

# Voorspelling van aandelenkoersen op basis van nieuwsberichten

Michael van Kempen (283858)

Bachelorscriptie Informatica & Economie

Begeleider: dr. ir. J. van den Berg

Juli 2006

1. Introductie .....	3
1.1 Het Vakgebied.....	3
1.2 Doelstelling .....	5
1.3 Methodologie .....	6
1.4 Structuur .....	7
2. Eerder werk .....	8
2.1 ‘Using News Articles to Predict Stock Price Movements’ .....	8
2.2 ‘Augmented Trading’ .....	8
2.3 ‘Stock Prediction – A Neural Network Approach’ .....	9
3. Theoretisch Kader .....	10
3.1 Methode.....	10
3.2 Data .....	13
4. Uitwerking en Implementatie.....	14
4.1 Nieuwsberichten.....	14
4.2 Prijzen.....	15
4.3 Events.....	16
4.3.1 Structuur .....	16
4.3.2 Bedrijfsspecifieke en marktspecifieke events .....	17
4.3.3 Groepen van events .....	18
4.3.4 Effect van een nieuwsbericht .....	20
4.3.5 Sentiment.....	21
4.4 Herberekening .....	23
4.5 De voorspelling .....	25
4.5.1 Effect van de markt .....	25
4.5.2 Effect van bedrijfsspecifieke events.....	26
4.5.3 Voorspelling van een aandeel .....	26
4.5.4 Voorspelling van de AEX-index .....	27
5. Resultaten .....	28
5.1 Voorspelling op donderdag 29 juni 2006 .....	28
5.2 Voorspelling op maandag 3 juli 2006 .....	29
5.3 Tekenreeksanalyse .....	30
6. Conclusie en verbeterpunten .....	32
7. Discussie .....	34
Referenties.....	35
Bijlage 1: Trefwoorden events .....	36
Bijlage 2: Trefwoorden bedrijven .....	37
Bijlage 3: Trefwoorden sentiment.....	42
Bijlage 4: Berekening van het effect.....	43
Bijlage 5: De AEX-index .....	44

# 1. Introductie

De hele dag word je geconfronteerd met informatie: in de krant, op televisie, op internet, in reclamefolders en zo zijn er nog vele voorbeelden te bedenken. Deze overweldigende hoeveelheid heeft een grote invloed op het dagelijks leven. Niet alle informatie is voor iedereen van even groot belang. Zo kunnen bepaalde nieuwsberichten jouw wereld op zijn kop zetten terwijl hetzelfde bericht bij andere mensen nauwelijks de gedachtegang zal onderbreken. Wanneer je informatie ‘tegenkomt’ is het eerste dat je doet kijken waar het over gaat. Vervolgens bepaal je of het voor jou interessant zou kunnen zijn en pas als dat het geval is ga je beter kijken naar de informatie en deze aandachtig doorlezen.

Net als de ‘gewone’ wereld wordt de aandelenwereld ook de hele dag overspoeld met allerlei nieuwsberichten. Deze nieuwsberichten worden geanalyseerd en er wordt gekeken naar wat het effect van het nieuws zal zijn. Als het een effect is dat van belang is voor de verwachte cashflows zal de koers veranderen. Hoe groot dit effect zal zijn is moeilijk te zeggen, maar hierdoor wordt het interessant om te onderzoeken of op basis van het meest actuele relevante nieuws de koers voorspeld kan worden. Dit is ook wat hier onderzocht zal gaan worden. Kan op basis van nieuwsberichten het effect op de aandelenkoersen en hiermee de koers zelf voorspeld worden? Om dit te onderzoeken zal een programma geschreven worden dat nieuwsberichten leest, verwerkt en met behulp hiervan een voorspelling van de aandelenkoersen maakt.

## 1.1 *Het Vakgebied*

Al zeer lange tijd proberen mensen zicht te krijgen op wat er in de toekomst gaat gebeuren zodat hierop ingespeeld kan worden. Dit begon duizenden jaren geleden met bijvoorbeeld het aanleggen van een voorraad voedsel om de winter door te komen. Naarmate de mens zich meer ontwikkelde is de behoefte aan het wegnemen van onzekerheid alleen maar groter geworden. Tegenwoordig willen we veel al graag van tevoren weten. Hoe is het weer volgende week in Spanje? Wat kunnen we verwachten van de komende verkiezingen, wie staat er voor in de peilingen? Met dit soort vragen houden veel mensen zich bezig.

Ook binnen het bedrijfsleven zijn voorspellingen zeer belangrijk. Zo is het handig om al van tevoren te weten wat de concurrent gaat doen, zodat daarop ingespeeld kan worden. Omdat het vaak zeer grote bedragen op kan leveren zijn bedrijven bereid veel te betalen voor een juiste voorspelling. Eén van de meest sprekende voorbeelden van een mogelijk zeer lucratieve voorspelling is een juiste prognose van de aandelenkoersen. Wie zou er nou niet graag de enige zijn die met relatief grote mate van zekerheid weet welke koersen gaan stijgen en gaan dalen zodat er aandelen gekocht kunnen worden die winst op gaan leveren.

Vanwege deze mogelijk hoge winsten die te behalen zijn met een juiste voorspelling van de aandelenkoersen, is er gezocht naar methoden om de prijs te voorspellen. De eerste methoden die gebruikt werden waren voornamelijk gebaseerd op intuïtie en de eigen visie op de markt. Naarmate meer mensen zich gingen bezighouden met aandelen, kwam er meer interesse voor andere methodes. Men is gaan kijken naar historische gegevens en wat deze te zeggen hebben over de toekomst. Door deze gegevens verder te analyseren en een trend te bepalen kon al een redelijke voorspelling gemaakt worden. Later is dit verder uitgebreid en verbeterd door onder andere de correlatiecoëfficiënten tussen aandelen mee te nemen in de voorspelling.

Het nadeel van statistische methoden zoals bijvoorbeeld trendanalyse is dat er op een hoog abstractieniveau gekeken. Er wordt gekeken naar wat de koersen in het verleden hebben gedaan en er wordt vanuit gegaan dat deze trend doorzet. De onderliggende processen blijven buiten beeld terwijl deze juist de koerswijzigingen veroorzaken. Hierdoor blijven onverwachte pieken en dalen in de koersen veroorzaakt door de sfeer op de markt en andere onverwachte gebeurtenissen erg moeilijk te voorspellen. Omdat men toch graag een voorspelling wil hebben die ook met deze effecten rekening houdt, is er de laatste jaren meer gekeken naar elementen die de prijs beïnvloeden. Een indeling van elementen die invloed hebben op de aandelenkoersen is volgens de NYSE Group [1] gegeven door:

- De financiële gezondheid van een bedrijf. Zo zal een aankondiging dat de winst lager uit zal vallen ervoor zorgen dat de koers zal gaan dalen. Ook berichten over een CEO die met pensioen gaat, rechtzaken of stakingen geven geen goed beeld over de toekomst. Daarentegen zal bijvoorbeeld het bericht dat het bedrijf bezig is een ander bedrijf over te nemen juist wel een positief effect hebben op de koers.
- Informatie over de gehele industrie. Bedrijven kunnen het zelf wel goed doen, terwijl het in de rest van die bedrijfstak erg slecht gaat. Hierdoor gaan investeerders twijfelen

over het vermogen om te blijven groeien, waardoor de koersen kunnen gaan dalen. Een bekend verschijnsel hierbij is dat er vaak een cyclische beweging in de gehele industrie is waar te nemen. Zo is het bijvoorbeeld in de elektronica-industrie bekend dat er een top zit rond de feestdagen aan het eind van het jaar.

- Economische trend. Zoals de informatie over de gehele bedrijfstak belangrijk is voor de koersen, zo is ook de informatie over de gehele economie belangrijk voor de verwachtingen en wordt daarom ook nauwlettend in de gaten gehouden. Belangrijke indicatoren hierbij zijn het Bruto Nationaal Product BNP, inflatie, rentestand, werkloosheid, begrotingstekorten of –overschotten, koopkracht.
- Mondiale en nationale gebeurtenissen. Bepaalde nieuwsberichten kunnen een grote invloed hebben op de aandelenkoersen. Zo zal bijvoorbeeld een bericht over belastingverlaging ervoor zorgen dat de koersen zullen gaan stijgen.

De overeenkomst tussen deze categorieën is dat alle informatie over deze onderwerpen uit het nieuws gehaald kan worden. Het is daarom ook logisch dat mensen geïnteresseerd zijn in het verband tussen het nieuws en het verloop van de aandelenkoersen. Er zijn ook al meerdere methoden verzonden om het nieuws te verwerken in een voorspelling van de koersen. Meer hierover volgt in hoofdstuk 2.

## **1.2 Doelstelling**

Dit onderzoek is gericht op het verband tussen het bekend worden van nieuws en het verloop van de koersen. Het uiteindelijke doel is om te onderzoeken of het mogelijk is de koers te voorspellen door te kijken naar het meest recente nieuws. Hiervoor moet per nieuwsbericht gekeken worden welk soort nieuws het is, vervolgens moet er een verwacht effect aan toegekend worden, een tijdsduur moet bepaald worden en het bericht moet verwerkt worden in de voorspelling van de prijs.

*“Met nieuws in de journalistiek wordt meestal bedoeld nog niet eerder geopenbaarde gestructureerde verslagen van specifieke gebeurtenissen met name op politiek of economisch gebied, in journalistieke media.” [3]*

Omdat per definitie nog niet bekend is wat het nieuws in de toekomst zal zijn en de voorspelling wel voor een aantal uren gemaakt moet worden, wordt een voorspelling gemaakt op basis van de nieuwsberichten die reeds bekend zijn. Omdat in de echte koers ook het ‘nieuwe’ nieuws zit verwerkt, is de verwachting dat naarmate de tijd vanaf het voorspeltijdstip verstrijkt, de voorspelling meer gaat afwijken van het werkelijke koersverloop.

### **1.3 Methodologie**

Om de aandelenkoers te kunnen voorspellen zal gebruik gemaakt worden van een programma dat de nieuwsberichten analyseert en deze verwerkt in een voorspelling. Een dergelijk onderzoek is reeds gedaan bij het werkcollege Informatiesystemen onder leiding van dr. F. Frasincaar “Adding time to OWL” [2]. In dit onderzoek is niet een voorspelling gemaakt van aandelenkoersen, maar van de olieprijs. Hier wordt aan nieuwsberichten uit een RSS newsfeed een effect en een duur toegekend. Op basis van deze gegevens wordt een voorspelling gemaakt. Het programma dat hiervoor is geschreven zal als basis gebruikt worden voor dit project.

Er wordt dus een simulatie gemaakt van de markt waarbij gebruik gemaakt zal worden van real-time data. Om te voorkomen dat het te groot wordt zal alleen gekeken worden naar de AEX-index en de fondsen die hierin zitten. Door ieder fonds apart te voorspellen kan daarna de AEX-index berekend worden op basis van deze voorspellingen. Het programma zal een aantal weken draaien en in die periode al het nieuws dat binnengehaald wordt via newsfeeds verwerken in de prijs. Aan het eind van deze periode zal de verwachte koers vergeleken worden met de werkelijke koers. Hierbij zal in eerste instantie gekeken worden naar de AEX-index, maar wanneer er grote afwijkingen zijn zullen ook de individuele koersen gecontroleerd worden.

