

# ZIEKENHUIS LEEGDUUR VOORSPELLEN

Met deze opdracht heb je de mogelijkheid sterk bewijsmateriaal te creëren voor de volgende leeruitkomsten in je portfolio:

- Student past de principes van data gedreven werken toe binnen de specifieke context van de zorg en supply-chain domeinen.
- Student past geavanceerde analytische methoden zoals Visioning, NLP en Machine Learning toe om complexe voorspellende vraagstukken op te lossen.
- Student implementeert en optimaliseert dataverwerkingsprocessen in Power BI, waarbij de student data voorbereidt, modelleert, visualiseert, analyseert en assets effectief beheert.
- Student communiceert effectief door best practices in data visualisatie, UX en taalgebruik toe te passen.

Voor deze opdracht gebruiken we een fictieve dataset die representatief is voor de verblijfsduur van patiënten in een ziekenhuis. Je moet voor deze casus over ziekenhuiscare met voorspellende modellen en effectieve visualisaties aantoont dat je een groot deel van de stof van onderwijsperiode 2 integraal beheerst. Let op! Voorspellende modellen is dus o.b.v. machine learning technieken. Je hebt hiervoor de keuze uit de volgende tools: Python (code), Azure ML (low-code) of RapidMiner (no-code).

Voor deze opdracht gebruiken we fictieve datasets die representatief zijn voor een ziekenhuis bezetting. Jullie dienen de data te analyseren, bepalen wat de “features” zijn, bepalen welke voorspelmodellen er gebruikt zouden kunnen worden, het beste model kiezen, het model valideren en een advies en conclusie o.b.v. de data aan de directie van het ziekenhuis te geven (COO en CDO). De CDO wil naast de adviespresentatie ook een gedetailleerd rapport hebben met daarin alle concepten, details, onderbouwing en reflecties op de technieken en gebruikte technieken. De bonus opdracht is om van het gekozen model een productie applicatie te maken. Daarbij passen jullie de kennis over data engineering en de predictive technieken toe.

In dit scenario ben je net in dienst als data analist en ben je toegewezen aan 3 losse opdrachten. Je manager verwacht dat je jezelf in je proeftijd bewijst door deze 3 vragen goed te beantwoorden.

## SITUATIE

**Introduction:** Healthcare organizations are under increasing pressure to improve patient care outcomes and achieve better care. While this situation represents a challenge, it also offers organizations an opportunity to dramatically improve the quality of care by leveraging more value and insights from their data. Health care analytics refers to the analysis of data using quantitative and qualitative techniques to explore trends and patterns in the acquired data. While healthcare management uses various metrics for performance, a patient's length of stay is an important one.

Being able to **predict the length of stay (LOS)** allows hospitals to optimize their treatment plans to reduce LOS, to reduce infection rates among patients, staff, and visitors

**Project Goal:** The goal of this project is to accurately predict the Length of Stay for each patient so that the hospitals can optimize resources and function better.

**Hypothesis Generation:** Understanding the problem in detail by assuming different factors that impact the outcomes of Length of Stay before any data exploration or analysis. Here the variables can be divided into two levels: Patient-Level and Hospital-Level.

#### Patient-Level:

- *Type of Admission – Patients can be admitted in three levels Urgent, Emergency, and Trauma. Patients admitted to urgent care are likely to stay fewer days. Whereas Trauma patients usually stay longer because they must be monitored until they are qualified to be discharged.*
- *Severity of Illness – Severity can be classified as Minor, Moderate, and Extreme. A patient recorded as minor will stay fewer days than a patient recorded as extreme.*
- *Visitors with Patient – Patients with more visitors are like to stay longer in the hospital.*
- *Age – Infants and older Patients usually take a longer time to recover so they stay longer than younger Patients.*
- *Admission Deposit – Patients who are likely to deposit a high amount of money at the time of admission might have severe conditions and stay longer.*

