UHASSELT

KNOWLEDGE IN ACTION

Faculteit Bedrijfseconomische Wetenschappen

master handelsingenieur in de beleidsinformatica

Masterthesis

Random Forest voor Aandelenvoorspellingen op Basis van Fundamentele Analyse

Yarne Steegmans

Scriptie ingediend tot het behalen van de graad van master handelsingenieur in de beleidsinformatica

PROMOTOR :

Prof. dr. Sigrid VANDEMAELE

BEGELEIDER :

Mevrouw Kelsey SCHEPERS

UHASSELT

KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

2024
2025

UHASSELT

KNOWLEDGE IN ACTION

Faculteit Bedrijfseconomische Wetenschappen

master handelsingenieur in de beleidsinformatica

Masterthesis

Random Forest voor Aandelenvoorspellingen op Basis van Fundamentele Analyse

Yarne Steegmans

Scriptie ingediend tot het behalen van de graad van master handelsingenieur in de beleidsinformatica

PROMOTOR :

Prof. dr. Sigrid VANDEMAELE

BEGELEIDER :

Mevrouw Kelsey SCHEPERS

Random Forest voor Aandelenvoorspellingen op Basis van Fundamentele Analyse

Yarne Steegmans

"Universiteit Hasselt, Faculteit Bedrijfseconomische wetenschappen, Diepenbeek, 3590, Limburg, België"

Abstract De opkomst van artificiële intelligentie (AI) en machine learning (ML) biedt nieuwe mogelijkheden om aandelenprijzen te voorspellen via datagedreven analyses. Er is veel onderzoek gedaan op dit gebied en uit verschillende studies blijkt dat machine learning effectief kan worden toegepast voor het voorspellen van aandelenprijzen. De meeste van deze methoden richten zich echter op voorspellingen op de korte termijn, waarbij vooral gebruik wordt gemaakt van prijsgegevens en technische indicatoren. In deze masterthesis wordt daarentegen onderzocht in hoeverre fundamentele bedrijfsgegevens en macro-economische factoren bijdragen aan het voorspellen van aandelenprijzen voor bedrijven binnen de NASDAQ-100 index op lange termijn. Centraal in dit onderzoek staat het gebruik van Random Forest, een ML-algoritme dat geschikt is voor het verwerken van complexe datasets met variabelen van uiteenlopend belang.

Het onderzoek bestaat uit 4 verschillende methoden om aandelenprijzen te voorspellen. De prestaties van elke methode worden beoordeeld aan de hand van verschillende metrieken, waaronder de Gemiddelde Absolute Fout (MAE), de Gemiddelde Kwadratische Fout (MSE), de Wortel van de Gemiddelde Kwadratische Fout (RMSE), de Wortel van de Gemiddelde Logaritmische Kwadratische Fout (RMSLE) en de Gemiddelde Absolute Percentagefout (MAPE), de determinatiecoëfficiënt (R^2) en de aangepaste determinatiecoëfficiënt (Adjusted R^2). Daarnaast wordt bij elke methode onderzocht welke inputvariabelen het meest bijdragen aan de voorspellende kracht van het model, aan de hand van een analyse van de *feature importance*.

De resultaten tonen aan dat de prestaties sterk variëren naargelang de gekozen methode. Bij willekeurige splitsing van de data, zowel op bedrijfsniveau als in geaggregeerde vorm, worden hoge nauwkeurigheidsscores behaald met R^2-waarden tot 0,987 en lage fouten. Echter, wanneer het model wordt geëvalueerd in meer realistische contexten, zoals bij een *leave-one-ticker-out* of een *rolling window*-benadering met chronologische training, neemt de voorspellende nauwkeurigheid af. In het meest opvallende scenario worden zelfs negatieve R^2-waarden gemeten, wat wijst op een beperkte generaliseerbaarheid van het model wanneer het wordt getest op onbekende datasets.

Uit de analyse van de *feature importance* blijkt daarnaast dat zowel fundamentele bedrijfsvariabelen als macro-economische factoren invloed uitoefenen op de aandelenrendementen, waarbij factoren zoals activa, reële bbp en werkloosheidsgraad consistent als belangrijk naar voren komen. Daarentegen blijken ROA, FCF en CPI een minder belangrijke rol te spelen in de voorspellende modellen.

Keywords: Stock Price Prediction · Random Forest · Artificial Intelligence · Machine Learning · Fundamental Analysis · Financial Forecasting

1 Introductie

De opkomst van artificiële intelligentie (AI) heeft geleid tot nieuwe tools in financiële voorspellende analyses (Katterbauer & Moschetta, 2021). Het voorspellen van aandelenprijzen is een uitdagende taak, met name vanwege de complexiteit van financiële markten (Pham et al., 2024; Lin & Lobo Marques, 2024). Veel financiële experts hebben geprobeerd het probleem van het voorspellen van de opwaartse en neerwaartse bewegingen van de beurs aan te pakken, maar hebben slechts beperkt succes gehad (Balasubramanian et al., 2024). Het is nu haalbaarder dan ooit om dat te doen met de snelle vooruitgang van technologie in rekenvermogen, opslagcapaciteit en nauwkeurigheid van algoritmen (Fathali et al., 2022; Balasubramanian et al., 2024).

Machine learning (ML), een subset van AI, is een waardevolle toevoeging aan traditionele voorspellingsmethoden omdat het in staat is om patronen en verbanden te ontdekken in grote en complexe datasets die voor menselijke analisten of klassieke statistische modellen moeilijk te identificeren zijn (Makridakis et al., 2018). ML-modellen, en in het bijzonder algoritmen zoals Random Forest (RF), kunnen omgaan met hoge complexiteit en niet-lineaire relaties tussen variabelen zonder dat vooraf expliciete modelstructuren nodig zijn (Bayani et al., 2024; Vijh et al., 2020). Door gebruik te maken van historische data kunnen deze algoritmen niet alleen nauwkeurige voorspellingen genereren, maar ook inzicht geven in welke variabelen het meest bijdragen aan deze voorspellingen (Chen et al., 2020).

Ondanks de toename van technieken voor het voorspellen van de aandelenmarkt, geloven sommige analisten dat de aandelenmarkt onvoorspelbaar is (Fathali et al., 2022). Dat leidt tot twee bekende theorieën, de efficiënte markthypothese (EMH) en de random-walk-hypothese (RWH). EMH stelt dat een aandelenprijs op elk moment alle bekende marktinformatie bevat. Omdat analisten optimaal gebruikmaken van alle bekende informatie, zijn prijsschommelingen onvoorspelbaar, omdat nieuwe informatie willekeurig ontstaat (Chopra & Sharma, 2021). Terwijl volgens de random walk-theorie aandelenprijzen een 'random walk' uitvoeren, wat betekent dat alle toekomstige prijzen geen patronen volgen, leidt dit ertoe dat een analist de markt onmogelijk zou kunnen voorspellen (Fathali et al., 2022; Lin & Lobo Marques, 2024). De geldigheid van de EMH- en RW-theorie is controversieel (Chopra & Sharma, 2021). Talloze onderzoekers hebben de onderliggende aannames van de EMH en de RWH weerlegd en aangetoond dat de markt gedeeltelijk kan worden voorspeld; marktinefficiënties kunnen ontstaan door factoren zoals marktpsychologie, transactiekosten en informatie-asymmetrieën (Lin & Lobo Marques, 2024; Chopra & Sharma, 2021; Kumar et al., 2024). Dit hangt samen met de drie niveaus van marktefficiëntie binnen de efficiënte markthypothese: zwakke, semi-efficiënte en sterke markten. In zwakke tot semi-efficiënte markten kunnen hulpmiddelen zoals fundamentele en technische analyse effectief zijn voor het voorspellen van aandelenprijzen (Colin-Jaeger & Delcey, 2020; Vuong et al., 2024). Fathali et al. (2022) stelt dat door kennis van de historische aandelengegevens en fundamentele of technische analyse van een bedrijf, er een betrouwbare voorspelling van de toekomstige aandelenprijs van het bedrijf kan worden gemaakt.

Fundamentele analyse is een methode om de waarde van een aandeel of bedrijf te bepalen aan de hand van financiële bedrijfsgegevens en macro-economische factoren (Balasubramanian et al., 2024; Huang et al., 2022). Deze methode is voornamelijk geschikt voor het voorspellen van de lange termijn (Fathali et al., 2022).

Technische analyse daarentegen bestudeert historische prijzen, grafiekpatronen, trends en indicatoren (Fathali et al., 2022; Huang et al., 2022). Technische analyse wordt, in tegenstelling tot fundamentele analyse, eerder voor kortetermijnvoorspellingen gebruikt (Balasubramanian et al., 2024; Fathali et al., 2022; Huang et al., 2022).

In deze thesis wordt Random Forest toegepast om langetermijnaandelenprijsvoorspellingen te maken op basis van fundamentele analyse om de voorspellende kracht van AI in financiële markten te onderzoeken. Het doel is om inzicht te verkrijgen in de prestaties van het model en de accuraatheid ervan te beoordelen. Dit onderzoek is ontstaan door de tekortkomingen uit de bestaande literatuur, zoals aangetoond door Chopra & Sharma (2021), die stellen dat huidig onderzoek voornamelijk gericht is op korte termijn voorspellingen en dat er een gebrek is aan studies naar de toepassing van AI bij langetermijninvesteringsbeslissingen. Daarnaast wijzen Xiao & Ke (2021) erop dat bestaande onderzoeken vaak beperkt blijven tot theoretische kaders en zelden een concrete implementatie in de praktijk bevatten. Bovendien merkt Lin & Lobo Marques (2024) op dat fundamentele analyse onderbelicht blijft in de literatuur, mede door de complexiteit van het ontwikkelen van modellen die de beweegredenen achter aandelenprijzen kunnen verklaren. Dit onderzoek beoogt deze tekortkomingen op te vullen door een Random Forest-analyse toe te passen, waarmee de lange termijn prestaties van aandelen kunnen worden voorspeld op basis van fundamentele analyse. Tijdens het onderzoek zal een antwoord worden gegeven op onderstaande onderzoeksvragen:

- Hoe accuraat kan een Random Forest-model aandelenprijzen voorspellen op basis van fundamentele analyse?
- Welke fundamentele bedrijfsvariabelen en macro-economische factoren hebben de grootste impact op de aandelenprijzen van bedrijven?

De structuur van deze paper is als volgt opgebouwd. In sectie 1 wordt gestart met een introductie waarin het onderzoeksprobleem wordt geschetst. Daarbij wordt ook het financiële denkkader besproken. In sectie 2 wordt de bestaande literatuur onder de loep genomen en wordt verklaard waarom Random Forest een geschikt model is voor dit onderzoek. Vervolgens behandelt sectie 3 het theoretisch kader. Hierin worden de kernconcepten en onderliggende theorieën toegelicht. Er wordt eerst ingegaan op artificiële intelligentie en de toepassing daarvan binnen financiële markten, waarbij specifiek wordt stilgestaan bij machine learning. Aansluitend wordt het voorspellingsmodel Random Forest besproken, inclusief een uitleg van de werking en een overzicht van de voor- en nadelen binnen financiële toepassingen. Sectie 4 beschrijft de gehanteerde onderzoeksaanpak voor zowel de literatuurstudie als het dataonderzoek, inclusief een toelichting van de gebruikte dataset. Daarnaast worden ook de methoden waarmee de prestaties van voorspellingsmodellen worden geëvalueerd, aan de hand van verschillende prestatie-indicatoren en de *feature importance* besproken. In sectie 5 worden vervolgens de onderzoeksresultaten besproken. Hier worden vier benaderingen besproken, waarbij telkens een Random Forest-analyse op de dataset wordt toegepast. Daarbij wordt voor iedere methode de accuraatheid van het model beoordeeld aan de hand van de prestatiemaatstaven, en worden de belangrijkste *features* per model besproken. In sectie 6 worden de belangrijkste tekortkomingen van het onderzoek besproken, evenals suggesties voor toekomstig onderzoek. De paper sluit af met een conclusie van de belangrijkste bevindingen in sectie 7.

2 Gerelateerde Literatuur

Uit de verkennende analyse van de bestaande literatuur over aandelenprijsvoorspellingen met machine learning blijkt dat Random Forest herhaaldelijk naar voren komt als het meest nauwkeurige model (Sadorsky, 2021; Nti et al., 2020; Subasi et al., 2021; Basak et al., 2019; Khan et al., 2022; Lohrmann & Luukka, 2019; Akhter & Raza, 2024; Vijh et al., 2020; Yuan et al., 2020; Purnomo et al., 2024; Sicheng, 2024; Sjahrunnisa et al., 2024; Zhong & Hitchcock, 2021; Huang et al., 2022). Bijgevolg zal Random Forest binnen dit onderzoek worden gebruikt als voorspellingsmodel. Bovendien geeft Sinha & Jain (2016) vier voordelen aan voor het gebruik van Random Forest bij de analyse van aandelenmarkten. Ten eerste levert Random Forest een hoge voorspellende nauwkeurigheid op door de uitkomsten van meerdere beslissingsbomen te combineren, waarbij elke boom is getraind op een andere subset van de data. Ten tweede is deze methode robuuster tegen overfitting dan individuele beslissingsbomen, wat resulteert in betrouwbaardere voorspellingen op onbekende datasets. Ten derde biedt Random Forest inzicht in de relatieve belangrijkheid van verschillende kenmerken (*features*), wat waardevolle informatie verschaft over de factoren die aandelenbewegingen beïnvloeden. Tot slot is Random Forest goed in staat om complexe niet-lineaire relaties tussen variabelen te modelleren, wat een belangrijk voordeel vormt ten opzichte van traditionele lineaire modellen (Sinha & Jain, 2016). In deze sectie wordt dieper ingegaan op bestaande studies die Random Forest toepassen voor het voorspellen van aandelenprijzen.

Het grootste deel van de bestaande literatuur richt zich voornamelijk op het gebruik van technische indicatoren als inputvariabelen voor Random Forest-modellen. Zo tonen Sadorsky (2021) aan dat technische data effectief kunnen worden ingezet voor het voorspellen van koersrichtingen, waarbij Random Forest betere prestaties levert dan logitmodellen. Eveneens wijzen Basak et al. (2019) en Vijh et al. (2020) op de betere nauwkeurigheid van Random Forest in vergelijking met andere machine learning-technieken zoals SVM en ANN, gemeten aan de hand van foutmaten zoals RMSE en MAE. Deze studies zijn echter voornamelijk gericht op kortetermijnvoorspellingen, doorgaans op dag- of weekbasis (Sadorsky, 2021; Basak et al., 2019; Vijh et al., 2020).

Een alternatieve benadering is te vinden bij studies zoals Akhter & Raza (2024) en Sjahrunnisa et al. (2024), waarin wordt geëxperimenteerd met de toevoeging van alternatieve inputvariabelen. Akhter & Raza (2024) maakt bijvoorbeeld gebruik van een sentimentanalyse om de voorspellende kracht van Random Forest te testen, terwijl Sjahrunnisa et al. (2024) laat zien dat macro-economische variabelen zoals rente en inflatie kunnen bijdragen aan verbeterde prestaties.

Bovendien richten de meeste bestaande studies zich op individuele aandelen of op regionale markten, zoals blijkt uit het werk van Akhter & Raza (2024) op de Pakistaanse KSE-100 en Sjahrunnisa et al. (2024) op de Indonesische markt. Akhter & Raza (2024) richt zich op de Pakistaanse KSE-100-index en gebruikt een Random Forest-classificatiemodel gebaseerd op technische indicatoren, waarbij een sliding window-aanpak en hyperparameteroptimalisatie leiden tot een precisie van 58%. Daarentegen onderzoekt Sjahrunnisa et al. (2024) de Indonesische markt en benadrukt het belang van externe macro-economische variabelen naast historische koersdata. Door factoren zoals rentetarieven en inflatie mee te nemen, verbeteren hun voorspellingsmodellen op maatstaven als MSE, MAPE en R^2 (Sjahrunnisa et al., 2024).

Daarnaast zijn er ook enkele studies die Random Forest gebruiken voor aandelenprijzen te voorspellen op basis van fundamentele analyse. Zo bestudeert Huang et al. (2022) de inzet van verschillende machine learning-algoritmen, waaronder Feed-forward Neural Network (FNN), Random Forest (RF) en Adaptive Neural Fuzzy Inference System (ANFIS), om aandelenrendementen op lange termijn te voorspellen op basis van fundamentele financiële data. Daarbij wordt gebruik gemaakt van een dataset met 22 jaar aan kwartaaldata (1996-2017) van 70 aandelen uit de S&P 100-index. De inputvariabelen bestaan uit 21 fundamentele financiële kenmerken, waaronder financiële ratio's, kasstromen, marges en groeicijfers. Het relatieve kwartaalrendement ten opzichte van de Dow Jones Industrial Average (DJIA) dient als doelvariabele. Uit de resultaten blijkt dat alle drie de modellen in staat zijn om winstgevende portefeuilles samen te stellen, waarbij het RF-model de hoogste prestaties leverde, met een gemiddeld kwartaalrendement van 1,63% boven de DJIA. De studie toont aan dat machine learning in staat is om effectieve langetermijnbeleggingsstrategieën te ontwikkelen (Huang et al., 2022).

