we conceptualise into the association rules remaining after the application of our method. We make and evaluate these conceptual leaps on the basis of expert opinions, which provide us with a level of confidence. However these conceptual leaps are inherently contestable. We rely on multiple strategies from Yin [14] to minimise this.

Yin [14] mentions three strategies: using multiple sources of evidence, establishing a chain of evidence and letting key informants review the report. In accordance with this we interview multiple experts from PSPs as well as issuers. To ensure we do not self-fulfil a prophesy (i.e. let the, we purposefully work with other experts in the evaluation phase compared the exploration phase. Additionally we provide extensive records of the gathered expert data, and interview and survey results (see Appendix B, D, E and F) and experts from the PSP where this research is conducted have extensively reviewed this report.

Internal Validity

Internal validity refers to which extend conclusions about *causal relations* can be made [14]. In this thesis we research the influence of payment characteristics (*independent variables*) on the authorisation decision of the issuer (*dependent variable*). Because the payment characteristics are determined before the payment is send to the issuer, it is evident that the one follows the other. However it is impossible to completely isolate the causal relationship.

Before the payment send by a PSP reaches the issuer, it passes at least the systems of an acquirer and the card network, but in practice it can be many more systems (e.g. from external processors, risk service providers, etc.). All these systems can in potential slightly alter the payment characteristics and already refuse the payment before it reaches the system of the issuer. In the first place, altering the studied payment characteristics not allowed and has major consequences for a parties which do so. We assume this does not happen.

In the second place, experts indicate that refusals by other parties (especially acquirers are relevant in this case) are known to be significantly less frequent. Additionally we specifically look for decision rules which are specific to the identifier pointing to the issuer (BIN) in payments processed via multiple acquirers. We assume that the possible small amount of acquirer refusals are diminished because of this. We do not find proof of these effects during the evaluation, which provides some confidence that this assumption holds. Nonetheless future research might improve our work by filtering out these refusals (e.g. many acquirers indicate this in the response code).

Another point worth mentioning is that we study causal relations without covering *statistical significance*. As Hämäläinen and Nykänen [90] note, traditional association rules do not capture the notion of statical dependence. Main reason is that calculating significance is computationally very complex, because significance is not a monotonic property [91].[1] To date no solution to calculate such measures is widely accepted [38]. Hence we do not take statistical significance into account as this is a problem field on its own. Nonetheless when acceptable solutions arise in future research, our work can improve on this.

In order to measure the causal relations we employ techniques from the field of association rule mining. This a well-established DM research field with many practical applications. We use an implementation of Borgelt [88] in this study, which is widely used [92–96]. Additionally we first test the implementation on a few limited datasets and manually verify if the rules found are indeed correct. This gives us a significant amount of trust in the internal validity in relation to the instrumentation.

External Validity

External validity refers to which extend our conclusions can be held true for other cases (i.e. *generalised*). In this thesis we apply our method on a dataset containing a large sample of of 11,5 million payments provided by a PSP. This dataset contains payments from thousands of merchants and multiple acquirers. We believe this is a very significant dataset in size. Although the results on different records with card payments (e.g. other BINs and or card networks) inherently differ, the data structure does not differ (much). Additionally other PSPs, merchants make use of similar data elements as these are imposed by the various standards in the payment industry. Hence we argue there is little doubt the method is similarly effective on different datasets.

During the evaluation we rely on two experts to judge the quality of the method and check if the method is able to find the relevant incidents (30 in total). The experts judging and the experts administering the incidents all belong to the same company. Additionally major part of the research we perform within the vicinity of this company and in cooperation with the consulted experts. Hence the judgement on the quality of the method can be biased towards the dominant views within the company. We document all arguments on which decisions are made, and many arguments (especially related to the DM design steps) are a product of deductive (mathematical) reasoning. Hence future researchers and practitioners can follow a similar approach on other cases to verify (or falsify) our findings.

11.2.2. Reliability

Reliability refers to the consistency in the measurement results (i.e. if we obtain the same results if we apply the same procedure) [14, 89]. Because we rely on precisely defined programmable scripts for the method itself (see Blocks 8.1, 8.2, 8.3, 8.4, **??**, 8.5 and 8.6) and make use of interview, survey and data gathering protocols (see Appendix B, D, E and F). Other researchers should be able to examine (or, more importantly, repeat) the research process which led to the conclusions for the research questions. This is one of the elements Yin [14] suggests to gain reliability.

Another element is to provide an evidential base on which the findings are based [14]. We especially base our DM findings on many works from other researchers which can be accessed via the bibliography and are referenced throughout the text. The claims we make related to the payment industry specific parts of our work, is substantiated by interviews, surveys and gathered incidents (see Appendix B, D, E and F). Hence we argue our research is repeatable and therefore has a high degree of reliability.

[1] For Boolean function monotonicity means that for every combination of inputs, switching one input from false to true can only make the output switch from false to true and not from true to false [91].

11.3. Reflection

In this section we reflect on the choices made in this thesis. First, we reflect on the choices made regarding the delineation of this thesis. Second, we reflect on the choices made regarding the method. Third, on the choices concerning the software, and fourth on the choices concerning the dataset.

11.3.1. Delineation

We choose to extract knowledge from payment data to solve moral hazard effects, which are mainly caused by an information asymmetry between issuers and other parties in the payment network. We argue for arming the affected parties with a 'friendly weapon', as governance is hampered by the social complexity of the problem. In line with this we focus on the creation of such a 'friendly weapon' instead of exploring governance solutions. This line of argumentation is quite black-and-white. Hence we very likely understate the potential for governance solutions. Governance solutions might contribute to further resolve the moral hazard effects on areas where the method we suggest is not effective.

The method can only resolve the information asymmetry as far as authorisation decisions can be reverse-engineered on the basis of the available data (input, output, and external data). Hence the method can only reverse-engineer a decision if it has been made on the basis of this data. The issuer might have additional data (e.g. the credit history of the cardholder) to base the decision on. This part of the decision still remains a black-box which requires other solutions. Additionally we found that there might be other factors at play not related to moral hazard (e.g. economical policies, outdated account details, etc.). Hence there is a lot more to systematic refusals than moral hazard only.

Another choice we make is to exclusively search for useful techniques in the field of DM. The research field is essentially identical to our problem of 'mining knowledge from data', however there can be other useful solutions from other fields we do not consider. Additionally there might be other solutions from the field of DM which we do not consider.

11.3.2. Method

We select association rule mining after comparing the DM functionalities and related techniques from Han *et al.* [38]. DM is a very broad and active research field and a categorisation at a specific point in time inherently is a simplification of reality. Although the categorisation is quite recently made by well-respected authors in the research field, there can be DM techniques we are unaware of which can improve this work.

Besides association rule mining, decision tree algorithms can also create ('white-box') rule-based models. We ultimately value association rule mining over decision tree algorithms. Mainly because in theory association rule mining guarantees to find all groups of systematically refused transactions, while the heuristic nature of decision tree algorithms can not guarantee this completeness. However association rule mining requires a confidence threshold. Hence we had to choose for a specific percentage when payments are systematically refused. We set this threshold at 70%. This is inherently contestable.

During the design of the method we start considering the individual rules as a hierarchical network of rules to compute measures such as improvement and UC.[2] Decision trees are also hierarchical, but typically depict only features related to one attribute per node. Eventually we frame the problem in a way which has many similarities with a binary classification problem. Because decision tree algorithms aim at solving such problems these algorithms might also find usable rules. Potential advantages are that the rules from this algorithm would be mutually exclusive (i.e. none overlapping). Mutually exclusive results enables users to directly assess the overall impact of the rules (e.g. to tie the overall impact to

[2]Not a tree, because there can be multiple root nodes

the financial records). Additionally decision tree algorithms are computationally more efficient. Hence it makes sense to explore decision tree algorithms further to improve on these advantages.

