DataDriven Business:

ProRail

- Latif Azaïm - 1851448

- Mathias Hendriks – 1740564

- Max Jansen – 1848848

- Yujian Jiang – 1835919

1. Inleiding

Voor het vak DataDriven Business is ProRail naar de HU gekomen met de vraag of hulp kunnen bieden voor het DataLab. Het DataLab van ProRail is een afdeling waar grote hoeveelheden data verzameld wordt, wat daarna geanalyseerd en verwerkt wordt om verbeteringen toe te passen op het spoorwegnet.

Het spoorwegnet is erg groot en er zullen altijd wel problemen opduiken van klein naar groot. Dit is voor iedereen erg vervelend. ProRail is daardoor veel bezig met het oplossen van de problemen en de kosten lopen daardoor ook op. En voor de reizigers is het ook vervelend. Die moeten omreizen en zullen zich ergeren als het herhalend optreed.

Daarom is het belangrijk dat er snel gehandeld moet worden. Daarvoor moet er natuurlijk wel duidelijk zijn wat de problemen zijn en hoe lang die gaan duren, zodat er vooruit gepland kan worden om de minsten vertragingen op te lopen.

Nu is er dus gevraagd of er hulp geboden kan worden bij het verduidelijken van de problemen. Er is gevraagd of het mogelijk is om te voorspellen hoe lang een storing gaat duren zodat er omheen gepland kan worden en de vertragingen zo beperkt mogelijk kunnen blijven.

2. Opdracht

De opdracht die wij gekregen hebben is om uit te zoeken of het mogelijk is om te voorspellen

hoe lang een storing gaat duren. Omdat het best een groot proces is dat we door lopen. Is het geheel eigenlijk opgedeeld in drie kleinere opdrachten die uiteindelijk een geheel moeten vormen.

Het eerste deel is de Business Understanding, zodat ons duidelijk wordt waar wij onze taken moeten vervullen en hoe het proces loopt. Daarna moeten wij door gebruik te maken van slimmen systemen, die kunnen leren van oude data, voorspellende modellen kunnen maken. Hiermee onderzoeken we of het mogelijk is om de hersteltijden te voorspellen van de storingen. Als laatste moet er een gebruikers applicatie gemaakt worden die gebruikt maakt van een voorspellend model, waar de gebruiker een storing kan invoeren en een verwachte hersteltijd terug krijgt.

Het DataLab heeft ons daarom toegang gegeven tot een verzameling data waarin alles van oude storingen in staan. Ook hebben wij inzicht gekregen over hoe zij te werk gaan en wat de verwachtingen zijn van ons.

3. Business Understanding

De eerste stap is om het proces te verduidelijken om het door te nemen. Dit is belangrijk omdat wij moeten weten waar wij onze applicatie en voorspellingen toe kunnen passen. Daarvoor hebben wij een BPMN1. (Business Process Model and Notation) gemaakt om een duidelijk inzicht te krijgen over hoe het proces loopt als er een storing optreed. Daaruit kunnen wij ook uitzoeken welke data op de momenten beschikbaar is. Het kan namelijk zo zijn dat als je een voorspelling wilt maken, dat je een deel van de benodigde data dan nog niet hebt. Daarom is het ook belangrijk om te weten hoe de data in elkaar zit en wat we gaan gebruiken.

4. Data

Daarna zijn we ons gaan focussen op de data. Dit is namelijk een van de belangrijkste onderdelen aangezien we moeten weten waar we mee werken. Ook moet er natuurijik zekerheid zijn over de data of alles wel juist is. Er staan vaak namelijk verkeerde en/of zelfs lege plekken in de data. Daar kun je natuurlijk niks mee.

Als eerste zijn we gaan onderzoeken wat er allemaal in de data zit. Op het eerste oog zijn wij breedschalig gaan kijken welke kolommen wij willen gebruiken. Hieruit is gekomen dat wij de volgende uitgekozen hebben:

- *stm\_oorz\_code*: Dit is de oorzaakcode die geclassificeerd wordt door de aannemer.

- *stm\_sap\_melddatum*: Dit is de datum van wanneer de melding gemaakt is.

- *stm\_sap\_meldtijd*: Dit is de tijd van wanneer de melding gemaakt is.

- *stm\_geo\_mld*: Dit is de geocode van waar het incident zich plaats vindt.

- *stm\_aanntpl\_dd*: Dit is de datum vanaf wanneer de aannemer aanwezig is.

- *stm\_aanntpl\_tijd*: Dit is de tijd vanaf wanneer de aannemer aanwezig is.

- *stm\_fh\_dd*: Dit is de datum wanneer het incident verholpen is.

- *stm\_fh\_tijd*: Dit is de tijd wanneer het incident verholpen is.

- *stm\_techn\_mld*: Dit is categorie waar de incident bij hoort.

- *stm\_prioriteit*: Dit is de prioriteitcode waarmee aangegeven wordt hoe hoog de prioriteit ligt.

- *stm\_fh\_duur*: Dit is hoe lang het hele incident heeft plaatsgevonden.

5. Modellen

6. Applicatie

7. Conclusie