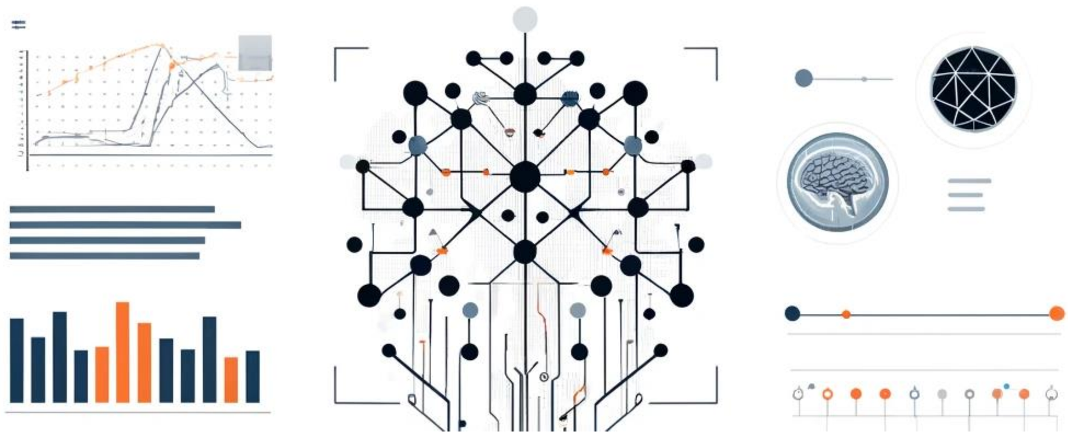
Project proposal

Project proposal - Improve the Road Safety in Breda

Ron L. Tabuchov
Mohamed K. A. M. Elshami
Daria Vladutu
Peter Husen

Data Science and Artificial Intelligence, Breda University of Applied Science



DISCOVER YOUR WORLD

Breda University OF APPLIED SCIENCES

Index

1 Introduction

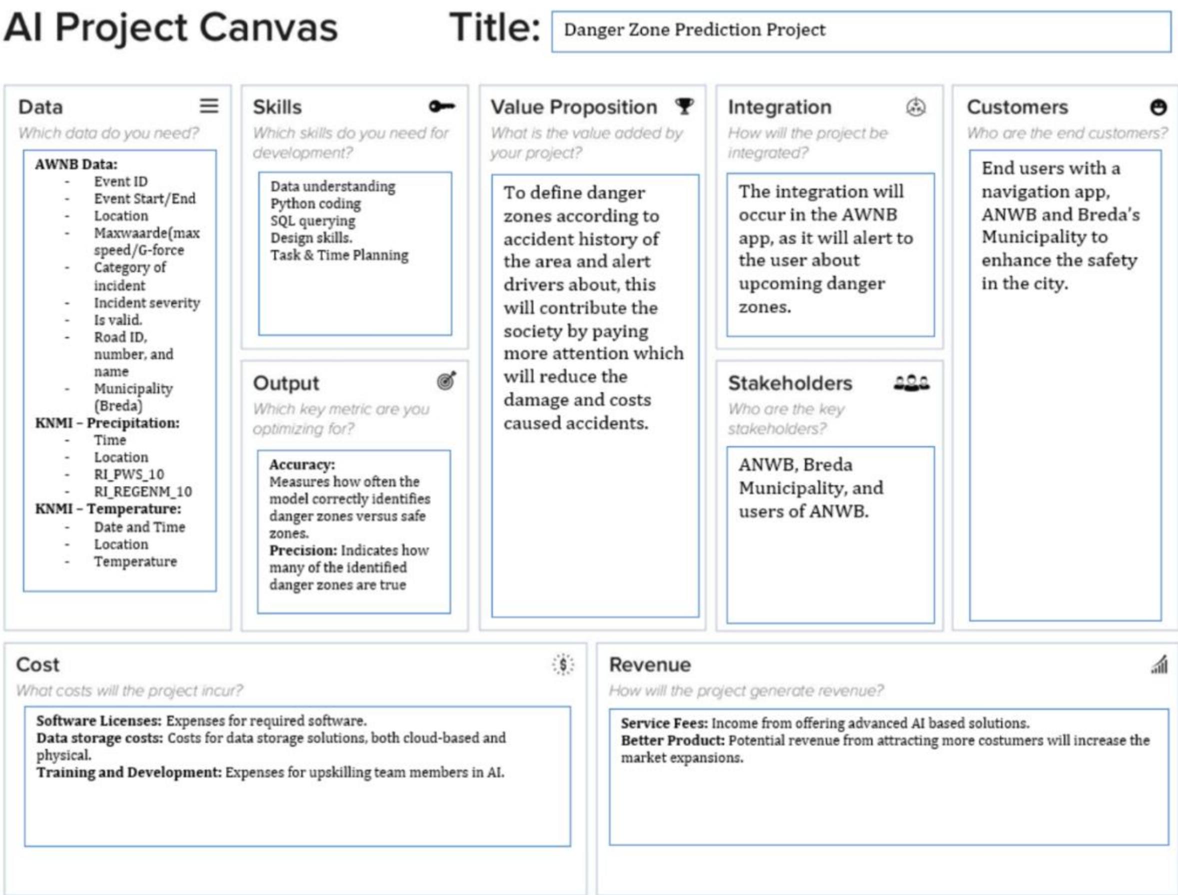
The purpose of this project is to improve road safety in Breda through innovative solutions and data-driven approaches such like identifying key risk factors, developing predictive models and implementing effective interventions to reduce traffic accidents.

2 Problem Statement

Identification and prediction of danger zones for traffic incidents in Breda.