We compare the results of our method on one arbitrary BIN, with the results of a common decision tree algorithm in Appendix G. We observe that while our method finds rules explaining all the (distinctive) features which describe a significantly large groups of refusals, the decision tree algorithm finds rules containing the least amount of attributes (with an unbounded number of attribute values per rule premise) to classify a group either as (a group of) authorisations or as (a group of) refusals. This is inherently different and less effective for finding groups of systematic refusals (as we also argue in Chapter 6). Although quite recently multivariate decision tree (i.e. multi-category classification) algorithms have been developed. To a large extend the method we propose can also be considered a multi-category classification method. Hence we recommend to contrast this thesis with work from this field. This can potentially lead to new insights.

11.3.3. Software

We use the programming language *R* for the implementation of our method. Especially the packages *arules*, *data.table*, and *ggplot2* provide numerous building blocks to set up our experiment. This saves a lot of time when implementing, however it also has its limitations.

For instance we do not have compete control over how association rules are mined and how the interest measures are calculated. Pushing some constraints deep into the mining can significantly improve the computational efficiency, as for instance shown by Bayardo *et al.* [78] (e.g. antecedent/-consequent constraint, but also the calculation and filter on improvement and UC). Another limitation is that we believe R is tailored at one-off (or n-off) analyses. As a consequence automating the analysis itself and creating an interactive tool for the end-user is something which we find not practical in R. We have preference R when doing incidental analysis or for prototyping, however once there is a large audience of end users and frequent updates (hence performance) is required, we recommend a general-purpose programming language, such as *Python* and *Java*, in combination with *JavaScript* libraries, such as *D3.js*.

11.3.4. Dataset

We choose to use a dataset containing a large sample of payments provided by a PSP. We believe this is a very significant dataset in size. Although the results on different records with card payments (e.g. other BINs and or card networks) inherently differ, the data structure is almost completely similar. Hence there is little doubt the method is effective on different datasets.

Besides different records, the records we also limit the analysis to specific attributes (i.e. dimensions). We base the dimensions on an extensive exploration of prior incidents with issuer refusals and expert interviews. However there might be additional interesting dimensions we currently do not use. For instance we now use the first 6-digits of a card number to distinguish different cards from different issuers, but we know issuers can use more digits to distinguish between cards. However we can not use more than 6-numbers to card network regulation related to the information security.

11.4. Recommendations for Future Research

In this thesis we explore several directions. During this exploration we choose not to explore certain directions which are open for future research. Directions we recommend to further explore are the following:

- **Computational performance**
  - The calculation of the UC interest measure we introduce requires significant computational resources. We suggest a number of directions to potentially improve the performance:
    - Use a more efficient implementation to find sub- and super-rules (e.g. use indexes or a 'divide and conquer' strategy)
    - Find all frequent itemsets on the complete set directly instead of iteratively per BIN
    - Constraint-based mining could be used to use the constraints of the iterative approach in the frequent itemset generation step (because in the non-iterative approach from the previous bullet point the minimum support can not be adjusted per BIN, which leads to more unnecessary calculations)
    - Use a smarter implementation to induce only strong association rules following a certain patterns (instead of all association rules and later filter on the rules with the right pattern)
  - Srikant *et al.* [97] propose different ways of pushing the constraints deep into the mining phase to improve the frequent itemset mining step. The main advantages are faster execution and lower memory utilisation. This direction can potentially improve the performance of the application we describe in this thesis.
- **Closed-pattern mining**
  - We find some seemingly odd decision rules which might be due to the current implementation of association rule mining. We use closed-pattern mining which basically means the algorithm searches for the most precise description of the data. [redacted].
- **Moral hazard in authorising payments**
  - We underpin the moral hazard assumption with the observation that issuers systematically refuse payments which are risky in terms of liability. However to formulate the assumption as an hypothesis and make claims, requires another type of research.
- **Other directions**
  - Our approach to trim uninteresting rules top-down (using the UC measure) and bottom-up (using the improvement measure) can be further refined. For instance by introducing another measure which describes if (and how) rules which are a more specific version of a rule (which meets the UC threshold as well) significantly (based on the improvement in confidence and the decrease in support) improve the more generic version of a rule.
  - When rules are induced with a minimum of 50% confidence all data is covered by association rules (and thus the rules are collectively exhaustive). If we combine this with the exclusion of rules using UC in combination with our suggestion to make all rules mutually exclusive (see "Mutually Exclusive Aggregation" part from Section 6.2.5), we could in theory build a classifier which is able to predict class labels for all data objects. Research on the feasibility

and performance of a classifier based on these ideas can be a valuable contribution to the field of DM.

- When multiple automatic strategies to act on a group of systematic refusals can be effective, Bandit algorithms can potentially be useful to automatically decide which one to use. Additionally Bandit algorithms can change the strategy automatically once an issuer has changed its decision logic.

12

Epilogue

During our work we largely follow the research approach of CRISP-DM. CRISP-DM is a widely used methodology by DM experts across industries [16]. We do not go into enough depth to make hard claims about CRISP-DM. However we believe that some elements in our approach are not well reflected in CRISP-DM. We believe these elements greatly contributed to our results and because of this we like to share our thoughts to possibly inspire future researchers and practitioners.

Figure 12.1 explains the additional elements from our approach with respect to CRISP-DM. The major difference of our approach is that we believe to give more attention to understanding the strengths and weaknesses of the different DM techniques. This understanding allows us to be able to select the most suitable technique to approach the DM problem. Besides, we are able to customise the technique on the basis of this understanding. Hence we believe that a better *technique alignment* can contribute significantly to the solution of any DM problem.

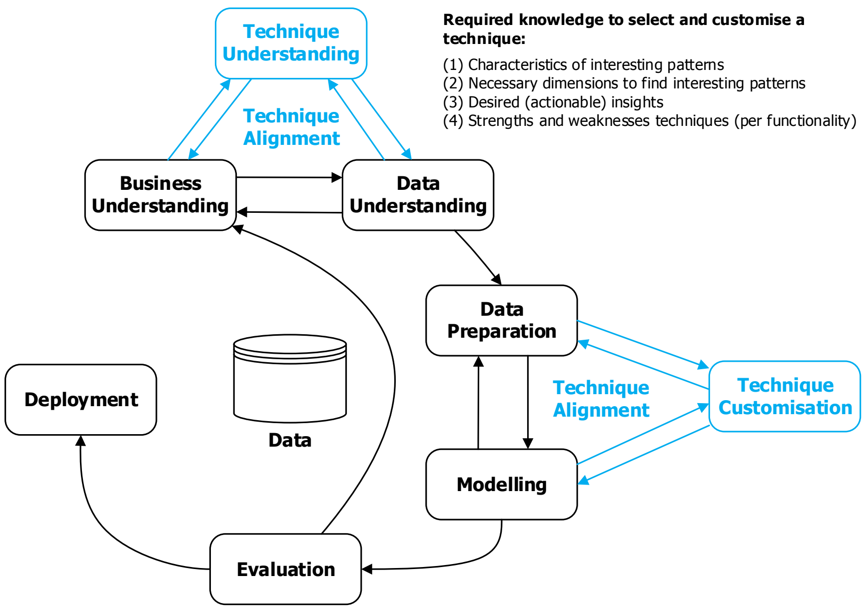

Figure 12.1: The extension on CRISP-DM to better align DM techniques with the business problem and the available data. We mark the extension in blue.