In een vergelijkbare lijn onderzoekt Zhong & Hitchcock (2021) hoe verschillende datatypes, waaronder fundamentele ratio's, technische indicatoren en tekstuele sentimentdata uit nieuwsartikelen, gecombineerd kunnen worden binnen verschillende machine learning-modellen voor het voorspellen van aandelen- en indexrendementen binnen de S&P 500. De dataset omvat 20 jaar aan data (2000-2019) van 518 bedrijven, waarbij het log rendement voor de volgende week (t+1) als doelvariabele wordt gebruikt. Random Forest werd, naast andere modellen zoals ARIMA, FFNN en LSTM, ingezet en presteerde beter dan de traditionele statistische modellen, met een accuraatheid van 60,35%. Deze resultaten bevestigen de toegevoegde waarde van Random Forest in het gebruiken van complexe, niet-lineaire interacties tussen variabelen. Bovendien toont de studie aan dat het combineren van verschillende datatypes en het gebruik van ensemblemethoden leidt tot nauwkeurigere voorspellingen dan enkelvoudige modellen (Zhong & Hitchcock, 2021).

Tot slot bestaat er een studie die zich richt op de NASDAQ-100, maar daarbij een andere methode han-

teert dan het onderzoek dat binnen deze thesis zal worden uitgevoerd. De studie van Chopra & Sharma (2021) vergelijkt zeven machine learning-algoritmes, waaronder Random Forest, op basis van hun prestaties bij het voorspellen van aandelenkoersen voor vier belangrijke indexen: NASDAQ, NYSE, NIKKEI en FTSE. De gebruikte inputvariabelen zijn historische marktgegevens die per dag zijn verzameld via Yahoo Finance en bestaan uit: openingsprijs, slotkoers, hoogste en laagste prijs van de dag, aangepaste slotkoers en handelsvolume. De resultaten tonen aan dat Random Forest, met name wanneer getraind op zogenoemde gelekte data, tot de best presterende modellen behoort met nauwkeurigheden tot 93% (Chopra & Sharma, 2021).

Hoewel Random Forest op grote schaal is toegepast in aandelenkoersvoorspelling, gebeurt dit in de bestaande literatuur vooral op basis van technische indicatoren en ligt de focus daarbij voornamelijk op kortetermijnvoorspellingen. Het gebruik van fundamentele analyse voor directe prijsvoorspelling blijft onderbelicht in de literatuur, wat dit onderzoek positioneert binnen een relatief onontgonnen gebied. Bovendien richt geen van de bestaande studies zich specifiek op de voorspelling van NASDAQ-100 bedrijven met Random Forest op basis van fundamentele analyse.

3 Theoretisch Kader

Deze sectie behandelt het theoretisch kader van belangrijke concepten die tijdens het onderzoek aan bod zullen komen. Eerst wordt het concept Artificiële intelligentie besproken, met een focus op machine learning. Vervolgens wordt Random Forest als voorspellingsmodel behandeld.

3.1 AI en Voorspellende Modellen in Financiële Markten

Artificiële intelligentie (AI) kan worden gedefinieerd als het opnemen van menselijke intelligentie in machines, waarbij elk computerprogramma met elementen van menselijke intelligentie wordt beschouwd als AI (Lin & Lobo Marques, 2024).

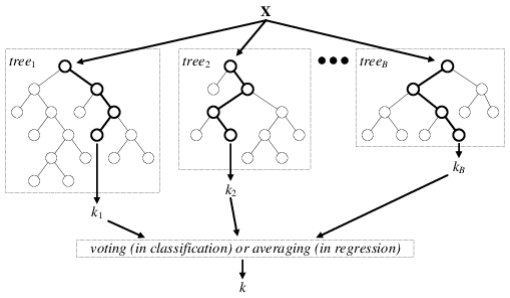
3.1.1 Machine Learning

Machine learning (ML) is een subset van artificiële intelligentie (AI) en omvat technieken waarmee computers kunnen leren van gegevens zonder expliciet te worden geprogrammeerd (Balasubramanian et al., 2024). Het doel van ML is om computers te trainen met behulp van beschikbare gegevens, zodat machines leren hoe ze beslissingen moeten nemen op basis van verwerkte informatie (Lin & Lobo Marques, 2024).

Machine learning-taken worden geclassificeerd als *supervised learning* en *unsupervised learning*. Bij *supervised learning* worden trainingsgegevens gebruikt waarin voorbeelden voorzien zijn, dit betekent dat de dataset bestaat uit invoergegevens en bijbehorende correcte uitvoerwaarden. Het model leert een patroon te herkennen in de gegevens en kan vervolgens voorspellingen doen op nieuwe, onbekende data. Bij *unsupervised learning* daarentegen worden gegevens gebruikt die geen vooraf bepaalde labels of categorieën bevatten. Het doel van *unsupervised learning* is patronen, structuren of relaties binnen de gegevens te ontdekken zonder expliciete menselijke begeleiding (Fathali et al., 2022).

3.1.2 Random Forest

Binnen deze thesis wordt gebruikgemaakt van Random Forest (RF), een *supervised* machine learning-algoritme dat inzetbaar is voor zowel classificatie- als regressietaken. RF behoort tot de *ensemble* learning-methoden, waarbij meerdere beslisbomen worden gecombineerd om robuuste voorspellingen te genereren (Sadorsky, 2021; Sonkavde et al., 2023). Het algoritme bouwt een reeks bomen op basis van willekeurige subsets van de trainingsdata zoals weergegeven in figuur 1. Elke boom wordt afzonderlijk getraind, waardoor overfitting op specifieke patronen wordt beperkt. Door de uitkomsten te aggregeren, bij regressie via het gemiddelde van alle aandelenvoorspellingen, ontstaat een nauwkeurigere schatting dan bij één enkele boom (Dube & Verster, 2024; Karabadji et al., 2023).



Figuur 1. Random Forest Model

Random Forest wordt als geschikt beschouwd voor financiële voorspellingen vanwege zijn vermogen om niet-lineaire relaties en complexe interacties tussen variabelen te detecteren met hoge voorspellende nauwkeurigheid (Park et al., 2021; Sinha & Jain, 2016). Een ander voordeel van Random Forest in financiële toepassingen is de mogelijkheid om inzicht te verkrijgen in de belangrijkste factoren die bijdragen aan een voorspelling (Sinha & Jain, 2016). Door middel van *feature importanc*e-analyse kan worden bepaald welke variabelen de grootste invloed hebben op de aandelenprijs (Sadorsky, 2021; Chen et al., 2020). Hoewel Random Forest goed presteert bij financiële data-analyse, heeft het ook beperkingen. Het model is rekenintensief bij grote datasets en moeilijker te interpreteren dan traditionele regressiemodellen, waarin coëfficiënten duidelijker inzicht geven in verbanden tussen variabelen (Cao, 2024; Zhao et al., 2019).

4 Methodologie

De kern van deze thesis is een empirisch kwantitatief data-onderzoek waarin de relatie tussen bedrijfsfundamentals en macro-economische variabelen wordt onderzocht om de aandelenprijzen van bedrijven van de NASDAQ-100 te voorspellen met Random Forest. Om dit succesvol te realiseren, wordt eerst een literatuurstudie uitgevoerd.

4.1 Literatuurstudie

Voor het verzamelen van wetenschappelijke literatuur wordt gebruikgemaakt van de *UHasselt Discovery database*, die via de universiteitsbibliotheek van UHasselt toegang biedt tot een brede waaier aan databanken zoals *ProQuest, ScienceDirect* en *Emerald Insight*. Deze centrale catalogus maakt het mogelijk om efficiënt relevante wetenschappelijke bronnen te vinden. Aanvullend wordt ook consensus gebruikt om bronnen te zoeken. Dit is een AI-gedreven academische zoekmachine die wetenschappelijk onderzoek uit verschillende databases kan zoeken. Daarnaast wordt er ook *citation chaining* toegepast, zowel in de vorm van *backward searching* (via de referentielijsten van relevante bronnen) als *forward searching* (door te kijken welke andere studies de hoofdbron citeren). Aangezien het onderzoek zich richt op een technologisch onderwerp, worden uitsluitend Engelstalige zoektermen gebruikt. De gebruikte zoektermen worden weergegeven in tabel 1. Bovendien worden de bronnen gescreend op basis van vooropgestelde inclusie- en exclusiecriteria (tabel 2). Er worden enkel bronnen gebruikt die niet ouder zijn dan 10 jaar, omdat technologie snel evolueert en oudere bronnen soms niet langer relevant zijn.

Kernwoorden met betrekking tot architectuur: (ai) OR (artificial intelligence) OR (ml) OR (machine learning) OR (rf) OR (Random Forest)
Kernwoorden met betrekking tot aandelenvoorspelling: (stock prediction) OR (stock price prediction) OR (stock price forecasting) OR (stock forecasting) OR (stock forecast) OR (financial forecast) OR (stock price forecast)
Kernwoorden met betrekking tot fundamenten en macro-economische factoren: (fundamentals) OR (fundamental analysis) OR (macro economics)

Tabel 1. Zoektermen

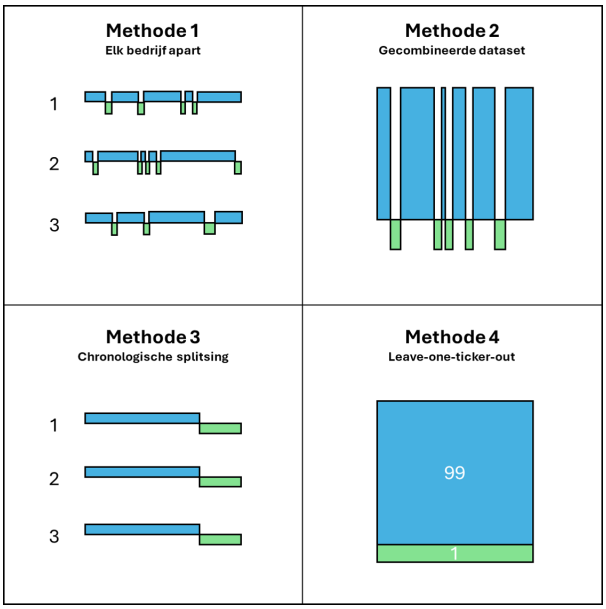
Inclusiecriteria	**Exclusiecriteria**
IN1: Er is online toegang tot de volledige tekst van het artikel	EX1: Het artikel is ouder dan 10 jaar
IN2: Het is een wetenschappelijk artikel	EX2: Het artikel is niet Engelstalig

Tabel 2. Inclusie- en exclusiecriteria

4.2 Dataonderzoek

De kern van deze thesis is een kwantitatieve data-analyse, op zoek naar de relatie tussen de fundamenten van bedrijven en macro-economische factoren, en de prestaties van hun aandelen op lange termijn. Het doel is om te achterhalen of Random Forest een betrouwbare aandelenvoorspelling kan maken. Voor de uitvoering van dit onderzoek wordt gebruikgemaakt van R Studio met de programmeertaal R, die geschikt is voor statistische analyses.

Tijdens het onderzoek worden vier verschillende methoden uitgevoerd en vergeleken om aandelenprijzen te voorspellen met behulp van Random Forest. De eerste methode bekijkt elk bedrijf apart. De data van een bedrijf wordt willekeurig opgedeeld in 80 procent voor training en 20 procent voor testen. Bij de tweede methode worden alle gegevens van alle 100 bedrijven binnen de NASDAQ samengevoegd tot één grote dataset. Ook hier wordt een willekeurige verdeling van 80 procent training en 20 procent testen toegepast voordat de Random Forest-analyse wordt uitgevoerd. De derde methode lijkt op de eerste, maar in plaats van een willekeurige verdeling wordt de data per bedrijf chronologisch gesplitst. De oudste gegevens worden gebruikt om het model te trainen en de meest recente gegevens om het model te testen. Tot slot bouwt de vierde methode voort op de tweede. Hierbij wordt niet willekeurig gesplitst, maar wordt telkens één bedrijf buiten de training gelaten. Dit bedrijf wordt dan gebruikt om te testen, terwijl de gegevens van de andere bedrijven worden gebruikt om het model te trainen. Deze aanpak staat bekend als de *leave-one-ticker-out* methode. In figuur 2 wordt een visueel overzicht van iedere methode weergegeven. Daarbij is blauw telkens de traindata en groen telkens de testdata.

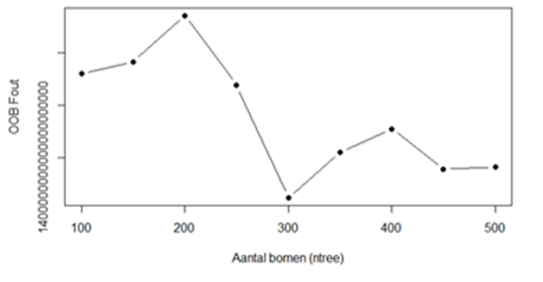


Figuur 2. Verschillende methoden

4.2.1 Hyperparameteroptimalisatie

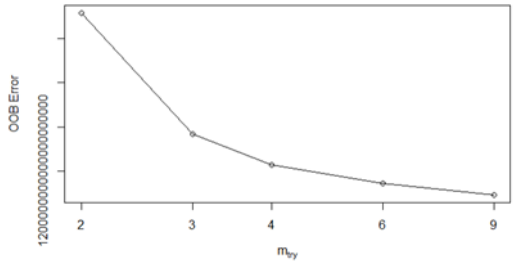
Daarnaast wordt in alle modellen aandacht besteed aan hyperparameteroptimalisatie, met als doel de accuraatheid van de modellen te verbeteren. Bij model 1 en model 3 gebeurt deze optimalisatie per bedrijf afzonderlijk. Voor elk bedrijf worden tijdens het trainingsproces de optimale waarden bepaald voor twee belangrijke instellingen van het Random Forest-model: het aantal bomen, aangeduid als ntree, en het aantal variabelen dat bij elke splitsing wordt overwogen, aangeduid als mtry (Probst et al., 2018).

Voor model 2 en model 4 zijn de instellingen ntree gelijk aan 300 en mtry gelijk aan 6. Deze waarden zijn gekozen op basis van een analyse van de *Out-Of-Bag* fout, een maatstaf die een inschatting geeft van de prestaties van het model op nieuwe, niet eerder gebruikte data (Probst et al., 2018). Uit de analyse blijkt dat de fout daalt naarmate er meer bomen worden toegevoegd, maar dat deze verbetering afvlakt rond de 300 bomen.



Figuur 3. Optimalisatie ntree

Voor mtry werd vastgesteld dat een waarde van 6 het beste resultaat oplevert, aangezien de fout tot dat punt afneemt en daarna nagenoeg gelijk blijft. Deze instellingen zorgen voor een goede balans tussen nauwkeurigheid en rekentijd (Probst et al., 2018).



Figuur 4. Optimalisatie mtry

4.3 Meten van Modelprestaties

Random Forest zal in het kader van dit onderzoek gebruikt worden om de relatie tussen de fundamenten van bedrijven en macro-economische variabelen, en de prestaties van hun aandelen op lange termijn te onderzoeken. Ter beoordeling van de nauwkeurigheid en geschiktheid van het model worden in deze sectie de evaluatiemethoden besproken die gebruikt zullen worden om de prestaties van het voorspellingsmodel te meten.

4.3.1 Train-Test Split

De Train-Test Split is een techniek in machine learning om de prestaties van een model te evalueren. Het doel is om een dataset op te splitsen in twee delen: een trainingsset en een testset. De trainingsset wordt gebruikt om het model te trainen, terwijl de testset wordt gebruikt om het model te evalueren op nieuwe data. Dit helpt bij het detecteren van overfitting, waarbij een model te goed presteert op de trainingsdata maar slecht presteert op nieuwe data. Het standaardpercentage voor de splitsing is vaak 80/20, waarbij het grootste deel van de data wordt gebruikt voor training en de rest voor testen. Het gebruik van een Train-Test Split zorgt ervoor dat de evaluatie van het model betrouwbaar is voor de prestaties op onbekende data (Gholamy et al., 2018).

4.3.2 Prestatiemaatstaven

Bij het evalueren van een Random Forest regressie-model is het daarnaast belangrijk om de prestaties te meten aan de hand van verschillende statistieken die inzicht geven in de nauwkeurigheid van de voorspellingen. De evaluatiemetrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Root Mean Squared Logarithmic Error (RMSLE), Mean Absolute Percentage Error (MAPE), R-squared (R^2) en Adjusted R-squared worden gebruikt om de nauwkeurigheid van het model bij regressieproblemen te beoordelen. In de volgende secties worden deze metrics afzonderlijk en in detail toegelicht (Chicco et al., 2021; Karch, 2019; Amin et al., 2023).

4.3.2.1 Mean Absolute Error (MAE)

De Mean Absolute Error (MAE) meet het gemiddelde van de absolute verschillen tussen de voorspelde waarden en de werkelijke waarden. Het geeft een direct interpreteerbare fout in dezelfde eenheid als de outputvariabele. Hoe lager de MAE, hoe beter het model presteert (Sadorsky, 2021; Amin et al., 2023).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

4.3.2.2 Mean Squared Error (MSE)

De Mean Squared Error (MSE) berekent het gemiddelde van de kwadratische verschillen tussen de voorspelde waarden en de werkelijke waarden. Doordat de fouten gekwadrateerd worden, straft MSE grotere fouten zwaarder

dan kleinere. Dit maakt het gevoelig voor outliers (Sadorsky, 2021; Chicco et al., 2021; Mir et al., 2022).

