Breda University OF APPLIED SCIENCES

PROJECT PRESENTATION

Daria Vladutu
Mohamed Elshami
Ron Lev Tabuchov
Peter Husen
Year 1 - Block D
Team 18

The Team

Daria Vladutu

Mohamed Elshami

Peter Husen

Ron Lev Tabuchov

Index

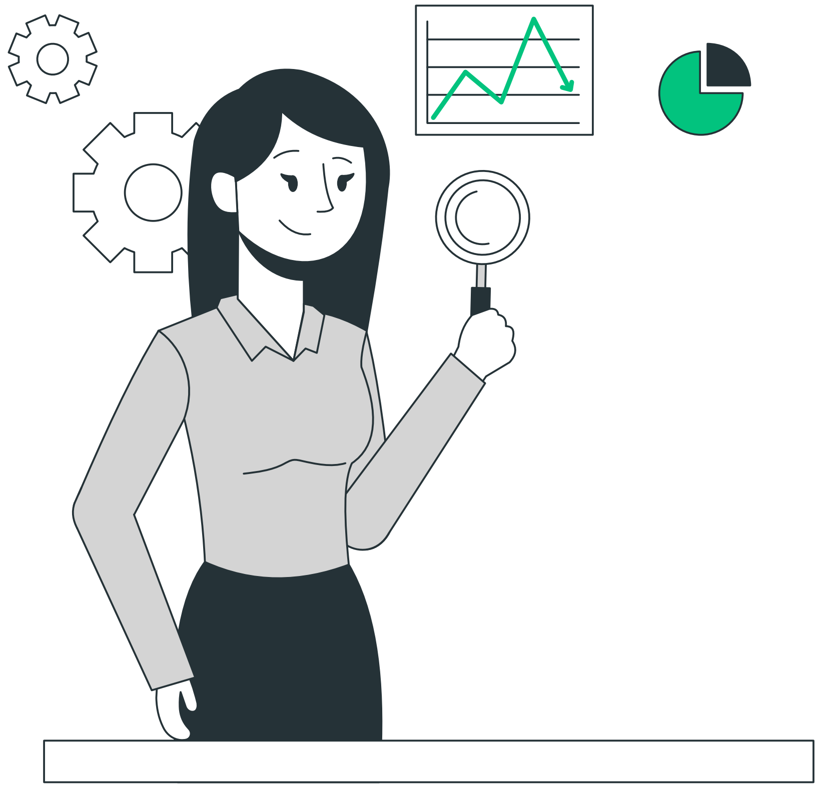
Problem Statement

Problem Statement

- **Improve road safety in Breda by predicting danger zones.**

- **Objective:** Develop a predictive model to identify the risk level of each street in Breda city as low, mid, or high risk.
- **Implementation:** Notify drivers about the risk level of the streets they are passing through, allowing them to adjust their behaviour and be more cautious.
- **Expected Impact:** This model will help reduce accidents and enhance overall safety in Breda city.

