



E-PLM 2.0



Experiment 4.3: Use case 1 – BouWatch data analyse

Eindpresentatie

www.smartindustry.nl

Het experiment

Doel van het experiment

- ❑ Ervaring opdoen met het verzamelen en gebruik van velddata
- ❑ Onderzoeken of analyse van een grote hoeveelheid data leidt tot het ontdekken van patronen en conclusies op het gebied van optimaliseren, product verbeteringen of vernieuwende diensten
- ❑ Onderzoeken hoe big data analyse in de praktijk toepasbaar is op velddata die verkregen is van sensor data, waarbij verstoringen van de sensor data niet bekend zijn

Werkwijze en methodiek

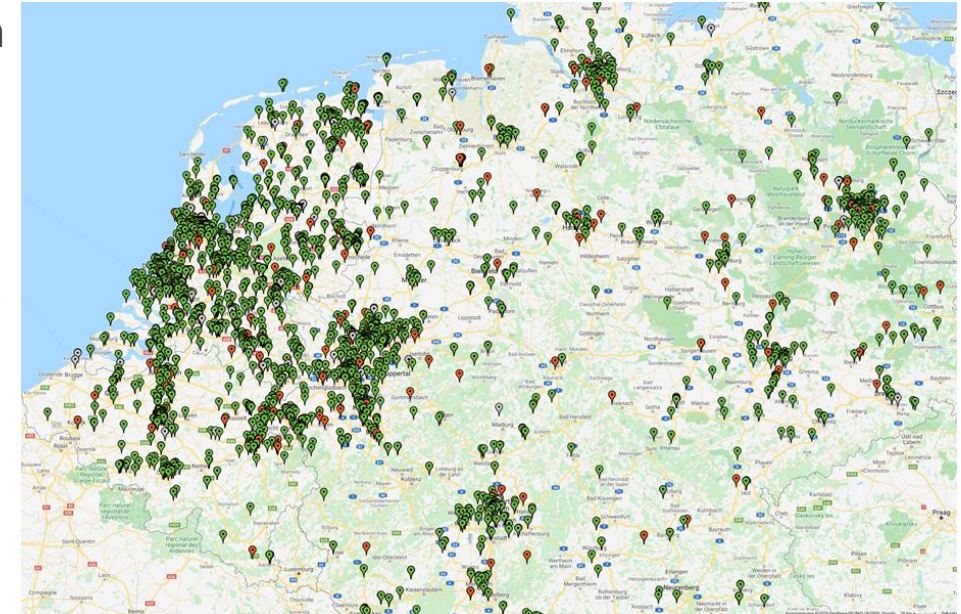
- ❑ De use case is gedeeld met het project team en in samenspraak verder gedetailleerd.
- ❑ Historische data is verzameld van een groot aantal systemen in het veld uit enkele verschillende databases. De data is onderzocht en waar nodig gefilterd voor compleetheid en correctheid.
- ❑ Met behulp van big data methoden zijn analyses uitgevoerd. De conclusies die de analyses suggereren zijn gevalideerd tegen de ervaring van de operationele systemen.
- ❑ De zelfde werkwijze is in drie rondes toegepast, waarbij de ervaring van elke ronde in de volgende ronde is meegenomen.



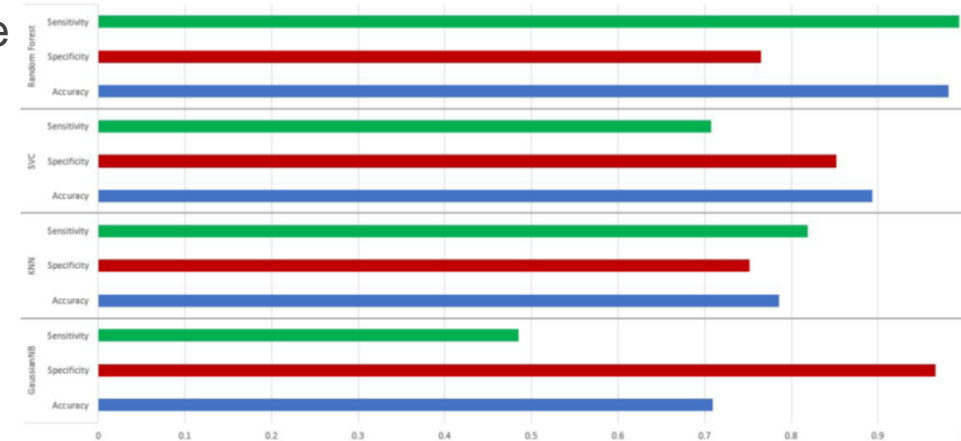
Resultaten

Tijdens het experiment is een grote hoeveelheid velddata verzameld. Om de betrouwbaarheid van de data op een hoog niveau te krijgen bleek lastiger dan van tevoren was voorzien. Daar liggen meerdere oorzaken aan ten grondslag. Dit heeft onder andere te maken met een scherpe definitie van de te meten waarden, een compleetheid van de data en vooral veel context informatie die te beperkt aanwezig was. Deze context informatie berust veel op menselijke waarnemingen en handmatige registratie, waardoor dit minder grootschalig en consistent beschikbaar was dan gehoopt. Hier volgt direct een belangrijke lesson learned uit dat een scherpe opzet van de verzameling van velddata de basis is.

Doordat in de use case bestond aan een drietal deel activiteiten kon er goed vergeleken worden tussen de uitvoer van deze drie activiteiten. De activiteiten hebben elkaar opgevolgd, waarbij er voor een volgende ronde duidelijk geleerd is van de eerdere ronde. In de eerste twee rondes is de analyse sterk handmatig geweest. De derde ronde hebben we eenvoudiger gemaakt door te focussen op een enkel onderwerp, en daarbij wel gebruik te maken van machine learning. Hoewel de exercitie om de machine learning goed toe te passen lastig bleek, heeft de combinatie van een eenvoudiger vraagstelling en machine learning een goede aanzet gegeven tot bruikbare resultaten.



Sensitivity, Specificity and Accuracy in the Partial Dataset



Lessons Learned & best practices

- Er bestaat een groot risico op vervuiling van velddata. In het experiment is een mix gebruikt van automatisch gegenereerde data en handmatige invoer. In beide typen komt vervuiling voor. De vervuiling kan worden beperkt door strakkere definities en protocollen te hanteren.
- Context informatie, die mogelijk niet automatisch vast te leggen is, moet wel in patronen vastgelegd worden. Gebruik keuze menu's, en zo min mogelijk met vrije tekst.
- Spendeer veel tijd/aandacht aan een methode om de data eenduidig op te schonen. Het eenduidig maken hiervan beperkt het risico dat het proces van opschonen van data zelf weer bias of fouten introduceert.
- Omdat zelfs bij een duidelijker definitie van velddata fouten en haperingen kunnen optreden is een goed ontworpen ETL (Extract-Transform-Load) cruciaal.
- De data analyse van de velddata, waarbij in het experiment gefocusseerd is op storingen, kan een goede trigger zijn om een probleem nader te analyseren. De data analyse laat een statistische trend zien. De interpretatie van dit resultaat vraagt echter een veel diepere root-cause-analysis.
- Bij machine learning met een neuraal netwerk is een goed gebalanceerde trainingsset van doorslaggevende invloed op de kwaliteit van het neurale netwerk.
- Meer lagen en nodes in een neuraal netwerk hoeft zeker niet te leiden tot een beter netwerk. De training van een simpeler netwerk bleek juist betere resultaten op te leveren. De complexiteit van het netwerk moet zich verhouden tot de complexiteit van het probleem.