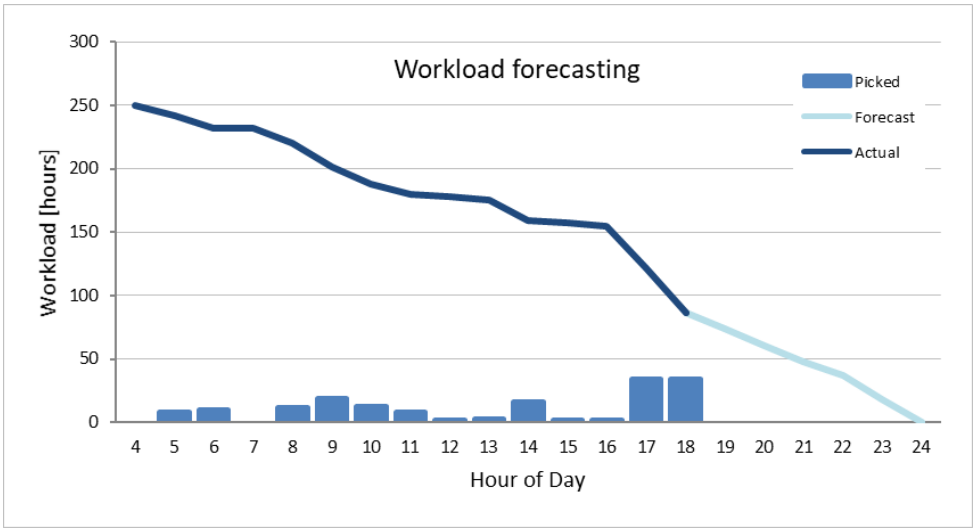
Working-hours estimation voor multi-orderpickingoperatie(s)

Wint Artificial Intelligence (AI) van de botte-bijl-methode?



De voorspelling van de werkvoorraad wordt gebruikt om het einde van de werkdag te voorspellen.

Rivaldo Sichtman (rivaldo.sichtman@actemium.com), Joost Vervoort (joost.vervoor@actemium.com)
Actemium whitepaper, Business Unit Logistics & Automation, Veghel, The Netherlands, October 2023

Samenvatting
In deze whitepaper onderzoeken we de meerwaarde van AI-modellen voor het voorspellen van de werkvoorraad voor een multi-orderpickingoperatie. Op basis van historische data van drie werkelijke klanten zijn verschillende modellen vergeleken, waaronder een eenvoudige benadering op basis van het aantal orderregels, lineaire regressiemodellen met en zonder outlier-reductie, en open-source neurale netwerken.

De belangrijkste bevindingen die wij deden:

- Diverse benaderingsmodellen vergeleken. Er zijn verschillende modellen onderzocht, variërend van eenvoudige regressie tot meer geavanceerde AutoML (machine learning). Interessant genoeg hebben eenvoudige modellen in vele gevallen beter gepresteerd. Een 3-feature regressiemodel presteert significant beter met behoud van inzicht in de modelwerking. AutoML (of andere geteste neurale netwerken) voorspellen niet significant beter, terwijl voor deze minder inzichtelijk is hoe de benadering tot stand komt.
- Klantspecifieke invloeden. Elk model is getest op zijn toepasbaarheid voor drie klanten. De resultaten tonen aan dat ondanks het feit dat bepaalde trends en patronen universeel kunnen zijn, de optimalisatie van een model in sommige cases sterk afhankelijk is van de specifieke klantcontext. Dit artikel illustreert dit aan de hand van 3 werkelijke klantcases.
- Kracht van visualisatie. Ieder model is een vereenvoudiging. Visualisatie van het model geeft transparantie en levert vertrouwen (of juist wantrouwen) in een model. Het biedt de kans om projectspecifieke features te ontdekken.
- Meerwaarde in de praktijk om werkvoorraad te voorspellen. Op basis van onze bevindingen zien we dat alleen voorspellen op basis van aantal orderregels – door ons de Basale Benadering (BB) genoemd – voor multi-orderoperaties onvoldoende betrouwbaar is. Een 3-feature model levert wel een betrouwbare voorspelling op van de hoeveelheid resterend werk.

Tenslotte, ongeacht type model en inzet, is het van belang dat de gebruiker op enigerwijze grip op en inzicht in de werking van het model heeft. Dit kan met visualisatie of een andere manier van presentatie. Op deze manier kan de gebruiker vertrouwen in het model opbouwen en uitleggen naar zijn collega's en blijft het niet een IT-tool die het goed doet op beurzen of in demo's.

Met dit artikel hopen we de wereld van AI en modellering dichter bij de praktijk van de logistiek te brengen. Voor vragen en/of opmerkingen kun je natuurlijk altijd contact opnemen met een van de auteurs.

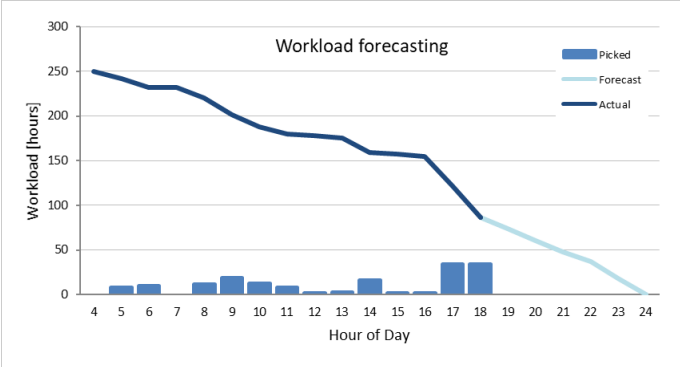
1 Inleiding

Binnen een logistiek proces is het pickingproces vaak een van de processen waarop veel optimalisatie wordt toegepast in de praktijk. Er zijn verschillende pickingmethoden, zoals wave, muli-order (sort while pick), batch (pick and sort) en verschillende pickingtechnieken (papier, scanning, voice, pick to light, goods to person, en anderen) die in dagelijkse pickingoperaties worden ingezet om maximale pickingprestaties te halen. Er is ook veel onderzoek gedaan naar optimale picking. Koster en anderen geven in [KOS07] een aardig overzicht van diverse onderzoeksresultaten in de literatuur.