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

4.3.2.3 Root Mean Squared Error (RMSE)

De Root Mean Squared Error (RMSE) is de wortel van de MSE en geeft de standaarddeviatie van de voorspellingsfouten weer. Doordat RMSE in dezelfde eenheid is als de doelvariabele, is het intuïtiever dan MSE, maar nog steeds gevoelig voor grote afwijkingen (Sadorsky, 2021; Chicco et al., 2021; Mir et al., 2022).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

4.3.2.4 Mean Absolute Percentage Error (MAPE)

De MAPE berekent het gemiddelde van de absolute procentuele fouten tussen de voorspelde en werkelijke waarden (Chicco et al., 2021; Mir et al., 2022).

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

waarbij $\bar{y}$ het gemiddelde is van de werkelijke waarden.

4.3.2.5 Root Mean Squared Logarithmic Error (RMSLE)

De RMSLE berekent de wortel van het gemiddelde van de kwadratische logaritmische afwijkingen. Dit helpt bij het verminderen van de invloed van grote foutwaarden door een logaritme te nemen van de voorspellingen en werkelijke waarden (Amin et al., 2023; Mir et al., 2022).

$$\text{RMSLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(1+\hat{y}_i) - \log(1+y_i))^2}$$

4.3.2.6 R-squared (R^2)

De R^2-score, of de *Coefficient of Determination*, meet welk deel van de variantie in de afhankelijke variabele verklaard wordt door het model. Een waarde van 1 betekent perfecte voorspellingen; een waarde van 0 betekent dat het model niet beter presteert dan het gemiddelde van de data (Sadorsky, 2021; Chicco et al., 2021).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

4.3.2.7 Adjusted R-squared (Aangepaste R^2)

De aangepaste R^2 corrigeert de R^2-score voor het aantal voorspellers (features) in het model. Dit voorkomt dat het R^2 onterecht toeneemt bij het toevoegen van irrelevante voorspellers, wat het model te optimistisch zou maken (Karch, 2019).

$$\text{Adjusted } R^2 = 1 - \left(1 - R^2\right) \cdot \frac{n-1}{n-p-1}$$

waarbij:

- n = aantal observaties
- p = aantal voorspellers (features)

4.3.3 Feature Importance

Feature importance is een kernaspect van Random Forest en laat zien welke kenmerken het meest bijdragen aan de voorspellingen van het model (Sinha & Jain, 2016). Een maat hiervoor is de %IncMSE (Percentage Increase in Mean Squared Error). Hierbij wordt één variabele willekeurig verstoord, waarna het effect op de voorspellingsfout wordt gemeten. Een sterke toename in de MSE duidt op een belangrijke variabele. Hoe groter deze toename, hoe belangrijker het kenmerk wordt beschouwd voor het model (Chen et al., 2020; Zhou & Hooker, 2019).

4.4 Dataverzameling

Om dit onderzoek uit te kunnen voeren, moet er data worden verzameld. De benodigde data wordt verzameld van bedrijven die deel uitmaken van de NASDAQ-100 index. De index bestaat uit 100 bedrijven en is marktkapitalisatiegewogen, wat betekent dat bedrijven met een grotere beurswaarde zwaarder wegen in de berekening van de indexwaarde (Subasi et al., 2021). De samenstelling van de index wordt periodiek herzien om ervoor te zorgen dat de opgenomen bedrijven blijven voldoen aan de criteria, zoals minimale marktkapitalisatie en liquiditeit. Een volledig overzicht van alle bedrijven binnen de NASDAQ-100 die zijn opgenomen in de dataset, is te vinden in de bijlagen in tabel 7 in sectie 8 (Nasdaq, 2025). De tabel bevat de tickers van de bedrijven in alfabetische volgorde, samen met de bijbehorende bedrijfsnaam en sector. Bovendien wordt het aantal rijen per dataset per bedrijf berekend, wat later van nut kan zijn om bepaalde bevindingen te kunnen analyseren. De laatste datum in de datasets voor elk bedrijf is altijd december 2024.

Voor de uitvoering van het onderzoek met Random Forest is het noodzakelijk zowel inputvariabelen als een doelvariabele te selecteren. De inputvariabelen, ook wel onafhankelijke variabelen genoemd, zijn de variabelen die het model gebruikt om voorspellingen te doen. De doelvariabele, of afhankelijke variabele, is de uitkomst die het model probeert te voorspellen. Binnen deze thesis zullen fundamentele bedrijfsgegevens en macro-economische variabelen worden gebruikt om aandelenprijzen te voorspellen. Voor het verzamelen van de macro-economische data wordt gebruikgemaakt van de database van de Federal Reserve Economic Data (Fred, 2025). De fundamentele bedrijfsgegevens en aandelenprijzen worden opgehaald via de API van Alphavantage (Alphavantage, 2025). In de volgende secties worden de keuzes voor de input- en doelvariabelen toegelicht.

4.4.1 Doelvariabele

In dit onderzoek wordt de marktkapitalisatie gebruikt als doelvariabele. Deze wordt berekend door de maandelijkse slotkoers van het aandeel te vermenigvuldigen met het aantal uitstaande aandelen op datzelfde moment. Er is bewust gekozen voor marktkapitalisatie, aangezien dit een betere weergave is van de werkelijke waarde van een bedrijf. Bovendien maakt het gebruik van marktkapitalisatie het mogelijk om bedrijven onderling te vergelijken binnen methodes zoals de 'leave-one-ticker-out'-methode of in een gecombineerde dataset waarin verschillende bedrijven worden samengenomen. Marktkapitalisatie is nauw verbonden met de aandelenprijs, aangezien deze eenvoudig kan worden herleid door de marktkapitalisatie te delen door het aantal uitstaande aandelen. Op die manier blijft het mogelijk om de aandelenprijs te reconstrueren.

Er wordt gekozen om de marktkapitalisatie te berekenen op basis van de maandelijkse slotkoers van het aandeel. Deze methode is bedoeld om de invloed van kortetermijnvolatiliteit te beperken en de onderliggende trend in de koersontwikkeling beter te capteren over een langere tijdsperiode (Paduri et al., 2024). Deze aanpak is bijgevolg geschikt voor het voorspellen van absolute aandelenprijzen in plaats van rendementen, hetgeen ook het doel is van dit onderzoek.

4.4.2 Inputvariabelen

De inputvariabelen waarmee het Random Forest-model zal worden getraind, bestaan zowel uit fundamentele bedrijfsfactoren als uit macro-economische variabelen. Ter identificatie van deze variabelen wordt de bestaande literatuur geraadpleegd.

4.4.2.1 Fundamentele bedrijfsfactoren

In studies waarin machine learning wordt toegepast voor het voorspellen van aandelenprijzen op basis van fundamentele analyse, blijkt dat zowel fundamentele ratio's, zoals winst per aandeel (EPS), rendement op eigen vermogen (ROE) en rendement op activa (ROA), als absolute financiële grootheden, zoals schulden, activa, winst en omzet, frequent worden gebruikt als inputvariabelen. Zo hanteren Huang et al. (2022) een brede set van 21 fundamentele financiële kenmerken, waaronder: PE, Book Value, Liability, Net Margin, Gross Margin, Operating Margin, EBIT Margin, ROE, ROA, Operating Cash Flow, Net Cash Flow, Capital Expenditure, PB, Debt to Equity, Asset Turnover, Inventory Turnover, Accounts Receivable Turnover, EPS, Dividenden Per Aandeel, Relative Return, Revenue Growth. Hun model focust zich sterk op bedrijfsinterne cijfers, afkomstig uit jaar- en kwartaalrapportages. Paduri et al. (2024) breiden dit uit door naast fundamentele ratio's ook balansposten, kasstroomoverzichten en uitgebreide operationele data op te nemen, aangevuld met prestatie-indicatoren zoals EPS en rendement op geïnvesteerd kapitaal. Zo nemen ze onder meer, uit de balans, posten op zoals aandelenkapitaal, eigen vermogen, schulden, reserves, totale activa, lopende investeringen en overige activa. Uit het kasstroomoverzicht verwerken ze kasstroom uit operaties, kasstroom uit investeringen en kasstroom uit financiering. Deze gegevens worden aangevuld met financiële prestatie-indicatoren zoals winst per aandeel (EPS), nettowinstmarge, rendement op eigen vermogen, rendement op geïnvesteerd kapitaal en marktkapitalisatie ten opzichte van de omzet. Een meer beknopte aanpak wordt gehanteerd door Zhong & Hitchcock (2021), die zich beperkt tot drie fundamentele ratio's: PE, PB en PS. Tan et al. (2019) vult dit aan door groeimaten en traditionele rendementsratio's op te nemen, zoals nettowinstgroei, bedrijfsinkomstengroei, ROA en ROE.

Naast studies over machine learning werd ook de bredere financiële literatuur geraadpleegd. Asness et al. (2019) introduceert het kwaliteitsconcept als overkoepelend kenmerk, waaronder winstgevendheid, groei en veiligheid vallen. Verschillende studies focussen daarnaast op traditionele ratio's zoals EPS, ROE en DER. Suyanto et al. (2021) vindt dat deze ratio's significant bijdragen aan aandelenwaardering. Lestari & Hadi (2024) bevestigt het belang van ROE, EPS en ROS. In contrast daarmee tonen Gursida (2017) aan dat in de kolensector vooral ROA en CR relevant zijn. Ook Digdowiseiso (2023) vindt dat bij defensieve aandelen enkel ROA een betekenisvolle rol speelt.

De bedrijfsgegevens zijn terug te vinden in de drie kernrapporten van een bedrijf: de resultatenrekening, de balans en het kasstroomoverzicht (Liu, 2020). Deze zullen worden afgehaald via de API van Alphavantage (Alphavantage, 2025). Er wordt gekozen om zowel absolute financiële grootheden in de dataset op te nemen, omdat deze variabelen directe informatie over de schaal van een bedrijf bevatten en beter aansluiten bij de voorspelling van de absolute marktkapitalisatie, als ratio's, omdat deze aanvullende informatie bieden over de rendabiliteit, waardering en groeidynamiek van een bedrijf, zonder dat zij afhankelijk zijn van marktprijzen.

Op basis van bestaande studies over aandelenvoorspellingen met Random Forest, waarin gebruik wordt gemaakt van fundamentele analyse, en rekening houdend met de inzichten uit de financiële literatuur, worden de volgende inputvariabelen geselecteerd. Concreet zullen uit de balans, het kasstroomoverzicht en de resultatenrekening per bedrijf de volgende variabelen worden opgenomen: totale activa, totale schulden, vrije kasstroom, omzet en nettowinst. Daarnaast zullen ratio's zoals de jaar-op-jaar omzetgroei, winstmarge, winst per aandeel (EPS), rendement op eigen vermogen (ROE), schuldgraad (DER) en rendement op activa (ROA) worden opgenomen, aangezien deze indicatoren waardevolle informatie verschaffen over de groeikracht, winstgevendheid en financiële structuur van bedrijven.

4.4.2.2 Macro-economische variabelen

Macro-economische variabelen geven inzicht in de toestand van een land of de wereldeconomie en hebben invloed op aandelenprijzen (Verma & Bansal, 2021). Belangrijke variabelen zijn rentevoeten, inflatie en economische groei (Sinha & Jain, 2016; B. & Deo, 2021; Chang et al., 2019; Verma & Bansal, 2021; Pramudito, 2023; Atanasov, 2021; Sjahrunnisa et al., 2024). Een stijging van de rente verhoogt de leenkosten voor bedrijven en consumenten, verlaagt de winstgevendheid en maakt obligaties aantrekkelijker, waardoor aandelenkoersen vaak onder druk komen te staan (Sjahrunnisa et al., 2024; B. & Deo, 2021; Chang et al., 2019). Inflatie verhoogt de kostenstructuur van bedrijven, vermindert de koopkracht van consumenten en verhoogt de onzekerheid op financiële markten, wat aandelenkoersen negatief beïnvloedt (Sjahrunnisa et al., 2024; Verma & Bansal, 2021; Chang et al., 2019). Economische groei, gemeten via het bruto binnenlands product (bbp), heeft doorgaans een positieve invloed op aandelenprijzen doordat hogere productie en consumptie leiden tot hogere winsten, terwijl een daling van het bbp duidt op verzwakking van de economie en lagere bedrijfsresultaten (Chang et al., 2019; Verma & Bansal, 2021; Pramudito, 2023). Tot slot beïnvloedt werkloosheid aandelenprijzen doordat lage werkloosheid vaak leidt tot hogere bestedingen en bedrijfswinsten, terwijl hoge werkloosheid het vertrouwen van beleggers kan schaden en aandelenkoersen doet dalen (Atanasov, 2021).

Naast de eerder besproken fundamentele bedrijfsfactoren, worden aan de dataset macro-economische variabelen toegevoegd. Deze macro-economische variabelen omvatten de rentevoet, inflatie, economische groei en werkloosheidsgraad. Door zowel fundamentele bedrijfsgegevens als macro-economische factoren te combineren, wordt gestreefd naar een vollediger model dat verschillende invloeden op aandelenprijzen in rekening brengt.

4.5 Data Preprocessing

Na het verzamelen van de macro-economische variabelen via de website van de Federal Reserve Economic Data (Fred, 2025), en het ophalen van de fundamentele bedrijfsdata en aandelenprijzen via de API van Alphavantage (Alphavantage, 2025), wordt de dataset klaargemaakt voor verdere analyse. Allereerst worden per dataset enkel de relevante kolommen geselecteerd met behulp van de select()-functie in R. Daarna worden de kolomnamen hernoemd om de interpretatie te vergemakkelijken en om ervoor te zorgen dat de datasets later probleemloos kunnen worden samengevoegd op basis van de datekolom.

Om deze samenvoeging correct te laten verlopen, wordt de date-kolom eerst omgezet naar het formaat `YYYY-MM-01`, waarbij de dag altijd op de eerste van de maand wordt gezet. Dit is voldoende, aangezien de analyse gebaseerd is op maandelijkse data en de specifieke dag binnen de maand voor deze studie niet relevant is. Vervolgens worden de datasets die op kwartaalbasis worden gerapporteerd zoals real_gdp, unemployment_rate, income_statement en cashflow_statement, omgezet naar maandelijkse data door de waarden over de drie maanden binnen elk kwartaal te dupliceren. Dat wordt gedaan om de NA-waarden op te vullen wanneer de kwartaaldata met de maanddata worden samengevoegd. Tot slot wordt de date-kolom omgezet naar een datumobject met behulp van as.Date(), en worden de overige waarden geformatteerd als numeriek met behulp van as.numeric() in R.

Alle datasets worden uiteindelijk samengevoegd op basis van de kolom 'date' en opgeslagen in een dataset genaamd 'Dataset'. Daarnaast worden enkele nieuwe kolommen toegevoegd, waarbij berekeningen op bestaande kolommen worden uitgevoerd om ratio's te berekenen. Dit wordt gedaan om de nauwkeurigheid van het model te verbeteren.

1. Marktkapitalisatie:

   $$\text{Marktkapitalisatie} = \text{Aandelenprijs} \times \text{Aantal aandelen}$$

2. Winstmarge:

   $$\text{Winstmarge} = \frac{\text{Netto winst}}{\text{Omzet}}$$

3. Jaar-op-jaar omzetgroei:

   $$\text{Jaar-op-jaar omzetgroei} = \left(\frac{\text{Omzet}}{\text{Omzet}_{t-12}}\right) - 1$$

4. Winst per aandeel (EPS):

   $$\text{EPS} = \frac{\text{Netto winst} - \text{Preferente dividenden}}{\text{Gemiddeld aantal uitstaande aandelen}}$$

5. Rendement op eigen vermogen (ROE):

   $$\text{ROE} = \frac{\text{Netto winst}}{\text{Gemiddeld eigen vermogen}}$$

6. Schuldgraad (Debt-to-Equity Ratio, DER):

   $$\text{DER} = \frac{\text{Totale schulden}}{\text{Totaal eigen vermogen}}$$

7. Rendement op activa (ROA):

   $$\text{ROA} = \frac{\text{Netto winst}}{\text{Gemiddeld totaal activa}}$$

4.6 Finale data

Wanneer alle data is verzameld en de berekeningen zijn voltooid, wordt de definitieve dataset samengesteld die gebruikt zal worden voor de uit te voeren analyse. Onderstaande tabel 3 geeft alle kolommen uit die dataset weer met een korte beschrijving. De uiteindelijke dataset bestaat uit 15 inputvariabelen, waarvan 4 macro-economische factoren, 5 fundamentele bedrijfsvariabelen uitgedrukt in absolute grootheden en 6 ratio's om relatieve prestaties tussen bedrijven en over tijd te kunnen vergelijken. De doelvariabele is de marktkapitalisatie.

