

Rapport – Prorail

PROJECTGROEP 4
CEYHUN ÇAKIR
DENNIS BESSELSSEN
MAARTEN DE JEU

Inhoudsopgaven

Inleiding	2
Probleemstelling	2
Doelstelling	2
Gap Analyse	3
Hoe lopen de processen en wat zijn de problemen? (IST)	3
Hoe ziet de gewenste situatie er uit? (SOLL)	3
Benodigde oplossing om van de (IST) naar (SOLL) te komen (GAP)	3
Stakeholder analyse	4
BPMN	6
Conclusie	7
Aanbevelingen	7

Inleiding

In dit document wordt er uitgelegd en getoond hoe we met behulp van data science een applicatie kunnen bouwen die een business probleem van ProRail kan oplossen. De probleemstelling die ProRail heeft is dat bij een melding van een probleem het vaak onduidelijk is wanneer het moment van functieherstel zal plaatsvinden.

Om dit probleem op te lossen gaan we ook wat vertellen over de methodologie die we hebben gebruikt. Aan het einde van dit document zullen we onze bevindingen en eventuele aanbevelingen concluderen.

Probleemstelling

Het probleem wat speelt binnen ProRail is dat bij een melding van een probleem onduidelijk is wanneer het moment van functieherstel zal plaatsvinden. Momenteel wordt er gewerkt op basis van prognoses van betrokken monteurs. Deze prognoses zijn niet altijd accuraat. Daardoor is het moeilijk in te zien wanneer het moment van functieherstel zal plaatsvinden

Doelstelling

Wij willen aan de hand van de gegeven dataset een voorspellend model maken waarmee we het moment van functieherstel kunnen voorspellen. Het is belangrijk dat het model herhaalbaar is en dat er goed gedocumenteerd onderzoek is gedaan naar de verschillende mogelijkheden om tot een oplossing te komen. Ook zou ProRail een adviesrapport willen waarin wordt verteld wat er uitgekomen is van het onderzoek. Uiteindelijk wil ProRail een voorspellend model wat makkelijk in gebruik kan worden genomen door een treinverkeersleider.

Gap Analyse

Hoe lopen de processen en wat zijn de problemen? (IST)

Momenteel krijgt de meldkamer een melding binnen van een storing. De storing wordt gecommuniceerd vanuit de meldkamer naar een aannemer. Wanneer een aannemer op de locatie van de storing arriveert, zal de aannemer een prognose maken en deze communiceren naar de meldkamer van ProRail. Op deze manier krijgt de treinverkeersleider een indicatie over het moment van functie herstel.

Het probleem dat nu binnen speelt binnen prorail is dat de prognose van de aannemer vaak geen accurate tijdsindicatie geeft over het moment van functieherstel. Dit zorgt vaak voor tijdsverlies bij het herstellen van het treinverkeer.

Hoe ziet de gewenste situatie er uit? (SOLL)

ProRail wil dat er door middel van een applicatie een accurate voorspelling wordt gemaakt over het moment van functie herstel. Deze applicatie zal onder andere data gebruiken die wordt geleverd door de aannemer, zoals bijvoorbeeld de prognose van de aannemer. De voorspelling wordt aan de treinverkeersleider getoont in een gebruiksvriendelijke applicatie.

Benodigde oplossing om van de (IST) naar (SOLL) te komen (GAP)

Wij hebben een Proof of Concept gemaakt van een voorspellend model wat aan een gebruiker wordt getoont door middel van een gebruiksvriendelijke interface. Dit model gebruikt de prognose van de aannemer om een accurate voorspelling te maken. Ook hebben we de uitkomsten van ons model vergeleken met de prognose van de aannemer. Uit dit onderzoek bleek ons model een stuk accurater te zijn dan de prognose van de monteur.

Stakeholder analyse

Stakeholders	Primair	Secundair
Interne stakeholders	Treinverkeersleiders,	Bestuur Prorail
Externe stakeholders	Spoorpartners	Vervoerders, Reizigers
Interface stakeholders		Rijkswaterstaat

Stakeholder:	Treinverkeersleiders
Belang bij project:	De primaire interesse van treinverkeersleiders is het krijgen van een duidelijk moment van functie herstel. Het verkrijgen van een duidelijk functie herstel moment zal de treinverkeersleiders helpen bij het opstarten van al het gehinderde verkeer.

Stakeholder:	Bestuur Prorail
Belang bij project:	De interesse van de top bij Prorail bij dit project is onder andere het verminderen van sancties bij verhinderingen. Door het soepeler oplossen van storingen zal Prorail ook een betere reputatie krijgen.