A

Article on the Method to Find the Distinctive Features of Rules

The Sandwich Method: Top-Down and Bottom-Up Pruning of Rules Mined from Large, Dense Databases

R.J.A. (Roy) van der Valk
Delft University of Technology
Jaffalaan 5, Delft, The Netherlands
hello@royvandervalk.nl

ABSTRACT

A very large amount of rules are found when mining for association rules in large, dense databases. This complicates the task of deriving actionable insights from these rules. This article presents a method which combines measures from prior work with a new 'top-down' measure to improve on this. The method is able to find the rule of main interest in a group of rules which party cover the same items. A case study in the payment industry shows that the logic of this method is intuitive and the method can provide experts with actionable insights from large, dense databases. The results signal that there is a large potential which justifies further research in this direction, for instance to find ways of dealing with the remaining rules describing partly overlapping data, to improve on performance, and extend the method with significance measures.

Keywords

Data mining, descriptive analytics, rule induction, association rule mining, rule interestingness

1. INTRODUCTION

In many scenarios important business decisions require knowledge about patterns or 'rules' underlying the data. For instance, when deciding which additional products to offer a customer, it is important to know which combination of items frequently leads to a sale. *Association rule mining* algorithms were initially developed for this type of market basket analysis [14].

Over the years, many areas other than market basket analysis use association rule mining [4]. To illustrate the concept, suppose a researcher has a dataset of resumes and wants to find out which factors companies value most when hiring job applicants. The following rule can be the result:

$$\text{technical study \& extracurricular activities} \Rightarrow \text{hired}$$
$$[\text{support} = 15\%, \text{confidence} = 70\%].$$

Standard association rule mining algorithms find all rules which meet a user-specified threshold for *support* (the proportion of a rule in the dataset) and a threshold for *confidence* (the probability of the rule being true) [2, 14].

Unlike market basket data, other datasets are often very dense [4]. Bayardo et al. [4] consider a dataset dense when it has any or all of the following properties:

- many frequently occurring items (e.g. sex=male);
- strong correlations between several items;
- many items in each record.

Mining for association rules on dense datasets using only support and confidence constraints leads to an extreme number of rules [15]. It is practically infeasible for experts to investigate all the rules and oversee the situation. As a result experts are unable to gain actionable insights from the results, which makes the results practically not very useful [15].

Many *interest measures* (or *pattern evaluation measures*) have been proposed to confront this challenge and determine the *interestingness* of a rule [1, 4, 7, 8, 16, 17]. Most of these measures provide useful information on the correlation between the left-hand-side (*antecedent*) and the right-hand-side (*consequent*) of the rule. Hence these measures only take the presence (or absence) of the items described by the rule into consideration.

This limits the effect of filtering rules on the basis of these measures. In a dense dataset there are strong correlations between several items. This leads to many rules which partially cover the same items and have comparable scores on these measures. Hence it is not possible to filter on the interesting rules using constraints on these measures. To cope with such scenarios this article aims to answer the question:

"How to find the rules of main interest in a group of rules which partly cover the same items?"

This article presents a method which determines the rule's interestingness relative to other rules. The method regards rules which partly cover the same items as an hierarchical network linking more generic and more specific versions of a rule together.[1] By traversing the network top-down the method determines a rule's interestingness relative to more specific rules. By traversing the network bottom-up the method determines a rule's interestingness relative to more generic rules. Basically, the method tries to find the most interesting rule by approaching it from two directions, hence the name 'sandwich method'.

[1] The network is not a tree, because there can be multiple root nodes

The outline of the article is as follows. Section 2 contains the related work on this topic. Section 3 builds on this knowledge to outline the new interest measure which allows for top-down pruning. Section 4 follows up on this section to explain the method in which this measure is used in combination with prior measures to find the rule of main interest in a group of rules which party cover the same items. This Section 5 contains the evaluation of this method. We discuss the work presented in this article and give suggestion for future work in Section 6. Finally, Section 7 contains the conclusions.

2. RELATED WORK

Standard association rule mining consists out of two steps [2]. First all *frequent itemsets* are mined which meet a *minimum support threshold*. Second strong association rules are generated from these frequent itemsets which meet a *minimum confidence threshold*. Typically performance mainly depends on the first step and the second step is much less costly [14].

Given rule A ⇒ B, confidence and support have the following formulas [14]:

$$support(A \Rightarrow B) = P(A \cup B) \tag{1}$$

$$confidence(A \Rightarrow B) = P(B|A) \tag{2}$$

Apriori is the common algorithm to mine for frequent itemsets [14]. Agrawal and Srikant [2] introduced the algorithm in 1994. The algorithm performs a *level-wise*, iterative search, where the k-itemsets are used to explore the $(k+1)$-itemsets. The algorithm is based on the Apriori property, which says that all nonempty subsets of a frequent items must also be frequent [2]. Over the last two decades many alternative algorithms have been proposed mainly aimed to improve the performance of the frequent itemset mining step [5, 13, 21].

Constraints can be applied to distinguish interesting rule from less interesting rules [14]. These constraints can play an important role in improving performance [4], however as Bayardo [3] argues this should not hinder the important role of *knowledge discovery* in the data mining process. Bayardo [3] argues that constraints should be *discovery preserving* and thus only filter out rules which are highly unlikely to ever be of interest to the analyst.

Many constraints have been suggested. For instance by constraining on *closed frequent itemsets* (filters out all subsets with the same support), or *maximal frequent itemsets* (filters out all subsets). Also filters have been suggested on specific *interest measures* (or *pattern evaluation measures*).

Most *interest measures* proposed are association, correlation and similarity measures on the relation between the antecedent (lhs) and the consequent (rhs) [1, 7, 8, 16, 17]. There is a general consensus that such measures should be *null-invariant*, meaning they should not be affected by transactions which do not contain the item(s) under study (*null-transactions*) [14, 19]. For instance χ^2 is not null-invariant [14, 19].

Han et al. [14] recommend to use the *Kulczynski* (*Kulc*) measure in combination with the *Imbalance Ratio* (*IR*). Kulc measures the arithmetic mean probability of one object having a certain attribute and another object having it too [6]. Kulc is null-invariant, because it is unaffected by unbalanced conditional probabilities, in other words a "balanced" skewness of the data [14]. The skewness is expressed numerically by IR [20]. Hence both measures provide a complete picture of the situation [20]. Kulc and IR have the following equations:

$$Kulc(A, B) = \frac{1}{2}(P(A|B) + P(B|A)) \tag{3}$$

$$IR(A, B) = \frac{|sup(A) - sup(B)|}{sup(A) + sup(B) - sup(A \cup B)} \tag{4}$$

These measures enable a user to constrain on rules with a minimum value on these measures. In this way rules with low predictive ability can be filtered out. To illustrate the usefulness of this concept, reconsider the rule from the introduction:

technical study & extracurricular activities ⇒ hired
[support = 15%, confidence = 70%].

This rule says that 70% of the applicants with a technical study and extracurricular activities got hired. Now suppose that in general 75% of the applicants got hired, than this rule offers no predictive advantage over the average. Kulc and IR provide insight in this predictive advantage. Bayardo et al. [4] note these measures still bear a closely related problem. For instance, consider the following controversial rule:

work experience & male ⇒ hired
[support = 16%, confidence = 79%].

Because 79% is above average this rule could be interpreted as the probability to be hired is increased because of the fact that the applicant is male (with accusations at the company's address as a consequence). However, consider the following rule is found as well:

work experience ⇒ hired
[support = 20%, confidence = 80%].

This shows that actually work experience is the distinguishing factor. Being male (instead of a female) with work experience even slightly reduces the probability of being hired. The *improvement* measure provides insight in this [4]. It shows how much a sub-rule improves the confidence of its super-rule. For instance for the second to last rule, the improvement would be -1%. Given rule $A \Rightarrow B$ with super-rule $A' \Rightarrow B$ the improvement measure has the following equation:

$$\begin{aligned} imp(A \Rightarrow B) = min(\forall A' \subset A, &conf(A \Rightarrow B) \\ &- conf(A' \Rightarrow B)) \end{aligned} \tag{5}$$

Bayardo et al. [4] argue that a negative improvement is typically undesirable in almost any data mining application, because the rule can be simplified by a more generic rule which applies to a larger population and is more predictive.