Kolomnaam	Beschrijving
real_gdp	Reële BBP
fed_funds_rate	Rentevoet
cpi	Consumentenprijsindex
unemployment_rate	Werkloosheidspercentage
assets	Activa
liabilities	Schulden
revenue	Omzet
net_income	Netto winst
freecashflow	Vrije kasstroom
profit_margin	Winstmarge
yoy_revenue_growth	Jaar-op-jaar omzetgroei
eps	Winst per aandeel
roe	Rendement op eigen vermogen
der	Schuldgraad
roa	Rendement op activa
marketcap	Marktkapitalisatie

Tabel 3. Dataset

Volgens onderstaand algoritme wordt de finale dataset opgesteld. Dit R-algoritme maakt een lege lijst genaamd Final_dataset en vult deze met datasets voor elk bedrijf in de nasdaq100-lijst. Voor elk symbool in de lijst wordt een korte vertraging van 0,5 seconden ingevoerd met Sys.sleep(0.5) om de belasting van het systeem te verminderen. Via de premium versie van alphavantage kunnen er 75 verzoeken per seconde worden verzonden, waardoor er dus een kleine vertraging moet worden ingesteld om geen error te krijgen. Vervolgens wordt de functie get_dataset(symbool) aangeroepen om de dataset zoals beschreven in tabel 3 voor het specifieke symbool op te halen. De opgehaalde dataset wordt opgeslagen in de Final_dataset-lijst, waarbij de naam van het symbool als sleutel wordt gebruikt. Hierdoor wordt elke dataset gekoppeld aan het bijbehorende symbool in de uiteindelijke lijst. De output van dit R-algoritme is een lijst genaamd Final_dataset, waarin voor elk symbool uit de nasdaq100-lijst de bijbehorende dataset, zoals weergegeven in tabel 3, wordt opgeslagen. Deze dataset zal gebruikt worden om de verschillende Random Forest-modellen op te bouwen.

```
Final_dataset <- list()

for (symbool in nasdaq100) {

  Sys.sleep(0.5)

  Dataset <- get_dataset(symbool)

  Final_dataset[[symbool]] <- Dataset

}
```

5 Resultaten en Discussie

In deze sectie wordt het algoritme voor elke methode besproken. Vervolgens worden de resultaten per model gepresenteerd en geanalyseerd, evenals de *feature importance*.

5.1 Methode 1: Bedrijfsspecifiek met willekeurige splitsing

Methode 1 richt zich op een geïndividualiseerde benadering waarbij elk bedrijf afzonderlijk wordt geanalyseerd. Voor elk aandeel wordt een apart Random Forest-model getraind en geëvalueerd. De dataset wordt daarbij willekeurig gesplitst in 80% training en 20% testing. Deze methode biedt gedetailleerd inzicht in de voorspelbaarheid van elk individueel aandeel, zonder beïnvloeding van andere bedrijven.

5.1.1 Algoritme

Het algoritme voor de eerste methode traint en evalueert per bedrijf uit de dataset een Random Forest-model voor het voorspellen van de '*marketcap*'-waarde. Voor elk symbool wordt eerst de bijbehorende data geselecteerd en opgesplitst in kenmerken X en doelvariabele Y. Vervolgens wordt de data gesplitst in een trainings- en testset met een vaste *random seed* voor reproduceerbaarheid. Daarna worden het optimale aantal bomen (ntree) en het optimale aantal kenmerken per split (mtry) bepaald. Met deze parameters wordt het Random Forest-model getraind, voorspellingen worden gedaan op de testset, en evaluatiestatistieken en *feature importance* worden berekend.

Algorithm 1 Random Forest-analyse: Bedrijfsspecifiek met willekeurige splitsing

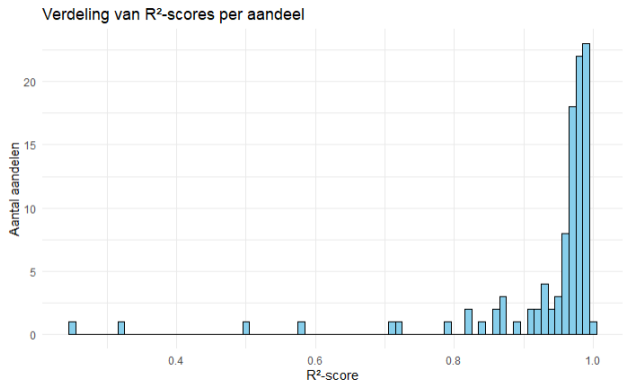
Input: `dataset` ← finale dataset met marketcap
Output: `final_metrics_df`, `final_incMSE_df`
Initialiseer lege dataframes: `final_metrics_df`, `final_incMSE_df`
for all `symbol` in `symbols` **do**
    `data_clean` ← `dataset[[symbol]]`
    `X` ← alle kolommen behalve 'marketcap'
    `Y` ← kolom 'marketcap'
    Zet random seed = 42
    `trainIndex` ← steekproef van 80% van `Dataset`
    `X_train`, `X_test` ← splits `X` op `trainIndex`
    `y_train`, `y_test` ← splits `Y` op `trainIndex`
    `optimal_ntree` ← `ntree(X_train, y_train)`
    `optimal_mtry` ← `mtry(X_train, y_train)`
    Train model: `rf_model` ← `randomForest(X_train, y_train, optimal_ntree, optimal_mtry)`
    `y_pred` ← `predict(rf_model, X_test)`
    `metrics_df` ← `metrics(y_test, y_pred)`
    `incMSE_df` ← `calculate_incMSE(rf_model)`
        `final_metrics_df` ← `rbind(final_metrics_df, metrics_df)`
        `final_incMSE_df` ← `rbind(final_incMSE_df, incMSE_df)`
end for

5.1.2 *Prestaties*

De prestatiemetrieken voor methode 1, zoals besproken in sectie 4.3.2, zijn opgenomen in de bijlagen, zie tabel 8. De gemiddelde MAE bedraagt 7,55 miljard, met een mediaan van 3,03 miljard. Dit suggereert nauwkeurige voorspellingen, ondanks enkele uitschieters. De gemiddelde MSE is 864 miljard miljard, en de RMSE gemiddeld 12,21 miljard met een mediaan van 3,98 miljard, wat opnieuw duidt op de invloed van enkele grote fouten.

De gemiddelde MAPE bedraagt 8,53%, met een mediaan van 6,75%, wat wijst op een acceptabele relatieve afwijking. De RMSLE-waarden zijn laag (gemiddeld 0,11; mediaan 0,09), wat aangeeft dat het model ook bij sterk variërende aandelenprijzen goed presteert.

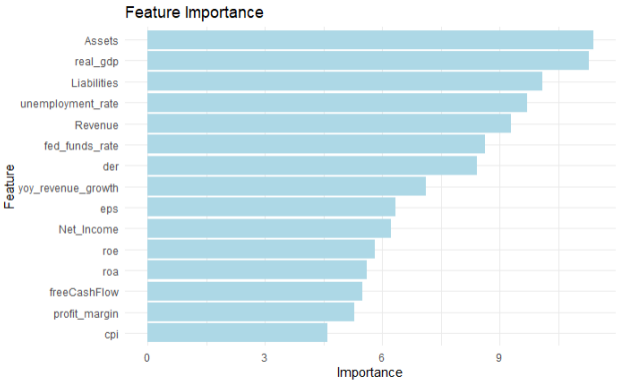
Het model verklaart gemiddeld 93% van de variantie (R^2), met een mediaan van 97%, en een aangepaste R^2 van gemiddeld 98%. Dit bevestigt de sterke generaliseerbaarheid van het model. Figuur 5 toont aan dat de meeste aandelen een R^2 boven 0,9 hebben. Enkele aandelen met een R^2 onder 0,6, vaak met weinig historische data, laten verminderde nauwkeurigheid zien.



Figuur 5. Verdeling van de R^2-scores per aandeel

5.1.3 *Feature Importance*

Figuur 6 geeft de feature importance weer van de inputvariabelen in het Random Forest-model. Uit de figuur blijkt dat vooral Assets en real_gdp de grootste bijdrage leveren aan de voorspellende kracht van het model, wat erop wijst dat zowel de schaal van een onderneming als de algemene economische groei cruciale factoren zijn bij het voorspellen van aandelenprijzen. Ook Liabilities, unemployment_rate en Revenue spelen een belangrijke rol, wat de invloed benadrukt van zowel financiële bedrijfsstructuur als macro-economische omstandigheden. Variabelen zoals freeCashFlow, roe, roa, profit_margin en cpi blijken minder bepalend, wat suggereert dat rendabiliteitsratio's en inflatie een beperktere impact hebben binnen deze methode.



Figuur 6. Feature Importance Model 1

5.2 *Methode 2: Geaggregeerde dataset met willekeurige splitsing*

In de tweede methode worden de gegevens van alle onderzochte bedrijven samengevoegd tot één grote, gezamenlijke dataset. Ook hier wordt een willekeurige verdeling gehanteerd waarbij 80% van de data wordt gebruikt voor training en de overige 20% voor testen. Deze aanpak maakt gebruik van een grotere hoeveelheid data, wat de robuustheid en generaliseerbaarheid van het model ten goede kan komen. Doordat het model leert van een grote hoeveelheid aan bedrijfsgegevens, kan het mogelijk beter omgaan met onbekende situaties.

5.2.1 *Algoritme*

Algorithm 2 Random Forest-analyse: Geaggregeerde dataset met willekeurige splitsing

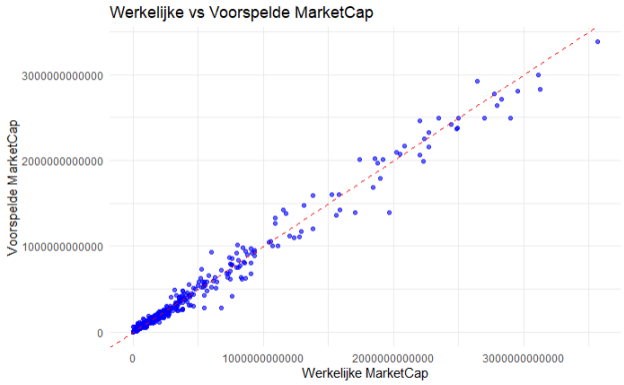
```
Input: Dataset ← finale dataset met marketcap
Output: metrics_df
all_data ← do.call(rbind, Dataset)
X ← all_data[, !names(all_data) %in% MarketCap]
y ← all_data[MarketCap]
set.seed(42)
train_indices ← sample(1:nrow(all_data), size = 0.8 * nrow(all_data))
X_train ← X[train_indices, ]
y_train ← y[train_indices]
X_test ← X[-train_indices, ]
y_test ← y[-train_indices]
rf_model ← randomForest(X_train, y_train, ntree = 300, mtry = 6, importance = TRUE)
y_pred ← predict(rf_model, X_test)
metrics_df ← metrics(y_test, y_pred)
```

5.2.2 *Prestaties*

De prestatiemetrieken die besproken werden in sectie 4.3.2 voor methode 2 zijn terug te vinden in tabel 4. Aangezien het model bij deze methode op de geaggregeerde data wordt getraind, moet er geen gemiddelde of mediaan worden berekend.

Metric	Value
MAE	9,79B
MSE	1,076,472,880,001,74B
RMSE	32,81B
MAPE	14,94%
RMSLE	0.2099
R^2	0.9867
Adjusted R^2	0.9866

Tabel 4. Evaluatiemetrics Model 2



Figuur 7. Werkelijke vs Voorspelde MarketCap

De Mean Absolute Error (MAE) bedraagt 9,79 miljard, wat aangeeft dat de voorspellingen gemiddeld met dit bedrag afwijken van de werkelijke marktkapitalisatie. De MSE is groot, namelijk 1.076 biljoen miljard, wat het effect benadrukt van grote fouten, vooral bij grote bedrijven. De RMSE, die de gemiddelde fout meet in dezelfde eenheid als de voorspelling, bedraagt 32,81 miljard. Deze relatief hoge waarde kan deels verklaard worden door het feit dat grote bedrijven zoals Apple en Nvidia met hun hoge marktkapitalisatie het model beïnvloeden.

De gemiddelde Mean Absolute Percentage Error (MAPE) bedraagt 14,94%, wat hoger is dan bij Methode 1. Dit geeft aan dat de relatieve afwijking van de voorspellingen gemiddeld bijna 15% bedraagt, wat redelijk is gezien de schaal van de dataset, maar minder nauwkeurig dan bij de bedrijfsspecifieke benadering. De Root Mean Squared Logarithmic Error (RMSLE) is 0,2099. Deze waarde wijst erop dat het model, hoewel het goed omgaat met logaritmisch geschaalde verschillen, wat minder robuust is voor voorspellingen van bedrijven met zeer uiteenlopende marktgroottes.

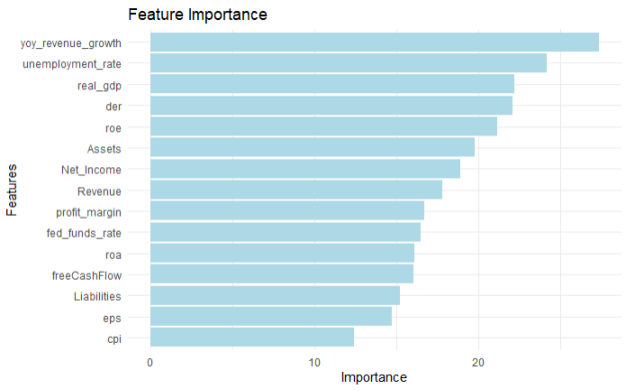
Ondanks de grotere spreiding in foutmaten blijft de verklaarde variantie goed. De R^2-waarde van 0,9867 geeft aan dat bijna 99% van de variantie in de marktkapitalisatie verklaard wordt door het model. Ook de aangepaste R^2 is bijna identiek (0,9866), wat bevestigt dat het model geen overbodige variabelen bevat die de voorspellingen kunstmatig verbeteren. Deze hoge determinatiecoëfficiënten tonen aan dat, ondanks de hogere foutenmarges, het model zeer krachtig is in het herkennen van onderliggende patronen over bedrijven heen. Het model blijkt dus sterk in het generaliseren van kennis uit de volledige dataset naar nieuwe, ongeziene data.

Grafiek 7 laat zien hoe goed het Random Forest-model de marktkapitalisatie van bedrijven kan voorspellen. Op de horizontale as staat de werkelijke marktkapitalisatie, terwijl de verticale as de voorspelde marktkapitalisatie toont. Elke blauwe stip in de grafiek staat voor één bedrijf. Hoe dichter een punt bij de rode lijn ligt, hoe beter de voorspelling van het model is. De rode lijn is de lijn waarop de voorspelling precies gelijk is aan de werkelijke waarde. Als alle punten op deze lijn zouden liggen, zou het model perfect accuraat zijn.

De grafiek toont aan dat de meeste punten dicht bij de rode lijn liggen. Dit betekent dat het model over het algemeen goede voorspellingen doet. Bij hogere marktkapitalisaties is er iets meer spreiding, wat aangeeft dat de voorspellingen daar iets minder precies zijn. Deze grafiek bevestigt dat het model betrouwbaar is in het voorspellen van de marktkapitalisatie.

5.2.3 Feature Importance

Uit de *feature importance* van Methode 2 blijkt dat yoy_revenue_growth veruit de belangrijkste voorspellende variabele is binnen het geaggregeerde model, gevolgd door unemployment_rate en real_gdp. Ook variabelen zoals der, roe, Assets en Net_income leveren een aanzienlijke bijdrage, wat het belang van financiële structuur en winstgevendheid bevestigt. Onderaan bevinden zich onder meer Liabilities, eps en cpi, die in dit model een relatief beperkte rol spelen. Deze verdeling suggereert dat het model vooral leunt op groeigerelateerde en macro-economische indicatoren bij het voorspellen van marktkapitalisatie over meerdere bedrijven heen.



Figuur 8. Feature Importance Model 2

5.3 Methode 3: Bedrijfsspecifiek met chronologische splitsing

De derde methode lijkt qua structuur op de eerste, maar wijkt af in de manier waarop de data wordt opgesplitst. In plaats van een willekeurige verdeling, wordt de dataset per bedrijf chronologisch gesplitst. De oudste 80% van de gegevens wordt gebruikt om het model te trainen, terwijl de meest recente 20% wordt ingezet om het model te testen. Deze aanpak weerspiegelt beter de realistische situatie waarin een voorspellingsmodel uitsluitend op basis van historische data beslissingen moet nemen over toekomstige ontwikkelingen. Een nadeel is dat eventuele veranderingen in bedrijfsstructuur of marktcondities in de tijd mogelijk minder goed worden meegenomen, waardoor de nauwkeurigheid van voorspellingen kan afnemen.

5.3.1 Algoritme

Dit algoritme bouwt voort op het algoritme van methode 1 en richt zich op het trainen van een Random Forest-model voor tijdreeksvoorspellingen met behulp van een *rolling window*-techniek. Voor elk symbool in de dataset worden de gegevens opgesplitst in trainings- en testsets binnen een *rolling window*. Het *"rolling window"* betekent dat het model steeds wordt getraind op een subset van de data (de trainingsset), terwijl het wordt getest op de volgende stap in de tijd (de testset). Het *window* beweegt zich stap voor stap door de dataset, waarbij telkens de meest recente data wordt gebruikt om het model te trainen, en de voorspellingen worden vergeleken met de werkelijke waarden in de testset.

Binnen elke stap worden de optimale hyperparameters van het Random Forest-model bepaald (het aantal bomen ntree en het aantal variabelen per split mtry). Het model wordt vervolgens getraind met de geselecteerde parameters, en de voorspellingen worden opgeslagen. Na het doorlopen van alle stappen wordt de modelperformance geëvalueerd door de prestatiemetrieken te berekenen, en wordt ook de verandering in de Mean Squared Error (incMSE) door de verschillende features geanalyseerd. De resultaten van de modelprestaties en de belangscores worden opgeslagen in respectievelijk de dataframes final_metrics_df en final_incMSE_df. Dit proces wordt herhaald voor elk symbool in de dataset.