Naast optimalisatie zijn er, in de wetenschap al veel eerder dan in de praktijk, vele pogingen gedaan om de werkvoorraad voor een logistiek proces te voorspellen.

Op basis van eenvoudige benaderingen, zie [KOS17], maar ook steeds meer met technieken die we scharen onder de verzamelterm Artificial Intelligence (AI). Vaak wordt in dat laatste geval gedoeld op methodes gebaseerd op neurale netwerken die getraind worden met trainingsdata. Er zijn legio publicaties te vinden over AI in logistiek, zie bijvoorbeeld [BIS06] en [YUY21]. Maar wat is hier nu de praktische meerwaarde van?

Wij onderzochten **de meerwaarde van AI**, maar ook "klassiekere" modellen **voor het voorspellen van de werkvoorraad van een multi-order-pickingoperatie voor 3 werkelijke[1] pickingoperaties**.



Figuur 1: Hoe de voorspelling van de werkvoorraad gebruikt kan worden om het einde van de werkdag te voorspellen (gegeven bepaalde personeelsinzet).

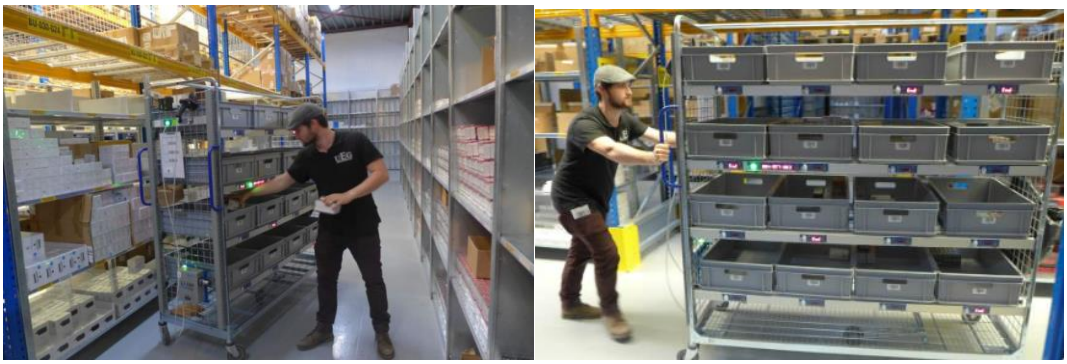
gevolgd door een benaderingsmodel voor de totale werkvoorraad van een (deel van de) dag. We sluiten af met een vergelijking van de verschillende benaderingsmodellen en de belangrijkste conclusies.

1.1 Multi-orderpicking

Als gezegd, in de praktijk bestaan diverse methodes en technieken voor orderpicking. De terminologie in literatuur en praktijk wisselt nog wel eens onder invloed van de markt en/of de maatschappij (goods to man werd bv. goods to person). Soms worden verschillende termen gebruikt voor eenzelfde proces. Wij doelen met multi-orderpicking op een pickmethode waarbij meerdere orders gelijktijdig worden verzameld en tijdens het picken worden gesorteerd (verdeeld) naar order. In de literatuur/praktijk wordt hiernaar ook wel gerefereerd als sort while pick.

In onze voorbeelden zijn we uitgegaan van multi-orderpickingoperaties met een pick to lightcart. Hiermee kan de orderpicker gelijktijdig items voor verschillende orders verzamelen tijdens één enkele ronde door het magazijn. Zie onderstaand een voorbeeld uit de praktijk.

Bij multi-orderpicking kan slim combineren van orders (intelligent batchen) de totale picktijd aanzienlijk verkorten. Hier heeft Actemium in de loop der jaren intelligentie voor ontwikkeld, maar daar gaat dit artikel niet over. Echter, zou de batch-intelligentie niet meegenomen moeten worden in het model voor de voorspelling van de werkvoorraad?



Figuur 2: Multi-orderpickproces

Het vervolg van dit artikel is als volgt opgebouwd. We starten met neerzetten van de context en definitie van multi-orderpicking, werkvoorraad en modellen voor voorspelling. Daarna presenteren we de resultaten van verschillende benaderingsmodellen voor 1 batch,

[1] De betreffende pickingoperaties en data zijn wel anoniem gemaakt.

1.2 Werkvoorraad en waarom werk-voorraad voorspellen?

Wij definiëren de werkvoorraad hiervoor nu als het aantal bruto[2] manuur dat nog besteed moet worden aan een proces, gegeven een bepaald aantal order(regel)s. In deze context willen we dus het resterend aantal uur voorspellen om de overgebleven order(regel)s te picken,

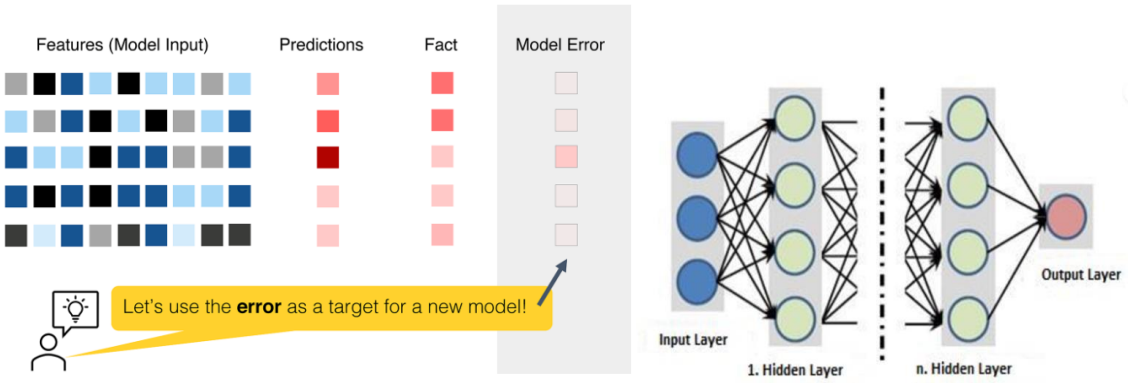
Waarom zouden we de werkvoorraad voor een pickingproces willen voorspellen? Inzicht in de resterende hoeveelheid werk geeft een logistiek manager bijvoorbeeld de mogelijkheid de verwachte eindtijd in te schatten ("Wanneer zijn we klaar?") of efficiënter capaciteit (mensen) te plannen over de dag ("Hoeveel mensen zijn er nog nodig?").