A larger minimum on improvement can also be justified on dense datasets, because due to the density slight improvements in confidence are often due to 'noise' in the dataset [4].

We extend the work from Bayardo et al. [4] and borrow the notion of comparing rules which are relatively super- and sub-rules to each other. However as we argue in the next section there another problem is yet unsolved. This applies to super-rules, which derive almost all there predictive ability from sub-rules.

3. UNIQUE CONFIDENCE

In order to illustrate the problem that is yet unsolved, reconsider the last rule in combination with another rule:

$$\text{work experience} \Rightarrow \text{hired}$$
$$[\text{support} = 20\%, \text{confidence} = 80\%].$$

$$\text{work experience \& referral} \Rightarrow \text{hired}$$
$$[\text{support} = 17\%, \text{confidence} = 85\%].$$

At first notice, both rules look interesting. Both rules offer a predictive advantage over the average hiring rate and the sub-rule improves the super-rule in terms of confidence. However, because the support of the sub-rule covers a large part of the support of the super-rule, it could well be that the super-rule derives all its predictive advantage from this sub-rule. In this case it could thus well be that work experience alone is not a distinguishing factor to be hired, but it only if the job applicant has work experience together with a referral, the job applicant has an advantage.

Table 1 illustrates this for a dataset containing 100 job applicants. A 20% support means that 20 applicants with work experience are hired. From dividing the rule's support by the rule's confidence of 80% derives that in total 25 applicants have work experience (called the *coverage* of a rule). Similarly we calculate the total applicants with work experience and a referral.

Table 1: Example to illustrate that without the 'work experience & referral' rule, the 'work experience' rule loses its confidence.

Rule	Total	Hired	Conf.
work experience	25	20	80%
work experience & referral	20	17	85%
work experience & no referral	5	3	60%

By subtracting this rule from the more generic rule regarding work experience, we observe that only 60% of the applicants are hired with work experience, but without a referral. This means a job applicant with work experience, but without a referral has a lower probability to be hired. It can be that also only 60% of the applicants are hired with a referral but no work experience. This means that only applicants with work experience in combination with a referral have an advantage.

So a job applicant having work experience or a referral alone does not have an advantage, only the combination gives an advantage. Hence if the question would be *"What are the distinguishing factors to be hired?"*, it makes sense to filter out the rule containing work experience or a referral alone. These rules give an deceiving image of the situation.

In order to generalise this, we introduce Unique Confidence (UC). UC describes the confidence of the part of the rule which is not covered by its sub-rules. From the example follows the equation, which explains UC for a single (hence sUC) rule $A' \Rightarrow B$ with sub-rule $A \Rightarrow B$:

$$sUC(A' \Rightarrow B) = \frac{sup(A' \Rightarrow B) - sup(A \Rightarrow B)}{\frac{sup(A' \Rightarrow B)}{conf(A' \Rightarrow B)} - \frac{sup(A \Rightarrow B)}{conf(A \Rightarrow B)}} \quad (6)$$

A rule can have multiple sub-rules, thus it is important to check the sUC for each sub-rule. For a specific rule the minimum UC based on all its sub-rules is the confidence which the rule keeps despite its sub-rules. Additionally the elements in the denominator are known as coverage [18], or simply the support of the antecedent (i.e. left hand side). Hence we alter the sUC equation as follows to an equation for UC:

$$UC(A' \Rightarrow B) = min(\forall A \subset A', \frac{sup(A' \Rightarrow B) - sup(A \Rightarrow B)}{sup(A') - sup(A)}) \quad (7)$$

We argue that rules with an UC lower than the minimum confidence are typically undesirable in almost any data mining application, because a more specific rule is present which causes the super-rule to lose its confidence (i.e. the rule would not meet the minimum confidence threshold if the data covered by the sub-rule is not taken into account). Especially in dense datasets, situations such as in the two example rules are likely to occur.

This measure exposes a scenario to which the improvement measure is prone. Bayardo et al. [4] note that an improvement greater than zero can also be justified on dense datasets. However as shown by the example, it can be the case that if rules are close in terms of support that a small improvement makes the super-rule uninteresting instead of the sub-rule. Hence filtering out the sub-rules on the basis of a small improvement only is ill advised, because it can lead to wrong conclusions about factors which are likely to be distinguishing in a certain outcome. However, once rules are filtered out based on UC, it makes sense to prune the rules which only marginally improve on a super-rule. This is the basis of the method we present in this article.

4. THE SANDWICH METHOD

Conceptually the method we propose is relatively simple. Figure 1 shows its workings on the illustrative rules from the previous sections combined with a 'work experience & referral & female ⇒ hired' rule with a slight improvement over its super-rule. Assume the figure contains only rules found with a confidence higher than 75%, thus the rule which offer a predictive advantage (similar to IR > 0,5) over the average hiring rate (75%). First we filter out the rule with a negative improvement. Second we filter out the rules with a UC which is lower than 75% (hence the improvement of the responsible sub-rule is not applicable any more). Last the rules we filter out the sub-rules with only a marginal improvement.

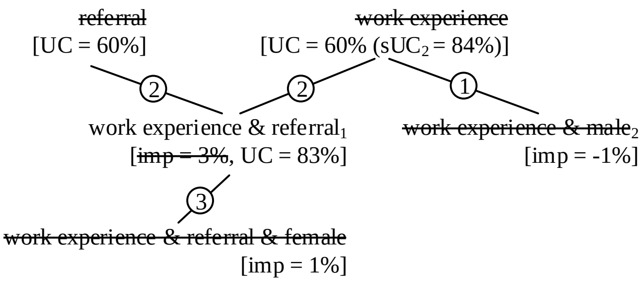

Figure 1: Example of rules with consequent '⇒ hired' filtered out using the sandwich method.

Another reason why rules with a negative improvement are uninteresting is because such a rule per definitions leads to a super-rule's sUC which is higher than the super-rule's confidence (also illustrated by sUC_2 of 83% which is higher than the rule confidence of 81%). Hence such a rule needs not be taken into account when calculating UC and constraint-based mining (as for instance suggested by Bayardo et al. [4]) can be safely applied to only find rules with a positive improvement as well as a IR $> 0{,}5$. Sometimes the user has a specific question (in the example: *"What are the distinguishing factors to be hired?"*), in which only rules with a certain consequent are interesting (in the example: '⇒ hired'). This constraint can also be applied during constraint-based mining. By enforcing constraints during the mining phase, the performance is significantly better than when the filtering is applied after the mining phase [4].

For efficient implementation it is possible to calculate UC during mining at the same time the improvement measure is calculated. For each sub-rule, when the improvement in confidence for a specific super-rule is determined, we can calculate the sUC for the super-rule based on the specific sub-rule. Then we can store the sUC value in the super-rule and update it once a lower sUC is found for another rule. Thus, once the mining is completed the value represents the UC based on all sub-rules and filter out the rules which have a UC which is lower than the minimum confidence threshold. This keeps the extra computation at a minimum. After the mining phase the user can filter out rules with only marginal improvement (on the remaining rules).

5. EVALUATION

This section provides the evaluation of the method at a large company active in payment processing. The method is applied on a provided pre-processed dataset containing 11.469.725 card payment transactions to find out which payments fail significantly more often than others. We cannot disclose the exact details of the dataset, besides that a transactions contains 12 items with information which relate to the type of payment (e.g. card information, sales channel, card security code, amount, currency, etc.) and the parties involved (e.g. shopper, merchant, banks, etc.). Many items are frequently occurring (e.g. card type=Maestro, sales channel=online, etc.) and several items have strong correlations (e.g. merchant's mainly process in a certain currency and via certain banks, etc.).