5.3.2 Prestaties

De prestatiemetrieken die besproken werden in sectie 4.3.2 voor methode 3 zijn terug te vinden in de bijlagen in tabel 9. Deze aanpak benadert een realistische toepassing in tijdsvolgorde, maar blijkt de prestaties van het model te beïnvloeden. De gemiddelde Mean Absolute Error (MAE) bedraagt 20,17 miljard, met een mediaan van 5,82 miljard. Dit grote verschil wijst opnieuw op de aanwezigheid van uitschieters, waarbij sommige bedrijven tot zeer grote afwijkingen leiden. De Mean Squared Error (MSE) ligt met gemiddeld 3.877 biljoen miljard aanzienlijk hoger dan bij de andere methoden. De RMSE bedraagt gemiddeld 25,7 miljard, met een mediaan van 7,9 miljard, wat wijst op een sterke spreiding in de nauwkeurigheid per bedrijf.

De MAPE-waarde van gemiddeld 10,01% (mediaan: 8,97%) toont aan dat het model redelijk accuraat is in relatieve termen, ondanks de grotere absolute fouten. De afwijking is hoger dan bij methode 1, maar vergelijkbaar met die van methode 2. De RMSLE bedraagt gemiddeld 0,13 met een mediaan van 0,11, wat suggereert dat het model relatief goed omgaat met logaritmische schaalverschillen.

De verklarende kracht van het model is lager dan bij de voorgaande methoden. De gemiddelde R^2 bedraagt 0,55 en de mediaan 0,60, wat aangeeft dat slechts iets meer dan de helft van de variantie in de aandelenprijzen verklaard wordt. De aangepaste R^2, die corrigeert voor het aantal gebruikte voorspellers, komt gemiddeld uit op 0,66, terwijl de mediaan zakt naar 0,42. Deze resultaten tonen aan dat het model in een tijdsgevoelige context minder goed in staat is om robuuste voorspellingen te genereren, mogelijk door veranderingen in marktomstandigheden die moeilijk op basis van historische gegevens te modelleren zijn.

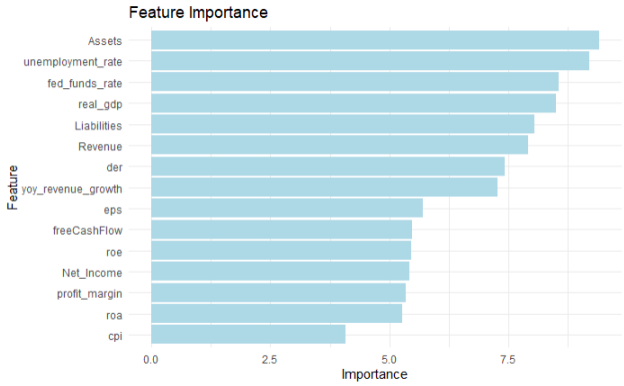
Deze methode kan ook door visuele weergaven betere inzichten bieden in de accuraatheid van het model. Hierbij wordt de voorspelde aandelenprijs vergeleken met de werkelijke aandelenprijs, en kan visueel worden beoordeeld hoe goed deze overeenkomen. De grafieken tonen over het algemeen sterke overeenkomsten. Een voorbeeld van zo'n grafiek, die de aandelenprijs van Netflix weergeeft, is te zien in figuur 9. Daarbij is de blauwe lijn de voorspelde marktkapitalisatie.

Figuur 9. Netflix voorspelling

5.3.3 Feature Importance

De *feature importance* van methode 3 wordt weergegeven in figuur 10. Ook hier blijkt dat Assets de belangrijkste voorspeller is, gevolgd door unemployment_rate, fed_funds_rate en real_gdp. Dit geeft aan dat zowel de schaal van het bedrijf als economische indicatoren zoals werkloosheid en rentestanden een grote rol spelen bij lan-

getermijnvoorspellingen. Middelmatig belangrijk zijn onder andere Liabilities, Revenue en der, terwijl groeimaten zoals yoy_revenue_growth en winstkenmerken zoals eps en freeCashFlow een bescheiden bijdrage leveren. Onderaan staan roa, cpi en profit_margin, die weinig invloed uitoefenen in dit tijdgevoelige model. De resultaten suggereren dat structurele en macro-economische variabelen belangrijker zijn dan rendabiliteitsratio's bij het voorspellen van toekomstige marktkapitalisatie.



Figuur 10. Feature Importance Model 3

5.4 Methode 4: Leave-one-ticker-out

De vierde methode bouwt voort op de geaggregeerde aanpak van methode twee, maar introduceert een belangrijke variatie: het *leave-one-ticker-out* principe. Hierbij wordt steeds één bedrijf volledig buiten de trainingsdata gehouden. Het model wordt getraind op de gegevens van alle andere bedrijven en vervolgens getest op het buitengehouden bedrijf. Deze methode simuleert een scenario waarin een model ontwikkeld wordt op basis van algemene marktkennis en deze toepast op een bedrijf waarvoor nog geen specifieke historische data beschikbaar is. Dit maakt het mogelijk om de generaliseerbaarheid van het model tussen verschillende bedrijven te evalueren. Een nadeel is dat het model mogelijk niet goed presteert op bedrijven met unieke eigenschappen die niet in de trainingsset zitten.

5.4.1 Algoritme

Het algoritme traint voor elke ticker in de dataset een Random Forest-model. Het begint door de gegevens van de huidige ticker te splitsen in test- en trainingsdata. Vervolgens wordt het model getraind met de trainingsdata, waarbij de marketcap-kolom wordt uitgesloten van de *features*. Na het trainen van het model worden voorspellingen gemaakt op de testdata en worden de prestaties van het model gemeten door de voorspellingen te vergelijken met de werkelijke waarden.

Voor elke ticker worden de resultaten van het model, inclusief de prestaties en de *feature importance*, verzameld in een tabel. Dit proces wordt herhaald voor elke ticker in de dataset. Aan het einde van het algoritme worden alle resultaten samengevoegd in een data.frame, dat een overzicht geeft van hoe goed het model voor elke ticker presteert en welke variabelen het meest invloedrijk waren bij het maken van voorspellingen.

Algorithm 4 Random Forest Leave-One-Ticker-Out

Input: Dataset ← finale dataset met marketcap
Output: results
results ← `data.frame()`
for each ticker i in *Dataset* **do**
    symbol ← names(*Dataset*)[i]
    test_data ← *Dataset*[*i*]
    ticker_name ← symbol
    train_data ← do.call(rbind, $Dataset_{\text{model\_4}}$[-i])
    X_{train} ← train_data[, all columns except "MarketCap]
    y_{train} ← train_data[target_variable]
    X_{test} ← test_data[, all columns except "MarketCap]
    y_{test} ← test_data[target_variable]
    rf_model ← randomForest(X_{train}, y_{train}, ntree = 300, mtry = 6, importance = TRUE)
    predictions ← predict(rf_model, X_{test})
    metrics_df ← metrics(y_{test}, predictions)
    incMSE_df ← calculate_incMSE(rf_model)
    imp_df ← rf_model.importance
    results ← rbind(results, metrics_df, incMSE_df, imp_df)
end for

5.4.2 Prestaties

De prestatiemetrieken die in sectie 4.3.2 zijn besproken, zijn voor methode 4 terug te vinden in de bijlagen, in tabel 10. De resultaten tonen een duidelijke verslechtering van de modelprestaties. De gemiddelde Mean Absolute Error (MAE) bedraagt 52,55 miljard, met een mediaan van 19,92 miljard. De gemiddelde MSE komt uit op 25.043 miljard miljard, met een mediaan van 644,77 miljard. De RMSE bereikt een gemiddelde van 73,29 miljard, wat duidt op grote spreiding in de voorspellingen. De mediaan van de RMSE ligt op 25,39 miljard, wat opnieuw wijst op enkele zeer grote uitschieters.

De foutmarges in relatieve termen zijn eveneens hoog. De gemiddelde Mean Absolute Percentage Error (MAPE) bedraagt 0,78, wat neerkomt op 78% afwijking van de werkelijke waarde. De mediaan ligt iets lager op 42%, wat nog steeds ver boven een aanvaardbaar niveau ligt voor financiële voorspellingen. De Root Mean Squared Logarithmic Error (RMSLE) komt gemiddeld uit op 0,58, met een mediaan van 0,49.

De verklaringskracht van het model is in deze methode vrijwel afwezig. De gemiddelde R^2-waarde is negatief: –17,69, wat betekent dat het model slechter presteert dan een simpele gemiddeldevoorspelling. De aangepaste R^2 ligt zelfs nog lager, met een gemiddelde van –28,87. Enkel de mediaan van R^2 (0,09) en adjusted R^2 (–0,01)

laten zien dat in enkele gevallen nog een minimale voorspellende waarde aanwezig is. Over het algemeen toont deze methode aan dat het model onvoldoende generaliseert naar volledig nieuwe bedrijven, wat duidt op een gebrek aan overdraagbaarheid van patronen tussen verschillende ondernemingen.

Om beter inzicht te krijgen in de zwakke prestaties van deze methode, kunnen de voorspelde waarden worden gevisualiseerd en vergeleken met de werkelijke waarden. Hieronder worden enkele grafieken weergegeven. Daaruit blijkt dat het model vaak in staat is om de juiste vorm van de werkelijke aandelenkoers te benaderen, maar dat de exacte waarden vaak afwijken. Dat is ook logisch, aangezien het hier gaat om een leave-one-ticker-out-methode waarbij alle data op een algemene manier worden geïnterpreteerd, terwijl in de realiteit verschillende bedrijven handelen tegen uiteenlopende waarderingsmultiples.

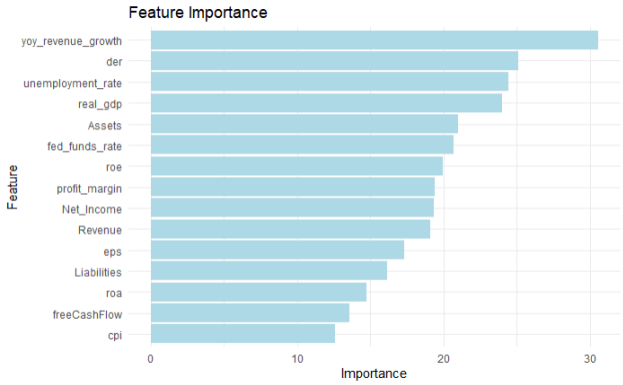
Figuur 11. Voorspelling MU

Figuur 12. Voorspelling META

Figuur 13. Voorspelling MDB

5.4.3 Feature Importance

De *feature importance* van methode 4 is sterk gelijkend aan die van methode 2 en laat zien dat yoy_revenue_growth de meest invloedrijke variabele is, gevolgd door der, unemployment_rate en real_gdp. Dit benadrukt het belang van groeidynamiek, financiële structuur en macro-economische omstandigheden bij het voorspellen van bedrijven waarvoor geen specifieke historische data beschikbaar is. Ook Assets en fed_funds_rate tonen een matige voorspellende waarde, terwijl kenmerken zoals roe, profit_margin en Net_income een beperkte bijdrage leveren. Onderaan bevinden zich freeCashFlow, cpi en roa, die nauwelijks impact hebben in dit model. De resultaten wijzen erop dat in situaties waarin het model moet generaliseren naar onbekende bedrijven, vooral groeigerelateerde en externe economische factoren bepalend zijn.



Figuur 14. Feature Importance Model 4

5.5 De rol van AI in handels- en investeringsstrategieën

In deze sectie zal de rol van artificiële intelligentie (AI) in handels- en investeringsstrategieën worden besproken. Er zal verder worden ingegaan op de toepassing van AI in risicomanagement, waar het helpt bij het nauwkeuriger inschatten en mitigeren van risico's. Daarnaast wordt de invloed van AI op beleggingsselectie behandeld, waarbij AI

in staat is om grote hoeveelheden data te analyseren en datagestuurde beslissingen te ondersteunen. Ten slotte zal worden ingegaan op de opkomst van robo-adviseurs, die dankzij AI steeds breder toegankelijk worden en een alternatieve benadering bieden voor het verstrekken van financieel advies.

5.5.1 Risicomanagement

Risicomanagement is een van de meest voorkomende toepassingen van AI binnen investeringsbeslissingen (El Hajj & Hammoud, 2023). AI-algoritmen kunnen grote hoeveelheden historische data analyseren en zo risico's opsporen die voor mensen lastig te herkennen zijn (Khattak et al., 2023). Dit zorgt ervoor dat risico's sneller en nauwkeuriger worden geïdentificeerd, waardoor beleggers en bedrijven sneller kunnen reageren op veranderingen in de markt (Wang, 2024).

Vooral binnen risicobeoordeling maken banken en andere financiële instellingen veel gebruik van AI-modellen zoals Random Forests. Deze modellen analyseren gegevens van kredietnemers, financiële cijfers en economische trends om in te schatten hoe groot het risico is dat iemand niet zal terugbetalen. Op basis van deze analyses kunnen instellingen beter onderbouwde beslissingen nemen over het verstrekken van kredieten en zo financiële risico's verkleinen (Sinha & Jain, 2016).

5.5.2 Portfolio Management

In het beheren van beleggingsportefeuilles wordt AI, en vooral methodes zoals Random Forests, steeds vaker gebruikt. Deze technieken helpen om te voorspellen wat beleggingen in de toekomst kunnen opleveren en of aandelen te duur of juist goedkoop zijn. Zo kunnen beleggers een goede mix van verschillende beleggingen samenstellen en tegelijk beter omgaan met risico's (Sinha & Jain, 2016). Ook kan AI bedrijven analyseren om te zien hoe gezond ze financieel zijn en hoeveel groeikansen ze hebben. Uit onderzoek blijkt dat beleggers die AI gebruiken bij hun keuzes vaak betere resultaten behalen. Dit zorgt ervoor dat AI steeds populairder wordt in de beleggingswereld (El Hajj & Hammoud, 2023; Das et al., 2024).

Een voorbeeld hiervan zijn robo-adviseurs. Dit zijn automatische systemen die mensen helpen bij het kiezen van beleggingen, het plannen van hun financiën en het beheren van hun vermogen. Ze werken grotendeels zelfstandig, zonder dat er veel menselijke hulp nodig is (Zhu et al., 2024). Dankzij slimme algoritmes kunnen roboadviseurs advies geven dat past bij de persoonlijke doelen en wensen van klanten. Omdat deze systemen grote hoeveelheden gegevens kunnen analyseren, geven ze vaak stabiele en goed onderbouwde adviezen. Daarnaast maken ze minder snel fouten die mensen soms wel maken, wat het vertrouwen in deze systemen vergroot (El Hajj & Hammoud, 2023; Wang, 2024; Zhu et al., 2024)

6 Tekortkomingen en aanbevelingen voor toekomstig onderzoek

De huidige literatuur legt sterk de nadruk op voorspellingen van aandelenprijzen op de zeer korte termijn, vaak binnen de context van dagen of uren. Daardoor is er weinig onderzoek dat kijkt naar voorspellingen over een langere periode, zoals meerdere jaren. Het zou dan ook waardevol zijn om toekomstige studies te richten op het uitbreiden van dit perspectief, door modellen te ontwikkelen en te analyseren die in staat zijn om aandelenkoersen over meerdere jaren te voorspellen, zoals in deze thesis werd gedaan.

Daarnaast blijft Random Forest in combinatie met fundamentele analyse vaak onderbelicht in de bestaande literatuur. Terwijl veel studies zich richten op technische indicatoren en sentimentanalyse als inputvariabelen voor machine learning-modellen, wordt de potentiële waarde van fundamentele gegevens nog onvoldoende benut. Verdere verkenning van deze combinatie kan niet alleen bijdragen aan academische inzichten, maar ook waardevolle toepassingen bieden voor financiële voorspellingsmodellen.

In toekomstig onderzoek zou het interessant kunnen zijn om de methodes die werden gebruikt binnen dit onderzoek ook toe te passen op andere machine learning modellen en de prestaties tussen de verschillende modellen te vergelijken. In deze thesis werd gewerkt met Random Forest, maar het zou waardevol zijn om te onderzoeken hoe andere modellen, zoals neurale netwerken of XGBoost, presteren bij het voorspellen van aandelenprijzen. Op die manier kan beter worden bepaald welk model het meest geschikt is voor dit soort voorspellingen.

Een belangrijke beperking van Random Forest (RF) is dat het sterk leunt op historische patronen, waardoor het slecht presteert bij onvoorziene, zeldzame en extreme gebeurtenissen, ook wel 'black swan'-gebeurtenissen genoemd, zoals geopolitieke crises, natuurrampen of plotselinge marktcrashes, die niet voorkomen in de trainingsdata. Tegelijkertijd laat deze studie externe marktfactoren, zoals geopolitieke ontwikkelingen, economische schokken of veranderingen in regelgeving, buiten beschouwing. Dergelijke invloeden zijn moeilijk te modelleren door hun onvoorspelbare aard en het ontbreken van gestructureerde data, maar ze kunnen wel een aanzienlijke impact hebben op aandelenkoersen en investeringsgedrag. Het uitsluiten van deze externe factoren kan ertoe leiden dat het model belangrijke context mist voor een realistische marktinschatting. Toekomstig onderzoek zou kunnen verkennen hoe dergelijke invloeden toch op een zinvolle manier geïntegreerd kunnen worden in het voorspelproces.