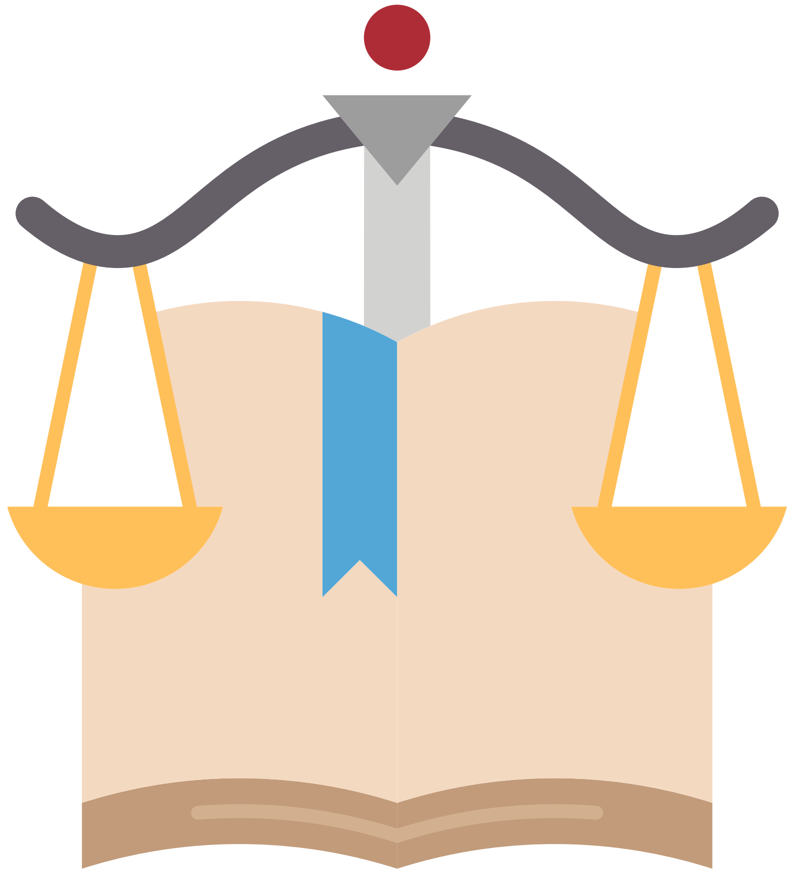
Business Value



Business Value for ANWB

- **Enhanced Service Offering:** Integrating the predictive model into the ANWB app will enhance its functionality and offer a unique selling point by providing real-time danger zone notifications.
- **Increased User Engagement:** Safety alerts will encourage frequent app usage and improve user retention by enhancing user satisfaction.
- **Brand Reputation**: Prioritizing user safety will strengthen ANWB's reputation as a reliable, safety-conscious organization and build community trust.
- **Data Insights:** The predictive model will generate valuable data on road safety, aiding in strategic planning and enhancing product offerings.
- **Economic Benefits:** Reducing accidents can lower emergency service and insurance claim costs, while improved features can drive revenue growth.
- **Long-Term Benefits:** The project aligns with ANWB's long-term goals of sustainability and safety, with potential scalability to other cities.