Een gangbare methode voor het inschatten van de resterende werkvoorraad (in uren) is door te kijken naar het aantal openstaande order(regel)s en dit te vermenigvuldigen met de gemiddelde tijd (vanuit ervaring) om een order(regel) af te handelen.
Gezien de toegenomen aandacht voor AI en machine learning kregen – en krijgen – wij in toenemende mate behoefte om concreet te onderzoeken of we de schat aan historische data konden inzetten voor betere voorspellingen.

1.3 Artificial Intelligence (AI): korte achtergrond en introductie

Onder andere met de komst van onder andere ChatGPT en steeds duidelijkere aanwezigheid van AI voor het brede publiek, krijgt Artifical Intelligence (AI) steeds meer aandacht.
AI als containerbegrip is eigenlijk niet heel nieuw. Het bestaat al lang. Van de eerste chatbot in de jaren 60 evalueerde de wereld van AI via schaakcomputers naar zelflerende neurale netwerken. In "AI: Alsmaar Intelligenter" [BUIJ20] biedt Stefan Buijsmans een zeer leesbaar overzicht van de ontstaansgeschiedenis, ontwikkelingen en vereenvoudigde beschrijving van de werking van AI. Die laatste verschijningsvorm – (zelf)lerende neurale netwerken – is waar men tegenwoordig meestal naar refereert als men praat over AI.

Eenvoudig gezegd wordt bij deze aanpak een black-boxmodel (neuraal netwerk) getraind met trainingsdata. De **training data** bestaat uit combinaties van **inputs** (ook wel features, wiskundig aangeduid met x_i) en bijbehorende **output(s)** (wiskundig aangeduid met y_j). Het model kan blijvend verfijnd worden als de trainingsdata set wordt aangevuld. Op basis van dit model, kan voor nieuwe inputs x_i een voorspelde waarde $\hat{y}_j$ gegenereerd worden. Achteraf kan de voorspelde waarde vergeleken worden met de werkelijke waarde. Het verschil is de **voorspelfout** (**modelfout**) e.



Figuur 3: Algemene weergave van benaderingsmodel (van https://www.evidentlyai.com) en idee om zelf te leren (links) en algemene weergave van een neuraal netwerk (rechts)

[2] Met bruto bedoelen we dat dit manuren besteed zijn aan een proces, inclusief kleine procesverstoringen (zoeken, wachten op aanvulling, korte stops), maar exclusief pauzes.

1.4 Aanpak vervolg

In dit artikel zullen we op basis van de historische data van bestaande operaties verschillende soorten modellen maken en de voorspelfout evalueren. We bouwen de complexiteit (en doorgrondbaarheid) van het model op. We starten met (lineaire) regressiemodellen, waarbij we het aantal features laten toenemen. Daarna trainen we een aantal open-source neurale netwerken (o.a. AutoML) en vergelijken de resultaten.

We starten met het voorspellen van de tijd voor 1 batch. Daarna gebruiken we dit model voor het voorspellen van de werkvoorraad voor een (deel van een) werkdag. Dit gebeurt door het samenstellen van openstaande orders via ons batching-algoritme, waarna we kunnen inschatten hoelang het zou duren om deze samengestelde batch te verzamelen. Op deze manier, zo hopen we, kunnen we de kwaliteit van de voorspelde werkvoorraad verhogen.

De vraag die we ons dan stellen is:

Biedt een model (neuraal netwerk of anders) meerwaarde dan de BB[3]-methode van aantal order(regel)s x gemiddelde tijd per order(regel)?

Ook benieuwd? Lees verder!

2 Benaderingsmodel voor 1 batch

Als eerste stap gaan we verschillende modellen evalueren voor het bepalen van de tijdsduur voor het picken van 1 batch. Van begin (lege kar; start batch) tot einde (alles verzameld; kar leeg).

[3] De BB-methode is een onofficiële afkorting door ons geïntroduceerd voor de Basale Benadering, ofwel de Botte-Bijl-methode.

2.1 Benadering op basis van het aantal orderregels (BB)

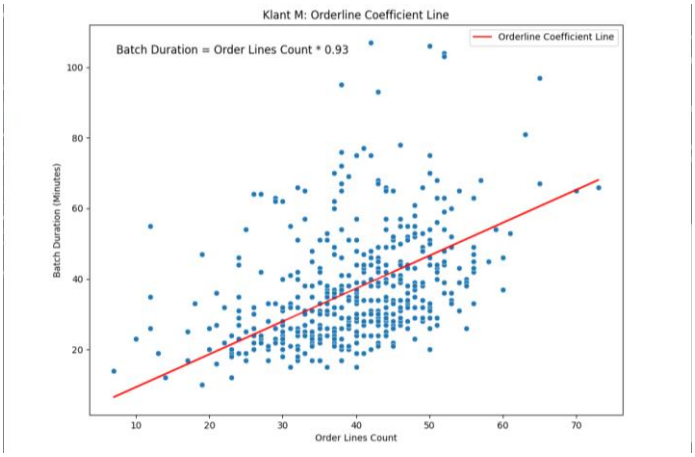
Het basale benaderingsmodel (BB), gebaseerd op het aantal orderregels, ziet er als volgt uit:

$$B_i = N_{OL,i} \cdot t_{avg,OL}$$

Hierin is:

B_i	De tijd voor batch i
$N_{OL,i}$	Aantal orderregels in batch i
$t_{avg,OL}$	Bruto Gemiddelde tijd [min] per orderregel. Deze is bepaald door de totale duur van alle batches (T_{totaal}) te delen door het totaal aantal orderregels ($N_{OL.totaal}$).