7 Conclusie

In deze thesis werd onderzocht wat de impact is van fundamentele bedrijfsfactoren en macro-economische variabelen op de aandelenprestaties van bedrijven uit de NASDAQ-100 index. Hiervoor werd gebruikgemaakt van

het machine learning-algoritme Random Forest. Het doel van deze studie was om te achterhalen hoe accuraat Random Forest aandelenvoorspellingen kan doen en welke factoren daarbij de grootste invloed hebben.

Om dit te onderzoeken, werden vier verschillende benaderingen toegepast. De eerste methode betreft een bedrijfsspecifieke aanpak met willekeurige gegevenssplitsing, waarbij per bedrijf een afzonderlijk model werd getraind en geëvalueerd. De tweede methode maakt gebruik van een geaggregeerde dataset waarin gegevens van alle bedrijven gecombineerd werden, eveneens met een willekeurige splitsing tussen trainings- en testgegevens. De derde methode volgt een bedrijfsspecifieke aanpak met een chronologische gegevenssplitsing, waarbij gebruik werd gemaakt van de rolling window-techniek om rekening te houden met de tijdsvolgorde in de data. De vierde methode is de leave-one-ticker-out-strategie, waarbij telkens één bedrijf volledig werd uitgesloten uit de trainingsdata en uitsluitend werd gebruikt voor de testfase.

De gebruikte dataset bevat zowel fundamentele bedrijfsfactoren als macro-economische variabelen. De inputvariabelen voor het Random Forest model zijn: real_gdp (Reële BBP), fed_funds_rate (Rentevoet), cpi (Consumentenprijsindex), unemployment_rate (Werkloosheidspercentage), assets (Activa), liabilities (Schulden), revenue (Omzet), net_income (Netto winst), freecashflow (Vrije kasstroom), profit_margin (Winstmarge), yoy_revenue_growth (Jaar-op-jaar omzetgroei), eps (Winst per aandeel), roe (Rendement op eigen vermogen), der (Schuldgraad), roa (Rendement op activa). De doelvariabele in deze studie is de marktkapitalisatie (Marketcap).

De prestaties van de modellen zijn geëvalueerd op basis van meerdere prestatiemaatstaven: MAE, MSE, RMSE, R^2, RMSLE, MAPE en Adjusted R^2. Tabel 5 geeft een overzicht van drie belangrijke metrics (R^2, RMSLE en MAPE) voor de vier onderzochte methoden. Hieruit blijkt dat Methode 1 en Methode 2 de meest nauwkeurige voorspellingen leveren, met respectievelijk een R^2 van 0,93 en 0,99, en een MAPE van 8,53% en 14,94%. Methode 3 behaalt een redelijke MAPE van 10,01%, maar de verklaringskracht is merkbaar lager (R^2 = 0,55). Methode 4, gebaseerd op de leave-one-ticker-out aanpak, presteert duidelijk het zwakst met een negatieve R^2 van –17,69 en een zeer hoge MAPE van 78,00%, wat duidt op slechte generaliseerbaarheid. Hoewel de grafieken van Methode 3 en 4 visueel soms een gelijkaardig verloop vertonen tussen voorspelde en werkelijke waarden, blijkt uit de statistieken dat deze overeenkomst vooral oppervlakkig is en niet leidt tot betrouwbare voorspellingen. Een mogelijke oorzaak van de zwakke prestatie van methode 4 is dat het model geen bedrijfsspecifieke informatie kan gebruiken bij het voorspellen van de marktkapitalisatie van een volledig onbekende onderneming.

Metriek	Methode 1	Methode 2	Methode 3	Methode 4
R^2	0,93	0,99	0,55	-17,69
RMSLE	0,11	0,21	0,13	0,58
MAPE	8,53%	14,94%	10,01%	78,00%

Tabel 5. Vergelijking van de prestaties van de verschillende methoden

Bij de analyse van de feature importance van elk model komt een gelijkaardig patroon naar voren. De feature importance-scores van alle modellen zijn weergegeven in tabel 6, waarbij de belangrijkste voorspeller een score van 15 heeft gekregen en de minst belangrijke een score van 1. Uit deze resultaten blijkt dat werkloosheidsgraad, assets, BBP en de jaarlijkse omzetgroei consequent tot de belangrijkste voorspellers behoren. Daarentegen worden CPI, ROA en FCF in vrijwel alle modellen als de minst relevante voorspellers aangeduid.

	M1	M2	M3	M4	Totaal
unemployment_rate	12	14	14	13	53
assets	15	10	15	11	51
real_gdp	14	13	12	12	51
yoy_revenue_growth	8	15	8	15	46
der	9	12	9	14	44
fed_funds_rate	10	6	13	10	39
revenu	11	8	10	6	35
liabilities	13	3	11	4	31
roe	5	11	5	9	30
net_income	6	9	4	7	26
eps	7	2	7	5	21
profit_margin	2	7	3	8	20
freecashflow	3	4	6	2	15
roa	4	5	2	3	14
cpi	1	1	1	1	4

Tabel 6. Scores per methode voor feature importance

Random Forest kan, mits een correcte toepassing, als een geschikt model worden beschouwd voor het voorspellen van aandelenkoersen op basis van fundamentele en macro-economische variabelen en kan worden ingezet voor risicomanagement, beleggingsselectie en roboadviseurs. Voor een praktische implementatie van deze bevinding zou het waardevol kunnen zijn om Random Forest te integreren in een breder voorspellingssysteem, waarin meerdere machine learning-algoritmen gezamenlijk bijdragen aan de uiteindelijke voorspelling. Op die manier kan worden geprofiteerd van de sterktes van verschillende modellen. Daarnaast kan de trainingsdata in de toekomst verder worden verfijnd door aanvullende informatiebronnen te integreren, zoals sentimentanalyses of relevante nieuwsartikelen, om zo de voorspellende kracht van het model verder te versterken.

Literatuur

Akhter, Z. & Raza, H. (2024). Stock price prediction using machine learning: Evidence from pakistan stock exchange. *NUML international journal of business & management*, 1-26.

Alphavantage. (2025). *Alphavantage.co*.

Amin, M. N., Iftikhar, B., Khan, K., Javed, M. F., AbuArab, A. M. & Rehman, M. F. (2023). Prediction model for rice husk ash concrete using ai approach: Boosting and bagging algorithms. *Structures*.

Asness, C., Frazzini, A. & Pedersen, L. (2019). Quality minus junk. *Review of Accounting Studies*.

Atanasov, V. (2021). Unemployment and aggregate stock returns. *Journal of Banking and Finance*.

B., A. & Deo, M. (2021). The influence of macroeconomic variables on the stock market performance. *SSRN Electronic Journal*.

Balasubramanian, P., P, C., Badarudeen, S. & Sriraman, H. (2024). A systematic literature survey on recent trends in stock market prediction. *PeerJ. Computer science*.

Basak, S., Kar, S., Saha, S., Khaidem, L. & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 552-567.

Bayani, S. V., Mohamed, I. A. & Venkatasubbu, S. (2024). Robustness and interpretability of machine learning models in financial forecasting. *European Journal of Technology*.

Cao, Y. (2024). Prediction and analysis of corporate financial distress based on random forest model and gbdt. *SHS Web of Conferences*.

Chang, B., Meo, M., Syed, Q. & Abro, Z. (2019). Dynamic analysis of the relationship between stock prices and macroeconomic variables. *South Asian Journal of Business Studies*.

Chen, R., Dewi, C., Huang, S.-W. & Caraka, R. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*.

Chicco, D., Warrens, M. & Jurman, G. (2021). The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*.

Chopra, R. & Sharma, G. D. (2021). Application of artificial intelligence in stock market forecasting: A critique, review, and research agenda. *Journal of risk and financial management*, 1-34.

Colin-Jaeger, N. & Delcey, T. (2020). When efficient market hypothesis meets hayek on information: beyond a methodological reading. *The journal of economic methodology*, 97-116.

Das, S. K., Anwar, S., Tulsyan, U., Gupta, Y., Vudatha, R., Hassan, S. & Gardezi, I. (2024). The role of ai in financial markets: Impacts on trading, portfolio management, and price prediction. *Deleted Journal*, 1000-1006.

Digdowiseiso, K. (2023). What drives the stock returns? examining the fundamental factors on the consumer defensive sector companies. *Calitatea*, 177-186.

Dube, L. & Verster, T. (2024). Interpretability of the random forest model under class imbalance. *Data science in finance and economics*, 446-468.

El Hajj, M. & Hammoud, J. (2023). Unveiling the influence of artificial intelligence and machine learning on financial markets: A comprehensive analysis of ai applications in trading, risk management, and financial operations. *Journal of risk and financial management*, 434.

Fathali, Z., Kodia, Z. & Ben Said, L. (2022). Stock market prediction of nifty 50 index applying machine learning techniques. *Applied artificial intelligence*.

Fred. (2025). *Federal reserve economic data (fred)*.

Gholamy, A., Kreinovich, V. & Kosheleva, O. (2018). Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation.

Gursida, H. (2017). The influence of fundamental and macroeconomic analysis on stock price. , 222-234.

Huang, Y., Capretz, L. & Ho, D. (2022). Machine learning for stock prediction based on fundamental analysis.

Karabadji, N. E. I., Korba, A. A., Assi, A., Seridi-Bouchelaghem, H., Aridhi, S. & Dhifli, W. (2023). Accuracy and diversity-aware multi-objective approach for random forest construction. *Expert Syst. Appl.*.

Karch, J. (2019). Improving on adjusted r-squared. *Collabra: Psychology*.

Katterbauer, K. & Moschetta, P. (2021). An innovative artificial intelligence and natural language processing framework for asset price forecasting based on islamic finance: A case study of the saudi stock market. *Econometric research in finance : ERFIN*, 183-196.

Khan, W., Ghazanfar, M. a., Azam, M. A., Karami, A., Alyoubi, K. & Alfakeeh, A. (2022). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*.

Khattak, B. H. A., Shafi, I., Khan, A. S., Flores, E. S., Lara, R. G., Samad, M. A. & Ashraf, I. (2023). A systematic survey of ai models in financial market forecasting for profitability analysis. *IEEE access*, 125359-125380.

Kumar, P., Hota, L., Tikkiwal, V. A. & Kumar, A. (2024). Analysing forecasting of stock prices: An explainable ai approach. *Procedia computer science*, 2009-2016.

Lestari, A. & Hadi, A. (2024). Dynamics of stock prices in the sharia capital market: The important role of financial fundamental analysis. *Asy Syar'iyyah: Jurnal Ilmu Syari'ah Dan Perbankan Islam*.

Lin, C. Y. & Lobo Marques, J. A. (2024). Stock market prediction using artificial intelligence: A systematic review of systematic reviews. *Social sciences & humanities open.*

Liu, W. (2020). How useful is it for banks to analyze financial statements. *American Journal of Industrial and Business Management.*

Lohrmann, C. & Luukka, P. (2019). Classification of intraday s&p500 returns with a random forest. *International Journal of Forecasting*, 390-407.

Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLoS ONE.*

Mir, A. A., Kearfott, K., Çelebi, F. & Rafique, M. Z. (2022). Imputation by feature importance (ibfi): A methodology to envelop machine learning method for imputing missing patterns in time series data. *PLoS ONE.*

Nasdaq. (2025). *Nasdaq-100 index quotes.*

Nti, I. K., Adekoya, A. F. & Weyori, B. A. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction. *Journal of big data*, 1-40.

OpenAI. (2023). *Chatgpt [large language model].*

Paduri, A., Darapaneni, N. & Khanpuri, A. (2024). Utilizing fundamental analysis to predict stock prices.

Park, H., Kim, Y. & Kim, H. Y. (2021). Stock market forecasting using a multi-task approach integrating long short-term memory and the random forest framework. *Appl. Soft Comput..*

Pham, H. V., Lam, H. P., Duy, L. N., Pham, T. B. & Trinh, T. D. (2024). An improved convolutional recurrent neural network for stock price forecasting. *IAES international journal of artificial intelligence.*

Pramudito, A. (2023). The effect of macroeconomics on stock prices through financial performance as an intervening variable. *Journal of World Science.*

Probst, P., Wright, M. N. & Boulesteix, A. (2018). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.*

Purnomo, K., Wijaya, R. A., Fajar, M. & Suri, P. A. (2024). Comparative analysis of bilstm deep learning model and random forest regressor performance on indonesian nickel mining company stock prices. *Procedia computer science*, 778-786.

Sadorsky, P. A. (2021). A random forests approach to predicting clean energy stock prices. *Journal of risk and financial management*, 1-20.

Sicheng, J. (2024). Predict stock market price by applying ann, svm and random forest. *SHS web of conferences.*

Sinha, R. & Jain, R. (2016). Beyond traditional analysis: Exploring random forests for stock market prediction.

Sjahrunnisa, A., Suciati, N. & Hidayati, S. C. (2024). Combination of historical stock data and external factors in improving stock price prediction performance. *Jurnal EECCIS*, 30-36.

Sonkavde, G., Dharrao, D. S., Bongale, A. M., Deokate, S. T., Doreswamy, D. & Bhat, S. K. (2023). Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications. *International journal of financial studies*, 94.

Subasi, A., Amir, F., Bagedo, K., Shams, A. & Sarirete, A. (2021). Stock market prediction using machine learning. *Procedia Computer Science*, 173-179.

Suyanto, S., Safitri, J. & Adji, A. P. (2021). Fundamental and technical factors on stock prices in pharmaceutical and cosmetic companies. *Finance, Accounting and Business Analysis*, 67-73.

Tan, Z., Yan, Z. & Zhu, G. (2019). Stock selection with random forest: An exploitation of excess return in the chinese stock market. *Heliyon.*

Verma, R. & Bansal, R. (2021). Impact of macroeconomic variables on the performance of stock exchange: a systematic review. *International Journal of Emerging Markets.*

Vijh, M., Chandola, D., Tikkiwal, V. A. & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia computer science*, 599-606.

Vuong, P. H., Phu, L. H., Nguyen, T. H. V., Duy, L. N., Bao, P. T. & Trinh, T. (2024). A bibliometric literature review of stock price forecasting: From statistical model to deep learning approach. *Science Progress.*

Wang, S. (2024). The application of artificial intelligence-based risk management models in financial markets.

Xiao, F. & Ke, J. (2021). Pricing, management and decision-making of financial markets with artificial intelligence: introduction to the issue. *Financial innovation*, 85-85.

Yuan, X., Yuan, J., Jiang, T. & Ain, Q. U. (2020). Integrated long-term stock selection models based on feature selection and machine learning algorithms for china stock market. *IEEE access*, 22672-22685.

Zhao, X., Wu, Y., Lee & Cui, W. (2019). iforest: Interpreting random forests via visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 407-416.

Zhong, S. & Hitchcock, D. (2021). S&p 500 stock price prediction using technical, fundamental and text data. *Statistics, Optimization & Information Computing*, 769–788.

Zhou, Z. & Hooker, G. (2019). Unbiased measurement of feature importance in tree-based methods. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1-21.

