2 Inleiding

Dit rapport is naar aanleiding van de specialisatie opdracht van de cursus Data Science. De specialisatie opdracht heeft als onderwerp ‘het maken van een data sience pipeline voor NEC Nijmegen’. We zullen op basis van een paar vragen uitwerkingen maken volgens de data science workflow op de data van NEC.

3 Werkwijze

De werkwijze die wij hebben gevolgd is volgens de data science workflow. We hebben het probleem bekeken en een doel bepaald die we wilden bereiken door middel van het opstellen van deelvragen.

Daarna hebben we de data verkend, zodat we wisten wat voor data we tot onze beschikking hadden. In dat proces hebben we de data omgevormd tot data die wij goed konden gebruiken. Daarna hebben we de data geanalyseerd door er statistische berekeningen op los te laten en machine learning toe te passen.

De laatste stap is het de visualisatie. Daarbij zullen de resultaten worden gecommuniceerd.



4 Probleem en Doelstelling

Tijdens deze analyse zullen we een kijkje gaan nemen naar het aantal afgelegde kilometers van een speler. In de analyse zullen we dan ook duidelijk maken op welke positie die speler speelt. Dit kan voordelig zijn op het moment dat er een nieuwe speler wordt aangetrokken door de club en er aan de hand van zijn afgelegde afstand wordt gekeken of deze op gelijke lijn ligt met de huidige spelers op die positie.

De vraag die we bij dit onderzoek dan ook zullen stellen is: Op welke positie wordt gemiddeld welke afstand afgelegd?

De deelvragen die we hierbij stellen om tot een goed antwoord op de hoofdvraag te komen zijn:

* Wat is de afstand die een **speler** aflegt? En op welke snelheid?
* Kun je de **positie** van een speler identificeren?

5 Conclusie

Wat is de afstand die een **speler** aflegt? En op welke snelheid?:

Uit de resultaten van de grafiek kan je zien hoeveel de 10 veldspelers en de 2 invallers hebben gelopen. De verschillen lopen uit van 2 kilometer tot maar liefst 12 kilometer. dit zou verklaart kunnen worden omdat de 2 invallers veel minder speeltijd hebben als de spelers die al op het veld staan. daarnaast speelde ze met 3 centrale verdedigers, die ook niet veel kilometers maken in een wedstrijd.

Kun je de **positie** van een speler identificeren?

we hebben de latidute en longitude in kunnen lezen in een grafiek, maar daaruit was moeilijk te zien welke speler waar stond. de starttijd was niet te bepalen waardoor we geen duidelijke beginpunt hadden. hierdoor hebben we geen duidelijk overzicht wie waar stond. uiteindelijk konden we dus niet de positie indentificeren.

6 GitHub

<https://github.com/ericb97/DSSpecialisatieOpdracht.git>