Figuur 4 illustreert de hypothetische lineaire relatie tussen het aantal orderregels en de tijdsduur van een batch voor een echte klant, die we in deze context "klant M" zullen noemen.



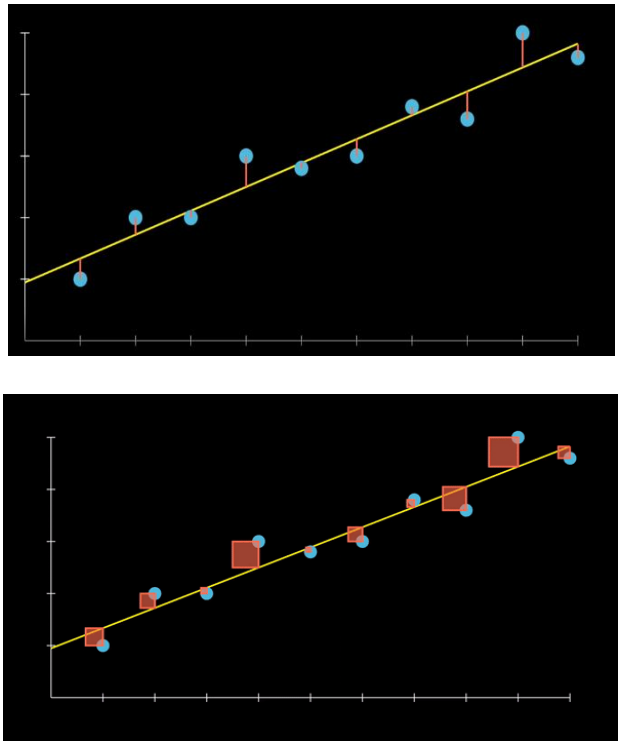
Figuur 4: Correlatie tussen het aantal orderregels en de tijdsduur van een batch

Op basis van de historische data zien we dat 1 orderregel gemiddeld 0,93 minuut kost. Het valt direct op te merken dat veel werkelijke datapunten (blauw) niet precies op de voorspelde lijn van het model (rood) liggen. Het verschil is de voorspelfout (ook wel error of residu).

Bij het evalueren van de prestaties van voorspellingsmodellen, is het cruciaal om een kwantificeerbare maatstaf te hebben die de nauwkeurigheid van de voorspellingen uitdrukt. Wij

hanteren hiervoor de (Root) Mean Squared Error ((R)MSE).

MSE is een maat voor de gemiddelde kwadratische afwijking tussen de daadwerkelijk waargenomen waarden en de voorspelde waarden. Het kwadraat van het verschil tussen de voorspelde en werkelijke waarden wordt genomen om positieve en negatieve afwijkingen gelijkwaardig te behandelen, en om grotere afwijkingen zwaarder te wegen dan kleinere.



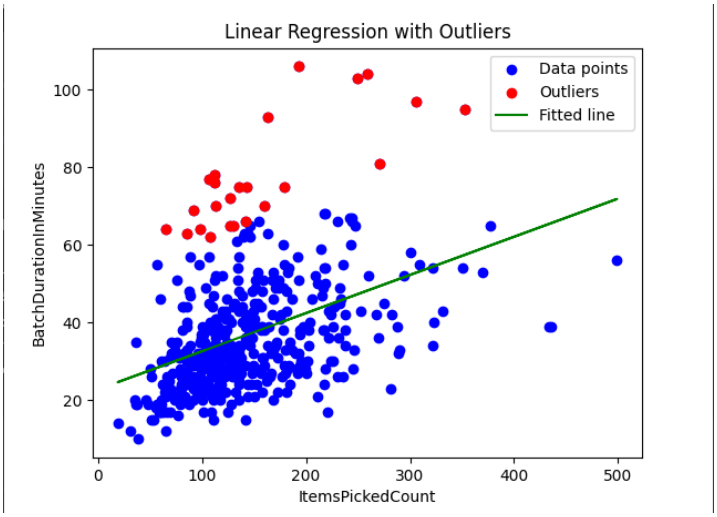
Figuur 5: Residuen (links) en visualisatie van de kwadratische afwijking voor verschillende modellen (rechts)

Vaak wordt de Root Mean Squared Error (RMSE), de wortel van de MSE, gehanteerd opdat de waarde in relatie staat tot de ordergrootte van een voorspelling. De RSME van bovenstaand model voor klant M bedraagt 14,6451 op een voorspelde waarde tussen de 20 en 50. We zitten er dus ongeveer 30 tot 75% naast met onze BB-methode. Noot: naast de RSME is eigenlijk de verdeling van de RSME ook interessant.

De modelresultaten worden verderop in overzicht getoond voor alle klantcases.

2.2 Outlier-reductie

Een manier om – in feite ieder – model te verfijnen is het toepassen van outlier-reductie. Outliers worden gedefinieerd als datapunten die significant afwijken van het model. Zie hieronder hetzelfde model als zojuist, maar met outlier-reductie. In statistische termen zijn outliers dit de observaties die meer dan een bepaald aantal standaarddeviaties (in onderstaande geval, 2.7) van het gemiddelde liggen.



Figuur 6: De in rood aangegeven datapunten vertegenwoordigen afwijkingen die meer dan 2.7 standaarddeviaties verwijderd zijn van de regressielijn, ook bekend als outliers.

Door het combineren van lineaire regressie met de verwijdering van outliers kan de voorspellende kracht van het model verder toenemen. Outliers kunnen een onevenredige invloed uitoefenen op de regressielijn en zo de betrouwbaarheid van de voorspellingen verstoren. Door deze outliers te identificeren en te verwijderen, kunnen we een nauwkeuriger en robuuster model creëren dat de werkelijke trends in de data beter weergeeft.