Zhu, H., Vigren, O. & Söderberg, I.-L. (2024). Implementing artificial intelligence empowered financial advisory services: A literature review and critical research agenda. *Journal of business research.*

8 Bijlagen

Symbool	Bedrijfsnaam	Sector	Rijen	Jaren
AAPL	Apple Inc.	Informatietechnologie	183	15.25
ABNB	Airbnb	Retail & consumentengoederen	42	3.50
ADBE	Adobe Inc.	Informatietechnologie	186	15.50
ADI	Analog Devices	Informatietechnologie	184	15.33
ADP	Automatic Data Processing	Informatietechnologie	183	15.25
ADSK	Autodesk	Informatietechnologie	189	15.75
AEP	American Electric Power Company	Energie & nutsbedrijven	168	14.00
AMAT	Applied Materials	Informatietechnologie	183	15.25
AMD	Advanced Micro Devices	Informatietechnologie	164	13.67
AMGN	Amgen	Gezondheidszorg	186	15.50
AMZN	Amazon	Retail & consumentengoederen	185	15.42
ANSS	ANSYS, Inc.	Informatietechnologie	177	14.75
APP	Applovin Corporation	Informatietechnologie	38	3.17
ARM	Arm Holdings	Informatietechnologie	15	1.25
ASML	ASML Holding	Informatietechnologie	243	20.25
AVGO	Broadcom	Informatietechnologie	80	6.67
AXON	Axon Enterprise	Informatietechnologie	162	13.50
AZN	AstraZeneca	Gezondheidszorg	243	20.25
BIIB	Biogen	Gezondheidszorg	189	15.75
BKNG	Booking Holdings	Retail & consumentengoederen	174	14.50
BKR	Baker Hughes Company	Energie & nutsbedrijven	84	7.00
Ccep	Coca-Cola Europacific Partners	Retail & consumentengoederen	243	20.25
CDNS	Cadence Design Systems	Informatietechnologie	174	14.50
CDW	CDW Corporation	Informatietechnologie	135	11.25
CEG	Constellation Energy	Energie & nutsbedrijven	31	2.58
CHTR	Charter Communications	Communicatiediensten	162	13.50
CMCSA	Comcast	Communicatiediensten	187	15.58
COST	Costco Wholesale	Retail & consumentengoederen	180	15.00
CPRT	Copart	Retail & consumentengoederen	161	13.42
CRWD	CrowdStrike	Informatietechnologie	64	5.33
CSCO	Cisco Systems	Informatietechnologie	180	15.00
CSGP	CoStar Group	Informatietechnologie	158	13.17
CSX	CSX Corporation	Industrie	172	14.33
CTAS	Cintas Corporation	Industrie	168	14.00
CTSH	Cognizant Technology Solutions	Informatietechnologie	181	15.08
DASH	DoorDash	Retail & consumentengoederen	45	3.75
DDOG	Datadog	Informatietechnologie	60	5.00
DXCM	DexCom	Gezondheidszorg	158	13.17
EA	Electronic Arts	Communicatiediensten	189	15.75
EXC	Exelon	Energie & nutsbedrijven	183	15.25
FANG	Diamondback Energy	Energie & nutsbedrijven	142	11.83
FAST	Fastenal	Industrie	185	15.42
FTNT	Fortinet	Informatietechnologie	146	12.17
GEHC	GE HealthCare Technologies	Gezondheidszorg	23	1.92
GFS	GlobalFoundries	Informatietechnologie	38	3.17
GILD	Gilead Sciences	Gezondheidszorg	183	15.25
GOOG	Alphabet	Communicatiediensten	108	9.00
HON	Honeywell International	Industrie	186	15.50
IDXX	IDEXX Laboratories	Gezondheidszorg	173	14.42
INTC	Intel	Informatietechnologie	186	15.50
INTU	Intuit	Informatietechnologie	185	15.42
ISRG	Intuitive Surgical	Gezondheidszorg	177	14.75
KDP	Keurig Dr Pepper Inc.	Retail & consumentengoederen	180	15.00

Symbool	Bedrijfsnaam	Sector	Rijen	Jaren
KHC	The Kraft Heinz Company	Retail & consumentengoederen	113	9.42
KLAC	KLA Corporation	Informatietechnologie	174	14.50
LIN	Linde plc Ordinary Shares	Industrie	71	5.92
LRCX	Lam Research	Informatietechnologie	159	13.25
LULU	Lululemon Athletica	Retail & consumentengoederen	172	14.33
MAR	Marriott International	Retail & consumentengoederen	180	15.00
MCHP	Microchip Technology	Informatietechnologie	188	15.67
MDB	MongoDB	Informatietechnologie	81	6.75
MDLZ	Mondelez International	Retail & consumentengoederen	186	15.50
MELI	MercadoLibre	Retail & consumentengoederen	174	14.50
META	Meta Platforms	Communicatiediensten	151	12.58
MNST	Monster Beverage Corporation	Retail & consumentengoederen	166	13.83
MRVL	Marvell Technology	Informatietechnologie	46	3.83
MSFT	Microsoft	Informatietechnologie	183	15.25
MSTR	MicroStrategy	Informatietechnologie	136	11.33
MU	Micron Technology	Informatietechnologie	174	14.50
NFLX	Netflix	Communicatiediensten	179	14.92
NVDA	NVIDIA	Informatietechnologie	179	14.92
NXPI	NXP Semiconductors	Informatietechnologie	66	5.50
ODFL	Old Dominion Freight Line	Industrie	167	13.92
ON	ON Semiconductor	Informatietechnologie	159	13.25
ORLY	O'Reilly Automotive	Retail & consumentengoederen	177	14.75
PANW	Palo Alto Networks	Informatietechnologie	145	12.08
PAYX	Paychex	Informatietechnologie	186	15.50
PCaR	PACCAR Inc.	Industrie	186	15.50
PDD	PDD Holdings Inc.	Retail & consumentengoederen	74	6.17
PEP	PepsiCo	Retail & consumentengoederen	180	15.00
PLTR	Palantir Technologies Inc.	Informatietechnologie	34	2.83
PYPL	PayPal Holdings	Informatietechnologie	110	9.17
QCOM	Qualcomm	Informatietechnologie	174	14.50
REGN	Regeneron Pharmaceuticals	Gezondheidszorg	177	14.75
ROP	Roper Technologies	Industrie	183	15.25
ROST	Ross Stores	Retail & consumentengoederen	176	14.67
SBUX	Starbucks	Retail & consumentengoederen	180	15.00
SNPS	Synopsys	Informatietechnologie	174	14.50
TEAM	Atlassian	Informatietechnologie	33	2.75
TMUS	T-Mobile US	Communicatiediensten	36	3.00
TSLA	Tesla	Retail & consumentengoederen	160	13.33
TTD	The Trade Desk	Communicatiediensten	92	7.67
TTWO	Take-Two Interactive	Communicatiediensten	171	14.25
TXN	Texas Instruments	Informatietechnologie	180	15.00
VRSK	Verisk Analytics	Informatietechnologie	165	13.75
VRTX	Vertex Pharmaceuticals	Gezondheidszorg	180	15.00
WBD	Warner Bros. Discovery	Communicatiediensten	173	14.42
WDAY	Workday	Informatietechnologie	146	12.17
XEL	Xcel Energy	Energie & nutsbedrijven	186	15.50
ZS	Zscaler	Informatietechnologie	79	6.58

Tabel 7: Dataset NASDAQ-100

Symbool	MAE	MSE	RMSE	MAPE	RMSLE	R^2	AdjustedR^2
AAPL	114.62B	34873850274106.40B	186.75B	8.60	0.11	0.98	0.96
ABNB	5.01B	68214402586.35B	8.26B	5.70	0.10	0.72	1.32
ADBE	7.28B	129130847116.19B	11.36B	6.78	0.08	0.98	0.97
ADI	1.88B	8417683612.71B	2.90B	5.29	0.07	0.99	0.99
ADP	3.09B	22990719425.38B	4.79B	5.26	0.08	0.98	0.96
ADSK	1.57B	6065684754.53B	2.46B	6.66	0.09	0.98	0.97
AEP	1.28B	3696234031.33B	1.92B	3.98	0.05	0.97	0.95
AMAT	4.59B	59196136681.39B	7.69B	7.52	0.09	0.98	0.97
AMD	8.64B	241486220414.40B	15.54B	16.89	0.22	0.97	0.94
AMGN	4.99B	47158514448.81B	6.87B	4.52	0.06	0.97	0.95
AMZN	46.38B	6232978038312.91B	78.95B	7.54	0.09	0.99	0.98
ANSS	0.96B	2143863288.69B	1.46B	5.68	0.07	0.98	0.97
APP	3.54B	20847221062.39B	4.57B	41.14	0.38	0.58	1.37
ARM	13.30B	204347838792.13B	14.30B	11.95	0.13	0.82	1.03
ASML	6.90B	143592362823.81B	11.98B	9.52	0.11	0.99	0.99
AVGO	12.38B	268881494781.96B	16.40B	5.84	0.07	0.99	-Inf
AXON	0.88B	2031602187.89B	1.43B	16.40	0.20	0.99	0.98
AZN	4.04B	27777966176.61B	5.27B	5.90	0.07	0.93	0.90
BIIB	3.18B	20671097217.79B	4.55B	8.02	0.10	0.95	0.92
BKNG	4.22B	34171404228.63B	5.85B	6.14	0.08	0.97	0.94
BKR	4.55B	43421845195.95B	6.59B	16.44	0.22	0.87	-1.02
Ccep	1.00B	1775956441.69B	1.33B	6.63	0.09	0.96	0.93
CDNS	1.44B	6677007329.31B	2.58B	4.97	0.06	0.99	0.99
CDW	1.01B	2025241222.53B	1.42B	6.44	0.08	0.97	0.94
CEG	3.10B	19825238741.30B	4.45B	6.41	0.08	0.92	1.05
CHTR	4.56B	34531881821.35B	5.88B	8.37	0.10	0.98	0.96
CMCSA	6.40B	59505536812.41B	7.71B	5.59	0.07	0.98	0.97
COST	5.70B	67601638834.40B	8.22B	4.92	0.06	0.99	0.99
CPRT	1.60B	7970079443.99B	2.82B	8.60	0.11	0.97	0.94
CRWD	5.03B	72315343339.39B	8.50B	14.54	0.19	0.87	1.54
CSCO	7.06B	80381323789.68B	8.97B	4.65	0.06	0.96	0.92
CSGP	0.74B	1306406548.98B	1.14B	5.18	0.07	0.99	0.98
CSX	2.30B	8887043369.07B	2.98B	5.14	0.06	0.97	0.95
CTAS	0.78B	1629400809.47B	1.28B	3.80	0.05	0.99	0.99
CTSH	1.90B	11297998327.90B	3.36B	5.41	0.08	0.91	0.85
DASH	2.82B	10411973580.24B	3.23B	11.69	0.14	0.89	1.13
DDOG	3.61B	16634561139.90B	4.08B	15.24	0.19	0.86	1.39
DXCM	1.78B	8357742009.82B	2.89B	11.36	0.14	0.97	0.94
EA	1.49B	3856405211.59B	1.96B	8.88	0.10	0.98	0.97
EXC	2.16B	11129916951.98B	3.34B	6.01	0.08	0.87	0.78
FANG	1.01B	2230485246.12B	1.49B	6.29	0.07	0.97	0.94
FAST	1.16B	2680455611.11B	1.64B	6.62	0.08	0.98	0.96
FTNT	1.87B	13115443154.80B	3.62B	9.53	0.13	0.97	0.93
GEHC	3.74B	14702269133.61B	3.83B	9.50	0.10	0.97	1.01
GFS	2.96B	13402979456.17B	3.66B	8.81	0.11	0.50	1.44
GILD	4.76B	39847464736.60B	6.31B	5.77	0.07	0.96	0.94
GOOG	52.25B	7263849475042.40B	85.23B	4.34	0.06	0.98	0.92
HON	3.58B	30065035189.24B	5.48B	4.69	0.06	0.98	0.97
IDXX	1.28B	4228570577.19B	2.06B	5.59	0.07	0.99	0.97
INTC	7.30B	92706779505.62B	9.63B	5.34	0.07	0.96	0.94
INTU	6.65B	131055703263.68B	11.45B	8.42	0.12	0.96	0.94
ISRG	4.22B	65158524611.41B	8.07B	5.90	0.07	0.99	0.97
KDP	1.32B	3366038743.43B	1.83B	5.02	0.06	0.99	0.98
KHC	3.16B	16388741818.32B	4.05B	6.58	0.09	0.97	0.92
KLAC	1.51B	4826613348.24B	2.20B	4.34	0.06	0.99	0.99

Symbool	MAE	MSE	RMSE	MAPE	RMSLE	R^2	AdjustedR^2
LIN	9.49B	143184177169.61B	11.97B	7.40	0.09	0.92	2.17
LRCX	3.41B	26796924734.66B	5.18B	8.12	0.09	0.98	0.96
LULU	1.72B	8967440993.72B	2.99B	9.27	0.11	0.97	0.95
MAR	2.47B	15277457748.25B	3.91B	8.08	0.12	0.96	0.92
MCHP	1.24B	4104991053.97B	2.03B	6.62	0.09	0.98	0.97
MDB	2.89B	13541695364.87B	3.68B	18.93	0.21	0.84	-1.61
MDLZ	2.97B	12781020917.93B	3.58B	4.91	0.06	0.96	0.93
MELI	2.93B	24443760753.17B	4.94B	11.18	0.14	0.98	0.96
META	44.14B	6697212289593.68B	81.84B	8.66	0.12	0.93	0.85
MNST	1.42B	3937755272.45B	1.98B	5.96	0.08	0.99	0.98
MRVL	3.98B	26534575075.41B	5.15B	9.23	0.12	0.79	1.31
MSFT	57.66B	7424484756238.27B	86.17B	5.47	0.07	0.99	0.99
MSTR	0.20B	95545387.81B	0.31B	10.50	0.14	0.71	0.36
MU	3.85B	31313603182.84B	5.60B	8.57	0.10	0.98	0.96
NFLX	8.56B	200077544123.56B	14.14B	12.49	0.16	0.97	0.95
NVDA	28.45B	10296573335796.50B	101.47B	9.52	0.12	0.99	0.98
NXPI	4.26B	25889115353.79B	5.09B	11.86	0.14	0.86	1.88
ODFL	1.24B	3982947260.01B	2.00B	8.97	0.12	0.98	0.97
ON	1.30B	3996515601.92B	2.00B	8.10	0.10	0.97	0.95
ORLY	1.21B	2999449327.31B	1.73B	5.11	0.07	0.99	0.98
PANW	2.18B	11710760460.93B	3.42B	6.34	0.07	0.99	0.98
PAYX	0.90B	2003173752.07B	1.42B	3.62	0.05	0.99	0.98
PCaR	2.10B	11386247872.49B	3.37B	7.90	0.10	0.93	0.88
PDD	57.19B	5130593470163.07B	71.63B	26.27	0.27	0.91	2.29
PEP	5.26B	47486130674.48B	6.89B	3.17	0.04	0.98	0.97
PLTR	10.70B	245267674962.47B	15.66B	18.62	0.26	0.94	1.04
PYPL	9.78B	186834493804.44B	13.67B	8.29	0.10	0.97	0.91
QCOM	7.75B	93409424403.84B	9.66B	7.68	0.10	0.94	0.89
REGN	4.79B	63459083149.67B	7.97B	8.60	0.11	0.95	0.92
ROP	1.16B	3341887408.38B	1.83B	4.68	0.06	0.99	0.99
ROST	1.57B	3779466382.24B	1.94B	6.14	0.07	0.98	0.96
SBUX	3.57B	22939189771.70B	4.79B	5.51	0.07	0.98	0.97
SNPS	1.05B	2638647870.95B	1.62B	3.58	0.05	1.00	0.99
TEAM	8.76B	121635384106.47B	11.03B	15.59	0.20	0.32	1.45
TMUS	0.80B	1040296735.45B	1.02B	17.05	0.22	0.25	1.65
TSLA	33.11B	4831724311695.18B	69.51B	14.40	0.19	0.96	0.93
TTD	2.28B	7972213085.62B	2.82B	16.10	0.20	0.97	0.82
TTWO	0.77B	1375997862.33B	1.17B	8.34	0.10	0.98	0.97
TXN	3.56B	24901835281.91B	4.99B	4.02	0.05	0.99	0.99
VRSK	1.05B	2487239161.24B	1.58B	4.76	0.06	0.97	0.95
VRTX	2.23B	9595944001.18B	3.10B	6.63	0.09	0.99	0.98
WBD	1.77B	7067664160.87B	2.66B	7.78	0.11	0.82	0.67
WDAY	2.44B	10451683177.37B	3.23B	6.73	0.08	0.98	0.95
XEL	0.85B	1304682404.68B	1.14B	3.76	0.05	0.99	0.98
ZS	1.70B	4137934972.11B	2.03B	11.80	0.14	0.95	-Inf
Gemiddeld	7.55B	863995564646.80B	12.21B	8.53	0.11	0.93	0.98
Mediaan	3.03B	15833099783.28B	3.98B	6.75	0.09	0.97	0.97

Tabel 8: Resultaten methode 1

Symbool	MAE	MSE	RMSE	MAPE	RMSLE	R^2	AdjustedR^2
AAPL	249.49B	90684936995179.0B	301.14B	8.75	0.11	0.6	0.29
ABNB	7.35B	69408648659.03B	8.33B	7.96	0.09	0.35	1.65
ADBE	26.01B	1149686557984.86B	33.91B	12.46	0.16	0.43	-0.0
ADI	7.18B	80654029168.27B	8.98B	7.61	0.1	0.55	0.19
ADP	6.57B	65862337389.97B	8.12B	6.45	0.08	0.35	-0.17
ADSK	4.97B	32113541948.12B	5.67B	10.43	0.11	0.47	0.06
AEP	3.02B	13346064366.82B	3.65B	6.43	0.08	0.31	-0.34
AMAT	13.73B	304086604945.72B	17.44B	11.49	0.14	0.73	0.51
AMD	24.22B	1074199504617.46B	32.77B	13.11	0.18	0.77	0.53
AMGN	9.9B	152438967210.13B	12.35B	6.94	0.09	0.61	0.31
AMZN	153.12B	29891100504299.9B	172.89B	10.45	0.12	0.85	0.72
ANSS	1.84B	5698203184.27B	2.39B	7.24	0.09	0.49	0.06
APP	5.72B	60614389660.31B	7.79B	18.72	0.29	0.74	1.2
ARM	9.19B	102818801494.93B	10.14B	6.67	0.07	0.2	1.11
ASML	32.83B	1512015630679.63B	38.88B	12.49	0.15	0.54	0.31
AVGO	115.95B	23856701701465.8B	154.46B	18.09	0.28	0.75	4.82
AXON	2.03B	6785360558.4B	2.6B	13.48	0.17	0.89	0.77
AZN	9.51B	290608862816.46B	17.05B	7.97	0.13	0.78	0.67
BIIB	5.0B	38595976487.38B	6.21B	15.86	0.18	0.33	-0.18
BKNG	8.51B	103079856796.71B	10.15B	8.58	0.1	0.8	0.62
BKR	6.26B	88164051201.46B	9.39B	13.49	0.19	0.58	-inf
Ccep	2.1B	6298140833.27B	2.51B	8.45	0.1	0.8	0.7
CDNS	5.25B	40868498459.06B	6.39B	8.55	0.11	0.91	0.83
CDW	1.87B	4974174359.24B	2.23B	6.73	0.08	0.67	0.15
CEG	12.65B	207731350253.14B	14.41B	19.31	0.25	0.29	1.43
CHTR	6.93B	104642485218.29B	10.23B	12.9	0.16	0.56	0.12
CMCSA	12.23B	277543413982.53B	16.66B	7.58	0.1	0.72	0.5
COST	23.35B	886148226869.55B	29.77B	7.94	0.1	0.91	0.83
CPRT	3.84B	41894094242.0B	6.47B	9.2	0.16	0.59	0.18
CRWD	12.41B	208744278741.94B	14.45B	18.27	0.23	0.29	3.12
CSCO	11.5B	207192027729.74B	14.39B	5.7	0.07	0.43	-0.05
CSGP	2.55B	9674313063.05B	3.11B	8.18	0.1	0.39	-0.26
CSX	3.66B	22483108400.42B	4.74B	5.5	0.07	0.42	-0.1
CTAS	3.32B	16387609837.14B	4.05B	6.65	0.08	0.89	0.78
CTSH	2.17B	6568799643.77B	2.56B	6.08	0.07	0.69	0.45
DASH	6.87B	58628933981.95B	7.66B	13.53	0.16	0.11	1.89
DDOG	3.63B	21086703036.91B	4.59B	9.33	0.12	0.24	2.68
DXCM	6.54B	66866063241.64B	8.18B	16.75	0.2	0.08	-0.9
EA	2.0B	5793997989.65B	2.41B	5.59	0.07	0.2	-0.4
EXC	2.51B	16560528391.18B	4.07B	5.9	0.09	0.36	-0.16
FANG	3.75B	43124440693.27B	6.57B	10.67	0.18	0.64	0.16
FAST	2.35B	9095912928.15B	3.02B	6.47	0.08	0.72	0.49
FTNT	4.76B	32595175683.42B	5.71B	9.44	0.11	0.36	-0.43
GEHC	2.47B	8060210425.2B	2.84B	6.29	0.07	0.27	1.24
GFS	2.35B	11190149686.76B	3.35B	10.45	0.14	0.18	1.64
GILD	5.09B	47194750518.61B	6.87B	5.62	0.08	0.54	0.18
GOOG	111.31B	19267385393793.7B	138.81B	6.01	0.08	0.83	0.29
HON	5.8B	55229407230.81B	7.43B	4.41	0.06	0.35	-0.14
IDXX	3.94B	26386015278.11B	5.14B	10.69	0.13	0.5	0.05
INTC	19.15B	587064997018.72B	24.23B	14.92	0.18	0.63	0.35
INTU	11.4B	225768193666.49B	15.03B	7.98	0.1	0.69	0.45
ISRG	12.92B	224202502385.26B	14.97B	10.88	0.12	0.91	0.84
KDP	2.26B	7330050532.91B	2.71B	4.69	0.06	0.59	0.25
KHC	2.17B	7542473263.53B	2.75B	5.23	0.07	0.39	-1.24
KLAC	6.77B	68571321337.62B	8.28B	9.4	0.11	0.88	0.77