We focus on evaluating the concept and hence we do not aim for an efficient implementation in terms of performance (see the last paragraph of Section 4 for a conceptual approach for this). We use the *arules* package in R and a minimum confidence of 70% and a minimum support of 5% to induct rules from closed frequent itemsets [11]. After mining we constrain on rules to only rules with the consequent '⇒ succeeded/failed'. This leads to 600.908 rules in total.

For these rules we calculate the interest measures. The package includes an implementation for the improvement measure [12]. We provide our own implementation for the Kulc, IR and UC measures similarly to Hahsler et al. [10, p. 28-30].

Table 2 shows the constraints on the Kulc, improvement and UC measures. About 16% of the rules offer no predictive advantage over the average (Kulc $> 0{,}5$). Additionally another 38% of the remaining rules are uninteresting, because the rules offer no improvement. Finally UC is able to mark another 4% of the rules as uninteresting because these rules are completely dependent on a specific sub-rule for their confidence. In total 58% of the rules is filtered out because of the constraints.

Table 2: Statistics about exclusion of constraints.

Constraint	Individual Exclusion (%)	Cumulative Exclusion (%)
Kulc $> 0{,}5$	16,0	16,0
improvement $> 0\%$	48,8	53,9 (+37,9)
UC $> 70\%$	14,5	58,0 (+4,1)

To validate if the constraints are indeed discovery preserving, we independently survey two data analysts from the payment company. Discovery preserving means only rules are filtered out which are unlikely to ever be of interest to the analyst. Each analyst we show 39 rules in a network similar to Figure 1. After the explanation of the interest measures (support, confidence, Kulc, IR, improvement and UC) is given, the analysts marks the rules which are unlikely to be ever of interest.

One the one hand, the method marks 16 rules as uninteresting, because these do not meet one of the constraints. There are 13 which do not meet the UC constraint, 7 which do not improve the confidence of a parent rule, and 4 do not meet both constraints. On the other hand, the analysts mark 33 rules as uninteresting.

Figure 3 provides an overview of the judgements of the analysts. Although the experts confirm the 16 rules filtered out by the UC and improvement constraints (judgement A and B), the analysts also indicate 17 to 18 rules are not interesting for other reasons. Another 3 to 4 rules are not interesting because they only marginally improve the confidence of a super-rule (judgement C). The thresholds the analysts use is lower than 0,05% or lower than 0,1% depending on the analyst. Additionally the analysts note that 14 rules are highly similar in terms of the items in the rule, but also in terms of support and confidence (judgement D). Hence the analysts note that probably these patterns explain datasets which almost completely overlap.

Table 3: Results of 2 expert surveys on 39 rules.

	Judgement	Rules
A:	Low UC	13
B:	No confidence improvement	7 (4 in A)
C:	Marginal confidence improvement	3 to 4
D:	Highly similar to another rule which probably almost completely overlaps	14
	Total:	33 to 34

On the basis for the expert surveys, we conclude that the method is able to filter out a significant number of additional uninteresting rules, which cannot be identified using existing interest measures. The method is discovery preserving, because only rules are filtered out which are unlikely to ever be of interest to the analyst. The rules which only marginally improve the confidence of a super-rule can also be filtered out using the method. However the answer to the question "What is marginal?", is dependent on the analyst.

Although the newly introduce UC measure filters out the smallest portion of rules, this does not necessarily mean it is the least useful. Because the method works top-down it filters out rules with relative high support on the consequent of interest (in this case '⇒ succeeded/failed'). In many situations experts are first considering the rules with the highest support, because these rules have the highest impact. Hence filtering out the deceiving high support rules can be of big value in preventing disastrous business decisions.

One challenge remains (see Judgement D in Table 3). This challenge relates to dealing with highly similar rules which likely explain largely overlapping data. It appears the correlations between some items is of such strength, that when small variations occur the results contain highly similar rules. These rules are not always a true sub- or super-rule of another rule. For instance given rules 'merchant=A & currency=Euro ⇒ failed' and 'merchant=A & channel=online ⇒ failed', it can be that both rules actually describe a similar scenario. This is the case when all Merchant A's transactions are in Euro (hence the Merchant A rule is omitted, because it is not a closed) and all but one of Merchant A's transactions are online (hence not omitted via closed frequent itemsets). Because the Merchant A rule is omitted because of closed frequent itemsets this nuance can not be observed. In the sample rules, rules prone to this challenge mainly reside in the far edges of the rules found (with low support/impact).

Despite the remaining challenge the method proves to be quite useful in practice by focussing on the high impact rules. Using this method the payment company is able to find all major groups of failed payments which are administered in their ticketing system. Additionally the company is able to effectively target the newly identified high impact groups of failed payments.

6. DISCUSSION AND FUTURE WORK

In essence the work we present has many similarities with classification problems. Basically we aim to classify which transactions lead to a certain consequence. We aim to use the classification model directly for end-user understanding of the data (i.e. descriptive analytics). Hence a white-box model is important. For instance decision tree algorithms can create rule-based models as well, however we purposely avoid heuristic methods as these inherently trade optimality, completeness, accuracy or precision for speed [3].

In certain scenarios speed has a relative low importance. For instance when a business intelligence application delivers weekly insights for a management meeting (i.e. descriptive analytics). For such problems non-heuristic methods can have preference. Association rule mining algorithms, such as Apriori, Eclat and FP-growth, guarantee completeness and can effectively deal with certain levels of computational complexity. As noted by Bayardo [3] these algorithms are now commonly used for classification purposes, which can deliver more accurate classifiers [9].

One of the major challenges in association rule mining is to present this 'completeness' in a way which is comprehensible for a user. In most cases (especially on dense datasets) many rules are found, which makes it hard for the user to get a clear and complete picture of the situation. Many works in the field of association rule mining aim at helping the user to distinguish the interesting rules and filter out the uninteresting ones. Our work similarly contributes to this challenge.

After the design of our method we find Liu et al. [15] to similarly prune uninteresting rules by comparing them to their sub-rules. Liu et al. [15] frame the problem quite differently. Liu et al. [15] focus on identifying non-actionable rules via a constraint using χ^2. Hence the top-down approach on pruning is not completely unique.

Our method has a major difference in that it uses a new interest measure by reusing the same confidence threshold used during the mining and entirely work using the existing measures of support and confidence. We argue this has several advantages. First we link the constraints from mining to the pruning phase, and we explain why UC should have the same threshold as confidence. We argue this makes the logic of the pruning very intuitive. Second by entirely relying on existing measures the added computational complexity of our method is minor, especially compared to the additional counting required by the method of Liu et al. [15]. Third χ^2 is not null-invariant, while confidence is. Fourth a threshold on χ^2 can be hard to determine.

In other words the question related to a χ^2 threshold is: "How much higher should the performance (e.g. hiring rate) of a rule be to be interesting, when compared to the average of its antecedent (i.e. lhs)?", is often hard question for experts to answer. We opt for an approach where the expert answers the question: "How low or high should the performance (e.g. hiring rate) be, to be interesting?". During the evaluation we find early proof that this method is highly intuitive and in alignment with the logic of experts, hence it can be easily explained to business stakeholders, which eases acceptance.

We agree with Liu et al. [15] there should be at least a predictive advantage compared to the average and we constrain on this, but stricter constraints we argue are ill advised because there is no guarantee that interesting rules are accidentally omitted. Similarly to Bayardo [3] we opt for constraints which are discovery preserving. Hence we choose to show the information regarding the predictive advantage (via Kulc and IR) to the user and allow the user to decide when a rule is interesting and when it is not, instead of forcing the user to make the decision upfront, without prior information.

