

Prognose aantal studenten

06/06/2024

Amir Khodaie

Academic Affairs, *Institutional Research*

Aanleiding en doel

Bericht op Voxweb d.d. 8 september 2023 (<https://www.voxweb.nl/nieuws/collegezalen-zijn-te-vol-studenten-worden-weggestuurd>)

Collegezalen zijn te vol, studenten worden weggestuurd

08 sep 2023 Vox

De opleidingsdirecteur van Artificial Intelligence, Johan Kwisthout, betreurt het dat niet iedereen het hoorcollege kon bijwonen. 'Onze insteek en inspanning is dat alle studenten de colleges kunnen bijwonen. De studie is in het afgelopen jaar flink gegroeid waarbij tot op het laatste moment onduidelijk is hoeveel studenten daadwerkelijk aan hun studie beginnen. In het voorjaar worden de roosters al gemaakt en het is lastig om een realistische inschatting te maken.'

Aanleiding en doel

- **Van meerdere kanten steeds meer behoefte aan / vraag naar voorspellingen. Op dit moment: pas bij start nieuwe studiejaar beeld van aantal studenten en nauwelijks iets van latere jaren. Behoeftte o.a. vanuit oogpunt van:**
 - Algemeen: steeds meer behoefte aan het kunnen anticiperen wat op ons afkomt.
 - Macro: toekomstverkenning Sterken, 'Kiezen voor kwaliteit', internationaliseringsdiscussie, gedifferentieerde groei.
 - Bedrijfsvoering opleiding en universiteit: huisvesting, zaalbezetting, roostering, aantal docenten, financiën.
- **Referentieramingen en prognoses DUO (https://duo.nl/open_onderwijsdata/hoger-onderwijs/aantal-studenten/prognose-studenten-ho.jsp): niveau sector en universiteit en niet de gewenste accuraatheid**
- **Doel: ontwikkelen van een zo accuraat mogelijk, betrouwbaar, breed gedragen voorspelmodel van het aantal studenten aan de Radboud Universiteit**

Tijdlijn: wat hebben we gedaan?

1. Voorjaar 2022: begonnen met eigen expertise met voorspellen instroom komend jaar

- Ratiomodel: relatie vooraanmeldingen – inschrijvingen is meer diffuus dan verwacht.

2. Najaar 2022: doorontwikkelen met externe partij (Sioux). Observaties/opbrengsten:

- Nieuwe tak van sport (voor RU als elders), dus veel pionieren/experimenteren/ontdekken.
- Significante verbetering aan het voorspelmodel.
- Veel nieuwe kennis opgeleverd bij de collega's.

3. 2023: zelf doorontwikkeld voortbordurend op samenwerking met Sioux

- Met opgedane kennis verder doorontwikkeld: andere data, extra variabelen, aanscherping methode.
- Verdere significante verbetering van het voorspelmodel.

Waar staan we nu?

WAT GAAT GOED?

- We hebben een stabiel, robuust prognosemodel voor aankomende studenten, op het niveau van de RU, faculteit en opleiding. Voor bachelor en master; en naar nationaliteitsgroep.
- Wat ons betreft: een model van dusdanige accuraatheid dat het ondertussen de organisatie is ingebracht.

WAT KAN NOG BETER?

- Voor sommige opleidingen is de accuraatheid van de voorspellingen (net) wat minder. Met name bij beleidsmatige interventies. Vraagt continue aandacht en doorontwikkeling.

Vervolgstappen en plannen

KORTE TERMIJN

- Optimaliseren voorspellingen van specifieke opleidingen (model-technisch).
- Uitbreiden met prognose totaal aantal studenten komend jaar.
- Uitbreiden met prognose voor jaren daarna (meerjarenvoorspelling).
- Optimalisatie modulering.