## **1.4 Structuur**

Na deze introductie zullen in het volgende hoofdstuk een aantal onderzoeken besproken worden die ook gericht zijn op het maken van voorspellingen op basis van nieuwsberichten. Vervolgens zal in hoofdstuk 3 stilgestaan worden bij de achterliggende theorie en benodigde gegevens. De uitwerking en precieze implementatie van deze theorie zal in hoofdstuk 4 uitgebreid uitgelegd worden. Hier staat dus de meer technische data van de berekeningen en voorspelmethode. Vervolgens zullen in hoofdstuk 5 de resultaten getoond worden. In hoofdstuk 6 worden deze nader besproken en wordt de conclusie eruit getrokken die antwoord zal geven op de hoofdvraag of het mogelijk is om aandelenkoersen te voorspellen op basis van nieuws-berichten. Ook zullen hierin punten aangehaald worden die nog verbeterd kunnen worden.

## **2. Eerder werk**

Zoals al eerder is vermeld, is al enig onderzoek gedaan naar het effect dat nieuws heeft op het koersverloop. Een aantal van deze onderzoeken zullen hier kort besproken worden. Hierbij komt vooral aan bod welke methode gebruikt wordt en wat de voor- en nadelen hiervan zijn

### ***2.1 'Using News Articles to Predict Stock Price Movements'***

In de paper 'Using News Articles to Predict Stock Price Movements' van Győző Gidófalvi [5] wordt een methode beschreven waarmee nieuwsberichten verwerkt kunnen worden in een voorspelling van de aandelenkoersen. Deze methode gaat uit van de  $\beta$  van specifieke aandelen. Deze wordt berekend door de procentuele verandering van een specifiek aandeel te vergelijken met de verandering van de betreffende index. Sommige aandelen hebben een  $\beta$  lager dan 1, wat betekent dat deze redelijk stabiel zijn in vergelijking met de index. Een  $\beta$  hoger dan 1 betekent dat de aandelen extremer reageren dan de index. Door uitkomende nieuwsberichten kan de werkelijke verandering onder, boven of op de verandering van de index uitkomen. Door gebruik te maken van de  $\beta$  kan een voorspelling gemaakt worden voor een zeer korte termijn (ongeveer 20 minuten).

Het voordeel van de gebruikte methode is dat per nieuwsbericht bepaald kan worden wat het effect is waardoor een gedetailleerde voorspelling ontstaat. Het nadeel is echter dat voor een voorspelling van de aandelenkoersen een voorspelling van de marktindex nodig is, omdat deze de basis vormt waar de individuele aandelen van afwijken.

### ***2.2 'Augmented Trading'***

In de master scriptie 'Augmented Trading' van Arthur Hugo van Bunningen [6] wordt een soortgelijke methode beschreven, maar hier wordt voornamelijk ingegaan op de analyse van



de nieuwsberichten. De nieuwsberichten worden doorzocht op bepaalde trefwoorden. Vervolgens worden synoniemen hiervan bekeken en dit blijft zich een bepaald aantal stappen herhalen. Aan de hand van de resultaten van deze analyse wordt het bericht in een bepaalde groep ingedeeld met een bepaald effect.

Een voordeel van deze methode is dat de gedetailleerde classificatie ervoor zorgt dat er een goede voorspelling gemaakt kan worden door naar vergelijkbare nieuwsberichten uit het verleden te kijken. Een nadeel is dat semantische analyse vaak erg kwetsbaar is en het vaak zal gebeuren dat de indeling in een bepaalde groep niet de juiste is.

### **2.3 'Stock Prediction – A Neural Network Approach'**

Een onderzoek dat wel binnen het vakgebied ligt, maar hier wat uit de toon valt is de master thesis 'Stock Prediction – A Neural Network Approach' van Karl Nygren [4]. Dit onderzoek is anders dan de andere twee omdat de methode niet is gebaseerd op nieuwsberichten, maar op een traditionelere methode: technische analyse van tijdseries. Het is echter vergeleken met de eerdere onderzoeken op dat gebied weer een volgende stap, omdat gebruik gemaakt wordt van neurale netwerken. Neurale netwerken zijn lerende systemen die, zoals de naam al suggereert, gebaseerd zijn op het menselijk brein. Door de neuronen en de verbindingen hiertussen te simuleren kan een systeem gemaakt worden dat patronen kan herkennen in grote hoeveelheden data.

Het voordeel van neurale netwerken is dat ze niet gebonden zijn aan vaste regels en formules in tegenstelling tot de traditionele methoden. De systemen leren dus daadwerkelijk van het verleden en kunnen zo bepalen welke informatie bruikbaar is voor een voorspelling en welke niet. Een ander voordeel is dat de systemen erg flexibel zijn en dus niet bij afwijkende data onbetrouwbare resultaten geven.

Een nadeel van neurale netwerken is dat de werking van het systeem onbekend blijft (de zogenaamde 'black box'). De te analyseren data wordt ingevoerd en de output is te bekijken, maar wat er precies gebeurt is niet te bekijken. Daarnaast hebben neurale netwerken een relatief lange tijd nodig om te 'leren'.

Toch is het interessant om hier te noemen, omdat de combinatie van neurale netwerken met nieuwsgelateerde voorspellingen zeer interessant kan zijn voor toekomstig onderzoek.

## 3. Theoretisch Kader

### 3.1 Methode

In het verleden werd het grootste deel van de voorspellingen gemaakt door historische gegevens te analyseren en statistische methoden hierop toe te passen. Technische analyse heeft als onderdeel van deze statistische methoden als nadeel dat de voorspelling die hiermee verkregen wordt erg grof is en er vanuit gaat dat de trend uit het verleden door zal zetten. Het is echter bekend dat aandelenkoersen aan grote veranderingen onderhevig zijn en niet goed zijn te voorspellen met behulp van technische analyse. Volgens de efficiënte markthypothese zou er zelfs geen enkele voorspelling mogelijk zijn.

Zoals in het eerste hoofdstuk al duidelijk is gemaakt zal de voorspelling zoals deze hier gemaakt wordt, niet gebaseerd zijn op technische analyse, maar op de verwerking van nieuwsberichten. De voorspelling wordt gemaakt door nieuwsberichten te analyseren en te verwerken in de koersen. Omdat er dagelijks tientallen nieuwsberichten uitkomen en niet al deze berichten een gelijk effect hebben op de koersen, zullen de berichten ingedeeld worden in groepen. Deze groepen hebben een bepaald standaard effect dat gebruikt zal worden voor de voorspelling. Details over de bepaling van het effect volgen in het volgende hoofdstuk.

Volgens de efficiënte markt hypothese zouden deze nieuwsberichten direct en goed verwerkt worden in de aandelenkoers. Hierbij gaat het bij een efficiënte markt in de zwakke vorm om de juiste prijs als alleen historische gegevens bekend zijn. Bij een efficiënte markt in de semi-sterke vorm om alle publiek gemaakte informatie. Bij een efficiënte markt in de sterke vorm gaat het om alle informatie en dus ook om de insiderinformatie. In dit onderzoek zal uitgegaan worden van een efficiënte markt in de semi-sterke vorm, want er wordt gekeken naar nieuwsberichten die voor iedereen toegankelijk zijn. Al zegt de efficiënte markt hypothese dat de aandelen direct en correct geprijsd zullen worden, gaan we er hier echter vanuit dat dit een veronderstelling is die in de praktijk niet altijd waar is en er enige tijd zit tussen het uitkomen van nieuws en de juiste verwerking daarvan in de koersen. Deze verwerking gaat gepaard met enige schommelingen totdat uiteindelijk de juiste prijs bereikt wordt. Hoe lang deze aanpassingsperiode duurt zal per groep verschillen.

De onderzoeken uit het vorige hoofdstuk zijn gericht op de verwerking van nieuwsberichten en dan vooral op het toekennen van een effect aan de berichten door deze te analyseren. Deze benadering zal hier ook toegepast worden, maar zal verder worden uitgebreid. We gaan er namelijk vanuit dat het directe effect van een bepaalde gebeurtenis niet het enige is dat effect heeft op de koersen. Zo is het bekend dat sentiment een belangrijke rol speelt op de aandelenmarkt. Om dit te kunnen modelleren zal het sentiment weergegeven worden als een versterking of een verzwakking van een effect. Als de markt positief gestemd is, zullen positieve berichten beter ontvangen worden en zal het effect versterkt worden. Bij een negatief sentiment zullen goede berichten daarentegen juist afgezwakt worden. Andersom geldt hetzelfde.

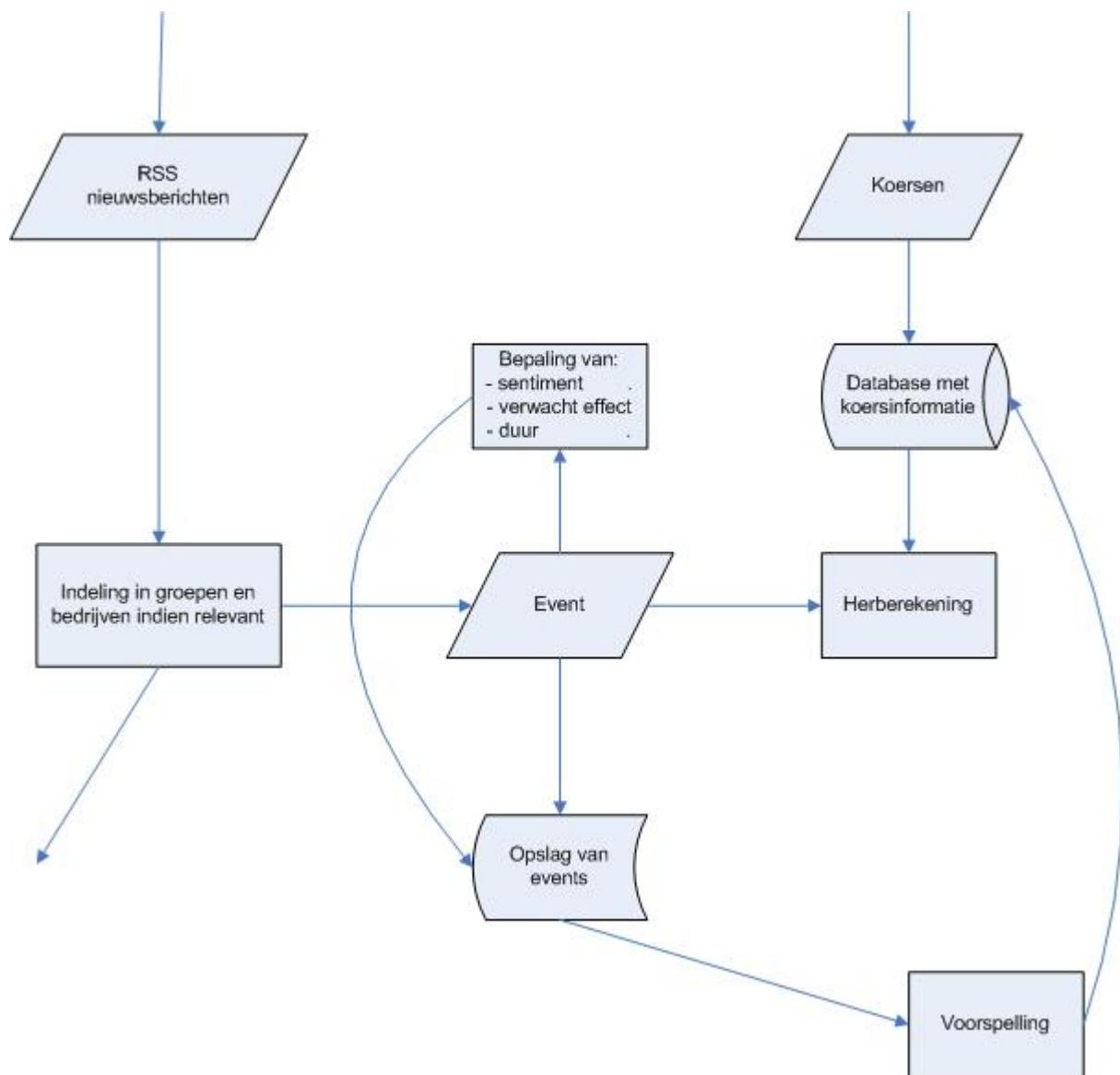
	Positief gestemde markt	Negatief gestemde markt
Positief nieuwsbericht	Versterkt	Verzwakt
Negatief nieuwsbericht	Verzwakt	Versterkt