We suggest a number of directions for future research. More research attention may be given to researching the quality of this method in a broader set of application domains. For instance also how this method might contribute to improving classifiers based on association rule mining. The measure we introduce determines the most generic rules which are not dependent on their sub-rules. In essence, we offer a way of controlling the generalisability of a model. This can potentially contribute to determining the desired trade-off between the bias and variance of a model.

Specifically the method can be improved on the ability to effectively deal with rules which explain overlapping data. For instance dropping the closure constraint for frequent itemsets leads to a completer network of rules, which in potential offers better pruning possibilities (e.g. to filter out highly similar rules based on overlapping data, see Judgement D in Table 3). However dropping this constraint might introduce additional complexities.

The sUC measure offers another pruning possibility. The sUC measure basically determines the negated counterpart of a rule. For instance given the rule 'work experience $\Rightarrow$ hired', it determines for the rule 'work experience & referral $\Rightarrow$ hired' the measures related to the rule 'work experience & no referral $\Rightarrow$ hired'. Instead of using improvement to determine if a sub-rule offers enough predictive advantage over its super-rule, sUC can determine if a rule offers a significant predictive advantage over its negated counterpart. On the contrary the improvement measure does not discriminate between two sub-rules with similar confidence, but different support. We argue it typically preferable to take support into account, because this provides a completer picture of the relationship between a rule and its sub-rules.

Besides pruning possibilities, research in this direction might also focus on the creation of mutually exclusive (i.e. none overlapping) results. Mutually exclusive results enables users to directly assess the overall impact of the rules (e.g. to tie the overall impact to the financial records). This, together with a tree being collectively exhaustive, are important advantages of decision tree algorithms. Hence association rule mining used for descriptive analytics can improve on this.

Another direction for improvement is to include measures of statistical significance to the existing interest measures. Significance measures can determine if an observation is statistically significant. For instance when determining the strength of the correlation or in determining the strength relationship between rules. Lastly, we do not pay much attention to performance in our implementation. Hence there is room for major improvement here.

7. CONCLUSIONS

In this article, we study the problem of mining association rules from large, dense databases. More specifically to mine rules directly for end-user understanding of the data (i.e. descriptive analytics). A major problem is that association rule mining finds a very large amount of rules of which many cover the same items in the data. To solve this problem we aim to answer the question:

"How to find the rules of main interest in a group of rules which partly cover the same items?"

We define a new measure to determine the relative interestingness of a rule with respect to other rules containing all but some items of the rule (sub-rules). Using this measure we prune rules which derive all their predictive advantage from sub-rules. This top-down pruning method, in combination with an existing bottom-up pruning method, forms the basis of the method we call the 'sandwich method'. After the pruning the rules of main interest remain.

The method adds only a small amount of computational complexity and thus allows for efficient implementation. We evaluate the method on a real-world dataset containing payments and gather expert opinions to determine the quality of the pruning. Experts note that all pruned rules are in fact uninteresting. No potentially interesting information is lost and the results show to be quite useful in practice by providing experts with actionable insights. We argue this justifies future research, for instance to find ways of dealing with the remaining rules describing partly overlapping data, to improve on performance, and extend the method with significance measures.

References

[1] C. C. Aggarwal and P. S. Yu. A new framework for itemset generation. In *Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*, pages 18–24. ACM, 1998.

[2] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In J. B. Bocca, M. Jarke, and C. Zaniolo, editors, *International Conference on Very Large Data Bases*, number 20, pages 487–499, Santiago de Chile, Chile, 1994. Morgan Kaufmann. ISBN 1-55860-153-8.

[3] R. J. Bayardo. The Hows, Whys, and Whens of Constraints in Itemset and Rule Discovery. In J.-F. Boulicaut, L. De Raedt, and H. Mannila, editors, *Constraint-Based Mining and Inductive Databases*, volume 3848 of *Lecture Notes in Computer Science*, pages 1–13. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. ISBN 978-3-540-31331-1. doi: 10.1007/11615576.

[4] R. J. Bayardo, R. Agrawal, and D. Gunopulos. Constraint-based rule mining in large, dense databases. *Data Mining and Knowledge Discovery*, 4(2-3):217–240, 2000.

[5] C. Borgelt. Frequent item set mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6):437–456, Nov. 2012. ISSN 19424787. doi: 10.1002/widm.1074.

[6] J. Bradshaw. YAMS-Yet another measure of similarity. *Euromug01, 13th–14th September*, 2001.

[7] S. Brin, R. Motwani, and C. Silverstein. Beyond market baskets: Generalizing association rules to correlations. In *Proceedings 1997 ACM SIGMOD International Conference Management of Data*, volume 26, pages 265–276, Tucson, AZ, 1997. ACM.

[8] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur. Dynamic itemset counting and implication rules for market basket data. In *ACM SIGMOD Record*, volume 26, pages 255–264. ACM, 1997.

[9] G. Chen, H. Liu, L. Yu, Q. Wei, and X. Zhang. A new approach to classification based on association rule mining. *Decision Support Systems*, 42(2):674–689, Nov. 2006. ISSN 01679236. doi: 10.1016/j.dss.2005.03.005.

[10] M. Hahsler, B. Grün, K. Hornik, and C. Buchta. Introduction to arules - A computational environment for mining association rules and frequent item sets. *The Comprehensive R Archive Network*, 2009.

[11] M. Hahsler, C. Buchta, B. Gruen, K. Hornik, and C. Borgelt. arules: Mining Association Rules and Frequent Itemsets, 2014.

[12] M. Hahsler, C. Buchta, B. Gruen, K. Hornik, and C. Borgelt. Calculating various additional interest measures, 2014.

[13] J. Han, J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data - SIGMOD '00*, volume 29, pages 1–12, New York, New York, USA, May 2000. ACM Press. ISBN 1581132174. doi: 10.1145/342009.335372.

[14] J. Han, M. Kamber, and J. Pei. *Data mining: concepts and techniques*. Morgan Kaufmann, 3rd edition, 2012. ISBN 978-0-12-381479-1.

[15] B. Liu, W. Hsu, and Y. Ma. Identifying non-actionable association rules. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 329–334. ACM, 2001.

[16] E. R. Omiecinski. Alternative interest measures for mining associations in databases. *Knowledge and Data Engineering, IEEE Transactions on*, 15(1):57–69, 2003.

[17] G. Piatetsky-Shapiro and W. Frawley. *Discovery, analysis, and presentation of strong rules*. 1991.

[18] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

[19] T. Wu, Y. Chen, and J. Han. Association Mining in Large Databases: A Re-examination of Its Measures. In J. N. Kok, J. Koronacki, R. Lopez de Mantaras, S. Matwin, D. Mladenič, and A. Skowron, editors, *Knowledge Discovery in Databases: PKDD 2007*, volume 4702 of *Lecture Notes in Computer Science*, pages 621–628. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. ISBN 978-3-540-74975-2. doi: 10.1007/978-3-540-74976-9.

[20] T. Wu, Y. Chen, and J. Han. Re-examination of interestingness measures in pattern mining: a unified framework. *Data Mining and Knowledge Discovery*, 21(3):371–397, Jan. 2010. ISSN 1384-5810. doi: 10.1007/s10618-009-0161-2.

[21] M. Zaki. Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering*, 12(3):372–390, 2000. ISSN 10414347. doi: 10.1109/69.846291.

B

Interview with an Issuer about Refusal Reasons

Confidential

C

Tool to Track Incidents with Issuer Refusals

Confidential

D

Labelling Incidents with Issuer Refusals

Confidential

E

Evaluation using Incidents with Issuer Refusals

F

Evaluation Survey

Confidential

G

Decision Tree for one BIN

Confidential

Bibliography

[1] D. Hock, *One From Many: Visa and the Rise of Chaordic Organization* (Berrett-Koehler Publishers, San Francisco, 2005).

