

# Vergelijkende studie van voorspellingsmodellen voor tijdreeksen

Onderzoeksvoorstel Bachelorproef 2019-2020

Emiel Declercq<sup>1</sup>

## Samenvatting

Artificiële intelligentie heeft zich de laatste jaren ontwikkeld als een van de belangrijkste domeinen in de hedendaagse technologie. Maar om ervoor te zorgen dat de mogelijkheden van artificiële intelligentie optimaal benut worden moet er ook rekening gehouden worden met de opkomst van nieuwere modellen die voor bepaalde toepassingen misschien veel beter zouden kunnen zijn dan de klassieker modellen. Dit zal onderzocht worden voor tijdsafhankelijke data.

## Sleutelwoorden

Machineleertechnieken en kunstmatige intelligentie — Machine learning — AI

## Co-promotor

Stijn Lievens<sup>2</sup>

Contact: <sup>1</sup> emiel.declercq@student.hogent.be; <sup>2</sup> stijn.lievens@hogent.be;

## Inhoudsopgave

1	Introductie	1
2	Stand van zaken	1
3	Methodologie	1
4	Verwachte resultaten	2
5	Verwachte conclusies	2
	Referenties	2
	Onderzoeksvoorstel	

## 1. Introductie

Artificiële intelligentie wordt steeds meer toegepast dus de optimale methodes bepalen om voorspellingen te doen is van vitaal belang. Verschillende datasets kunnen er volledig anders uitzien deze kunnen dan ook op verschillende manieren ingedeeld worden. Voor dit onderzoek zal gefocust worden op data die tijdsgebonden is. Door het gebruik van dit type data zullen de modellen rekening moeten houden met de tijdsafhankelijkheid tussen de verschillende waarden.

## 2. Stand van zaken

Er zijn heel wat methoden die kunnen toegepast worden om een voorspelling te maken van tijdsgebonden data. Voor deze paper zullen enkel polynomiale vergelijkingen, ARIMA en LSTM getest worden. De meest primitieve manier om een trend te voorspellen is het fitten van een polynomiale vergelijking op de data. Daarnaast kan ook de ARIMA-methode (Brownlee, 2018) gebruikt worden ofwel het Autoregressive Integrated Moving Average. Deze methode combineert autoregressie en voortschrijdend gemiddelde.

Autoregressie modeleert de volgende stap in een sequentie als een lineaire functie van de waarden uit voorgaande tijdspannes. De methode van het voortschrijdend gemiddelde modeleert de volgende stap in de sequentie als een lineaire functie van de resterende fouten van een gemiddeld proces bij voorgaande tijdspannes. Er moet ook opgemerkt worden dat er een verschil is tussen een model met een voortschrijdend gemiddelde en het voortschrijdend gemiddelde van de dataset zelf. Ook neurale netwerken kunnen toegepast worden bij het maken van voorspellingen bij time series. LSTM (Long Short Term Memory) is een vaak gebruikt modeltype om time series te voorspellen. Dit model zal het verloop van de volgende waarden voorspellen op basis van de ingevoerde waarden rekening houdend met de chronologie waarin ze voorkomen. Hierbij zal de invloed van oudere waarden minder relevant worden naargelang er meer waarden ingevoerd worden.

Op zowel de ARIMA als de LSTM modellen bestaan er varianten om multivariate time series te voorspellen, bij ARIMA worden deze benoemd als VARMAX modellen wanneer er meerdere afhankelijke variabelen zijn en ARIMAX modellen indien er meerdere onafhankelijke variabelen zijn. Ook bij polynomiale regressie kunnen multivariate times series voorspeld worden. Ook voor time series data waar een duidelijke seasonability zichtbaar bestaat er een variant op het ARIMA model genaamd SARIMA.

## 3. Methodologie

Om na te gaan welke methodes de beste resultaten behalen zullen zowel polynomiale regressie, ARIMA en LSTM toegepast worden op 2 datasets, 1 waarbij seasonability van toepassing is en 1 waar geen duidelijke jaarlijkse structuur uit af te leiden valt. Om deze methodes te scoren zullen de

laatste waarden weggelaten en voorspeld worden waardoor uit de foutmarge tussen de voorspellingen en de werkelijke waarden afgeleid zal kunnen worden welke methode de meest accurate voorspellingen zal kunnen maken.

#### 4. Verwachte resultaten

Er valt te verwachten dat polynomiale regressie het zwakste resultaat zal behalen. Dan verwacht ik dat LSTM best zal scoren gevolgd door ARIMA.

#### 5. Verwachte conclusies

Er valt te verwachten dat de voorgestelde technieken goede resultaten zullen behalen. Vooral voor LSTM liggen mijn verwachtingen vrij hoog omdat ik reeds artikels over indrukwekkende voorspellingen heb gelezen met deze methode. Ik heb een pak minder vertrouwen in polynomiale regressie uit eigen ervaring aangezien het moeilijk is om een goed passende curve te vinden. ARIMA zal waarschijnlijk ook goede resultaten behalen.

#### Referenties

Brownlee, J. (2018). 11 Classical Time Series Forecasting Methods in Python (Cheat Sheet). Verkregen 29 juni 2020, van <https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/>