Breda faces significant challenges with road safety. Traffic accidents result in personal injury and disruptions to the urban transport system. Current methods for identifying danger zones rely on historical data and manual analysis, which can be time-consuming and less effective.

Benefits implementing the model will reduce traffic accidents, enhance public safety, and improve traffic management as It provides city planners with data-driven insights to design safer infrastructure and implement targeted safety measures and alert drivers upon arrival to a certain risk zone or road.



3 Data Description

To effectively identify and predict high-risk zones for traffic incidents in Breda, we will utilize multiple relational datasets from reputable sources, each offering crucial attributes that contribute to the predictive modeling process. The datasets will be joined using SQL tables, enabling efficient data integration and querying. Below is a detailed description of the datasets and their significant features:

a. ANWB Dataset (Driver Characteristics)

The ANWB dataset provides comprehensive information about driver behaviors and characteristics, which are essential for understanding the human factors in traffic incidents. Significance: These attributes help in analyzing patterns related to driver demographics and behaviors, which can influence accident risk levels.

b. Open Meteo Breda Dataset (Weather Conditions)

The Open Meteo Breda dataset includes detailed weather data, which is a critical factor in road safety. The dataset contains:

• Timestamp: Date and time of the weather observation.

• Precipitation: Amount of rainfall or snowfall, which can affect road conditions.

• Temperature: Air temperature, influencing road surface conditions.

• WindSpeed: Wind speed, which can impact vehicle control.

Significance: Weather conditions have a direct impact on driving safety. Analyzing this data helps in understanding how adverse weather contributes to higher risk levels on the road.

c. KNMI Dataset (Rain Data)

The KNMI dataset provides detailed rain data, specifically focusing on:

• Timestamp: Date and time of the rain measurement.

• RainIntensity: Measured intensity of rainfall at different times.

• RainDuration: Duration of rain events.Significance: Rain intensity and duration are crucial for assessing the risk levels on the road. Heavy and prolonged rain can lead to slippery surfaces and reduced visibility, increasing the likelihood of accidents.

Created Features

To enhance the predictive power of our models, we have created an additional feature based on the provided datasets. Risk Level: Categorizes rain intensity into risk levels (e.g., low, medium, high) to quantify the impact of different rain conditions on road safety.

4 Methodology

Objective: Develop a machine learning model to identify and predict high-risk zones for traffic incidents in Breda using relational datasets.

To identify and predict high-risk zones for traffic incidents in Breda, we will be using a diverse set of relational datasets provided by ANWB (Driver Characteristics), KNMI (Weather), and BRON (Accidents). They include important features such as driving speed at the time of incidents and their locations, precipitation information, temperature information, etc.

Following the CRISP-DM cycle, in the data pre-processing phase we will be cleaning the data, merging the datasets on common keys and creating new features that are trivial to our problem statement. Exploratory data analysis has been performed on almost all datasets, however, we will be more thorough with this first in order to build a stable plan of action.

For the chosen model, we have concluded that a Multi-layer Perceptron would be best suited due to its ability to capture complex relationships even when working with larger datasets, they are versatile which will be useful when experimenting with hyperparameters and with the way we might change our approach to the problem itself (for example, solving it using a regression model vs. classification).

Training will be realised using the train, test, split method and with a validation set (70, 20, 10). Hyperparameter tuning will also be included in this part of the process.

Evaluation will be done using methods such as MSE, confusion matrix, precision, recall, and of course the f1-score, plotting the training and validation accuracies, etc.

Deployment will be concluded with a final wireframe, demo video, and a survey conducted on 10 students.

Possible model architecture:

```
# Create a sequential model
model = Sequential()

# Add 3 dense layers of 128, 64 and 32 neurons each
model.add(Dense(128, input_shape=(12288,), activation='relu'))
model.add(Dense(64, activation='relu'))
model.add(Dense(32, activation='relu'))

# Add a dense layer
model.add(Dense(6, activation='softmax'))

# Compile your model
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])

# Train your model
history_no_earlystopping = model.fit(X_train, y_train, epochs=20, validation_data=(X_val, y_val))
```

Risk Assessment:

High Risk: The AI system falls under high-risk applications as it directly impacts public safety and infrastructure management.

Risks: Potential bias in predictions, data privacy concerns, and misinterpretation of model outputs leading to incorrect decision-making.

Legal Obligations and Approach to Addressing Them:

GDPR Compliance:

Data Anonymization: Ensure all personal data is anonymized to protect privacy.

Data Minimization: The data that will be primarily used has been collected. If necessary, more datasets will be added to the data lake.

Consent Management: Obtain explicit consent from individuals if personal data is used.

Transparency and Explainability:

Model Explainability: Use techniques like SHAP (SHapley Additive exPlanations) to explain model

predictions.

Documentation: Provide clear documentation on how the model works, its inputs, and its outputs.

Bias and Fairness:

Bias Detection: Regularly check for bias in model predictions against protected attributes (e.g., age, gender).

Fairness: Implement fairness constraints to ensure the model does not disproportionately affect any specific group.

High-Risk AI Regulations (EU AI Act):

Risk Assessment: Conduct a thorough risk assessment to identify and mitigate potential risks.

Conformity Assessment: Ensure the AI system undergoes a conformity assessment before deployment.

5 Project Timeline

Project Timeline and high-level plan can be accessed in the link below:

https://github.com/BredaUniversityADSAI/2023-24d-fai1-adsai-teamwork-t18/blob/main/Improve the safety in Breda - Plan.docx

6 References

- European Parliament. (2021). Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206
- European Union. (2016). General Data Protection Regulation (GDPR). https://eur-lex.europa.eu/eli/reg/2016/679/oj