PROJECT PROPOSAL

BLOCK 1D-2023-2024

Accident predicting algorithm for road safety

Names of the Students:

Sally Ibrahim, Petar Paskalev, Sasha Stacie, Mihail Fratu, Jonas Vos

Date 24/05/2024

Introduction

Globally, road traffic accidents remain a severe public health challenge, claiming approximately 1.2 million lives each year according to the World Health Organization. In the Netherlands, the situation is particularly alarming: 2022 saw an increase of over 20 percent in road fatalities and serious injuries compared to the previous year, with totals reaching 745 deaths and 8300 serious injuries. Research from the Institute for Road Safety Research predicts that these numbers will continue to rise in the coming years.

In response to this growing concern, our project proposes the development of a Google Maps extension designed to enhance road safety through advanced predictive analytics. This innovative extension aims to evaluate and predict the risk levels of roads during a journey, categorizing them into low, medium, or high risk. By leveraging a combination of historical and real-time data, including weather conditions, light conditions, road surface, area types, accidents, speed limit, and the severity of past incidents. Our system will provide drivers with risk assessments. These assessments are intended to assist drivers in making safer travel decisions and avoiding potentially hazardous routes.

The key objectives of this project are:

- To develop a machine learning model that accurately predicts road risk levels categorized into low, medium, and high.
- To integrate this predictive model seamlessly into a Google Maps extension, providing users with intuitive visual cues about road safety.
- To enhance the decision-making process for drivers, enabling them to choose safer routes and reduce the likelihood of accidents.

Applying machine learning techniques to this problem is particularly significant for several reasons:

- **Predictive Accuracy**: Machine learning algorithms excel at identifying patterns and making predictions based on complex datasets. By analyzing historical data and real-time inputs, our models can provide highly accurate risk assessments that would be impossible with manual methods.

- **Dynamic Updates**: Machine learning models can continuously learn and adapt from new data, ensuring that the risk predictions remain relevant and up to date with changing conditions.

- **Scalability**: The use of machine learning allows our system to process vast amounts of data efficiently, making it scalable and applicable to diverse geographic regions and road conditions.

Problem Statement

Defining the Problem:

The fundamental challenge our project addresses is the proactive prediction and prevention of road accidents, a critical shift from the current reactive safety measures that are prevalent today. Traditional road safety systems often respond to incidents after they have occurred, rather than preventing them. This reactive approach results in numerous avoidable accidents each year, highlighting a significant gap in current road safety strategies.

Context and Background:

Road safety is a complex issue influenced by a multitude of factors including weather conditions, road types, times of day, and the severity of past incidents. Each of these elements can significantly alter the risk profile of driving routes and times. Our project proposes to analyse these variables using advanced data analytics to identify potential accident hotspots. By integrating this predictive model into widely used navigation systems such as Google Maps, we aim to transform these insights into actionable, real-time safety recommendations. This proactive approach will not only inform drivers of imminent risks but also equip local authorities with the data needed to implement targeted interventions.

Importance of Solving the Problem:

Addressing this problem is crucial for several reasons:

- **Enhanced Driver Safety:** By providing drivers with advance warnings about high-risk areas and potentially dangerous conditions, we can significantly decrease the likelihood of accidents.

- **Economic Benefits:** Reducing the frequency of road accidents lowers the economic burden on municipal authorities and healthcare systems due to fewer emergency responses and medical treatments. It also benefits insurance companies by decreasing the number of claims related to road accidents.

- **Saving Lives:** Most importantly, the predictive insights generated by our system have the potential to save lives by preventing accidents before they happen, contributing to safer roadways for everyone.

Project Goals and Objectives

- **Develop an Accurate Predictive Model:** Construct a machine learning model that can accurately analyse risk factors and predict road safety levels, classifying them into categorical risk levels (low, medium, high).

- **Seamless Integration with Google Maps:** Embed the predictive model within Google Maps to deliver real-time, navigational risk assessments directly to the user interface, ensuring that drivers receive timely warnings.

- **Data-Driven Decision Making:** Empower municipal authorities and policymakers with data insights that facilitate better road safety planning and infrastructure management.

Data Description

Our project utilizes two crucial datasets that collect detailed records on driving incidents and environmental conditions, enabling a comprehensive analysis aimed at enhancing road safety.