Legal & AI Act



Legal

Intended purpose

Integrated feedback

Human oversight

Cybersecurity

Output logs

Legal

Safety component, decision influencing

High risk system

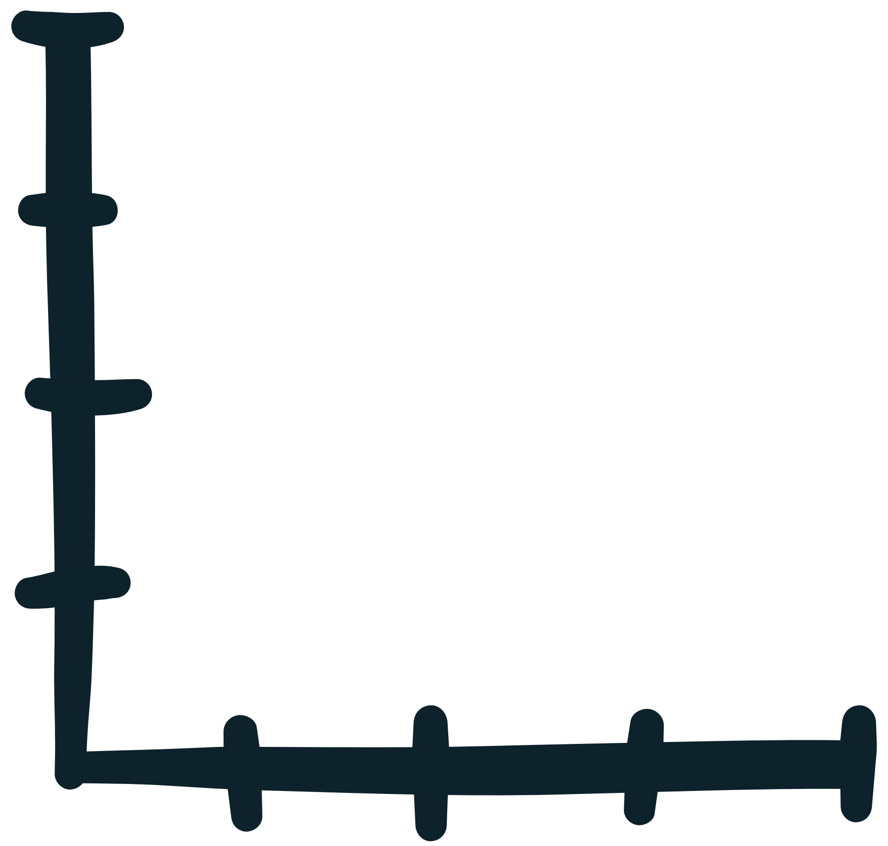
Representative data

Supportive tool, not overwriting

Document findings

Transparency

Data



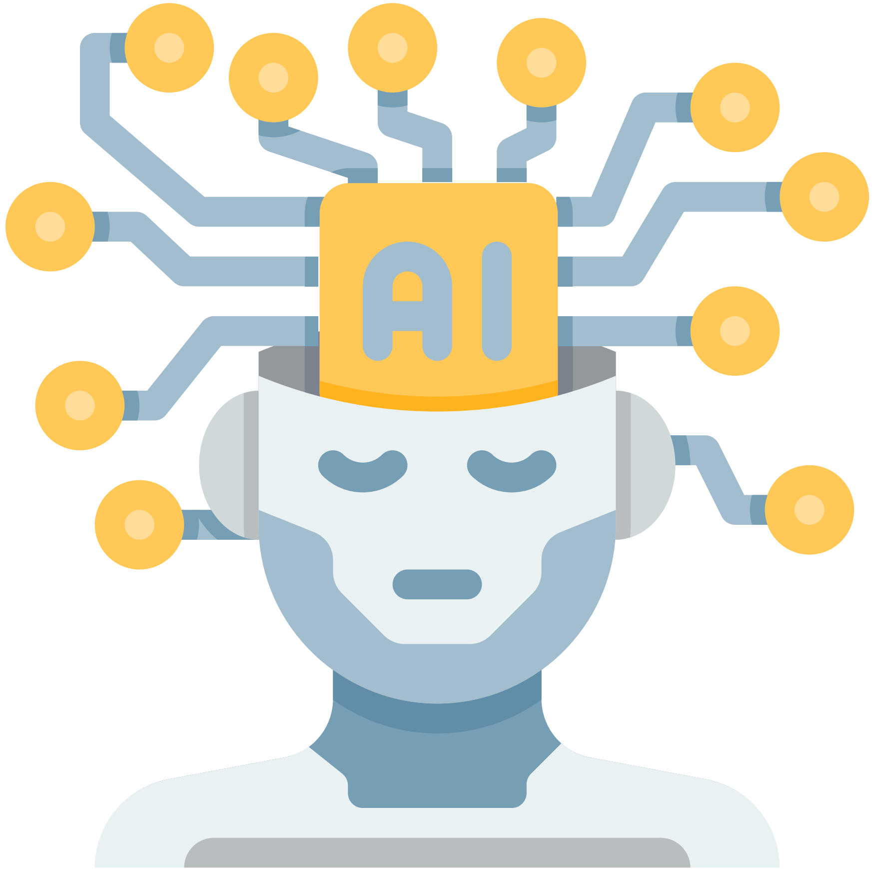
Data analysis

Describe EDA for all 3 models

Data preparation

Describe preprocessing steps for all 3 models

Machine Learning



Machine Learning

The purpose of using machine learning in improving road safety in Breda is to develop predictive models that can identify the risk level of the roads.

By leveraging various machine learning algorithms, the system can analyze large datasets to classify roads based on their risk levels, and provide real-time warnings to drivers.

