

Portfolio Assignment 101

Dit verslag is een terugblik op het vak Data Science. Omdat dit nieuw gebied is voor de meeste studenten hebben we in de eerste les als groepje uitgezocht wat Data Science nou precies is. Wij kwamen uit op de onderstaande definitie:

Data Science combineert vaardigheden vanuit wiskunde, statistiek, programmeren, AI en machine learning samen met domeinkennis om te komen tot inzichten uit data die op het eerste oog niet zichtbaar zijn. Op deze inzichten moet men kunnen handelen.

Na alle lessen die ik heb gehad en opdrachten die ik heb gemaakt ben ik tot de conclusie gekomen dat dit een goede definitie is van Data Science. Dit is omdat alle vaardigheden die genoemd zijn ook de vaardigheden zijn die ik geleerd heb. Hieronder zal ik een aantal voorbeelden van deze vaardigheden toelichten.

Wiskunde

Wiskunde is een vaardigheid wat best vaak terugkomt in de opdrachten. Denk bijvoorbeeld aan verschillende formules zoals de Standard Deviation zoals te zien is in figuur 1 en de Root Mean Square Error in figuur 2.

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}}$$

- σ = population standard deviation
- \sum = sum of...
- X = each value
- μ = population mean
- N = number of values in the population

Figuur 1 – Standard Deviation formule

$$RMSE = \sqrt{\frac{\sum_{i=1}^N \|y(i) - \hat{y}(i)\|^2}{N}},$$

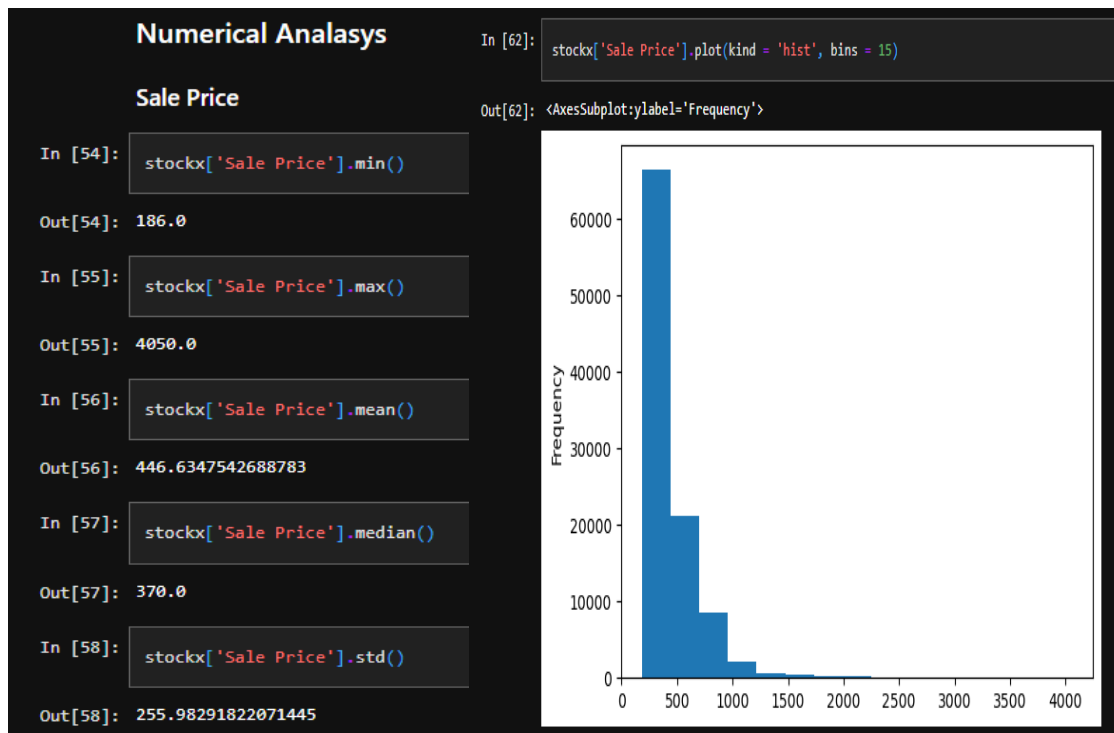
Figuur 2 – RMSE formule

Ik vond dit best interessant om te leren vooral omdat ik hier eigenlijk nooit eerder over geleerd hebt. Wiskunde bij Data Science is dus vooral het leren en begrijpen van formules zodat je je voorspellingen nog beter kan maken.

