Report Data Science

Sa’ad Abdullah 2154528 [s.abdullah@student.avans.nl](mailto:s.abdullah@student.avans.nl)

Om deze rapport te beginnen moeten we het eerst hebben over het bedrijfsleven. Het bedrijfsleven hedendaags is vaak uitgerust met de nieuwste technieken om zoveel mogelijk winst makkelijk binnen te halen. Deze bedrijven hebben vaak behoefte aan analyses en terugkoppeling. Hier komt het stukje Business Intelligence binnen. Bedrijven gebruiken dat met als doel veel informatie te verkrijgen van data die op eerste blik onbruikbaar lijkt te zijn.

Data Science is een van de tools die hierbij helpen. Data Science is het manipuleren van Data met gebruik van verschillende tools om uiteindelijk verbanden binnen data te kunnen vinden.

Een analyse uitvoeren op bestaande data en verbanden in vinden is cruciaal voor een bedrijf. In de gemaakte portfolio opdrachten zitten verschillende elementen die zijn gebruikt om verbanden te leggen tussen verschillende kolommen. Neem als voorbeeld opdracht 19. Hierin gebruiken we verschillende tools en formules om een machine verbanden te laten zoeken binnen een data set. Deze groepeert voor ons automatisch het in groepen die wij weer kunnen onder verdelen groepen waar we voorspellende uitkomsten aan kunnen toe hangen.

Ook wordt er binnen het portfolio klein beetje gebruik gemaakt van Machine Learning wat weer hoort bij AI (Artificieel Intelligence). Hier wordt een machine een patroon aangeleerd om de verbanden tussen verschillende data sets eigenlijk voor te spellen. De feedback die hiervan afkomt is handig om zelf dan conclusies erover te trekken.

Report Data Scientist

Het doel van een Data Scientist is het aanleggen van een weg naar inzichten die voor bedrijven handig zijn, doormiddel van data. Om te beginnen kijkt een Data Scientist zelf naar de data set zelf. Hij analyseert eerst alle kolommen om erachter te komen met wat voor data hij werkt.

Hij kijkt eerst voor soort data een kolom bevat (numerieke, categoriaal, etc..). Hij berekent de gemiddelde en de min en de max voor de numerieke data (Assignment 4) (Universiate Analysis).

Zodra de data goed in kaart is gebracht voor de data scientist begint hij met verbanden zoeken binnen verschillende kolommen (Bivariate analysis). Denk aan het vinden van correlatie binnen numerieke data bijvoorbeeld (Assignment 9). Ook wil hij de correlatie bepalen tussen categorische en numerieke data natuurlijk. Je wilt wel weten of verschil in de categorieën bepalend zijn voor sommige factoren.

Zodra een data scientist klaar is met verschillende kolommen te vergelijken met elkaar gaat hij aan de slag met het kijken of de computervoorspellingen kan maken binnen de dataset (Assignment 15) (Multivariate Analysis). Hier wordt er gebruik gemaakt van de zo gehete decision trees. Deze ‘bomen’ worden gebruikt om beslissingen die de computer maakt binnen een analyse in kaart te brengen. Hierin kan een data scientist kijken in welke mate het lukt en of deze data voorspelbaar kan zijn.

Het laatste gedeelte werd er gebruik gemaakt van clusters. Hierin laat je een systeem keuzes maken om de verschillende groepen aan te wijzen binnen een data set. Ook hierin kijkt de data scientist of dit soepel verloopt of niet. De uitkomsten hiervan kan gebruikt worden voor verschillende doel eindes.

Tot slot is een data scientist een IT’er die graag gebruik maakt van wiskundige formules en algoritmes om verbanden en inzichten kunnen te verkrijgen over verschillende datasets die interessant zijn voor een bedrijf/stakeholder.