Naast het sentiment gaan we er ook van uit dat de dynamiek van de economie in zijn geheel invloed heeft op het prijsverloop. Zo zal een aantrekkende economie over het algemeen voor een stijging van de koersen zorgen en een zwakke economie voor meer onzekerheid en dus koersdalingen. Ook hier speelt het sentiment weer dezelfde rol als bij de bepaling van het effect van een nieuwsbericht. Hoe dit precies is geïmplementeerd zal ook in het volgende hoofdstuk uitgelegd worden.

Op het moment dat een voorspelling gemaakt moet worden zijn een aantal dingen bekend: Ten eerste is de koers van het aandeel op dat moment bekend. Ten tweede is er een lijst van nieuwsberichten bekend waarvan de aanpassingsperiode nog niet is verstreken. Met behulp van deze twee onderdelen kan een voorspelling gemaakt worden.

De voorspelling van de AEX-index vereist een andere benadering. Deze is niet direct afhankelijk van het nieuws, maar indirect via de fondsen waaruit de index is opgebouwd. Omdat elk fonds een bepaald gewicht heeft binnen de index, geven de voorspelde waarden van de fondsen, vermenigvuldigd met het gewicht van het desbetreffende fonds, de voorspelling van de AEX-index.

Het verloop van de nieuwsberichten door het programma wordt gegeven door het volgende schema:



(Figuur 1)

Een nieuwsbericht wordt gelezen uit een RSS newsfeed. Het bericht wordt geclassificeerd in een bepaalde groep en bij een bepaald bedrijf. Het kan voorkomen dat het nieuwsbericht niet relevant is voor verdere verwerking, omdat het bijvoorbeeld over een bedrijf gaat dat niet in de AEX-index is opgenomen. In dit geval wordt het nieuwsbericht niet verder verwerkt. Als het bericht wel relevant is, wordt er een event van gemaakt. Events vormen de basis van het systeem en worden in het volgende hoofdstuk uitgebreid uitgelegd. In deze fase worden ontbrekende waarden die nodig zijn voor een voorspelling ingevuld en wordt het event opgeslagen. Een proces dat los staat van deze keten is het ophalen van koersinformatie. Alle

koersen van de fondsen en de AEX-index worden opgeslagen in een database. De combinatie van deze 2 datastromen is voldoende om een voorspelling te kunnen maken. Tot slot is er na afloop van een event een herberekening nodig om het effect te bepalen dat het event daadwerkelijk heeft gehad. Meer over al deze stappen volgt in hoofdstuk 4.

### **3.2 Data**

Als input zijn data nodig. Er zijn 2 datastromen nodig. Ten eerste de data die informatie bevat over de aandelenkoersen. Omdat er erg veel verschillende aandelen zijn en het onderzoek daardoor te groot zou worden, is gekozen om het onderzoek te beperken tot de fondsen van de AEX-index. De prijzen van deze fondsen en die van de index zelf zijn gemakkelijk te vinden op [www.euronext.nl](http://www.euronext.nl). Ook is hier een geschiedenis te vinden van het afgelopen jaar. Het voordeel van deze bron is dat de prijzen zeer regelmatig worden bijgewerkt (meerdere malen per uur).

De tweede datastroom is de stroom met de nieuwsberichten. Voor nieuwsbronnen zijn zogenaamde RSS-feeds. De afkorting RSS staat voor Really Simple Syndication. Dit zijn op xml gebaseerde bestanden die de nieuwsberichten bevatten in ruwe vorm, dus zonder opmaak zoals bij html het geval is. Het voordeel van deze RSS-feeds is dat de informatie reeds is ingedeeld en er extra informatie aan toegekend is. Zo zijn vaak de titel en het artikel zelf gescheiden van elkaar. Ook is er vaak een publicatiedatum bij geplaatst en een link naar het originele artikel op internet. Zo zijn er nog enkele onderdelen te verzinnen die hieraan toegevoegd kunnen worden. Door deze indeling zijn de berichten veel gemakkelijker te verwerken in toepassingen dan wanneer ze van een html-site gelezen zouden moeten worden. De RSS-feed die is gebruikt voor de nieuwsberichten is afkomstig van [www.beurs.nl](http://www.beurs.nl) [7].

## 4. Uitwerking en Implementatie

Er zal hier aan de hand van een aantal stappen uitgelegd worden hoe nieuwsberichten verwerkt worden in het programma en hoe hier een voorspelling van gemaakt wordt. Hiervoor zal ten eerste uitgelegd worden hoe de nieuwsberichten zijn opgebouwd. Vervolgens zal worden ingegaan op het up-to-date houden van de prijzen. Hierna zullen de events uitgelegd worden die de basis vormen voor de voorspelling die hierna volgt. Tot slot zal uitgelegd worden hoe de voorspelling en de herberekening van het echte effect van een event plaatsvinden.

### 4.1 Nieuwsberichten

De nieuwsberichten worden van internet uit een RSS-feed gelezen. Het voordeel hiervan is dat de nieuwsberichten altijd recent zijn. Hierdoor kunnen berichten die van invloed zijn op de prijs direct verwerkt worden. De nieuwsberichten komen van de website [www.beurs.nl](http://www.beurs.nl) [7] en bestaan uit 4 onderdelen:

- Titel.
- Datum. De datum en tijd waarop het nieuwsbericht is gepubliceerd.
- Beschrijving. Een korte beschrijving van de inhoud van het nieuwsbericht. Deze bestaat uit de eerste regels van het volledige artikel.
- URL. Een link naar het volledige artikel dat te vinden is op internet.

Om deze onderdelen te kunnen lezen wordt gebruik gemaakt van Informa [8]. Dit is een pakket dat binnen Java aangeroepen kan worden en waarmee RSS-feeds gemakkelijk gelezen kunnen worden.

Wanneer een nieuwsbericht binnen komt, wordt eerst gekeken of deze niet reeds voorkomt in het systeem. Hiervoor worden de links naar het volledige artikel vergeleken met de links van berichten die al zijn opgeslagen. Wanneer het nieuwsbericht nog niet eerder is voorgekomen wordt het artikel verder verwerkt. Hiervoor worden de 4 onderdelen die hierboven beschreven staan opgeslagen, waarbij de tekst van het volledige artikel ook wordt opgeslagen via de link die in het nieuwsbericht zat.

Daarnaast wordt een status aan het nieuwsbericht toegevoegd die de volgende waarden kan aannemen:

0. Nieuw nieuwsbericht. Nog niet aanwezig in het systeem.
1. Verwerkt in het systeem en in gebruik voor de voorspelling
2. Verwijderd. Deze status wordt toegekend indien het nieuwsbericht als ‘niet relevant’ wordt aangemerkt.

De initiële waarde van de status is 0. In een latere fase van verwerking wordt status 1 of 2 aan het bericht toegekend.

## **4.2 Prijzen**

Net als de nieuwsberichten worden ook de prijzen van internet gelezen. Het gaat hier om de koersen van de AEX-index en alle fondsen waaruit deze index is opgebouwd. Deze prijzen zijn afkomstig van de site [www.euronext.com](http://www.euronext.com) [9] en worden hier twee maal per uur bijgewerkt tot de meest actuele stand. Het programma leest deze prijzen en bewaart ze in het systeem. Wanneer de meest recente prijs op de site gelijk is aan de laatste die is toegevoegd, wordt de update genegeerd.

Er zijn een aantal stromen die van internet gelezen moeten worden. Ten eerste de prijzen van alle fondsen uit de AEX-index. Deze prijzen worden gesorteerd per fonds en in het juiste fonds opgeslagen. Daarnaast wordt van elk fonds bijgehouden hoe groot het gewicht van dat fonds is binnen de AEX-index. Ook wordt een prijsgeschiedenis van de AEX bijgehouden.

## 4.3 Events

Zodra een nieuwsbericht verwerkt wordt in het systeem en dus beschikbaar komt voor de voorspelling, wordt het nieuwsbericht omgevormd tot 'event'. De events zijn de basis van het programma en bevatten dezelfde informatie als de nieuwsberichten, maar worden verder uitgebreid met variabelen die nodig zijn voor de voorspelling.

### 4.3.1 Structuur

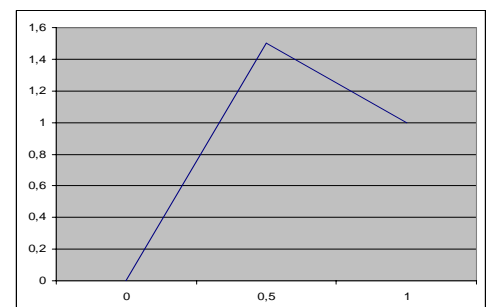
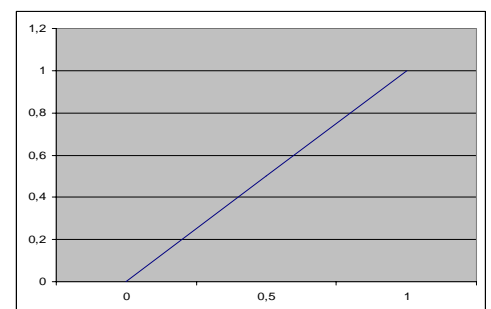
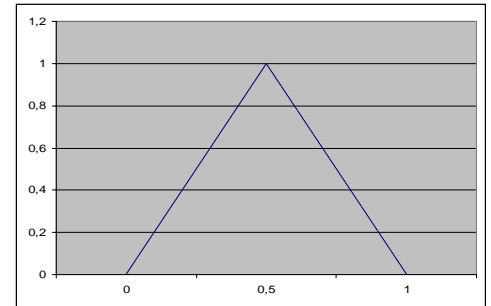
Een event bestaat uit de volgende onderdelen:

- Titel. Dit is dezelfde titel als uit het nieuwsbericht gelezen wordt.
- Beschrijving. De eerste regels van het volledige artikel. Deze wordt ook gekopieerd uit het nieuwsbericht.
- Artikel. De tekst van het volledige artikel.
- Link. Een link naar het artikel zoals dat op internet is te vinden.
- Groep. De events zijn in een aantal mogelijke groepen verdeeld. De variabele groep geeft aan in welke groep het desbetreffende event is ingedeeld. Meer hierover in paragraaf 4.3.3
- Bedrijf. Geeft aan op welk bedrijf het event van toepassing is.
- Startdatum. Deze variabele geeft de datum en tijd waarop het nieuws bekend is geworden en vanaf welk moment het effect hiervan dus loopt. Deze wordt gelijk gesteld aan het tijdstip waarop het nieuwsbericht is gepubliceerd.
- Duur. Per groep events wordt een waarde toegekend aan de tijd die het zal kosten voordat het nieuws in de koers is aangepast. Meer hierover volgt in paragraaf 4.3.4.
- Einddatum. Deze variabele geeft de datum en tijdstip aan waarop het effect van een nieuwsbericht volledig verwerkt is in de koers. Dit is dus de startdatum met hierbij opgeteld de duur van het effect.
- Effect. Het effect van het event. Meer hierover in paragraaf 4.3.4.
- Sentiment. Een waarde die het versterkende of verzwakkende effect van het sentiment uit het bericht bepaalt. Meer hierover in paragraaf 4.3.5.
- Type. Het aanpassen van de koers naar aanleiding van een event gaat gepaard met schommelingen. Daarnaast hebben ook niet alle events een blijvend effect. Sommige



berichten zullen tijdelijk voor onrust zorgen, maar daarna zal de koers zich herstellen tot het oude niveau. Om dit te kunnen modelleren wordt hier gebruik gemaakt van 3 types die een vereenvoudigde voorstelling geven van dit effect:

- Type 1: Piek. Het nieuws zorgt voor een tijdelijke piek met een grootte van het bepaalde effect. Het eindigt met een koers die op het oude niveau ligt. De waarde van de piek heeft de waarde van het effect.
- Type 2: Stijging/ daling. Het nieuws heeft een blijvend effect, maar er is weinig of geen sprake van een ‘schommelende aanpassing’. Hier wordt de prijs dus aangepast volgens een lineaire lijn. De waarde van het effect is de koerswijziging tussen het begin en het eind van het event.
- Type 3: Combinatie. Het effect is hier blijvend, maar de aanpassing gaat wel gepaard met schommelingen in de prijs. Het is dus een combinatie van type 1 en type 2. De waarde van het effect is het verschil tussen het begin en het einde van het event. De piek ligt halverwege deze periode en heeft een waarde van 1,5 x het effect.



### 4.3.2 Bedrijfsspecifieke en marktspecifieke events

Ieder bedrijf uit de AEX-index heeft een lijst van events die op het bedrijf van toepassing zijn. Sommige berichten zijn echter niet op een specifiek bedrijf van toepassing, maar op de markt in zijn geheel. Zo zullen bijvoorbeeld berichten van de Centrale Bank over een rentewijziging niet direct in verband gebracht worden met Ahold. Toch zal dit bericht een effect hebben op de koers van Ahold. Er is dus onderscheid te maken tussen bedrijfsspecifieke events die direct van toepassing zijn op het bedrijf en marktspecifieke events die indirect de koersen beïnvloeden. De marktspecifieke events kunnen echter op ieder bedrijf een ander effect

hebben. Daarom worden ze aan alle bedrijven toegekend, in tegenstelling tot de bedrijfsspecifieke events die aan de lijst van slechts één bedrijf worden toegevoegd (zie figuur 2)

	ABN Amro	Ahold	ASML Holding	...
Wijziging winstverwachting		Bedrijfsspecifiek		
Overname/ fusie/ afsplitsing/ samenwerking				
Overige bedrijfsspecifieke events				
Verandering consumentenvertrouwen				
Rentewijziging	Marktspecifiek			
Werkloosheid				
Stand van de AEX-index				
Stand van buitenlandse beurzen				
Economie algemeen				

(Figuur 2)

### 4.3.3 Groepen van events

Nieuwsberichten kunnen over allerlei onderwerpen gaan. Door een RSS-kanaal te kiezen dat alleen beursgerelateerde nieuwsberichten bevat, wordt een selectie gemaakt uit het totale aanbod aan nieuwsberichten. De nieuwsberichten die overblijven moeten geanalyseerd worden, waarbij gekeken wordt naar het soort nieuws en de betrokken partij. Hiervoor zijn een aantal groepen gemaakt van mogelijke gebeurtenissen. Ieder nieuwsbericht wordt in één van deze groepen ingedeeld of verwijderd als het niet relevant is voor de voorspelling. Door deze groepen te gebruiken kunnen de types gebeurtenissen beter vergeleken worden met elkaar en kan beter het effect worden bepaald. Hoe dit precies geïmplementeerd is, volgt hierna in paragraaf 4.3.4.

Na bestudering van de nieuwsberichten van enkele dagen zijn een aantal groepen gemaakt. Twee hiervan zijn al in het vorige hoofdstuk gepasseerd. Per groep zijn een aantal mogelijke events gedefinieerd. Deze indeling is als volgt:

- Bedrijfsspecifieke events:
- Wijziging winstverwachting
  - Overname/ fusie/ afsplitsing/ samenwerking
  - Overige bedrijfsspecifieke events
- Marktspecifieke events:
- Verandering consumentenvertrouwen
  - Rentewijziging
  - Werkloosheid
  - Economie algemeen (macro-economische cijfers, begrotingstekort)
- Koersinformatie:
- Stand van de AEX-index
  - Stand van buitenlandse beurzen

Het indelen in de juiste groep wordt gedaan door in het volledige artikel te bekijken hoe vaak bepaalde trefwoorden voorkomen, waarbij iedere groep en ieder bedrijf een lijst trefwoorden heeft (deze is te vinden in Bijlage 1: Trefwoorden events). De groep waarvan de trefwoorden het meeste voorkomen, wordt als suggestie meegegeven. Als trefwoorden in de titel van het nieuwsbericht voorkomen, krijgen deze voorrang bij de bepaling van de groep. Om dit te illustreren het volgende nieuwsbericht

***DAMRAK MIDDAG: AEX kabbelt rond slotstand van maandag***

*Op Euronext Amsterdam kabbelt de AEX-index rond de slotstand van maandag. De rest van Europa vertoont hetzelfde beeld. Rentevrees lijkt volgens handelaren opnieuw het credo. De markt kijkt vandaag onder andere uit naar de publicatie van de Amerikaanse index voor producentenprijzen over april. In Amsterdam kwamen ...*

(Bron: [7])

Omdat dit bericht een samenvatting is van allerlei gebeurtenissen die gedurende de dag zijn gebeurd, zou dit bericht in een aantal groepen ingedeeld kunnen worden wanneer naar de trefwoorden gekeken wordt. In de titel staat echter het woord ‘damrak’ wat een van de trefwoorden is voor het event ‘stand van de AEX-index’. Omdat het in de titel staat heeft dit

voorrang bij het bepalen van de groep. Hier zal dus als suggestie voor de groep ‘stand van de AEX-index’ gegeven worden.

Bij het bepalen van het bedrijf geldt hetzelfde. Ook hier wordt gekeken naar trefwoorden in het bericht, waarbij de titel voorrang heeft op de rest van de tekst (zie bijlage 2: Trefwoorden bedrijven). Omdat deze methode niet altijd de correcte groep en bedrijf als uitkomst zal geven, moet de keuze bevestigd worden door de gebruiker.

#### **4.3.4 Effect van een nieuwsbericht**

Wanneer een event toegevoegd wordt aan de lijst, wordt een verwachting gemaakt van het effect dat het bericht zal hebben op de koersen. De grootte van dit effect wordt berekend door te kijken naar vergelijkbare berichten uit het verleden. Als bijvoorbeeld een wijziging van de winstverwachting wordt doorgevoerd, wordt gekeken naar alle berichten over dit onderwerp in het verleden en het effect dat deze op de koersen hebben gehad. Het gemiddelde van deze waarden wordt gebruikt als verwachting van het effect. Er wordt bij deze berekening alleen gekeken naar het effect dat een wijziging van de winstverwachting al eerder heeft gehad op het bedrijf waar op dat moment naar gekeken wordt. Wanneer er nooit sprake is geweest van een vergelijkbare situatie, dus wanneer het bedrijf nooit een winstverwachting heeft gewijzigd, wordt gekeken naar het effect dat bij andere bedrijven plaats heeft gevonden na bekendmaking.

Omdat het effect bij aanvang van het programma nog niet bekend is, is aan ieder event een initiële waarde meegegeven. Zodra er één event in dezelfde groep is toegevoegd, zal deze waarde niet meer gebruikt worden. Om te zorgen dat events een juiste voorspelling krijgen, wordt na afloop van het event gekeken wat het effect daadwerkelijk is geweest. Meer hierover in paragraaf 4.4.

Omdat een effect een bepaalde duur heeft totdat het nieuws volledig verwerkt is in de prijzen, wordt in deze fase ook een duur toegekend aan het event. Omdat het moeilijk of zelfs onmogelijk is om de verwachte duur van een event uit de tekst van het artikel te halen, wordt

deze op een vaste waarde gesteld. Deze waarde verschilt per groep. Een overzicht van de initiële waarde van het effect, de duur van het effect en het type gesorteerd per soort event:

Soort event	Initiële waarde van het effect (in %)	Duur van een event (in uren)	Type
Wijziging winstverwachting	1,5	8	3
Overname/ fusie/ afsplitsing/samenwerking	1,5	12	2
Overige	0,5	4	1
Verandering consumptenvertrouwen	0,3	4	1
Rentewijziging	-0,2	4	3
Werkloosheid	-0,2	4	2
Economie algemeen	0,2	4	3
Stand van de AEX-index	0,3	4	2
Stand van buitenlandse beurzen	0,3	4	2

#### 4.3.5 Sentiment

Het sentiment wordt weergegeven als een variabele die het effect van een event kan versterken of verzwakken. De waarde van deze variabele ligt tussen 0 en 2. Om deze uit te rekenen wordt net als bij de bepaling van de groep en het bedrijf gekeken naar trefwoorden. In dit geval zijn het twee groepen woorden. Ten eerste de groep met positieve woorden als ‘winnen’, ‘stijgen’, ‘toename’, ‘herstel’ enz. Ten tweede de groep met negatieve woorden als ‘verlagen’, ‘dalen’, ‘verliezen’, ‘tekort’ enz. De complete lijst woorden is te vinden in Bijlage 3: Trefwoorden sentiment.

Hier een voorbeeld van een nieuwsbericht met positieve en negatieve woorden geselecteerd.

**[17:50] DAMRAK SLOT: Damrak pakt winst op kalme beursdag**

Op Euronext Amsterdam zijn de toonaangevende aandelenindices donderdag met winsten geëindigd. De handel verliep kalm op Hemelvaartsdag van het jaar 2006. Een handelaar wees erop dat er in de middag langzaam wat optimisme in de markt sloop, mede dankzij cijfers uit Amerika. Zo groeide de Amerikaanse economie in het eerste kwartaal van dit jaar minder hard dan verwacht. Dat nam de angst voor inflatie, en bijgevolg een hogere rente, weg. Desondanks was de omzet lager dan op een normale beursdag, zo merkte de handelaar op tegen Betten Beursmedia News. 'Om nu te zeggen dat het een inspirerende dag was? Nee, dat niet', aldus de handelaar. De AEX-index sloot 1,0% hoger op 440,28 punten. De Midkap won eveneens 1,0% tot 567,37 punten. Frankfurt won 1,8%, Parijs steeg 1,6% en Londen pakte 1,5%. De lange rente in Europa daalde met...