Symbool	MAE	MSE	RMSE	MAPE	RMSLE	R^2	AdjustedR^2
LIN	10.05B	155586165116.53B	12.47B	4.79	0.06	0.67	3.28
LRCX	10.56B	149340035265.62B	12.22B	12.65	0.15	0.78	0.54
LULU	4.9B	36091015991.99B	6.01B	12.13	0.14	0.39	-0.15
MAR	4.8B	38590612478.94B	6.21B	7.9	0.1	0.68	0.41
MCHP	5.83B	71154906471.12B	8.44B	13.78	0.17	0.13	-0.52
MDB	4.17B	26548496586.42B	5.15B	17.17	0.22	0.29	-inf
MDLZ	4.86B	37370543166.89B	6.11B	5.34	0.07	0.33	-0.19
MELI	7.55B	97521493203.71B	9.88B	12.21	0.16	0.7	0.44
META	100.54B	15480437460062.2B	124.42B	14.58	0.19	0.94	0.88
MNST	3.18B	12781134348.85B	3.58B	6.08	0.07	0.65	0.32
MRVL	9.98B	228462450564.66B	15.11B	13.0	0.21	0.03	2.24
MSFT	179.58B	45530240384994.7B	213.38B	7.46	0.09	0.84	0.71
MSTR	1.01B	2550631966.43B	1.6B	22.67	0.33	0.79	0.49
MU	9.67B	206608470490.86B	14.37B	10.54	0.15	0.68	0.39
NFLX	30.89B	1591379290116.77B	39.89B	17.47	0.22	0.78	0.59
NVDA	213.56B	105314564968886.0B	324.52B	17.28	0.22	0.95	0.91
NXPI	4.33B	28422607432.14B	5.33B	7.27	0.09	0.63	2.59
ODFL	4.77B	99972736786.71B	10.0B	9.62	0.18	0.62	0.26
ON	3.34B	19247310760.46B	4.39B	10.34	0.14	0.35	-0.33
ORLY	2.94B	13022028777.73B	3.61B	5.54	0.07	0.87	0.77
PANW	7.79B	103113645917.26B	10.15B	9.71	0.13	0.89	0.74
PAYX	2.55B	10326138687.88B	3.21B	5.73	0.07	0.1	-0.59
PCaR	3.92B	39243842689.6B	6.26B	8.39	0.14	0.71	0.5
PDD	106.49B	18120272624526.3B	134.61B	14.79	0.2	0.28	6.07
PEP	8.39B	123164725401.11B	11.1B	3.52	0.05	0.15	-0.57
PLTR	8.88B	137601725017.23B	11.73B	12.92	0.18	0.93	1.04
PYPL	16.12B	498837255840.99B	22.33B	22.9	0.25	0.08	-2.86
QCOM	16.11B	378439726166.84B	19.45B	10.08	0.12	0.55	0.16
REGN	8.29B	111449978528.31B	10.56B	9.4	0.12	0.64	0.33
ROP	2.45B	8954176004.89B	2.99B	4.78	0.06	0.81	0.65
ROST	2.99B	13151776796.73B	3.63B	8.07	0.1	0.76	0.55
SBUX	7.93B	109945054703.42B	10.49B	7.3	0.1	0.48	0.05
SNPS	5.14B	46378038064.84B	6.81B	7.75	0.1	0.85	0.71
TEAM	6.06B	64269051882.61B	8.02B	10.96	0.14	0.42	1.35
TMUS	0.24B	108218936.71B	0.33B	6.02	0.09	0.16	1.66
TSLA	132.22B	26071511033538.2B	161.47B	18.72	0.22	0.27	-0.51
TTD	4.08B	24884732212.75B	4.99B	9.73	0.12	0.65	-2.17
TTWO	2.0B	7016919500.75B	2.65B	8.13	0.11	0.73	0.48
TXN	10.31B	145618067031.98B	12.07B	6.49	0.08	0.43	-0.06
VRSK	1.93B	5361204838.41B	2.32B	6.08	0.07	0.66	0.33
VRTX	7.32B	77987934391.79B	8.83B	8.08	0.1	0.9	0.81
WBD	4.15B	28182969692.39B	5.31B	16.49	0.22	0.37	-0.19
WDAY	5.58B	46568979470.02B	6.82B	9.61	0.12	0.65	0.22
XEL	1.85B	4793715635.24B	2.19B	5.25	0.06	0.53	0.18
ZS	3.61B	18975721042.68B	4.36B	12.33	0.15	0.23	12.55
Gemiddeld	20.17B	3877497943328.19B	25.7B	10.01	0.13	0.55	0.66
Mediaan	5.82B	62441720771.46B	7.9B	8.97	0.11	0.6	0.42

Tabel 9: Resultaten methode 3

Symbool	MAE	MSE	RMSE	MAPE	RMSLE	R^2	AdjustedR^2
AAPL	607.49B	536501675866083.0B	732.46B	0.68	0.58	0.44	0.39
ABNB	35.43B	2075592951189.36B	45.56B	0.38	0.52	-6.9	-11.46
ADBE	43.83B	4052935387523.65B	63.66B	0.33	0.49	0.5	0.45
ADI	9.68B	243367610856.98B	15.6B	0.22	0.25	0.76	0.74
ADP	12.57B	242888799202.48B	15.58B	0.23	0.32	0.69	0.66
ADSK	5.51B	73654940451.85B	8.58B	0.25	0.33	0.8	0.78
AEP	22.4B	969234132548.39B	31.13B	0.7	0.57	-7.35	-8.17
AMAT	30.51B	1438635235502.0B	37.93B	0.69	0.55	0.27	0.2
AMD	34.98B	3202734569353.29B	56.59B	2.25	1.18	0.51	0.46
AMGN	26.61B	2611950544302.47B	51.11B	0.29	0.36	-1.22	-1.41
AMZN	259.88B	143324033687216.0B	378.58B	0.43	0.71	0.65	0.61
ANSS	3.71B	69812659062.48B	8.36B	0.18	0.24	0.23	0.16
APP	23.59B	626898661642.41B	25.04B	2.16	1.14	-4.68	-8.54
ARM	74.18B	6734065492932.74B	82.06B	0.57	0.96	-4.14	72.94
ASML	36.45B	4562755682425.59B	67.55B	0.45	0.55	0.58	0.56
AVGO	88.63B	24693434758386.4B	157.14B	0.26	0.4	0.39	0.25
AXON	5.39B	56476372930.7B	7.52B	1.93	1.07	-0.33	-0.47
AZN	33.1B	1492735691908.48B	38.64B	0.5	0.45	-1.17	-1.31
BIIB	22.47B	894478155962.13B	29.91B	0.55	0.5	-1.18	-1.36
BKNG	22.79B	1148863338230.1B	33.89B	0.32	0.43	-0.35	-0.47
BKR	29.58B	1328713407525.09B	36.45B	1.09	0.78	-3.51	-4.5
Ccep	18.29B	817074355409.05B	28.58B	1.14	0.78	-15.32	-16.4
CDNS	6.5B	103561126514.06B	10.18B	0.72	0.61	0.79	0.78
CDW	17.57B	352080696960.73B	18.76B	1.36	0.86	-3.75	-4.35
CEG	67.47B	4955487111917.71B	70.4B	2.19	1.16	-17.17	-35.34
CHTR	107.15B	21510016700346.6B	146.66B	2.07	1.1	-13.01	-14.45
CMCSA	393.13B	225991949848778.0B	475.39B	2.47	1.24	-70.98	-77.29
COST	314.58B	112346726436696.0B	335.18B	3.84	1.55	-11.77	-12.93
CPRT	5.03B	53244288152.57B	7.3B	0.4	0.37	0.78	0.76
CRWD	11.73B	228877315979.59B	15.13B	0.25	0.36	0.41	0.23
CSCO	27.48B	1431896530089.17B	37.84B	0.17	0.2	0.28	0.21
CSGP	3.79B	48853977231.61B	6.99B	0.38	0.39	0.67	0.63
CSX	11.34B	422088867145.84B	20.54B	0.23	0.27	-0.16	-0.27
CTAS	4.63B	35995850696.34B	6.0B	0.37	0.36	0.88	0.86
CTSH	24.22B	919725695083.34B	30.33B	0.7	0.58	-7.75	-8.55
DASH	21.75B	589999470590.28B	24.29B	0.69	0.57	-1.61	-2.96
DDOG	7.52B	82510412312.99B	9.08B	0.25	0.43	0.19	-0.09
DXCM	5.51B	71768786184.92B	8.47B	0.34	0.5	0.77	0.75
EA	11.59B	213339245390.39B	14.61B	0.77	0.62	-0.19	-0.29
EXC	43.31B	5541872107441.61B	74.44B	1.31	0.83	-102.75	-112.06
FANG	30.58B	1490082651465.2B	38.6B	2.59	1.26	-11.61	-13.11
FAST	5.07B	41432829307.61B	6.44B	0.25	0.26	0.58	0.54
FTNT	5.39B	54016745163.36B	7.35B	0.37	0.37	0.84	0.82
GEHC	19.79B	662651051301.64B	25.74B	0.57	0.49	-59.35	-188.66
GFS	20.06B	451569841515.24B	21.25B	0.7	0.55	-21.47	-36.8
GILD	44.17B	6510344768393.17B	80.69B	0.47	0.46	-4.37	-4.86
GOOG	216.63B	89517193191025.7B	299.19B	0.17	0.25	0.69	0.64
HON	29.78B	1825290009298.89B	42.72B	0.45	0.45	-0.27	-0.39
IDXX	4.47B	47240564737.89B	6.87B	0.2	0.27	0.81	0.79
INTC	115.6B	18026820165030.4B	134.26B	0.74	0.57	-7.23	-7.95
INTU	39.43B	2695293413219.77B	51.92B	0.72	0.87	0.07	-0.01
ISRG	18.14B	825208040259.31B	28.73B	0.26	0.39	0.56	0.52
KDP	13.56B	375586562480.87B	19.38B	0.49	0.43	-0.35	-0.47
KHC	53.41B	3703235459717.91B	60.85B	1.18	0.81	-4.52	-5.37
KLAC	8.3B	123155358679.65B	11.1B	0.43	0.4	0.82	0.8

Symbool	MAE	MSE	RMSE	MAPE	RMSLE	R^2	AdjustedR^2
LIN	40.39B	2285835681479.94B	47.81B	0.26	0.35	-0.47	-0.87
LRCX	19.43B	617577061598.17B	24.85B	0.86	0.66	0.44	0.38
LULU	4.68B	48427640981.53B	6.96B	0.3	0.35	0.83	0.81
MAR	15.77B	391924394557.79B	19.8B	0.48	0.42	-0.12	-0.22
MCHP	6.63B	121249326051.56B	11.01B	0.31	0.32	0.55	0.51
MDB	3.44B	17538219413.13B	4.19B	0.39	0.4	0.81	0.77
MDLZ	21.91B	977250000778.44B	31.26B	0.37	0.38	-3.14	-3.51
MELI	6.02B	119261267043.29B	10.92B	0.23	0.34	0.86	0.84
META	153.38B	32960786265897.8B	181.55B	0.39	0.63	0.71	0.68
MNST	4.45B	36779326380.7B	6.06B	0.17	0.23	0.87	0.86
MRVL	12.69B	207716179637.33B	14.41B	0.27	0.28	-0.14	-0.71
MSFT	432.78B	444395584010810.0B	666.63B	0.35	0.49	0.43	0.37
MSTR	6.42B	79688276950.55B	8.93B	3.06	1.37	-30.28	-34.19
MU	67.37B	8706725132282.21B	93.31B	1.84	1.04	-6.87	-7.62
NFLX	47.13B	4842436910800.78B	69.59B	0.46	0.74	0.46	0.41
NVDA	169.27B	148072105115326.0B	384.8B	0.53	0.77	0.53	0.48
NXPI	9.01B	138957274852.09B	11.79B	0.22	0.25	0.11	-0.16
ODFL	7.66B	80862523299.19B	8.99B	1.08	0.76	0.65	0.61
ON	15.12B	305973440823.72B	17.49B	1.74	1.01	-1.46	-1.71
ORLY	12.15B	353930132282.81B	18.81B	0.35	0.35	-0.41	-0.54
PANW	8.83B	186176180580.32B	13.64B	0.33	0.35	0.78	0.75
PAYX	2.12B	7517465320.48B	2.74B	0.09	0.11	0.95	0.95
PCaR	26.07B	1628580045922.5B	40.36B	0.85	0.67	-12.31	-13.49
PDD	470.01B	498370065412475.06B	705.95B	1.01	0.75	-7.26	-9.4
PEP	56.8B	5651338072472.96B	75.18B	0.36	0.36	-1.34	-1.55
PLTR	8.85B	135025399552.06B	11.62B	0.29	0.32	0.57	0.21
PYPL	57.58B	7520847692901.21B	86.72B	0.39	0.65	-0.24	-0.44
QCOM	27.89B	1759241924787.05B	41.94B	0.24	0.28	-0.22	-0.33
REGN	21.79B	1447379749694.87B	38.04B	0.52	0.75	-0.9	-1.08
ROP	7.39B	150823754717.92B	12.28B	0.28	0.29	0.51	0.46
ROST	4.09B	34203027258.22B	5.85B	0.17	0.21	0.78	0.76
SBUX	22.12B	1111844699350.14B	33.34B	0.29	0.39	0.07	-0.02
SNPS	10.25B	184261449092.17B	13.57B	0.72	0.58	0.69	0.66
TEAM	12.29B	228598667674.65B	15.12B	0.24	0.37	-2.16	-4.96
TMUS	26.65B	1534773108227.76B	39.18B	8.34	2.0	-1297.2	-2270.85
TSLA	173.11B	97725413867463.2B	312.61B	0.48	0.84	0.19	0.11
TTD	3.94B	26051722344.59B	5.1B	0.41	0.42	0.9	0.88
TTWO	13.43B	262690581016.46B	16.21B	2.43	1.18	-2.13	-2.44
TXN	20.97B	724359041010.62B	26.91B	0.32	0.37	0.72	0.7
VRSK	2.65B	11981612994.21B	3.46B	0.16	0.22	0.88	0.86
VRTX	11.7B	249669928333.78B	15.8B	0.35	0.69	0.74	0.72
WBD	27.81B	1813554694461.33B	42.59B	1.46	0.93	-48.5	-53.23
WDAY	8.21B	121634416735.62B	11.03B	0.23	0.36	0.7	0.66
XEL	13.35B	203123708788.37B	14.25B	0.66	0.52	-1.06	-1.25
ZS	5.85B	59934836718.02B	7.74B	0.32	0.37	0.53	0.42
Gemiddeld	52.55B	25043.21B	73.29B	0.78	0.58	-17.69	-28.87
Mediaan	19.92B	644.77B	25.39B	0.42	0.49	0.09	-0.01

Tabel 10: Resultaten methode 4