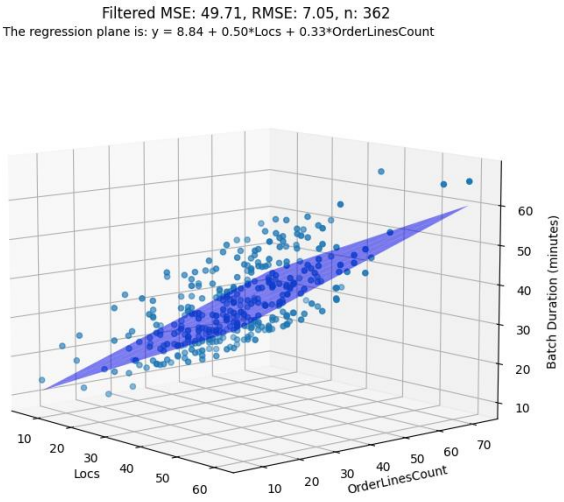
2.3 Benadering op basis van de tijd voor het opbouwen/afbreken van batches het aantal orderregels en aantal stops

Als volgende stap gaan we het lineaire regressie-model uit de vorige paragraaf uitbreiden met extra features. Op basis van ervaring voegen we toe:

1. Vaste tijd die het kost om een batch te starten (printen, kratten plaatsen, starten) en te beëindigen (leegmaken kar, bijkletsen met de teamleider (bv. over de manco's)). Deze schaalt immers niet mee met orderregels in de batch.

2. Niet alleen het aantal orderregels, maar ook het aantal locaties. Meer locaties in een batch betekent meer lopen en zoeken. Meer orderregels betekent meer verdelen na het picken.

Met deze toevoeging snappen we nog steeds wat ons white-boxmodel doet en dit zou de werkelijkheid toch beter moeten benaderen. Zie onder het resultaat voor klant M weergegeven in 3D.



Figuur 7: Een 3D weergave van het lineaire regressiemodel van de batchduur als functie van het aantal locaties tegenover het aantal orderregels

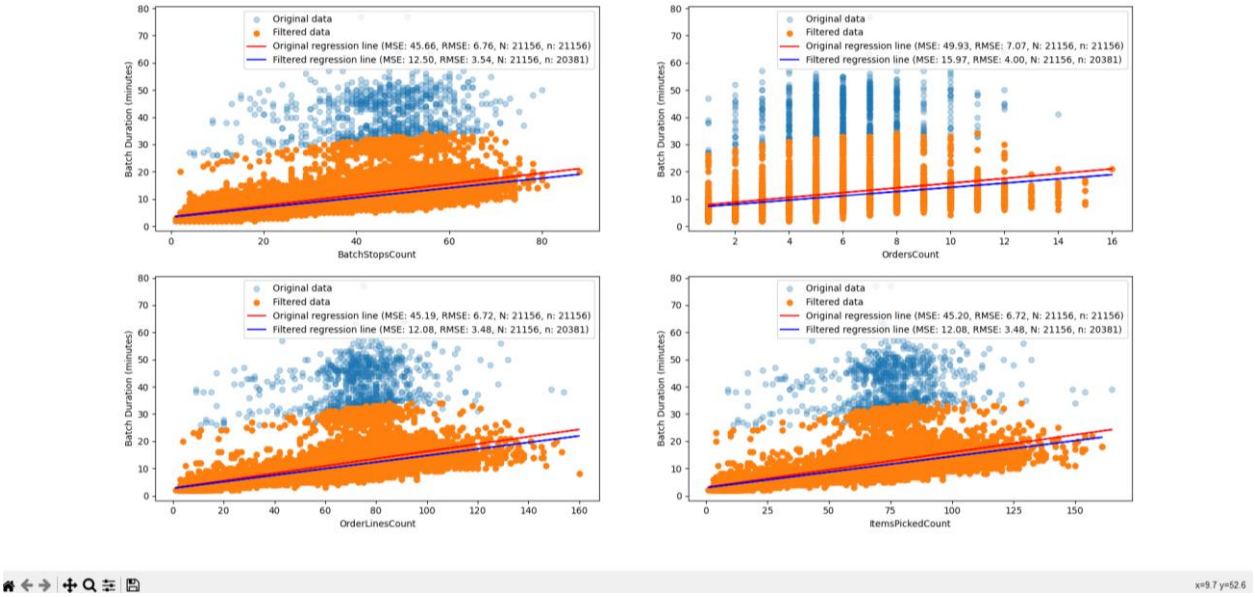
We zien hierin dat de RSME daalt naar 7,05. Goed nieuws. Volgens dit model kost een batch opbouwen/afbreken 8,84 minuten. Het picken kost daarna 0,5 minuut per locatie (lopen+zoeken) en 0,33 minuut per orderregel (sorteren). De relatieve voorspelfout voor deze klant is nu 10-25%. 3x zo goed dus.
Naast deze set van features zijn diverse andere 2-feature sets onderzocht. Bovenstaande bleken het best te presteren voor alle 3 de klanten.

2.4 Klantspecifieke features meenemen voor verdere verfijning regressie-model

Bij het zoeken naar betere regressie modellen – en vooral het visualiseren daarvan – bleek dat soms onverklaarbare groepen van uitschieters optraden. Dat konden niet allemaal outliers zijn. En dan komt de proces- en domeinkennis om de hoek. In het algemeen van orderpicking, maar soms ook van die specifieke klanten. Op basis van ervaring introduceren we extra kenmerken of 'features' om de specifieke nuances en patronen van elke klant beter te vangen.

Als voorbeeld hiervan hebben we de volgende extra kenmerken geïntegreerd in het model voor Klant M:

- Hoogte boven 2 meter: Een feature die aangeeft of een item zich boven de 2 meter hoogte bevindt, wat de pick-tijd kan beïnvloeden omdat de operator op een trap moet klimmen om te picken.



Figuur 8: Data van klant F die meerdere clusters weergeeft (linksboven: #locaties tegen batchduur, linksonder: #orderregels tegen batchduur, rechtsonder: #items gepickt tegen batchduur)

- Meervoud van 10: Een feature die de hoeveelheid van een item identificeert als een veelvoud van 10, omdat dit vaak omverpakkingen betreft; en als 1 pick gezien kan worden.