(Bron: [7])

Aantal positieve woorden ( $n_{pos}$ ): 13

Aantal negatieve woorden ( $n_{neg}$ ): 8

Door te kijken naar het aandeel van de positieve woorden in het geheel aan trefwoorden, wordt een indicatie gegeven van de verdeling van het aantal trefwoorden. Vervolgens moet dit kengetal omgezet worden naar een sentiment. Dit sentiment kan in het meest extreme geval het effect van een event verdubbelen. Wanneer er even veel positieve als negatieve woorden zijn, is er geen sprake van een positief of een negatief sentiment. In dit geval moet het sentiment dus gelijk zijn aan 1.

Het eerste kengetal dat berekend wordt is het aandeel van de positieve of negatieve woorden in het totaal aan trefwoorden. Hierbij wordt bij een positief verwacht effect gekeken naar het aandeel van positieve woorden en bij een negatief verwacht effect naar het aandeel negatieve woorden in het totaal. De formule die hierbij hoort is:

$$\begin{aligned} f &= n_{pos} / n && \text{als het verwachte effect} \geq 0 \\ f &= n_{neg} / n && \text{anders} \end{aligned} \quad (\text{Formule 1})$$

Hierbij is  $f$  de fractie positieve of negatieve woorden uit  $n$ .  $n$  is het totale aantal trefwoorden ( $n_{pos} + n_{neg}$ ).

Om deze fractie om te rekenen tot een waarde voor het sentiment worden de volgende formules gebruikt:

$$S_{event} = 2 * f \quad (Formule 2)$$

Waarbij  $S_{event}$  het sentiment is van het event.

Een bijzonder geval wordt gevormd wanneer het verwachte effect positief is, maar in het nieuwsbericht alleen negatieve woorden staan. Het sentiment zou dan gelijk zijn aan 0. Als dit geval zich voordoet, wisselt het verwachte effect van teken (vermenigvuldigd met -1). Hetzelfde geldt voor een negatieve verwachting uit een nieuwsbericht met alleen positieve woorden.

Deze formules uitgewerkt in het voorbeeldartikel geven dan:

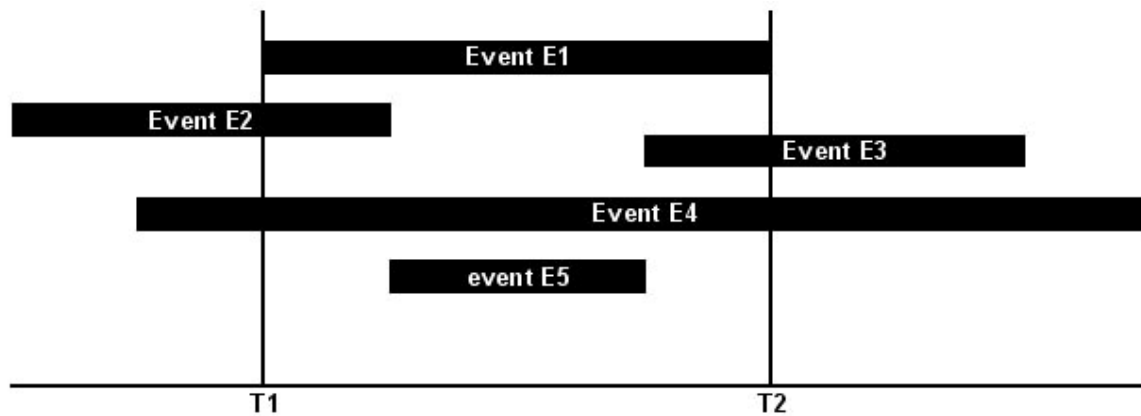
$$f = 13 / 21 = 0,62$$

$$S_{event} = 2 * 0,62 = 1,24$$

Dit betekent dus dat het effect door het sentiment 24 % versterkt wordt.

#### **4.4 Herberekening**

Het effect dat bij het creëren van het event is bepaald aan de hand van vergelijkbare events uit het verleden hoeft niet het juiste effect te zijn. Sterker nog: het werkelijke effect kan zelfs sterk van deze prognose afwijken. Het is daarom noodzakelijk dat achteraf bepaald wordt wat daadwerkelijk het effect is geweest. Hiervoor wordt gekeken naar de werkelijke koerswijziging gedurende de tijdsduur van het event. Het is echter mogelijk, en zelfs zeer aannemelijk dat er meerdere events op hetzelfde tijdstip invloed uitoefenen op de koers. Figuur 3 geeft hiervan een visuele voorstelling.



(Figuur 3)

Stel dat event E1 het event is waarvan we het werkelijke effect willen weten. E1 is begonnen op tijdstip T1 en geëindigd op tijdstip T2. De koerswijziging tussen deze twee tijdstippen kan bepaald worden uit de koersgeschiedenis. Deze koerswijziging zou gelijk moeten zijn aan de som van alle effecten die in die periode hebben gespeeld. Om de events die deze effecten veroorzaken te vinden, moet gezocht worden naar events waarvan het starttijdstip voor T2 en het eindtijdstip na T1 ligt. Event E2 t/m E5 uit figuur 2 voldoen hieraan en hebben dus allemaal een invloed gehad op de koers in deze periode. Om te kunnen zien hoe groot het effect was van deze events wordt per event gekeken hoe groot het effect in de (deel)periode was. Een voorbeeld hiervan:

Stel: event E2 is een event van type 2 (constante stijging/ daling) en heeft een effect van 3%. De duur van het event is 3 tijdseenheden. Op het moment dat de periode begint die hier bekeken wordt (T1) is al 2/3 van deze duur verstreken. Dit betekent dat nog 1/3 van het effect moet plaatsvinden en dat E2 dus nog voor een stijging van  $1/3 * 3 \% = 1\%$  zorgt (voor de precieze berekening zie Bijlage 4: Berekening van het effect).

Deze methode wordt voor ieder event toegepast. Door al deze (deel)effecten bij elkaar op te tellen, krijg je het effect van alle events in de periode met uitzondering van het event dat herberekend wordt. Door deze waarde van de koerswijziging af te trekken, krijg je de waarde van het effect van het herberekende event.



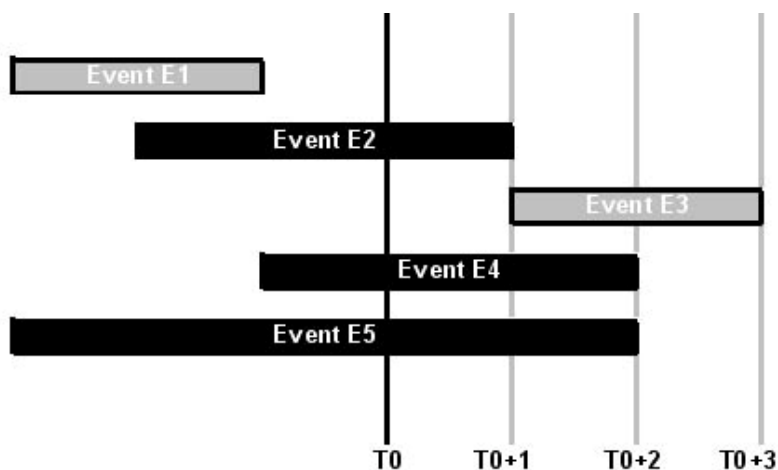
## 4.5 De voorspelling

De voorspelling bestaat uit een aantal onderdelen. Ten eerste wordt een voorspelling gemaakt van de markt in het algemeen. Vervolgens wordt een voorspelling gemaakt per aandeel. Wanneer deze stappen zijn doorlopen, kan hieruit een voorspelling van de AEX-index gemaakt worden.

### 4.5.1 Effect van de markt

De eerste stap die genomen wordt is een voorspelling van de markt in het algemeen. Hiervoor wordt eerst het marktsentiment bepaald. Dit is het sentiment uit alle uitgekomen nieuwsberichten, dus niet slechts op één bericht zoals daar in paragraaf 4.3.5 sprake van is. Omdat sentiment een periodiek verschijnsel is en het sentiment van 3 weken geleden niets zegt over het sentiment van deze dag, worden alleen events bekeken vanaf 1 week voor het moment van voorspelling. De bepaling van de waarde voor het marktsentiment gebeurt op dezelfde wijze als in paragraaf 4.3.5 is uitgelegd. In alle berichten wordt gekeken naar het aantal positieve en negatieve trefwoorden. Deze aantallen worden bij elkaar opgeteld en het sentiment wordt bepaald volgens formule 1 en 2.

In de volgende stap wordt het effect op de koersen bepaald dat door de ‘algemene berichten’ wordt veroorzaakt. Hier gaat het dus om de categorieën: marktspecifieke events en koersinformatie. Om dit te berekenen worden alle events genomen die op het moment van voorspelling ( $T_0$ ) nog een invloed uitoefenen op de koers, figuur 4 illustreert dit.



(Figuur 4)

Hier worden dus alle events genomen waarvan het starttijdstip vóór en het eindtijdstip na tijdstip T0 ligt. Door van alle events het (deel)effect te berekenen en deze bij elkaar op te tellen, kan een voorspelling van de markt gemaakt worden (voor meer informatie over deze berekening: zie Bijlage 4: Berekening van het effect). Vanaf tijdstip T0 wordt een voorspelling gemaakt voor de komende uren. Hiervoor wordt ieder uur opnieuw berekend wat de som is van de (deel)effecten van de events. Dit totale effect wordt vervolgens vermenigvuldigd met het marktsentiment dat in de vorige paragraaf is uitgelegd.

#### **4.5.2 Effect van bedrijfsspecifieke events**

De voorspelling van een aandeel gebeurt grotendeels volgens hetzelfde principe als de voorspelling van de markt. Er wordt gekeken naar alle events die op het moment van de voorspelling invloed uitoefenen op de koers (zie figuur 4). Er wordt nu echter niet gekeken naar events over de economie in het algemeen, maar alleen naar de bedrijfsspecifieke events. Van deze events is bekend wat het effect is en wat het sentiment is van het betreffende nieuwsbericht. Om het (deel)effect te berekenen worden deze twee waarden met elkaar vermenigvuldigd en opgeteld bij het totaaleffect. De uiteindelijke waarde die hieruit komt na verwerking van alle relevante events is dus het effect dat al de bedrijfsspecifieke events op de koers hebben.