#### Hospital-Level:

- *Ward Type – Patients allocated in ICU might stay longer than the general ward as their condition is more severe.*
- *Department – Patients under surgery are likely to stay longer than gynaecology as their recovery time is longer.*

Een ziekenhuis is naast een zorginstelling ook een bedrijf dat zo optimaal mogelijk de kosten moet beheersen en capaciteit en bezetting zo optimaal mogelijk moet plannen en voorspellen. Overbezetting zorgt voor teruglopende zorg en afzeggingen van reeds geplande zorg en daarmee het vergroten van wachtlijst. Denk maar terug aan de effecten van Corona op de zorg.

Een mooie kans om jezelf te bewijzen! Je hebt 3 opdrachten gekregen en een bonus opdracht.

1. **Data Analyse:** Analyseren, beschrijven, visualiseren, feature engineering en transformeren van de data om dit bruikbaar te maken voor Machine Learning.
2. **Model building:** Het testen van verschillende modellen en hieruit het beste model kiezen en beschrijven waarom. Tevens de juist validatie met uitleg geven (confusion matrix, AOU, nauwkeurigheid model etc.).
3. **Organisatie en Productie aanbeveling:** Aangeven hoe het model in productie gebracht zou kunnen worden (infra- en applicatie beschrijving). Dit in relatie tot welke inzichten er uit de data en de voorspellingsanalyse er te halen zijn. Tevens zou de CDO graag een advies hebben of er nieuwe Artificial Intelligentie (AI) tools er gebruikt kunnen worden en wat hier de voordelen van zijn.
4. **Bonus opdracht:** het model daadwerkelijk in productie plaatsen.

## 1. DATA ANALYSE

Je hebt van de organisatie al de dataset gehad die bestaat uit 2 delen, een training en een test set (train.csv en test.csv). Als analist wordt er van jou verwacht dat je weet wat dit betekent in het kader van het selecteren en maken van machine learning modellen. Natuurlijk verwacht de CDO dat je nog wel kritisch kijkt naar deze data en hier EDA op toepast naast het beschrijven van de dataset. De uitkomsten hiervan moet gedocumenteerd worden in het op te leveren rapport voor de CDO en je moet aangeven in het adviesrapport of er aan de data en de data kwaliteit in het kader van Data Governance nog iets gedaan moet worden.

### VOORWAARDEN

#### Technisch

1. Geef aan hoe je de EDA hebt uitgevoerd en met welke tools.
2. Geef aan hoe je structureel transformatie en EDA kan uitvoeren (dus niet eenmalige handmatige handelingen).

#### Informatie behoefte business

1. Wat zijn de verschillende EDA uitkomsten op gebruikelijke categorieën (bv. lege velden etc.)? Doe hiervoor onderzoek want er zijn er een aantal.
2. Beschrijf het data cleaning en preparation proces indien van toepassing.
3. Beschrijf en visualiseer de conclusie en aanbeveling op business, technisch en data niveau.

#### Hoe in te leveren

De adviespresentatie over data en het gedetailleerde rapport en andere gebruikte bestanden toevoegen aan het zip bestand dat je dient in te leveren op 14 januari 2024 op Brightspace.

## 2. MODEL BUILDING

O.b.v. de gegeven data (eventueel na een eigen transformatie) wordt er verwacht dat je verschillende algoritmes gaat test om te bepalen wat het beste model zou kunnen zijn. Je moet dit onderbouwen met overzichtelijke informatie. Daarbij hoort dus ook een gedegen validatie en een confusion matrix, AUC, nauwkeurigheidspercentage en eventueel andere statistisch informatie of conclusies. De CDO verwacht van jou dat je hiervoor beschikt over de juiste kennis.

### VOORWAARDEN

#### Technisch

1. Geef aan hoe je de *feature engineering* hebt gedaan.
2. Licht toe welke tools je onderzocht of gebruikt hebt voor het maken van een model.
3. Leg uit hoe je verschillende visualisaties hebt gemaakt.