This proactive approach aims to reduce the number of accidents and enhance overall road safety by informing and guiding drivers towards safer driving behaviors

Machine Learning

The project employs several types of machine learning models, each serving a specific role:

- **Random Forest:** Used for classifying roads based on risk levels by analyzing features such as weather conditions, road names, and incident counts.
- **K-means Clustering:** Utilized for grouping road segments into clusters based on geographic and incident data to identify patterns in accident occurrences.
- **Decision Tree Model:** Employed for understanding the decision-making process and predicting the risk levels of roads based on selected features.
- **Deep Neural Network:** Implemented to capture complex relationships in the data and enhance the accuracy of risk level predictions.
- **Recurrent Neural Network:** Used for time-series analysis to predict future road incident severities based on historical data.

Machine Learning

Model Comparison

The **Decision Trees** model (3rd iteration), tuned using RandomizedSearchCV, caps its overall test **accuracy of 97.38%**. The high precision, recall and F1-score across all classes indicate that the model is effective at distinguishing classes.

The **DNN model** (2nd iteration) shows impressive results with **98.86% accuracy** and overall high metrics, especially for class 0 - low ris, and 2 - high risk. However, due to its added complexity as a neural network model with L2 regularization, it is not better feasibility-wise than a standard machine learning model.

The 3rd iteration of **Random Forest** model, with the highest accuracy out of all basic Machine Learning models peaking to **98.84% accuracy**, exhibits robust precision, recall, and F1-score across all classes. Furthermore, the cross-validation scores are high, with an average **score of 98.94%** across all five folds.

Machine Learning

Model Architecture - Unique Steps

```
# Define weights for severity
severity_weights = {'HA1': 1, 'HA2': 2, 'HA3': 3, 'HB1': 1, 'HB2': 2, 'HB3': 3, 'HC1': 1, 'HC2': 2,
'HC3': 3,'HC4': 2, 'HC5': 3, 'HC6': 4, 'HC7': 3, 'HC8': 4, 'HC9': 5, 'HC10': 4, 'HC11': 5, 'HC12':
6,'HC13': 2, 'HC14': 3, 'HC15': 4, 'HC16': 3, 'HC17': 4, 'HC18': 5, 'HC19': 4, 'HC20': 5, 'HC21': 5}

# Calculate the 25th and 75th percentiles
q25 = road_incident_data['severity_score'].quantile(0.25)
q75 = road_incident_data['severity_score'].quantile(0.75)

# Adjust bins to avoid duplicates
bins = [road_incident_data['severity_score'].min() - 1, q25, q75,
road_incident_data['severity_score'].max()]

labels = ['low', 'mid', 'high']
```

Machine Learning

Balancing classes with SMOTE:

Class distribution in y_train before SMOTE:
risk_level
mid   371
low   364
high   245
Name: count, dtype: int64