Voor klant H voegden we toe:

- Sticker aantal tijdens het picken: Een feature die de hoeveelheid stickers aangeeft die geprint moeten worden tijdens het pickproces, wat de efficiëntie en snelheid van de orderverwerking kan beïnvloeden.

Zie in Figuur 8 een sprekend voorbeeld van klant F waarbij volgens klassieke outliers een hele wolk (blauwe punten) aan observaties zou moeten vervallen. Wij herkenden, na visualisatie, echter 3 wolken (clusters) van punten, waardoor een relatief lineair verband met orderregels (of gepickte aantal) leek te bestaan.

Na verdergaand onderzoek stelden we vast dat deze klant pauzeerde tijdens de batch. Het onderste cluster (de meeste) zijn batches zonder pauze, het middelste cluster bleken een kleine pauze van 15 minuten te bevatten en het bovenste cluster – je snapt hem waarschijnlijk al – 30 minuten pauze.

Voor een zuiver model voor deze bruto batchduur (zonder grote pauzes) moet de pauzeduur dus van de observaties worden afgetrokken.

2.5 Model met AutoML

De ontwikkeling van machine learning-technieken heeft in de afgelopen decennia geleid tot significante vooruitgang op het gebied van voorspellende modellering. De opkomst van geautomatiseerde machine learning, ook wel bekend als AutoML, symboliseert deze evolutie. ChatGPT meldde ons: "AutoML refereert naar technieken die automatisch goed presterende modellen ontdekken voor voorspellende doeleinden, en dat met minimale menselijke betrokkenheid. Het hoopt de kloof te overbruggen tussen domeinspecifieke kennis en het modelleren van taken, en vergemakkelijkt zo de creatie en implementatie van krachtige, voorspellende algoritmen." Dat belooft veel.

2.5.1 AutoML: Het Proces / Een inleiding

Traditioneel machine learning vereist de volgende acties: data-verzameling, data-opschoning, selectie van kenmerken (bijvoorbeeld aantal stops, regels, maar ook de hoogte van de picklocatie van een artikel kan impact hebben op de tijd), modelselectie, parameter tuning, en validatie van het model. AutoML streeft ernaar om dit proces te vereenvoudigen en meer toegankelijk te maken door deze taken deels automatisch te doen.

2.5.2 AutoML: Waarom TPOT?

We hebben verschillende AutoML-frameworks geëvalueerd om te bepalen welke het beste paste bij onze behoefte, waaronder die van gerenommeerde aanbieders zoals Autogluon, ondersteund door Amazon, H2O.ai, en Azure Machine Learning, gefaciliteerd door Microsoft.

Klant H		
Modellen	**Modelomschrijving**	**RMSE**
1 Feature - BB	F = #OL*0,24	3,52
2 Feature -OrderlineRegressie	F = (tbub) + #OL * 0,6285	2,612
3 Feature -PickitemLocationRegressie	F = (tbub) 1,92 + 0,0488 * PickItemQuantitySum 0,1831 * TotalUniqueLocation	2,162
5 Feature - MultifeatureRegressie + 1 process-specifieke feature	F = (tbub) 1,80 + 0,024 * TotalOrderLinesCount 0,011 * TotalQuantityBomPrintedCount 0,0424 * PickItemQuantitySum 0,162 * TotalUniqueLocation	2,147
AutoML	n,v,t,	2,7469

Figuur 9: Resultaten van benaderingsmodellen voor 1 batch voor klant H

Na evaluatie en vergelijking van de genoemde oplossingen, bleek TPOT (open-source, ontwikkeld door de universiteit van Pennsylvania) het best presterende AutoML-framework voor onze datasets in de verschillende testscenario's.

2.5.3 AutoML en TPOT

TPOT maakt gebruik van een genetisch algoritme om de beste parameters en modellen te vinden voor een specifieke dataset. Genetische algoritmen zijn geïnspireerd op de natuurlijke selectie en volgen het "survival of the fittest"-principe om tot de beste oplossing te komen. TPOT genereert automatisch machine learning pipelines en evalueert ze op basis van de root mean squared error. De best presterende pipelines worden gecombineerd om nieuwe pipelines te creëren, die opnieuw worden geëvalueerd. Dit iteratieve proces wordt voortgezet totdat een optimale pipeline is gevonden.

2.6 Resultaten

De resultaten van de diverse modellen hebben we samengevat in Figuur 9. Zie hierin een aantal benaderingsmodellen voor klant H met per model de RSME. Hierin zijn de beste white-box- en beste black-box (AutoML)-modellen opgenomen.
We zien dat het 3-feature-model (met vaste bouw/afbreek tijd) navenant de inspanning voor modelvorming het beste presteert. Zie in Figuur 10 het vergelijk van alle 3 klantcases, waarbij de model-inhoud is weggelaten en alleen de modelomschrijving en RSME is getoond.

3 Benaderingsmodellen voor de werkvoorraad

Nu we de batchduur kunnen benaderen – zij het met wisselend succes – kunnen we deze gebruiken om de werkvoorraad te voorspellen voor een hele (of een deel van de) werkdag.

3.1 2 modellen voor werkvoorraad

We hebben 2 modellen onderzocht voor de 3 klanten.

1. De BB-methode. Hierbij vermenigvuldigen we het aantal order(regel)s van (een deel van) de dag met de gemiddelde tijd per order(regel). Hiervoor hebben we geen batchbenadering nodig. In formulevorm:

BB-methode (de praktijk)	$\hat{W} = N_O \times t_{gem,O}$ of $N_{OL} \times t_{gem,OL}$

   Hierin is $\hat{W}$ de benadering van de werkvoorraad in [min] op basis van aantal orders (N_O) en gemiddelde tijd per order $t_{gem,O.}$ Of analoog op basis van aantal orderregels (N_{OL}) en gemiddelde tijd per orderregel ($t_{gem,OL}$).