#### Informatie behoefte business

1. Beschrijf de uitkomsten en conclusie van de *feature engineering*.
2. Maak een feature heatmap (visualisatie).
3. Welke tool(s) heb je gebruik om een model te testen en wat zijn de conclusies.
4. Wat zijn de resultaten van de verschillende modellen en wat is de beste (resultaat in tabel vorm)?
5. Een goede validatie en een confusion matrix, AUC, nauwkeurigheidspercentage en eventueel andere statistisch informatie of conclusies.

#### Hoe in te leveren

De adviespresentatie over het beste model en het gedetailleerde data rapport voor de CDO en andere gebruikte bestanden toevoegen aan het zip bestand dat je dient in te leveren op 14 januari 2024 op Brightspace

### 3. ORGANISATIE EN PRODUCTIE AANBEVELING

Je hoeft de oplossing niet te bouwen binnen de organisatie. Hiervoor hebben ze experts en een ICT afdeling. Je moet echter wel beschrijven welke tools, producten en applicatie/infrastructuur landschap er nodig is. Het geheel moet centraal opgezet worden en schaalbaar zijn.

Tevens zou de CDO graag een advies hebben of er nieuwe Artificial Intelligentie (AI) tools er gebruikt kunnen worden en wat hier de voordelen van zijn.

De organisatieadviezen zijn wat uitdagender voor jouw als analist maar als BIMer kan je hier wel een rol in spelen. Je hebt immers het boek *de intelligente organisatie* en kan daarmee, in combinatie met andere bronnen, je advies onderbouwen. Dit in relatie tot welke inzichten er uit de data en de voorspellingsanalyse te halen zijn. Denk bv. aan:

- Verbeteren planningsproces
- Slimme personeelsplanning en personeelsbeheer (Smart Staffing & Personnel Management)
- Geavanceerd risico- en ziektebeheer (Advanced Risk & Disease Management)
- Real-time waarschuwingen: klinische beslissingsondersteuning (CDS - Clinical Decision Support)
- Verbetering van de benadering van patiënten (Enhancing Patient Engagement)

#### VOORWAARDEN

##### Technisch

1. Zorg dat je duidelijk beschrijf hoe de tools zijn toegepast en dat het geheel reproduceerbaar is voor een andere medewerker

##### Informatie behoefte business

1. Welke tool(s), producten en applicatie/infrastructuur landschap er nodig om dit in te richten?
2. Welke adviezen zijn er te geven op organisatieniveau (onderbouw je aanbevelingen en gebruik literatuur als *de intelligente organisatie*) over de vervolgstappen?
3. Welke AI tools zouden het proces verder en slimmer kunnen automatiseren?

##### Hoe in te leveren

De adviespresentatie over de te nemen vervolgstappen en het gedetailleerde data rapport en andere gebruikte bestanden toevoegen aan het zip bestand dat je dient in te leveren op 14 januari 2024 op Brightspace.

#### 4. BONUS OPDRACHT

Beschrijf hoe het model daadwerkelijk in productie kan plaatsvinden. Dus hoe moet de implementatie daadwerkelijk geregeld worden? Onderbouw je advies op verschillende bronnen waaronder in ieder geval het boek 'de intelligente organisatie'.

#### EXTRA TOELICHTING

Binnen het eLearning materiaal van predictive zitten verschillende artikelen, boeken, code en data files die je kan gebruiken om de technische kant van dit onderwerp te onderzoeken.

Binnen het eLearning materiaal van Visualizing Information is materiaal aanwezig die helpt bij het maken van de meest effectieve visualisaties.

In het boek *de intelligente organisatie* van Daan van Beek is beschreven hoe AI modellen tot waarde gemaakt kunnen worden in organisaties.

Als je in Google of ergens anders zoekt op artikelen met als term "ligduur voorspellen voor ziekenhuizen" dan kom je veel bruikbare artikelen tegen