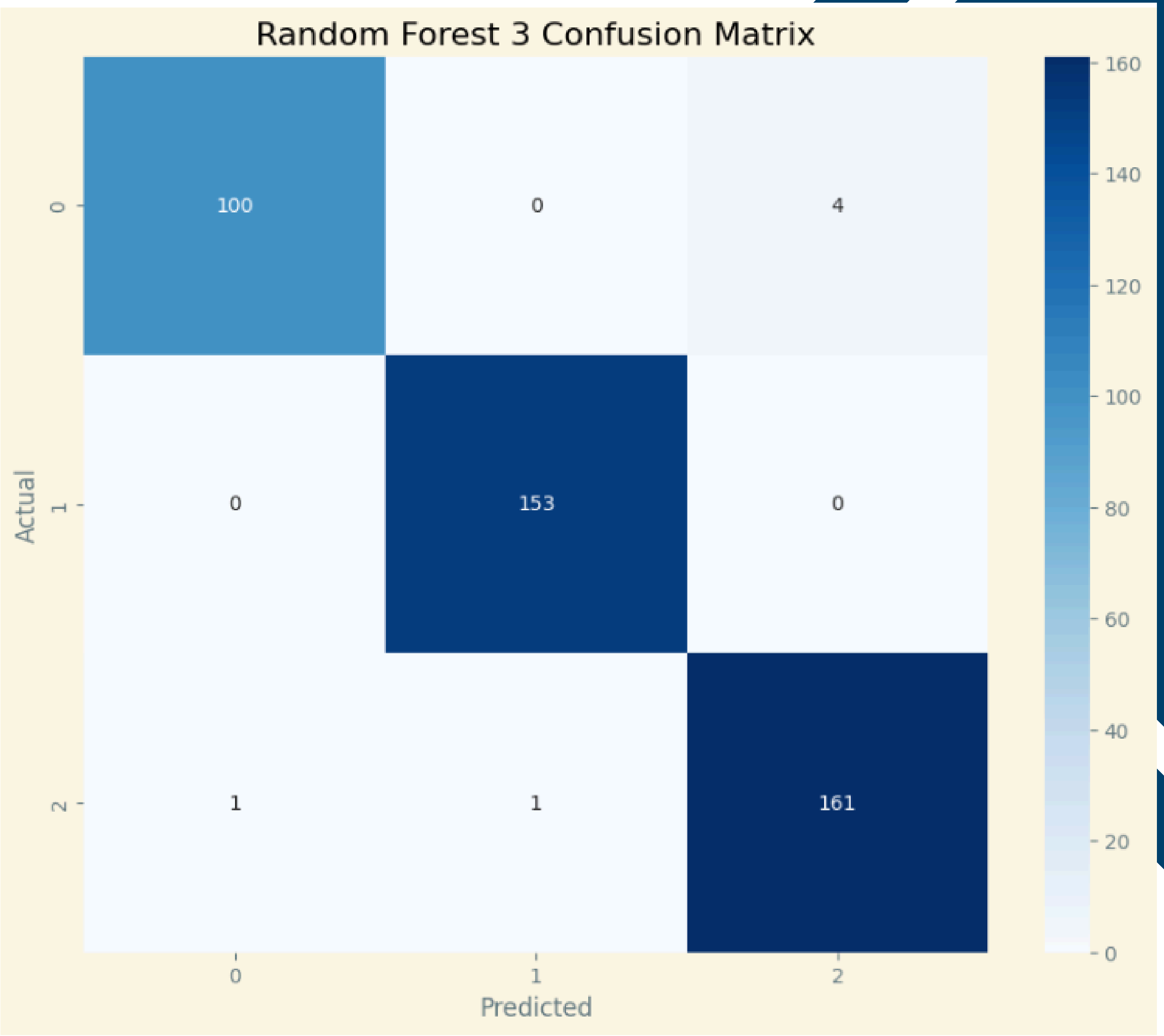
Class distribution in y_train after SMOTE:
risk_level
low   371
mid   371
high   371
Name: count, dtype: int64

Machine Learning

Model Evaluation

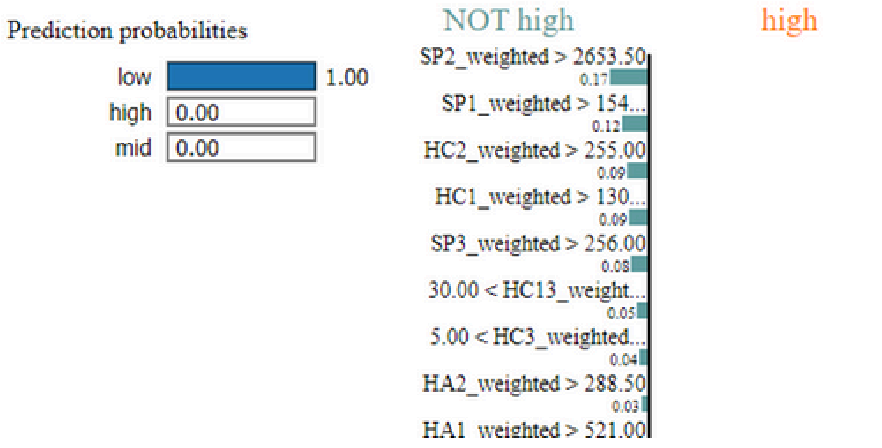
```
RandomForest Test Accuracy for Risk Level Classification: 0.9857142857142858
              precision    recall  f1-score   support

        high       0.99      0.96      0.98       104
         low       0.99      1.00      1.00       153
         mid       0.98      0.99      0.98       163

    accuracy                           0.99       420
   macro avg       0.99      0.98      0.98       420
weighted avg       0.99      0.99      0.99       420
```

```
Accuracy for each fold for 5-fold cross-validation: [0.97797357 0.99115044 0.99115044 0.98672566 1.        ]
Cross-Validation Scores: [0.97797357 0.99115044 0.99115044 0.98672566 1.        ]
Mean Cross-Validation Score: 0.9894000233909009
```