MIDDELLANGE TERMIJN

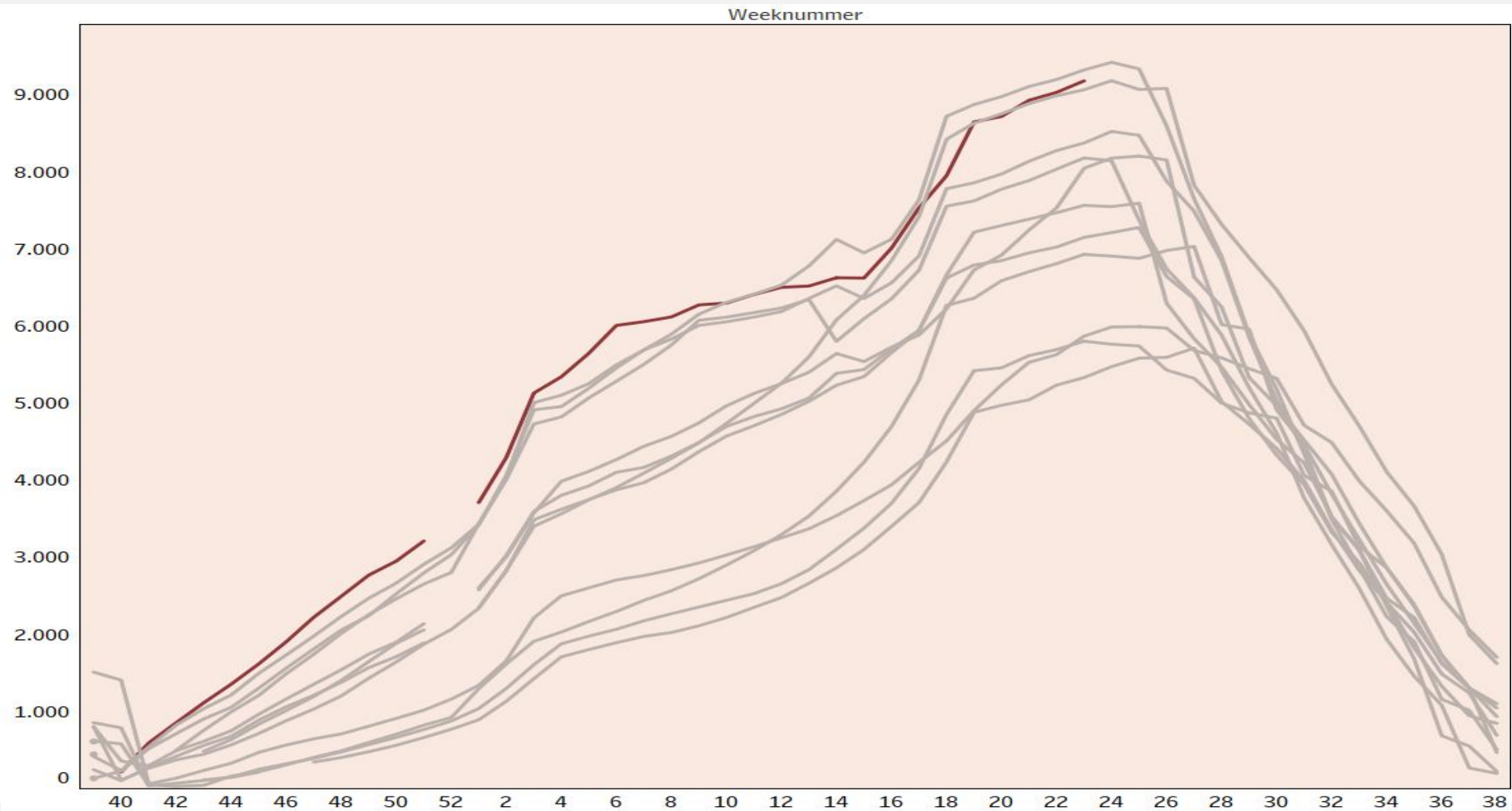
- Uitbreiden naar aanpalende, specifieke wensen, zoals roostering collegezalen.
- Uitbreiden met voorspellingen op specialisatieniveau.
- Examentype uitbreiden >> toevoegen voorspellingen premasters

LANGE TERMIJN

- Het (deels) beschikbaar stellen prognosemodel aan andere instellingen (Npuls).
- Hier meer informatie.

MODELLEN EN METHODE

RATIOMODEL





Methode 1: cumulatieve vooraanmeldingen

- **Data**
 - Cumulatieve databestand van studielink (gewogen vooraanmelders + inschrijvingen)
- **Methode**
 - Methode die lijkt op de SIOUX methode
 - De gewogen vooraanmelders (+ inschrijvingen) worden doorgetrokken naar het einde van het jaar (1 oktober) dmv sarima.
 - Op basis van deze data wordt via gradient boosting een voorspelling gedaan voor het aantal studenten

Methode 1: cumulatieve vooraanmeldingen

- **Voordelen**

- Presteert relatief goed (wel iets minder dan de andere methode)
- Relatief constante voorspellingen omdat beslissingen gedaan worden door een beslisboom algoritme.

- **Nadelen**

- Past zich niet heel sterk aan aan nieuwe informatie.
- Verder in het jaar betekent niet altijd dat voorspellingen beter worden (in tegenstelling tot de andere methode)



Methode 2: individuele vooraanmeldingen

- **Data**

- We maken aanvullend gebruik van de individuele informatie per vooraanmelder.
- Nationaliteit is een goede voorspeller, alsmede het ratio gewogen vooraanmelders t.o.v. ongewogen vooraanmelders uit het cumulatieve vooraanmeldbestand.
- (Woon)afstand tot de universiteit is voor NL-studenten een goede voorspeller.

- **Methode**

- We voorspellen hoe groot de kans is (tussen 0 en 1) dat een student zich inschrijft o.b.v. historische gegevens, d.m.v. een beslisboom-algoritme (XGBoost)
- De som van de kansen voorspelt het aantal studenten t/m de week waar we de inschrijvingen van hebben.
- Deze lijn wordt door middel van een algoritme (SARIMA) doorgetrokken tot 1 oktober (per week waar we in zitten), waardoor je steeds een voorspelling krijgt van het aantal studenten dat we op 1-oktober verwachten.