[2] Adyen, *Corporate Presentation*, Tech. Rep. (Amsterdam, 2014).

[3] D. S. Evans and R. Schmalensee, *Innovation and evolution of the payments industry,* in *Moving Money: The Future of Consumer Payments*, edited by R. E. Litan and M. N. Baily (Brookings, Washington DC, 2009) Chap. 3, pp. 36–76.

[4] J. H. Davis, *Do You Have Another Credit Card, Mr. President?* (2014).

[5] Adyen, *International Payment Service Provider,* (2014).

[6] R. P. DeGennaro, *Credit Card Processing: A Look Inside the Black Box,* Economic Review **91**, 27 (2006).

[7] R. Anderson and T. Moore, *The Economics of Information Security,* Science **314**, 610 (2006).

[8] J. Klick and G. Mitchell, *Government Regulation of Irrationality: Moral and Cognitive Hazards,* Minnesota Law Review **90**, 1620 (2006).

[9] B. Hölmstrom, *Moral Hazard and Observability,* The Bell Journal of Economics **10**, 74 (1979).

[10] R. J. Arnott and J. E. Stiglitz, *The Basic Analytics of Moral Hazard,* (1988).

[11] K. M. Eisenhardt, *Agency Theory: An Assessment and Review,* Academy of Management Review **14**, 57 (1989).

[12] S. Okasha, *Philosophy of Science: A Very Short Introduction*, Vol. 67 (Oxford University Press, New York, USA, 2002).

[13] R. Wieringa and A. Morali, *Technical Action Research as a Ralidation Method in Information Systems Design Science,* in *Design Science Research in Information Systems. Advances in Theory and Practice* (Springer, 2012) pp. 220–238.

[14] R. K. Yin, *Case Study Research: Design And Methods,* (1989).

[15] M. H. H. Oei, *Acceptance of Operational Business Intelligence in Organisations: Developing a Framework Describing the Context and Powers at Play Involved in Achieving OpBI Acceptance in Organisations,* (2014).

[16] A. I. R. a. L. Azevedo, *KDD, SEMMA and CRISP-DM: a parallel overview,* (2008).

[17] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, *CRISP-DM 1.0 Step-by-step data mining guide,* (2000).

[18] J. de Lange, A. Longoni, and A. Screpnic, *Moving Beyond the Web*, Tech. Rep. (Innopay, Amsterdam, 2012).

[19] R. Belliotti, *Helping Airports Understand the Payment Card Industry Data Security Standard* (Transportation Research Board, 2010).

[20] A. Kjos, *The Merchant-Acquiring Side of the Payment Card Industry: Structure, Operations, and Challenges,* (2007).

[21] M. Bourreau and M. Verdier, *Private cards and the bypass of payment systems by merchants,* Journal of Banking & Finance **34**, 1798 (2010).

[22] S. Chakravorti, *Externalities in Payment Card Networks: Theory and Evidence : Review of Network Economics,* Review of Network Economics **9** (2010).

[23] D. B. Humphrey, *Retail payments: New contributions, empirical results, and unanswered questions,* Journal of Banking & Finance **34**, 1729 (2010).

[24] Z. Wang, *Market structure and payment card pricing: What drives the interchange?* International Journal of Industrial Organization **28**, 86 (2010).

[25] B. Alexandrova-Kabadjova, E. Tsang, and A. Krause, *Market structure and information in payment card markets,* International Journal of Automation and Computing **8**, 364 (2011).

[26] M. Verdier, *Interchange Fees in Payment Card Systems: A Survey of the Literature,* Journal of Economic Surveys **25**, 273 (2011).

[27] R. Kemp, *Mobile payments: Current and emerging regulatory and contracting issues,* Computer Law & Security Review **29**, 175 (2013).

[28] L. Moussawi-Haidar, W. Dbouk, M. Y. Jaber, and I. H. Osman, *Coordinating a three-level supply chain with delay in payments and a discounted interest rate,* Computers and Industrial Engineering **69**, 29 (2014).

[29] F. J. Liébana-Cabanillas, J. Sánchez-Fernández, and F. Muñoz Leiva, *Role of gender on acceptance of mobile payment,* Industrial Management and Data Systems **114**, 220 (2014).

[30] J. Zhong, A. Dhir, M. Nieminen, M. Hämäläinen, and J. Laine, *Exploring Consumer Adoption of Mobile Payments in China,* in *Proceedings of the 17th International Academic MindTrek Conference: "Making Sense of Converging Media", MindTrek 2013* (2013) pp. 318–325.

[31] K. Ravikumar and A. Udhayakumar, *Secure multiparty electronic payments using ECC algorithm: A comparative study,* in *Proceedings - 2014 World Congress on Computing and Communication Technologies, WCCCT 2014* (2014) pp. 132–136.

[32] S. S. Ahamad, S. K. Udgata, and M. Nair, *A Secure Lightweight and Scalable Mobile Payment Framework,* (2014).

[33] K. K. R. Choo, *New payment methods: A review of 2010-2012 FATF mutual evaluation reports,* Computers and Security **36**, 12 (2013).

[34] E. Kim, Y. S. Lee, S. Y. Lee, J. W. Choi, and M. S. Jung, *A study on the information protection modules for secure mobile payments,* in *2013 International Conference on IT Convergence and Security, ICITCS 2013* (2013).

[35] S. Gold, *The evolution of payment card fraud,* Computer Fraud and Security **2014**, 12 (2014).

[36] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, *Data mining for credit card fraud: A comparative study,* Decision Support Systems **50**, 602 (2011).

[37] C. Phua, V. Lee, K. Smith, and R. Gayler, *A comprehensive survey of data mining-based fraud detection research,* arXiv preprint arXiv:1009.6119 (2010).

[38] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*, 3rd ed. (Morgan Kaufmann, 2012).

[39] G. Piatetsky-Shapiro and W. Frawley, *Knowledge discovery in databases* (1991) pp. 229–248.

[40] S.-H. Liao, P.-H. Chu, and P.-Y. Hsiao, *Data mining techniques and applications–A decade review from 2000 to 2011,* Expert Systems with Applications **39**, 11303 (2012).

[41] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, second edi ed. (Morgan Kaufmann Publishers, San Francisco, USA, 2005).

[42] M. J. A. Berry and G. S. Linoff, *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, second edi ed. (Wiley Publishing, 2004).

[43] R. Agrawal and R. Srikant, *Fast algorithms for mining association rules,* in *International Conference on Very Large Data Bases*, 20, edited by J. B. Bocca, M. Jarke, and C. Zaniolo (Morgan Kaufmann, Santiago de Chile, Chile, 1994) pp. 487–499.

[44] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, *Discovering frequent closed itemsets for association rules,* in *Proceedings of the 6th International Conference on Database Theory (ICDT '97)* (Springer, 1999) pp. 398–416.

[45] M. J. Zaki, S. Parthasarathy, M. Ogihara, and W. Li, *New Algorithms for Fast Discovery of Association Rules,* in *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97)*, Vol. 97 (1997) pp. 283–286.

[46] L. Bauer, S. Garriss, and M. K. Reiter, *Detecting and resolving policy misconfigurations in access-control systems,* ACM Transactions on Information and System Security (TISSEC) **14**, 2 (2011).

[47] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, second edi ed., Vol. 2 (Springer, 2009).

[48] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning,* (2012).

[49] V. Chandola, A. Banerjee, and V. Kumar, *Anomaly detection: A survey,* ACM Computing Surveys (CSUR) **41**, 15 (2009).