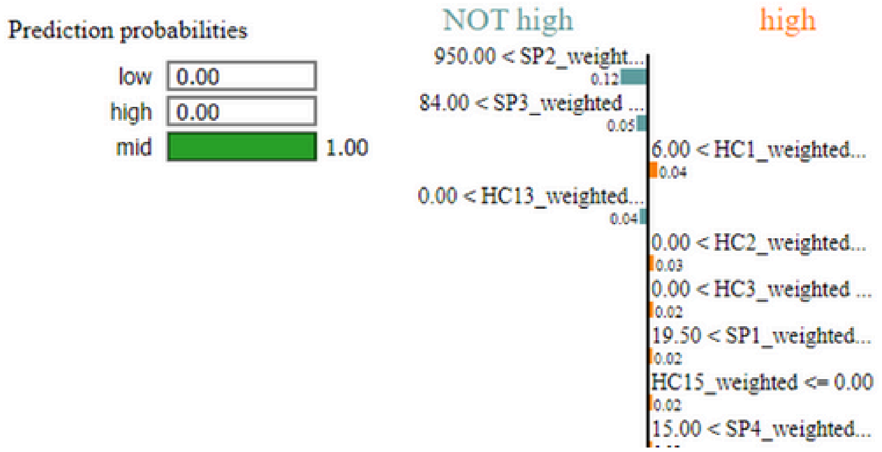


Machine Learning

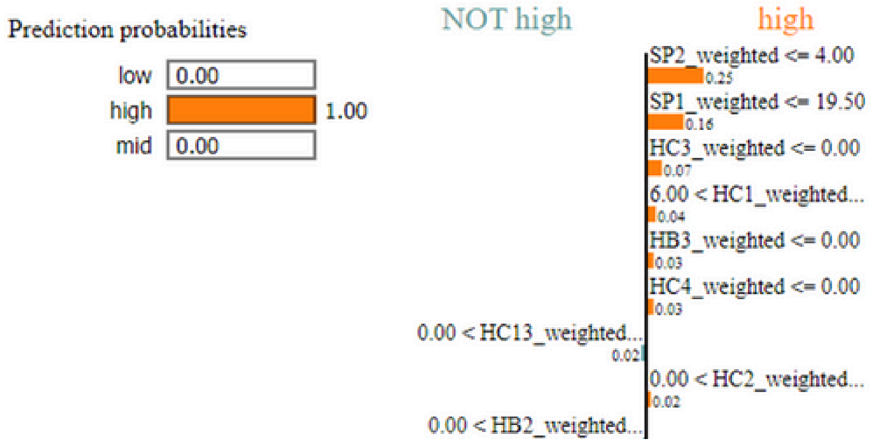
XAI



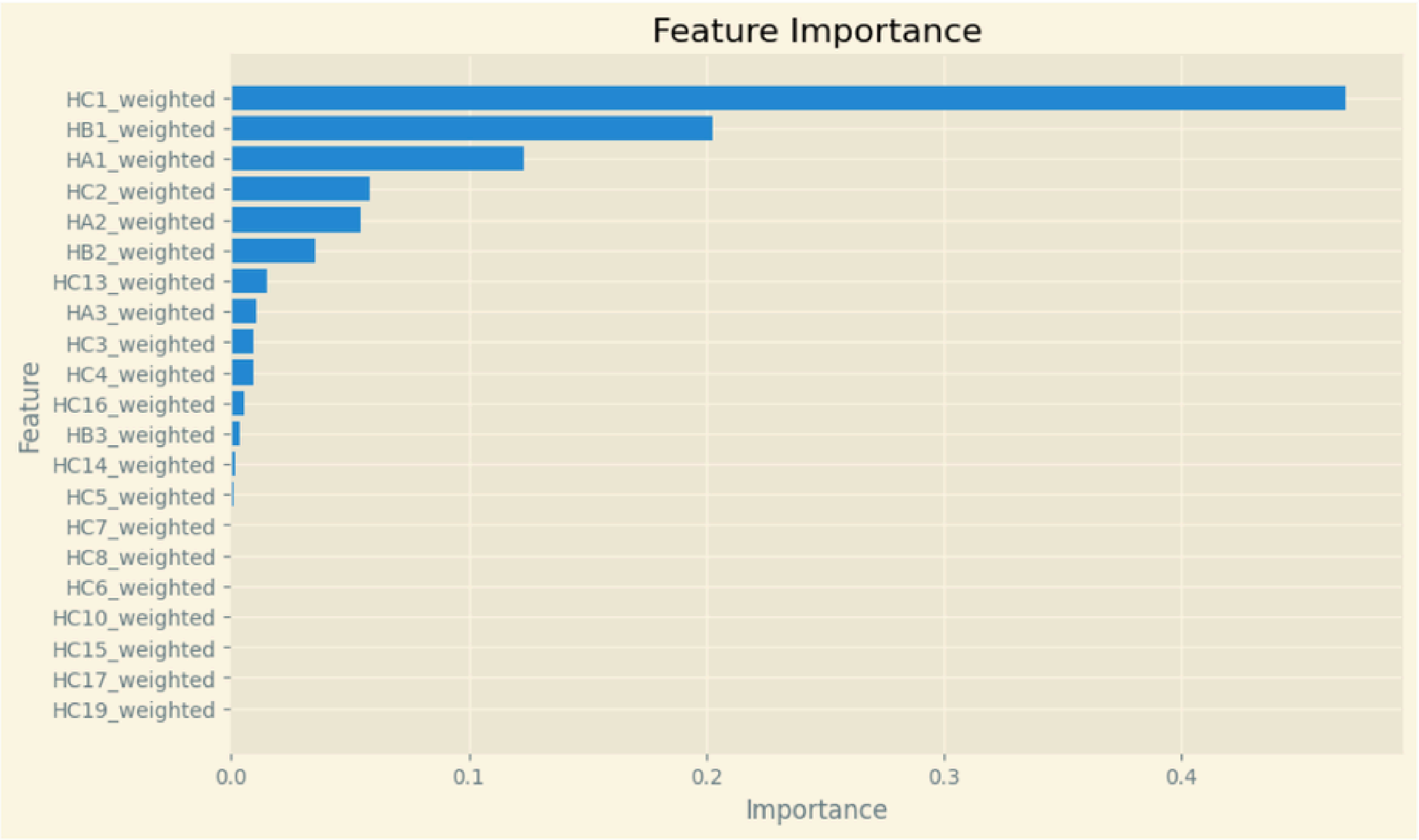
Feature	Value
SP2_weighted	4733.00
SP1_weighted	22638.00
HC2_weighted	419.00
HC1_weighted	23663.00
SP3_weighted	382.00
HC13_weighted	72.00
HC3_weighted	21.00
HA2_weighted	289.00