#### **4.5.3 Voorspelling van een aandeel**

Wanneer het effect dat de markt op de koersen heeft en het effect dat het bedrijfsspecifieke nieuws heeft bekend zijn, kan de voorspelling van de koers gemaakt worden. Hiervoor moet eerst de meest recente koers gevonden worden. Een voorspelling bestaat uit 8 waarden, beginnend op tijdstip T0, het moment waarop de voorspelling gemaakt wordt. De recente koers wordt hier ingevuld. De volgende waarde is T0 + 1 uur. Om hier een waarde bij te krijgen worden de onderdelen gebruikt die hiervoor zijn beschreven. Van zowel de marktvoorspelling als de aandeelvoorspelling is voor ieder uur een effect berekend. Dit is het effect van dat tussen tijdstip T0 en T0 + x uren plaats vindt. Wanneer het effect van de markt bijvoorbeeld 0,3 % zou zijn na 3 uur en het effect van het bedrijfsspecifieke nieuws -0,1 %,

wordt de voorspelling van 3 uur na T0 dus  $0,3 - 0,1 = 0,2$  % en is dus de verwachting dat de koers van T0 met 0,2 % gestegen zal zijn na 3 uur. Dit wordt voor de komende 7 uur vanaf tijdstip T0 gedaan. Het geheel van deze waarden vormt dus een voorspelling van de komende 7 uur waarbij gebruik gemaakt wordt van de informatie die op tijdstip T0 bekend is. Omdat er regelmatig nieuwe nieuwsberichten worden uitgebracht, moet ook regelmatig een nieuwe voorspelling gemaakt worden. In het geschreven programma wordt iedere vier uur een voorspelling gemaakt.

#### **4.5.4 Voorspelling van de AEX-index**

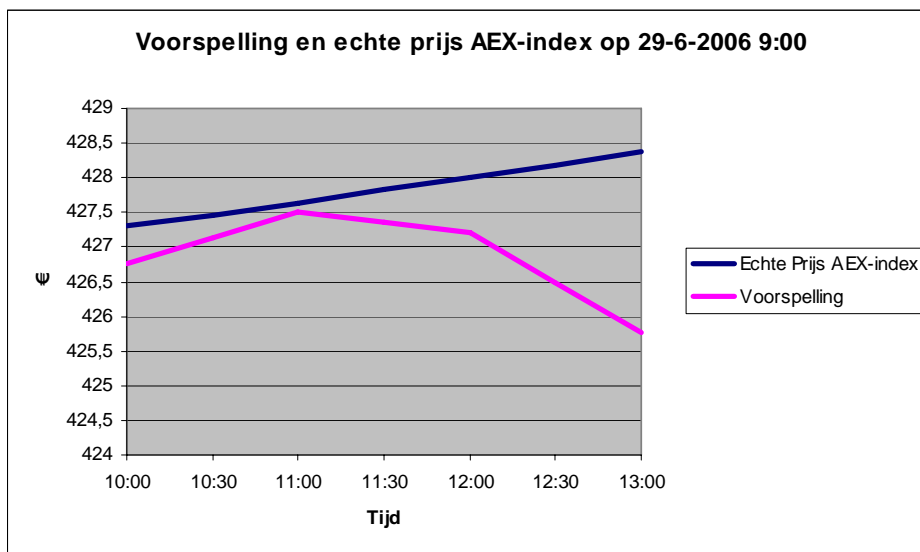
De AEX-index is een index die is gebaseerd op de koersen van een aantal aandelen. In totaal zijn er 24 fondsen. Al deze fondsen hebben een bepaald gewicht binnen de AEX-index. Door dit gewicht van ieder fonds te vermenigvuldigen met de koers, wordt de AEX-index berekend. Wanneer deze gewichten bekend zijn, kan ook een voorspelling gemaakt worden van de AEX-index door de laatste waarden van het gewicht te vermenigvuldigen met de verwachte koers voor de komende uren. Hierdoor kan ook voor de AEX-index ieder uur een voorspelling gemaakt worden voor de komende 7 uren. Een overzicht van de fondsen en gewichten is te vinden in Bijlage 5: De AEX-index.

## 5. Resultaten

Het programma heeft nadat het geschreven was, een aantal weken gedraaid en heeft in die tijd nieuwsberichten verzameld en deze verwerkt in een voorspelling. Ieder uur is een voorspelling gemaakt. Om te zien in hoeverre de voorspellingen kloppen zullen een aantal voorspellingen vergeleken worden met de daadwerkelijke koers.

### 5.1 Voorspelling op donderdag 29 juni 2006

De eerste voorspelling die bekeken zal worden is gemaakt op donderdag 29 juni 2006. Op deze dag is er niet veel bijzonder gebeurd, de AEX-index is gedurende de dag licht gestegen en de nieuwsberichten die uitgekomen zijn, bevatten geen schokkend nieuws. Op basis van deze nieuwsberichten is de voorspelling gemaakt die hieronder grafisch staat weergegeven.



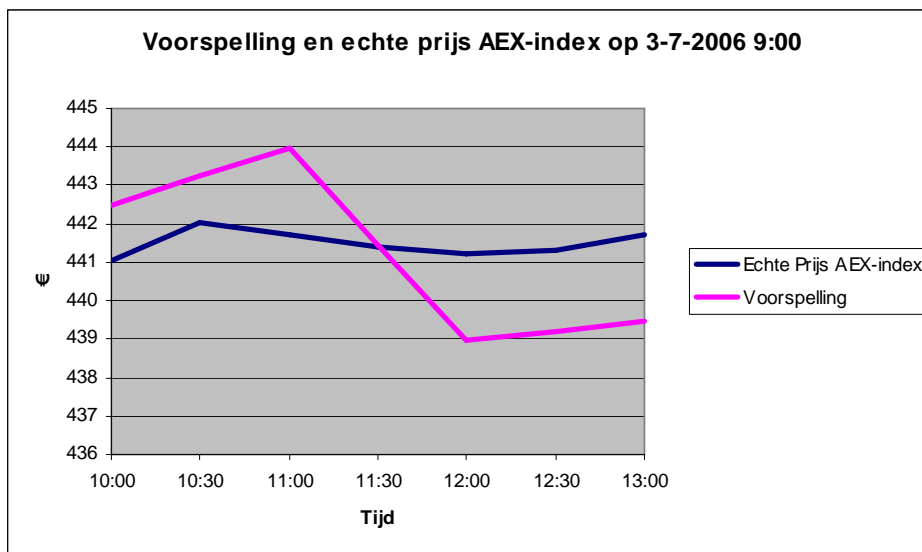
(Figuur 5)

Uit figuur 5 blijkt dat het echte koersverloop de voorspelling volgt, maar dat deze later meer van elkaar gaan afwijken en dat er dus bepaalde factoren zijn die voor een ander verloop zorgen. Dit wordt zoals verwacht veroorzaakt door 'nieuw' nieuws dat uitgekomen is ná het moment waarop de voorspelling is gemaakt. De voorspelling die hier is gemaakt is namelijk

alleen gebaseerd op het nieuws dat om 9:00 bekend was. Daarna is echter nieuws uitgekomen dat niet in de voorspelling zit verwerkt, maar wel in het werkelijke koersverloop.

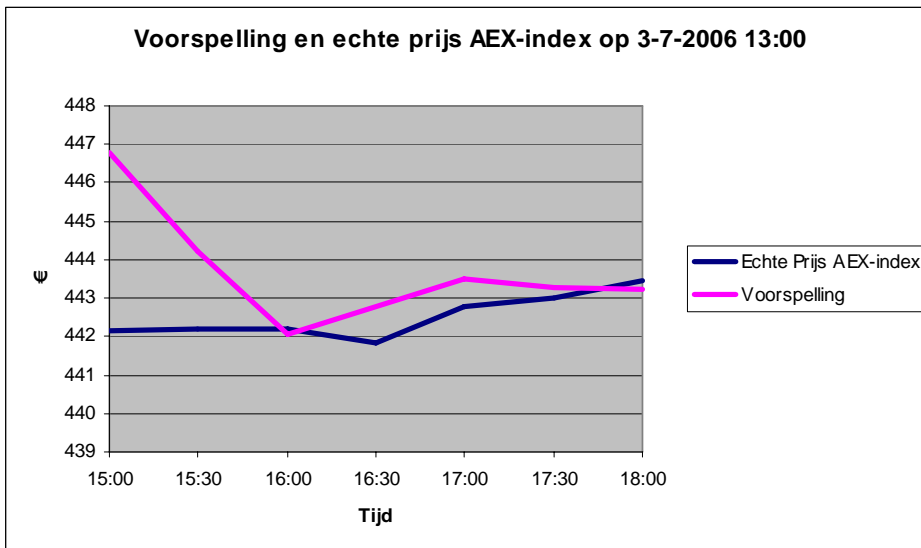
## 5.2 Voorspelling op maandag 3 juli 2006

De tweede voorspelling die bekeken wordt is gemaakt op maandag 3 juli. De voorspelling ziet er grafisch als volgt uit:



(Figuur 6)

Opvallend aan deze voorspelling is dat deze qua schommelingen redelijk goed in de buurt komt van het werkelijke koersverloop. De steilheid van de lijnen aan het begin en aan het eind van de grafiek komen vrijwel overeen met het werkelijke koersverloop. Alleen het stuk hier tussenin vertoont een scherpere daling dan in werkelijkheid het geval was

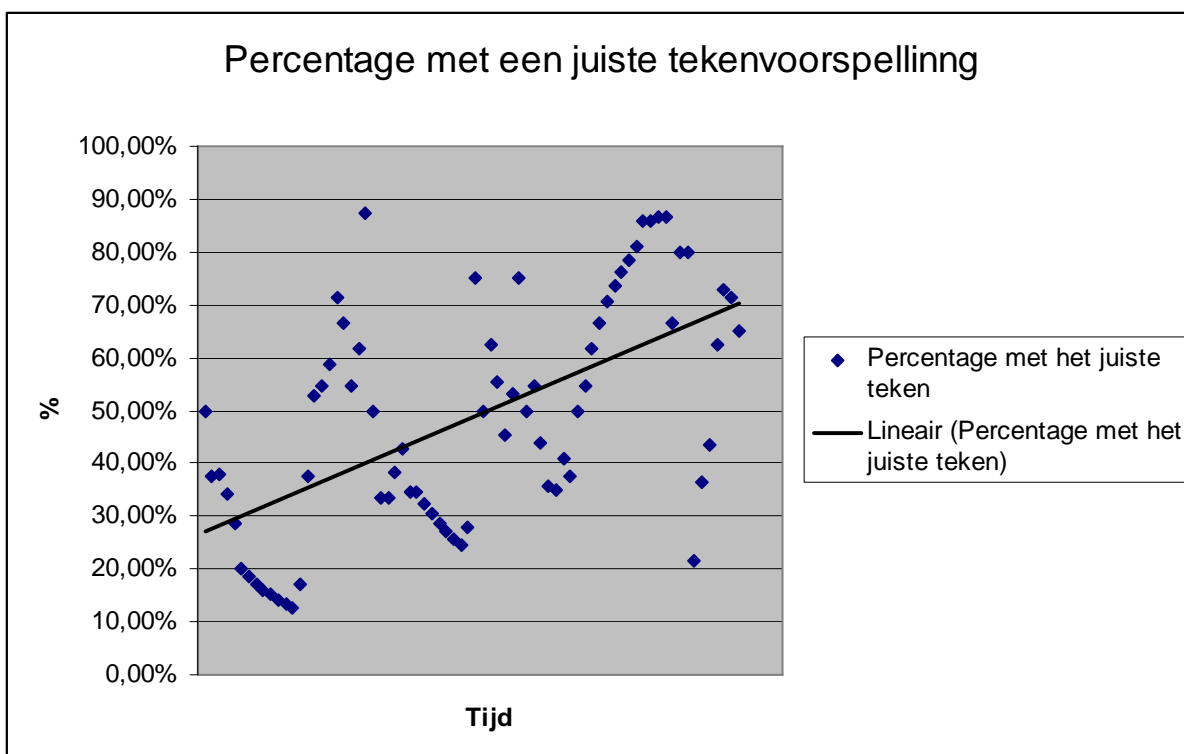


(Figuur 7)

De voorspelling die om 13:00 uur is gemaakt geeft een eigenaardige piek. Er is hier geen specifiek fonds en dus ook geen nieuwsbericht te vinden dat deze piek veroorzaakt. De piek is hier dus veroorzaakt door een combinatie van kleine effecten uit alle fondsen. Dit is een situatie die wel voor kan komen wanneer er meerdere fondsen zijn die een prognose hebben gekregen die een klein beetje te hoog is. Op het moment dat het event is afgelopen en wordt herberekend wat het echte effect was, zal een betere waarde worden verkregen. Deze waarde wordt voor toekomstige voorspellingen gebruikt. Het programma leert dus van zijn fouten. Vanaf 16.00 uur is de groep events die voor de piek zorgt afgelopen en ligt het niveau van de voorspelling dicht bij het werkelijke niveau.