2. De methode waarbij het batchmodel gebruiken. We verdelen de werkvoorraad naar batches en schatten per individuele batch de duur in. De werkvoorraad in formulevorm:

Batchvoorspelling met best-fit model	$\widehat{W} = \sum_{b=Nb}^{b=1} \widehat{T_b}$

   Hierin is $\hat{W}$ de benadering van de werkvoorraad in [min] of basis van N_b batches, waarbij per batch de

Klant H		Klant F		Klant M	
Modellen	**RMSE**	**Modellen**	**RMSE**	**Modellen**	**RMSE**
1 Feature - BB	3,52	1 Feature - BB	6,78	1 feature - BB	14,6451
2 Feature -OrderlineRegressie	2,612	2 Feature -OrderlineRegressie	3,25	2 feature - OrderlineRegressie	11,691
3 Feature -PickitemLocationRegressie	2,162	3 Feature -OrderlineLocatieRegressie	3,18	3 feature - PickitemLocationRegressie	10,842
5 Feature - MultifeatureRegressie + 1 process-specifieke feature	2,147	5 Feature - LocationOrderPickitemOrderlineRegressie	3,17	6 feature - MultifeatureRegressie + 2 process-specifieke feature	10,573
AutoML	2,7469	AutoML	6,71	AutoML	12,826

Figuur 10: Resultaten van benaderingsmodellen voor 1 batch voor klant H, F en M

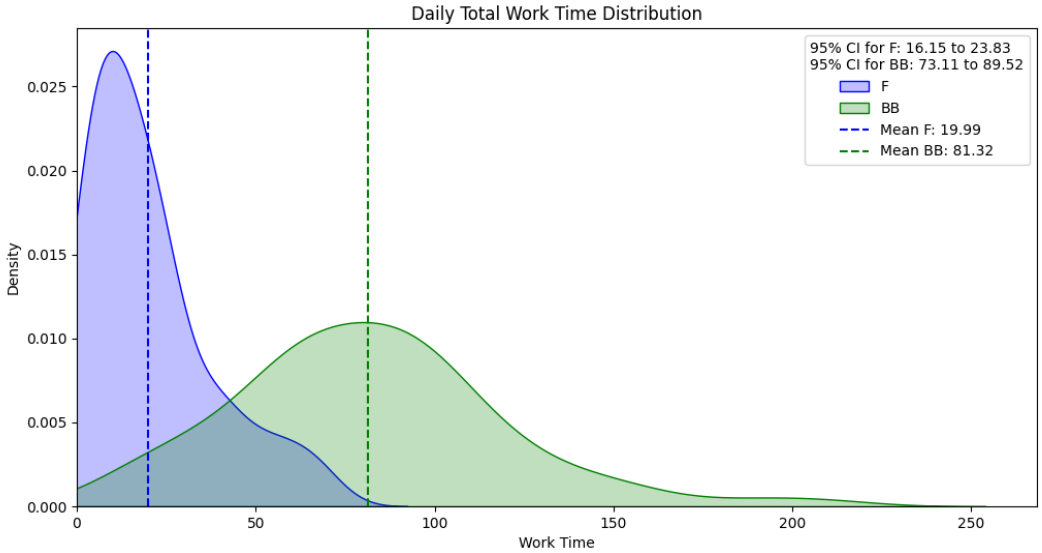
duur T_b benaderd wordt in [min] met een batch-model.

3.2 Resultaten

Om kwaliteit van verschillende modellen voor de werkvoorraad te vergelijken, gebruiken we de verschillende modellen voor het bepalen van de werkvoorraad voor de laatste 3 uur werk van een dag. Voor de drie klanten vergelijken - voor een groot aantal verschillende dagen - de voorspelde werkvoorraad met de werkelijke werkvoorraad. Voor elke dag kunnen we het verschil tussen beide methoden en de werkelijke waarde uiteindelijk uitdrukken in minuten.

3.2.1 Werkvoorraad voorspelling Klant F

Voor klant F zijn er twee resultaten doordat het uitsluiten van de batches die in de pauze vallen leidt tot een verbetering van de voorspelling, zie Figuur 12. Hierin zien we dat de BB-benadering een flinke afwijking heeft van ordegrootte 50% van de werkelijke waarde. Met een batch-model is de voorspelfout gemiddeld ongeveer 10%. 5x beter 😉 Voor nog beter zicht op de kwaliteit van de benadering hebben we de voorspelfout (RSME) van duizenden voorspelde dagen in dichtheidsplot uitgezet, zie Figuur 11.



Figuur 11: Density-plot van voorspelfout voor BB-methode (groen) vergeleken met het 5-feature-model inclusief pauzecorrectie

Klant F Modeltype	**Model**	**AVG Werkelijke workload [min]**	**AVG Voorspelde Workload [min]**	**AVG RSME [min]**
BB $\hat{W} = N_{OL}$ x $t_{gem,OL}$	W = #OLx0,64	185,73	266,16	80,85
Batch-model $\widehat{W} = \sum_{b=Nb}^{b=1} \widehat{T_b}$	5-featuremodel	185,73	170,57	22,27
Batch-model $\widehat{W} = \sum_{b=Nb}^{b=1} \widehat{T_b}$	5 feature-model met pauze-correctie	185,73	179,13	19,99

Figuur 12: Modellen voor de werkvoorraad van klant F vergeleken

3.2.2 Voorspelmodel van klant M en klant H

Zie in Figuur 13 en Figuur 14 de modelresultaten voor klant M en klant H.

Het valt op dat voor klant H het batchmodel de voorspelfout sterk verbetert: het batchmodel presteert ca. 3x beter dan de BB-methode. Echter voor klant M presteert het batchmodel BB-methode net zo goed (of eigenlijk net zo slecht) als de BB-methode. Dit is in lijn met de kwaliteit van het batchmodel voor klant M. Dit is met name het gevolg van een te kleine trainingsdataset. In alle drie de casussen - Klant F, Klant H en Klant M - blijkt het batch-model nauwkeuriger te zijn in het voorspellen van de dagelijkse werklast. De kwaliteit van het batchmodel voor 2 van de 3 klanten is 300% tot 500% nauwkeuriger. Voor 1 klant bleek de historische dataset onvoldoende groot (of onvoldoende verrijkt) voor een betrouwbaar model.