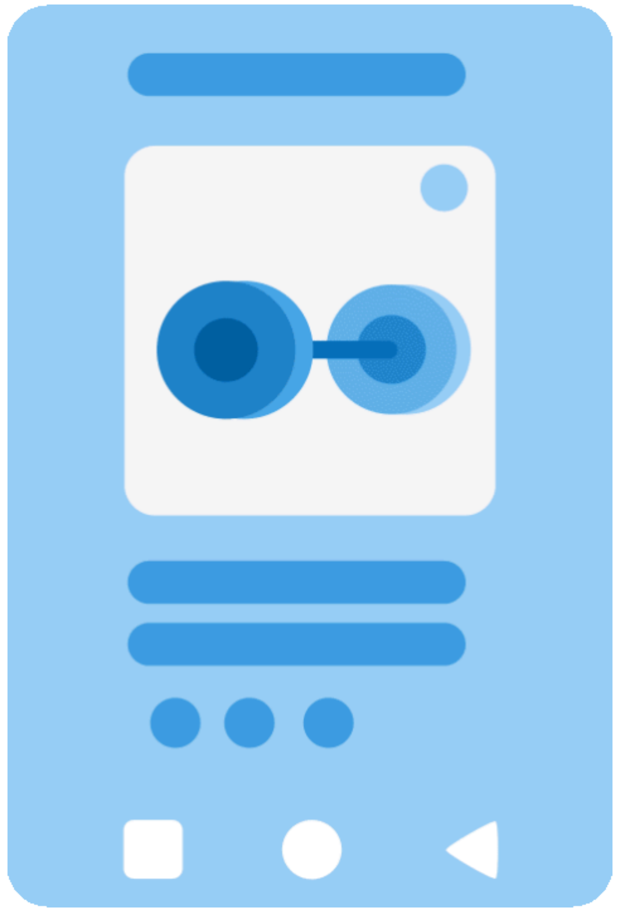
Feature	Value
SP2_weighted	1225.00
SP3_weighted	120.00
HC1_weighted	283.00
HC13_weighted	16.00
HC2_weighted	25.00
HC3_weighted	1.00
SP1_weighted	6618.00
HC15_weighted	0.00



Feature	Value
SP2_weighted	2.00
SP1_weighted	7.00
HC3_weighted	0.00
HC1_weighted	219.00
HB3_weighted	0.00
HC4_weighted	0.00
HC13_weighted	6.00
HC2_weighted	1.00



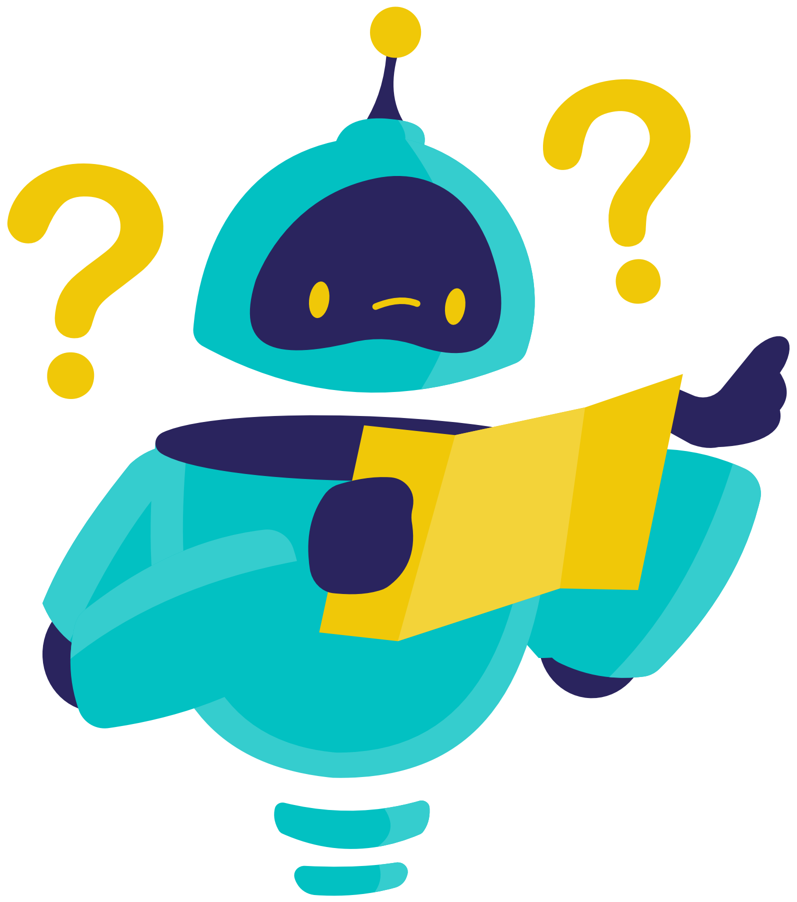
Deployment - Interface Design



XXXXX

XXXXX

Questions?



Thank you for your attention!