### 5.3 Tekenreeksanalyse

Omdat het voorspellen van de exacte waarden van de koersen vrijwel onmogelijk is, is een tekentoets een meer geschikte maatstaf voor de beoordeling van de resultaten dan een vergelijking van de koersen. Wanneer een tekentoets gebruikt wordt, wordt namelijk alleen gekeken naar het feit of volgens de voorspelling de koersen zouden stijgen of dalen en of dit overeenkomt met wat de werkelijke verandering is geweest. Omdat iedere dag enkele voorspellingen zijn gemaakt die elk uit een aantal waarden bestaan voor de koersen, is ervoor gekozen om per dag te bepalen welk deel van de voorspellingen de juiste richting had voorspeld (stijgen of dalen). Als resultaat komt hier een reeks percentages uit die aangeven welk deel van de voorspellingen juist was. In een grafiek ziet deze reeks er als volgt uit:



(Figuur 8)

Zoals uit figuur 8 blijkt is er een grote spreiding tussen de percentages, maar er zit duidelijk een stijgende lijn in (hier aangegeven door de zwarte lineaire trendlijn). Dit betekent dat het systeem daadwerkelijk leert van het verleden en dus beter wordt in het voorspellen of een koers gaat stijgen of dalen.

## 6. Conclusie en verbeterpunten

Uit de resultaten die hiervoor zijn een aantal conclusies te trekken. Ten eerste is gebleken dat delen van de voorspellingen redelijk dicht bij de juiste liggen of in ieder geval de juiste beweging volgen. De effecten zijn echter over het algemeen nog wat te extreem. Er zijn regelmatig pieken te vinden die verklaard kunnen worden door nieuwsberichten die een te hoog of een te laag effect hebben gekregen. Dit kan een gevolg zijn van een te korte ‘leerperiode’ waarin het programma heeft kunnen leren hoe groot het effect is van bepaalde groepen. Het kan ook veroorzaakt worden door een verkeerde groepsindeling (niet gedetailleerd genoeg). Een mogelijke oplossing hiervoor is om aan een event een bepaalde maximumwaarde toe te kennen. Het nadeel hiervan is echter dat in extreme situaties waar nieuwsberichten wel een effect kunnen hebben dat zo groot is, de prognose juist te gematigd is.

Door van een grotere afstand te kijken en de koersverandering in te delen in de categorieën ‘stijgen’ of ‘dalen’ in plaats van een exacte waarde eraan toe te kennen, is te zien dat het systeem leert van het verleden en steeds beter wordt in het voorspellen van de verandering. De methode werkt dus nog niet voldoende in het voorspellen van de exacte koersen, maar bij de voorspelling of een koers gaat stijgen of dalen worden de resultaten steeds beter.

Om dit onderzoek te verbeteren zijn een aantal gebieden die aandacht nodig hebben. Eén van de belangrijkste verbeterpunten is de indeling van de groepen met events. Tijdens de periode waarin het programma draaide, bleek al snel dat de huidige indeling vaak niet voldoende gedetailleerd is; Er zijn meer mogelijke events nodig dan de negen die er nu zijn.

Ten tweede kan de methode om de nieuwsberichten in een groep in te delen verbeterd worden. Hiervoor kan de woordenlijst met synoniemen uitgebreid worden. Ook kan er een lerende functie bij gemaakt worden die aan de hand van de bevestiging of wijziging van de gebruiker leert welke woorden bij een groep hoort.

Op het gebied van de voorspelling zijn ook enkele dingen te verbeteren. Zo kan de selectie van vergelijkbare berichten verbeterd worden. Op dit moment worden alle berichten uit



dezelfde groep bekeken, waaruit een gemiddelde wordt bepaald. Dit kan verbeterd worden door bijvoorbeeld recente berichten zwaarder mee te laten tellen of door andere vergelijkbare variabelen hiervoor te gebruiken. Een andere mogelijkheid tot uitbreiding is om de bepaling van de groep niet een keuze te laten zijn, maar om er een fuzzy variabele aan mee te geven. Dit houdt in dat een nieuwsbericht tot meerdere groepen kan behoren.

Tot slot heeft de duur van de events nog enige aandacht nodig. Het is waarschijnlijk dat de aanpassingsperiode korter is dan in dit onderzoek vanuit is gegaan. Dit moet nog goed worden nagekeken of het misschien niet een kwestie is van minuten in plaats van uren.

## 7. Discussie

Het is nog maar de vraag of we ooit in staat zullen zijn de koersen precies te kunnen voorspellen. Er gebeuren te veel dingen om alles in de koersen te kunnen verwerken. Door nieuwsberichten te gebruiken worden de belangrijkste gebeurtenissen al uit de grote massa informatie gefilterd. Maar wordt op deze manier niet te veel informatie weggegooid? Zeggen deze koude, objectieve nieuwsberichten genoeg over een wereld die wordt beheerst door menselijk handelen en emoties? Door te proberen het sentiment van het nieuwsbericht te bepalen heb ik geprobeerd iets van die emotie terug te halen en deze in de voorspelling te verwerken.

Om dit uit te breiden zijn meerdere mogelijkheden, maar vooral ‘nieuwe’ technieken als neurale netwerken en genetische algoritmen, oftewel de intelligente lerende systemen, hebben mijn speciale interesse. Door nieuwsberichten hieraan te koppelen en eventueel nog andere bronnen die verklaringen kunnen bieden voor koerswijzigingen, kunnen waarschijnlijk verrassende inzichten verkregen worden en verbanden ontdekt worden waar nog niet eerder aan was gedacht.

De vraag is echter wel hoe groot het nut zou zijn van een perfecte voorspelling. Als er één persoon is die precies weet wat de koers gaat doen, kan deze persoon daar een groot voordeel mee behalen. Op een gegeven moment krijgen anderen echter door dat er een perfecte voorspelling bestaat en gaan ze deze ook gebruiken, maar als iedereen een perfecte voorspelling tot zijn of haar beschikking heeft en iedereen daarom gelijk reageert op nieuws, verdwijnt het voordeel. Het enige effect dat de perfecte voorspelmethode dan heeft is dat het nieuws alleen maar nog sneller (vrijwel direct) in de koers verwerkt wordt.

## Referenties

- [1] [http://www.nyse.com/pdfs/nyse\\_chap\\_04.pdf](http://www.nyse.com/pdfs/nyse_chap_04.pdf)
- [2] Soylomez, A., Kempen, M. van, Wijngaarden, H. van, *Adding time to OWL*, Erasmus Universiteit Rotterdam, 2006  
(te vinden op: [http://home.tiscali.nl/mycall84/Addingtimetoowl\\_team\\_3.pdf](http://home.tiscali.nl/mycall84/Addingtimetoowl_team_3.pdf))
- [3] <http://nl.wikipedia.org/wiki/Nieuws>
- [4] Nuygren, K., *Stock Prediction – A Neural Network Approach*, Royal Institute of Technology (Zweden), Maart 2004  
(te vinden op: [www.f.kth.se/~f98-kny/thesis.pdf](http://www.f.kth.se/~f98-kny/thesis.pdf))
- [5] Gidófalvi, G., *Using News Articles to Predict Stock Price Movements*, Department of Computer Science and Engineering University of California, San Diego, juni 2001  
(te vinden op: <http://citeseer.ist.psu.edu/517027.html>)
- [6] Bunningen, A. van, *Augmented trading: From news articles to stock price predictions using syntactic analysis*, Department of Computer Science, University of Twente, januari 2004  
(te vinden op: <http://citeseer.ist.psu.edu/vanbunningen04augmented.html>)
- [7] <http://www.beurs.nl/rss/nieuws.rss>
- [8] <http://informa.sourceforge.net/>
- [9] [http://www.euronext.com/trader/composition/0,5372,1732\\_6868,00.html?selectedMep=2&idInstrument=15520&isinCode=NL0000000107](http://www.euronext.com/trader/composition/0,5372,1732_6868,00.html?selectedMep=2&idInstrument=15520&isinCode=NL0000000107)

## Bijlage 1: Trefwoorden events

### Wijziging van de winstverwachting:

*kwartaalwinst*  
*kwartaalcijfers*  
*jaarcijfers*  
*resultaten*  
*winst*  
*bedrijfsresultaat*  
*omzet*  
*winstvoorzicht*

### Overname/ fusie/ afsplitsing/ samenwerking:

*overnemen*  
*overname*  
*neemt over*  
*uitbreiding*  
*fusie*  
*afsplits*  
*opdelen*  
*verkoop*  
*boekwaarde*

### Verandering consumentenvertrouwen:

*consumentenvertrouwen*  
*vertrouwen*  
*bestedingen*

### Rentewijziging

*rente*  
*ecb*  
*europese centrale bank*  
*trichet*  
*interest*  
*fed*

### Werkloosheid:

*werkloosheid*  
*werklozen*  
*beroepsbevolking*  
*uitkering*  
*werkzoekend*

### Economie algemeen:

*BBP*  
*bruto binnenlands product*  
*overheidstekort*  
*overheidsoverschot*  
*begroting*  
*inflatie*  
*koopkracht*  
*dagblad*

### Stand van de AEX-index:

*EUR*  
*AEX*  
*aex*  
*beurs*  
*Amsterdam*  
*slot*  
*eindig*  
*damrak*

### Stand van buitenlandse beurzen:

*new york*  
*dow jones*  
*wall street*  
*sluit*  
*eindig*  
*hoofdfondsen*  
*londen*  
*parijs*  
*frankfurt*

## Bijlage 2: Trefwoorden bedrijven

### **ABN AMRO:**

*abn amro holding*  
*abn amro*  
*abn*  
*amro*  
*groenink*  
*rasmussens*  
*saigol*  
*schmittmann*  
*bobins*  
*barbosa*  
*kloosterman*  
*sobti*  
*rutgers*  
*overmars*  
*ter avest*  
*russell*

### **Aegon:**

*aegon*  
*cnv verzekeringen*  
*tkp pensioen*  
*van der werf*  
*e. koning*  
*edgar koning*  
*overmeer*  
*romijnsen*  
*varenkamp*