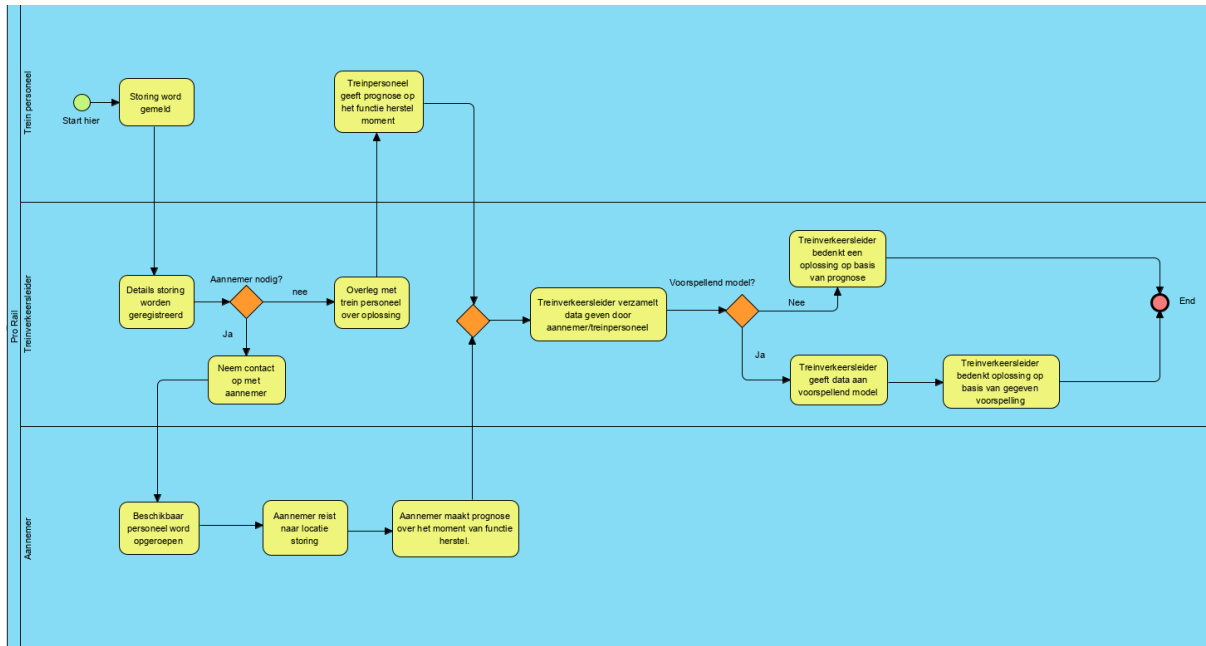
Stakeholder:	Spoorpartners
Belang bij project:	De interesse van spoorpartners bij dit project is een vermindering in hinderend verkeer. Als storingen soepeler worden opgelost, hebben spoorpartners daar ook profijt van.

Stakeholder:	Vervoerders
Belang bij project:	De interesse van de vervoerders is vooral gericht op het verkorten van storingen. Vervoerders kunnen sancties opgelegd krijgen als resultaat van storingen.

Stakeholder:	Reizigers
Belang bij project:	De interesse van de reizigers bij dit project is vooral het verkrijgen van inzicht bij een storing. Reizigers vinden het prettig om een accuraat inzicht te krijgen bij storingen.

Stakeholder:	Rijkswaterstaat
Belang bij project:	Afhankelijk van de storing, wordt rijkswaterstaat betrokken. Wanneer ProRail een goede voorspelling kan maken over het functieherstel moment, kan dit ook handig zijn voor Rijkswaterstaat.

BPMN



In het bovenstaande BPMN model is te zien dat het verschil tussen de huidige situatie en de gewenste situatie bestaat uit één extra stap in het proces waar het moment van functie herstel wordt bepaalt. Deze extra stap is het invullen van gegevens in het voorspellende model. Dit hoeft alleen gedaan te worden door de treinverkeersleider.

Hieruit is te concluderen dat het werkproces van de verschillende betrokken partijen weinig tot niet zal veranderen.

Conclusie

Als eerst willen we vaststellen dat de [RMSE](#) van de prognose ~45,5 minuten is. Dit zou je kunnen interpreteren als de gemiddelde afwijking van de prognose van de aannemer, vergeleken met de realiteit.

Zelf hebben wij meerdere machine learning modellen getest. Wij zullen hier de meest interessante vindingen benoemen.

Een linear regression model met een RMSE van ~38.8 minuten. Dit is al een sterke verbetering over het gebruik van alleen de prognose van de monteur. Dit linear regression model gebruikt als enige input de prognose van de monteur. De reden dat dit model een betere voorspelling geeft vergeleken met de prognose met de aannemer, is vanwege het feit dat aannemer een storing vaak pessimistisch inschat om zichzelf meer speling te geven bij het oplossen van de storing.

Ook hebben wij een decision tree regression model toegepast waar wij velden zoals: Prognose, geocode, oorzaak code, tijd op de dag & dag in het jaar; hebben ingevuld. Dit model kwam met een RMSE van ~36.9 minuten. Dit betekent dat er een kleine winst te behalen is door het invullen van meer datapunten. Helaas zorgt dit voor minder transparency, en wordt het ook lastiger voor de gebruiker om dit model te gebruiken. Dit gaf ons ook inzicht in het feit dat de huidige dataset weinig velden bevat die een significante invloed hebben op de mogelijke voorspellingen.

Samenvattend:

Ons linear regression model is makkelijk in gebruik, en geeft een betere voorspelling dan de aannemer. De huidige dataset bevat weinig significante datavelden die gebruikt kunnen worden bij het maken voor een voorspelling.

Aanbevelingen

Het advies wat wij geven aan ProRail is dat ze ons lineair regressie model gaan gebruiken. Dit model is makkelijk in gebruik en geeft accurate voorspellingen. Dit was voornamelijk te concluderen uit de onderzoeken die wij deden in dit project.

Ook geven wij als advies dat als ProRail een beter model wilt gebruiken, de dataset uitgebreid moet worden. Dit komt voornamelijk doordat er niet genoeg relevante datapunten in de huidige dataset zitten.