Statistiek

Als je over Data Science praat heb je het ook al eigenlijk meteen over Statistieken. Statistiek is voor mijn gevoel een best breed begrip maar volgens Wat is statistiek? - WisMon, 2023 is statistiek een tak van de wiskunde die zich bezighoudt met het verzamelen en analyseren van gegevens. Ook zou je met statistiek kun je gegevens beschrijven. Denk hierbij bijvoorbeeld aan het gemiddelde, de mediaan, de standaardafwijking of het weergeven van de gegevens in een diagram.

Dit is iets wat bij Assignment 6 terugkwam waarbij ik een Numerical Analasys moest doen zoals in figuur 3 en 4 te zien is.

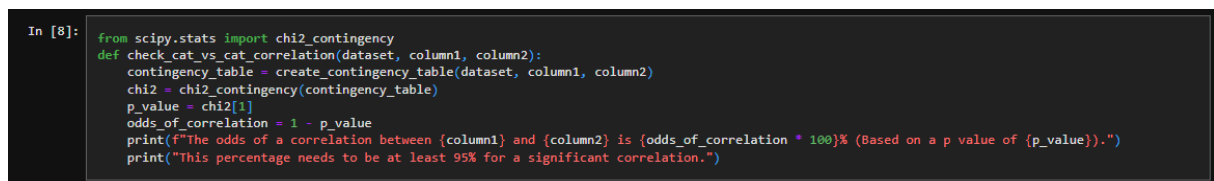


Figuur 3 – Numerical Analysis

Figuur 4 – Histogram

Programmeren

Als Data Scientist ontkom je er eigenlijk niet aan om te programmeren. Mijn gemaakte opdrachten zijn gemaakt door het gebruik van Python, een programmeertaal. Binnen deze taal wordt er gebruikt gemaakt van packages die het leven van een Data Scientist veel makkelijker maken. Een aantal voorbeelden hiervan zijn het berekenen van de correlatie tussen kolommen bij Assignment 14 zoals te zien is in figuur 5 en de hierboven getoonde figuren 3 en 4.



Figuur 5 – Berekenen van correlatie

AI en Machine Learning

Een belangrijk onderdeel van Data Science is ook Machine Learning en Artificial Intelligence, in de onderstaande figuur is ook te zien hoe populair deze begrippen zijn geworden. Hierbij is de blauwe lijn Machine Learning en de rode Artificial Intelligence.



Figuur 6 – Google Trends Machine Learning en Artificial Intelligence

Bij de gemaakte Assignments heb ik ook gebruik gemaakt van AI en Machine Learning. Een voorbeeld hiervan is het maken van voorspellingen op basis van andere gegevens. Hierbij moesten er twee datasets getraind worden waaruit een bepaalde accurate voorspelling kwam zoals te zien is in figuur 7.

```
predictionsOnTrainset = dt_classification.predict(stockx_train[features])
predictionsOnTestset = dt_classification.predict(stockx_test[features])

accuracyTrain = calculate_accuracy(predictionsOnTrainset, stockx_train.Brand)
accuracyTest = calculate_accuracy(predictionsOnTestset, stockx_test.Brand)

print("Accuracy on training set " + str(accuracyTrain))
print("Accuracy on test set " + str(accuracyTest))
```

Accuracy on training set 0.9997999113893296
Accuracy on test set 0.9998332610798012

Figuur 7 – Voorspellingen

Conclusie

Door het maken van alle Assignments is het voor mij veel duidelijker geworden wat Data Science nou precies is en hoe het gebruikt kan worden. Persoonlijk vind ik het ook best interessant om te doen en om erover te leren. Dit zou dus mogelijk ook iets zijn wat ik later als beroep ga doen.

Bronvermelding

Wat is statistiek? - WisMon. (2023, 31 januari). <https://www.wismon.nl/kennisbanvragen/wat-is-statistiek>