### **ASML:**

*asml*  
*meurice*  
*wennink*  
*van den brink*  
*fuchs*

### **Ahold:**

*ahold*  
*dahan*  
*hommen*  
*moberg*  
*wakkie*  
*rishton*  
*stop & shop*  
*giant food stores*  
*tops markets*  
*albert heijn*  
*gall & gall*  
*etos*  
*foodservice U.S.*  
*foodservice US*  
*albert*

### **Akzo Nobel:**

*akzo nobel*  
*akzo*  
*pharma*  
*coatings*  
*chemicals*  
*wijers*  
*frohn*  
*wilderbeek*  
*darner*

### **Buhrmann:**

*buhrmann*  
*corporate express*  
*graphic systems*  
*office products*  
*koffrie*  
*george dean*  
*g. dean*  
*hoffman*  
*waller*

**DSM:**

*dsm  
nutrition  
pharma  
performance materials  
industrial chemicals  
unlimited.dsm  
elverding  
zuidam  
sijbesma  
gerardu*

**Fortis:**

*fortis  
lippens  
votron  
verwilt  
mittler  
clijsters  
de boeck  
dierckx  
de mey  
van harten*

**Getronics:**

*getronics  
pink roccade  
pinkroccade  
de meijer  
kleiwegt  
oosterhoff  
van rooijen  
sonja kok  
s. kok  
enneking  
van blokland  
olde olthof*

**Hagemeyer:**

*hagemeyer  
kalff  
eustace  
bourigeaud  
t. de raad  
theo de raad  
a. baan  
adri baan  
van gelder*

**Heineken:**

*heineken  
l'arche  
amstel  
brand  
lingen's blond  
murphy's irish red  
vos  
wieckse  
maarten das  
m. das  
m das  
vuursteen  
hoyer  
van boxmeer  
graafland  
bolland*

**KPN:**

*kpn  
scheepbouwer  
demuyneck  
Smits*

**ING Groep:**

*ing groep*  
*ing group*  
*ing bank*  
*ing verzekering*  
*postbank*  
*nationale nederlanden*  
*fundix*  
*regio bank*  
*regiobank*  
*rvs*  
*westland utrecht*  
*westlandutrecht*  
*ing private banking*  
*tilmant*  
*cees maas*  
*c. maas*  
*c maas*  
*boyer de la giroday*  
*harryvan*  
*leenaars*  
*mcinerney*  
*van der noordaa*  
*de vaucleeroy*

**Numico:**

*numico*  
*nutricia*  
*milupa*  
*dumex*  
*mellin*  
*cow & gate*  
*bennink*  
*huët*  
*britton*  
*mareel*  
*mehra*  
*puri*

**Philips:**

*philips*  
*kleisterlee*  
*dutiné*  
*dutine*  
*huijser*  
*sivignon*

**Reed Elsevier:**

*reed elsevier*  
*elsevier*  
*hommen*  
*tierney*  
*van de aast*  
*davis*  
*armour*  
*enstrom*  
*prozes*  
*redlynch*  
*stomberg*  
*van lede*  
*zelnick*  
*de boer-kruyt*  
*elliott*  
*david reid*  
*d. reid*  
*d reid*

**Rodamco:**

*rodamco*  
*laglas*  
*pragt*  
*van rooijen*  
*dallinga*

**Royal Dutch Shell:**

*shell*  
*nederlandse aardolie maatschappij*  
*van der veer*  
*routs*  
*voser*  
*brinded*  
*cook*

**SBM Offshore:**

*sbm offshore*  
*sbm*  
*ihc caland*  
*miles*  
*blanchelande*  
*keller*  
*van der zee*

**TNT:**

*tnt*  
*tpg*  
*bakker*  
*van dalen*  
*koorstra*  
*kulik*  
*lombard*  
*schalekamp*

**Tomtom:**

*tomtom*  
*navigatiesyste*  
*goddijn*  
*vigreux*  
*pauwels*  
*geelen*  
*ribbink*  
*gretton*  
*wyatt*  
*van de kraats*

**Unilever:**

*unilever*  
*andr lon*  
*andrelon*  
*axe*  
*becel*  
*bertolli*  
*blue band*  
*boursin*  
*calv *  
*calve*  
*cif*  
*conimex*  
*dove*  
*glorix*  
*hertog*  
*iglo*  
*knorr*  
*lipton*  
*ola*  
*omo*  
*rexona*  
*robijn*  
*slim fast*  
*slimfast*  
*sun*  
*sunil*  
*unox*  
*vaseline intensive care*  
*cescau*  
*van der graaf*  
*kugler*  
*markham*  
*burgmans*  
*collomb*  
*singh*  
*manwani*  
*ogg*



**Vedior:**

*vedior*  
*preisig*  
*vervoort*  
*wilkinson*  
*valks*  
*salle*  
*miles*

**Wolters Kluwer:**

*wolters kluwer*  
*wolters*  
*kluwer*  
*skolar*  
*slin-eguide*  
*medi-span*  
*provation*  
*lippincott*  
*ovid*  
*boucher communications*  
*mckinstry*  
*beerkens*  
*detailleur*  
*mccauley*  
*cartwright*  
*becker*  
*grainger*

## Bijlage 3: Trefwoorden sentiment

### Positieve woorden:

*positief*  
*positieve*  
*winnen*  
*winst*  
*won*  
*optimis*  
*stijg*  
*steeg*  
*stegen*  
*hoger*  
*verhoging*  
*verhogen*  
*verhoog*  
*toename*  
*omhoog*  
*hoge*  
*voldoende*  
*overschot*  
*groei*  
*verbeter*  
*toename*  
*herstel*  
*boven verwachting*  
*indrukwekkend*  
*bekend*  
*breidt*  
*uitbreiden*

### Negatieve woorden:

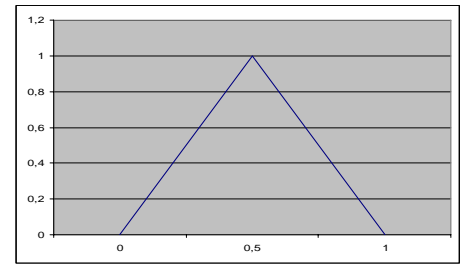
*negatief*  
*negatieve*  
*lager*  
*verlaging*  
*verlagen*  
*verlaag*  
*daling*  
*dalen*  
*daler*  
*daalt*  
*daalde*  
*verloor*  
*verlies*  
*verliezen*  
*verloren*  
*vrees*  
*angst*  
*omlaag*  
*boosdoener*  
*ongerust*  
*tekort*  
*probleem*  
*problemen*  
*verslechter*  
*afname*  
*minder*

## Bijlage 4: Berekening van het effect

- $d_{start}$  Tijd in uren tussen het moment waarop de voorspelling gemaakt wordt (en de koers dus bekend is) en het starttijdstip van het event dat berekend wordt  
 $d_{eind}$  Tijd in uren tussen het moment waarop de voorspelling gemaakt wordt (en de koers dus bekend is) en het eindtijdstip van het event dat berekend wordt  
 $t$  De duur van het event (in uren)  
 $f_{start}$  Waarde van het effect bij aanvang van de periode  
 $f_{eind}$  Waarde van het effect bij beëindiging van de periode  
 $f$  Het effect van het event

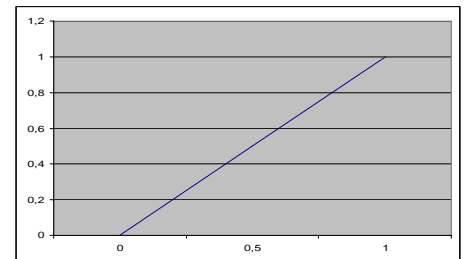
### Type 1 events:

$$\begin{aligned}
 f_{start} &= 0 && \text{als } d_{start}/t \leq 0 \\
 f_{start} &= 2 \cdot f \cdot d_{start}/t && \text{als } 0 \leq d_{start}/t \leq 0,5 \\
 f_{start} &= \text{effect} - 2 \cdot (d_{start}/t - 0,5) \cdot f && \text{als } 0,5 < d_{start}/t \leq 1 \\
 f_{eind} &= 2 \cdot f \cdot d_{eind}/t && \text{als } 0 \leq d_{eind}/t \leq 0,5 \\
 f_{eind} &= \text{effect} - 2 \cdot (d_{eind}/t - 0,5) \cdot f && \text{als } 0,5 < d_{eind}/t \leq 1 \\
 f_{eind} &= 0 && \text{als } d_{eind}/t > 1
 \end{aligned}$$



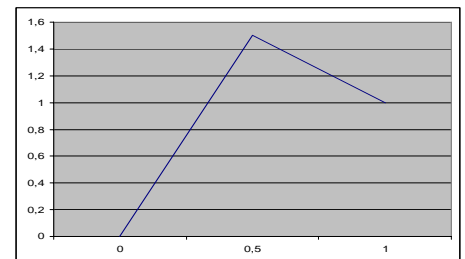
### Type 2 events:

$$\begin{aligned}
 f_{start} &= 0 && \text{als } d_{start}/t \leq 0 \\
 f_{start} &= d_{start}/t \cdot f && \text{als } 0 \leq d_{start}/t \leq 1 \\
 f_{eind} &= d_{eind}/t \cdot f && \text{als } 0 \leq d_{eind}/t \leq 1 \\
 f_{eind} &= f && \text{als } d_{eind}/t > 1
 \end{aligned}$$



### Type 3 events:

$$\begin{aligned}
 f_{start} &= 0 && \text{als } d_{start}/t \leq 0 \\
 f_{start} &= 2 \cdot 1,5 \cdot f \cdot d_{start}/t && \text{als } 0 \leq d_{start}/t \leq 0,5 \\
 f_{start} &= 1,5 \cdot f - (d_{start}/t - 0,5) \cdot f && \text{als } 0,5 < d_{start}/t \leq 1 \\
 f_{eind} &= 2 \cdot 1,5 \cdot f \cdot d_{eind}/t && \text{als } 0 \leq d_{eind}/t \leq 0,5 \\
 f_{eind} &= 1,5 \cdot \text{effect} - (d_{eind}/t - 0,5) \cdot f && \text{als } 0,5 < d_{eind}/t \leq 1 \\
 f_{eind} &= f && \text{als } d_{eind}/t > 1
 \end{aligned}$$



Het (deel)effect van het event in de periode tussen het moment waarop de voorspelling gemaakt wordt en het starttijdstip van het event dat berekend wordt, is gelijk aan  $f_{eind} - f_{start}$

## Bijlage 5: De AEX-index

Fondsen uit de AEX-index:

<b>Fonds</b>	<b>Gewicht</b>
ABN AMRO HOLDING	2.07
AEGON	1.77
AHOLD KON	1.68
AKZO NOBEL	0.31
AwSML HOLDING	0.53
BUHRMANN	0.20
DSM KON	0.22
FORTIS	1.43
GETRONICS	0.13
HAGEMEYER	0.56
HEINEKEN	0.27
ING GROEP	2.25
KPN KON	2.44
NUMICO	0.20
PHILIPS KON	1.46
REED ELSEVIER	0.79
RODAMCO EUROPE	0.07
ROYAL DUTCH SHELL	2.71
SBM OFFSHORE	0.15
TNT	0.48
TOMTOM	0.06
UNILEVER	1.86
VEDIOR	0.18
WOLTERS KLUWER	0.34