4 Conclusie

In deze whitepaper hebben verschillende modellen geëvalueerd. Eerst voor het bepalen van 1 batch en later voor een werkvoorraad voor de laatste 3 uur werk van een dag. Dit is getoetst voor drie klanten: Klant H, Klant M en Klant F.

Hierna volgen de belangrijke bevindingen die wij willen delen.

- **Diverse benaderingsmodellen vergeleken.** Er zijn verschillende modellen onderzocht, variërend van eenvoudige regressie tot meer geavanceerde AutoML (machine learning). Interessant genoeg hebben eenvoudige modellen in de meeste gevallen beter gepresteerd. Een 3-feature regressiemodel presteert significant beter met behoud van inzicht in de modelwerking. AutoML (of andere geteste neurale netwerken) voorspellen niet significant beter, terwijl voor deze minder inzichtelijk is hoe de benadering tot stand komt.
- **Klantspecifieke invloeden.** Elk model is getest op zijn toepasbaarheid voor drie klanten. De resultaten tonen aan dat ondanks het feit dat bepaalde trends en patronen universeel kunnen zijn, de optimalisatie van een model in sommige cases sterk afhankelijk is van de specifieke klantcontext. Dit artikel illustreert dit aan de hand van 3 werkelijke klantcases.
- **Kracht van visualisatie.** Ieder model is een vereenvoudiging. Visualisatie van het model geeft transparantie en levert vertrouwen (of juist wantrouwen) in een model. Het biedt de kans om projectspecifieke features te ontdekken.
- **Meerwaarde in de praktijk om werkvoorraad te voorspellen.** Op basis van onze bevindingen zien we dat alleen voorspellen op basis van aantal orderregels – door ons de Basale Benadering (BB) genoemd – voor multi-orderoperaties niet voldoende betrouwbaar is. Een 3-feature-model levert wel een betrouwbare voorspelling op van de hoeveelheid resterend werk.

Tenslotte, ongeacht type model en inzet, is het van belang dat de gebruiker op enigerwijze grip op en inzicht in de werking van het model heeft. Dit kan met visualisatie of een andere manier van presentatie. Op deze manier kan de gebruiker vertrouwen in het model opbouwen en uitleggen naar zijn collega's en blijft het niet een IT-tool die het goed doet op beurzen of in demo's.

5 Referenties

[BIS06] C.M. Bishop, (2006), *Pattern recognition and machine learning*, (1st Edition), Springer

[BUIJ20] Buijsman, S (2020), *AI: Alsmaar Intelligenter. Een kijkje achter de beeldschermen*, De bezige bij, 20 juli 2020

[KOS07] De Koster, R., Le-Duc, T., and Roodbergen, K.J. (2007), Design and control of warehouse order picking: a literature review. European Journal of Operational Research 182(2), 481-501.

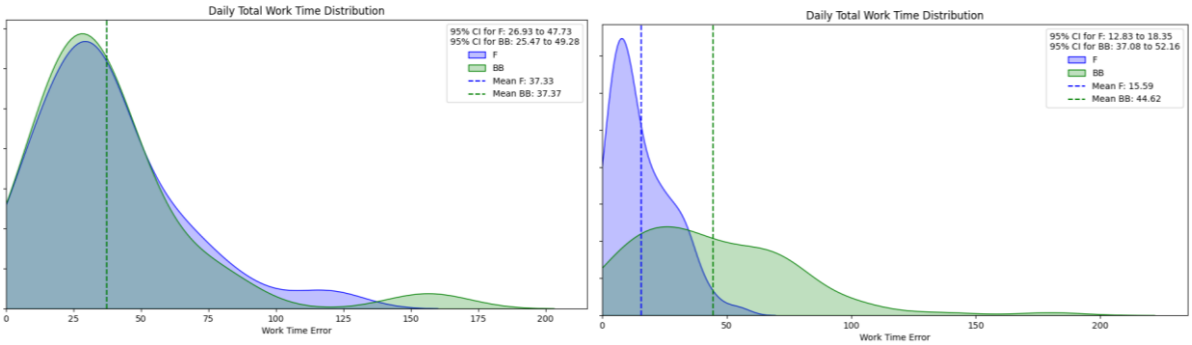
[LAR17] J.A. Larco, R. De Koster, K.J. Roodbergen, J. Dul, *Managing warehouse efficiency and worker discomfort through enhanced storage assignment decisions,* International Journal of Production Research, 55 (21) (2017), pp. 6407-6422

[YUY21] Yuya Okadome, Toshiko Aizono, *Adversarial data-selection based work-hours estimation method on a small dataset in a logistics center*, Computers & Industrial Engineering, Volume 164,
2022

Klant M Modeltype	Model	AVG Werkelijke workload [min]	AVG Voorspelde Workload [min]	AVG RSME [min]
BB $\hat{W} = N_{OL} \times t_{gem,OL}$	W = #OLx0.9323	199.11	200.10	37.37
Batch-model $\widehat{W} = \sum_{b=Nb}^{b=1} \widehat{T_b}$	8-feature-model	199.11	185.36	37.33

Klant H Modeltype	Model	AVG Werkelijke workload [min]	AVG Voorspelde Workload [min]	AVG RSME [min]
BB $\hat{W} = N_{OL} \times t_{gem,OL}$	W = #OL*0,24	185,57	227,55	44,62
Batch-model $\widehat{W} = \sum_{b=Nb}^{b=1} \widehat{T_b}$	7-feature-model	185,57	178,58	15,59

Figuur 13: Modellen voor de werkvoorraad van klant M en klant H vergeleken



Figuur 14: Voorspelfout voor BB-model (groen) en batch-model (blauw) voor Klant M (links) en klant H (rechts) vergeleken.