Methode 2: individuele vooraanmeldingen

- **Voordelen**

- Lijkt het beste te presteren.
- Bottom-up aanpak (zonder cross-training).
- Vanaf 1 mei zijn de voorspellingen (voor BA) redelijk accuraat (MA wat later).
- Andere methodes (o.a., Holt-Winters Time-Series Analysis) presteren minder goed dan de huidige methode.

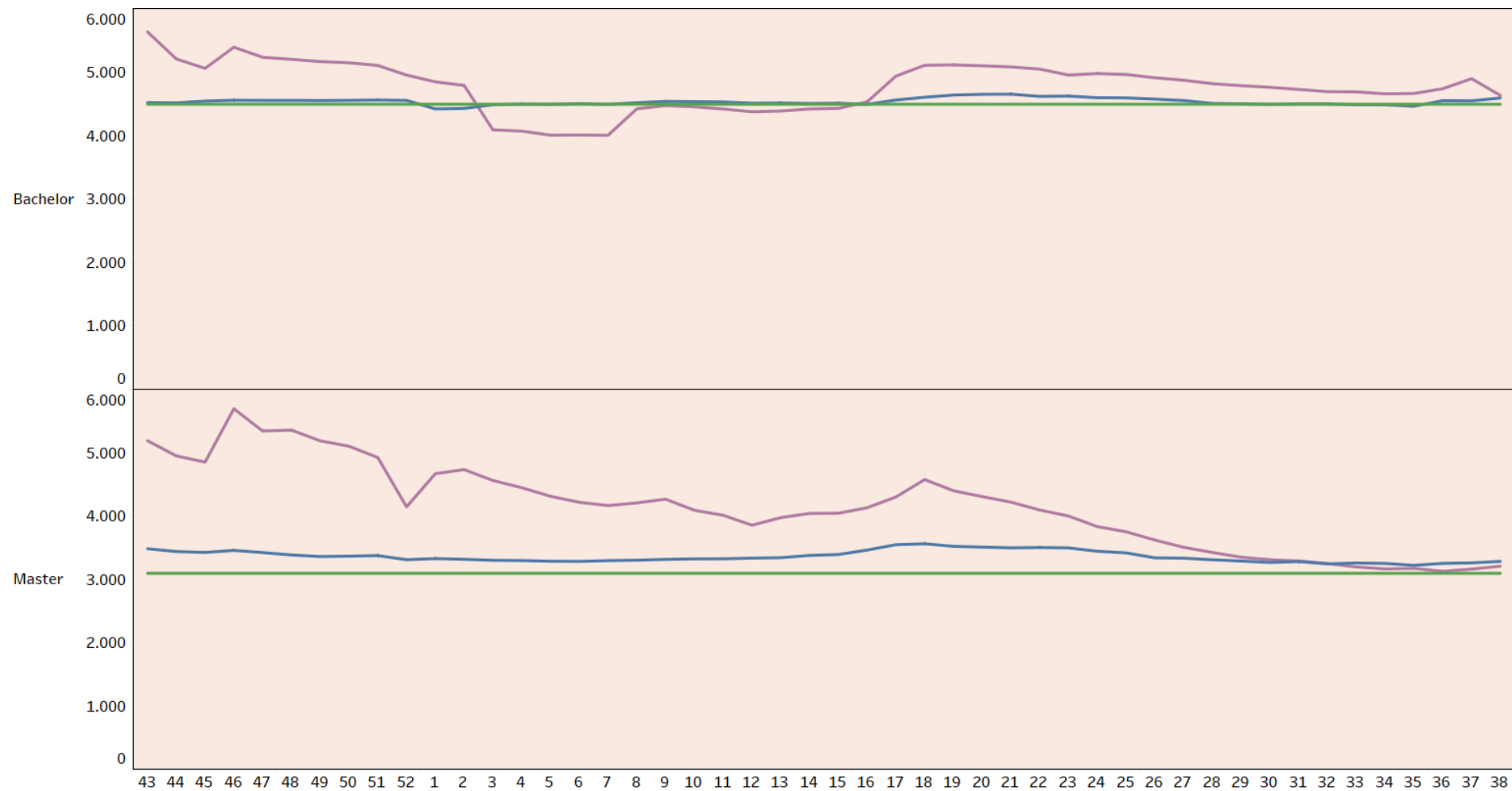
- **Nadelen**

- Over het algemeen blijft het moeilijk om te voorspellen of een student zich gaat inschrijven (begin van het jaar) en het doortrekken van de lijn.
- De verhouding vooraanmelding en daadwerkelijk inschrijven is niet altijd hetzelfde en varieert van jaar tot jaar.



Combinatiemodel

- De uiteindelijke voorspelling wordt een gewogen gemiddelde van de drie methodes
- Voordelig omdat alle drie de modellen op een andere manier reageren op de data (bijvoorbeeld met een constante over-/ondervoorspelling)
- Algoritme geschreven die uitzoekt welke methode beter is dan de rest, en anders wordt er een gewogen gemiddelde genomen van (één van) beide of alle methodes.



Dashboard

<https://www2.ru.nl/tableau/instroomprognoseradboud.html>