[50] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, *Anomaly-based network intrusion detection: Techniques, systems and challenges,* computers & security **28**, 18 (2009).

[51] C. Spence, L. Parra, and P. Sajda, *Detection, synthesis and compression in mammographic image analysis with a hierarchical image probability model,* in *Mathematical Methods in Biomedical Image Analysis, 2001. MMBIA 2001. IEEE Workshop on* (IEEE, 2001) pp. 3–10.

[52] R. Fujimaki, T. Yairi, and K. Machida, *An approach to spacecraft anomaly detection problem using kernel feature space,* in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (ACM, 2005) pp. 401–410.

[53] V. J. Hodge and J. Austin, *A survey of outlier detection methodologies,* Artificial Intelligence Review **22**, 85 (2004).

[54] H. D. Kuna, R. García-Martinez, and F. R. Villatoro, *Outlier detection in audit logs for application systems,* Information Systems **44**, 22 (2014).

[55] S. Pattanaik and P. P. Ghosh, *Role of Data Mining in E-Payment systems,* (2010), arXiv:1003.1816 .

[56] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, *The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature,* Decision Support Systems **50**, 559 (2011).

[57] H. Wang and S. Wang, *A knowledge management approach to data mining process for business intelligence,* Industrial Management & Data Systems **108**, 622 (2008).

[58] R. Škrinjar and P. Trkman, *Increasing process orientation with business process management: Critical practices',* International journal of information management **33**, 48 (2013).

[59] V. B. Vukšić, M. P. Bach, and A. Popovič, *Supporting performance management with business process management and business intelligence: A case analysis of integration and orchestration,* International journal of information management **33**, 613 (2013).

[60] C. M. Kahn and W. Roberds, *Credit and identity theft,* Journal of Monetary Economics **55**, 251 (2008).

[61] D. B. Douglass, *An Examination of the Fraud Liability Shift in Consumer Card-based Payment Systems,* Economic Perspectives **33** (2009).

[62] S. J. Murdoch and R. Anderson, *Verified by Visa and MasterCard SecureCode: or, How Not to Design Authentication,* in *Financial Cryptography and Data Security* (Springer, 2010) pp. 336–342.

[63] J. Green, *Exclusive Bankcard Profitability Study And Annual Report,* Cards & Payments , 36 (2008).

[64] K. Fitzgerald, *Exclusive Bankcard Profitability Study And Annual Report*, Tech. Rep. (SourceMedia, New York, 2011).

[65] W. Bolt and S. Chakravorti, *Digitization of Retail Payments,* in *The Oxford Handbook of the Digital Economy*, edited by J. Peitz, Martin; Waldfogel (Oxford University Press, New York, 2012) Chap. 5, pp. 108–135.

[66] A. Martin, *The Card Game: How Visa, Using Card Fees, Dominates a Market,* (2010).

[67] W. F. Baxter, *Bank interchange of transactional paper: Legal and economic perspectives,* Journal of Law and Economics **26**, 541 (1983).

[68] VISA, *Annual Report 2013*, Tech. Rep. (San Francisco, 2013).

[69] L. Constantine, *Priceless* (Kaplan Publishing, New York, 2009).

[70] K. J. Arrow, *Essays in the Theory of Risk-Bearing,* (1974).

[71] R. Schmalensee, *Payment Systems and Interchange Fees,* The Journal of Industrial Economics **50**, 103 (2002).

[72] J.-C. Rochet and J. Tirole, *Cooperation among Competitors: Some Economics of Payment Card Associations,* The RAND Journal of Economics **33**, 549 (2002).

[73] D. Evans and R. Schmalensee, *The Economics of Interchange Fees and their Regulation: An Overview,* (2005).

[74] J. Farrell, *Efficiency and Competition between Payment Instruments,* Review of Network Economics **5**, 26 (2006).

[75] PCI Security Standards Council, *PCI Security Standards Documents: PCI DSS, PA-DSS, PED Standards, Compliance Guidelines and More,* (2013).

[76] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, *Top 10 algorithms in data mining,* Knowledge and Information Systems **14**, 1 (2007).

[77] C. Ordonez and K. Zhao, *Evaluating association rules and decision trees to predict multiple target attributes,* Intelligent Data Analysis **15**, 173 (2011).

[78] R. J. Bayardo, R. Agrawal, and D. Gunopulos, *Constraint-based rule mining in large, dense databases,* Data Mining and Knowledge Discovery **4**, 217 (2000).

[79] C. Borgelt, *Software for Frequent Pattern Mining,* (2014).

[80] S. Brin, R. Motwani, and C. Silverstein, *Beyond market baskets: Generalizing association rules to correlations,* in *Proceedings 1997 ACM SIGMOD International Conference Management of Data*, Vol. 26 (ACM, Tucson, AZ, 1997) pp. 265–276.

[81] R. Agrawal, T. Imielinski, and A. Swami, *Mining Association Rules Between Sets of Items in Large Databases,* in *SIGMOD '93 Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, Vol. 22 (ACM, 1993) pp. 207–216.

[82] R Foundation, *The R Project for Statistical Computing,* (2014).

[83] RStudio Inc, *RStudio,* (2013).

[84] M. Hahsler, C. Buchta, B. Gruen, K. Hornik, and C. Borgelt, *arules: Mining Association Rules and Frequent Itemsets,* (2014).

[85] J. Conway, D. Eddelbuettel, T. Nishiyama, S. K. Prayaga, and N. Tiffin, *RPostgreSQL: R interface to the PostgreSQL database system,* (2013).

[86] M. Dowle, T. Short, S. Lianoglou, and A. Srinivasan, *data.table: Extension of data.frame,* (2014).

[87] H. Wickham, *ggplot2: elegant graphics for data analysis* (Springer, 2009).

[88] C. Borgelt, *Frequent item set mining,* Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **2**, 437 (2012).

[89] M. van der Velde, P. Jansen, and N. Anderson, *Guide to management research methods* (Wiley-Blackwell, 2004).

[90] W. Hämäläinen and M. Nykänen, *Efficient Discovery of Statistically Significant Association Rules,* in *2008 Eighth IEEE International Conference on Data Mining* (IEEE, 2008) pp. 203–212.

[91] J. Pei and J. Han, *Can We Push More Constraints into Frequent Pattern Mining?* in *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2000) pp. 350–354.

[92] J. S. Yoo, S. Shekhar, J. Smith, and J. P. Kumquat, *A Partial Join Approach for Mining Co-Location Patterns,* in *Proceedings of the 12th annual ACM International Workshop on Geographic Information Systems* (ACM, 2004) pp. 241–249.

[93] G. Grahne and J. Zhu, *Fast Algorithms for Frequent Itemset Mining Using FP-trees,* Knowledge and Data Engineering **17**, 1347 (2005).

[94] P. Carmona-Saez, M. Chagoyen, A. Rodriguez, O. Trelles, J. M. Carazo, and A. Pascual-Montano, *Integrated Analysis of Gene Expression by Association Rules Discovery,* BMC Bioinformatics **7**, 54 (2006).

[95] T. Quack, V. Ferrari, B. Leibe, and L. Van Gool, *Efficient Mining of Frequent and Distinctive Feature Configurations,* in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on* (IEEE, 2007) pp. 1–8.

[96] A. Tajbakhsh, M. Rahmati, and A. Mirzaei, *Intrusion Detection Using Fuzzy Association Rules,* Applied Soft Computing **9**, 462 (2009).

[97] R. Srikant, Q. Vu, and R. Agrawal, *Mining Association Rules with Item Constraints,* in *Proceedings of the Thirds International Conference Knowledge Discovery and Data Mining, KDD*, Vol. 97, edited by D. Heckerman, H. Mannila, D. Pregibon, and R. Uthurusamy (AAAI Press, 1997) pp. 67–73.