1. ANWB - Safe Driving Dataset: This dataset comprises detailed information on various driving incidents, making it indispensable for examining patterns, behaviours, and outcomes associated with road safety. It includes data on the location, timing, and characteristics of each incident. Specific measurements such as speeds at different points during an incident and the maximum values reached, whether in terms of speed or g-force, provide insights into the dynamics and severity of the incidents. Additionally, the dataset categorizes incidents into types such as speeding or accident severity, and assesses their severity, which is crucial for identifying risk factors and tailoring prevention strategies. A validity marker in the data ensures that analyses are based on verified and reliable data points.

2. Breda Accident Data (2017-2023): This dataset offers a comprehensive overview of road and weather conditions associated with recorded accidents over six years. It includes parameters such as road type, lighting conditions, weather status at the time of the accident, and more. The longitudinal nature of this dataset allows for trend analysis and seasonal impact assessment on road safety.

Significance of Data Attributes

The combination of detailed incident data and environmental conditions from these datasets enables a multidimensional analysis of road safety. This approach allows us to:

- **Understand temporal and spatial trends**: Analysing when and where incidents occur most frequently helps in designing targeted interventions and improving road safety measures in high-risk areas.
- **Evaluate driver behaviour and incident context**: By examining the conditions under which incidents occur, we can identify key risk factors and develop specific preventive measures tailored to those contexts.
- **Assess the impact of environmental factors**: Integrating weather data with incident records provides insights into how external conditions affect road safety, facilitating the development of guidelines and solutions that are adaptive to environmental changes.

Methodology

Our project employs a structured methodology to develop a predictive model that enhances road safety through Google Maps integration. Each step of our approach is carefully designed to ensure the reliability, accuracy, and applicability of our model in real-world scenarios.

Data Collection and Integration

The foundation of our predictive model is built on comprehensive data sourced from two key datasets:

- **Breda Accidents Data (2017-2023):** Provides extensive details on accidents, including severity, environmental conditions, and road characteristics. This dataset is crucial for modelling the impact of various factors on accident likelihood and severity.
- **ANWB Traffic Incidents Data:** Supplies geographical data points like average longitude and latitude, enriching our model's geographical prediction capabilities.

We merge these datasets to form a unified database that covers a diverse range of factors influencing road safety, thereby facilitating a more accurate prediction of accident areas.

Data Preprocessing

Effective preprocessing is critical to prepare the data for analysis and modelling:

1. **Handling Missing Values:** We implement advanced imputation techniques to address missing data, ensuring no valuable information is lost and maintaining the integrity of our analyses.

2. **Outlier Detection:** Using statistical methods, we identify and rectify anomalies in data that could potentially skew the model's outcomes, ensuring robustness in predictions.

3. **Encoding and Transformation:** Categorical data are encoded using one-hot encoding to better represent discrete features in a format suited for machine learning algorithms, enhancing model accuracy.

4. **Feature Scaling:** We standardize and normalize the data to ensure uniformity in the model's input values, which is crucial for algorithms that are sensitive to the scale of input data.

5. **Balancing the Dataset:** Employing techniques such as SMOTE (Synthetic Minority Over-sampling Technique), we address class imbalance to ensure equitable representation of all risk categories, which enhances the model's ability to generalize.

Model Development

We employ a variety of algorithms to explore different aspects of predictive accuracy and computational efficiency:

- **Neural Networks:** Due to their deep learning capabilities, neural networks are used for their strength in capturing complex patterns in large datasets, which is ideal for the multifaceted nature of road safety data.
- **Decision Trees and Ensemble Methods:** Decision trees provide a transparent model structure that is easy to interpret, while ensemble methods like Gradient Boosting and Random Forests enhance predictive accuracy and model stability through techniques like bagging and boosting.

Model Evaluation and Optimization

- **Performance Metrics:** We assess the model using accuracy, precision, recall, F1-score, and AUC-ROC curve metrics to ensure comprehensive evaluation from various perspectives.
- **Validation Techniques:** Through k-fold cross-validation, we ensure that the model is not only accurate but also robust across different subsets of data, minimizing overfitting.
- **Model Tuning:** Hyperparameter tuning is performed using grid search and random search methods to find the optimal settings for each model, maximizing performance.

Deployment and Legal Compliance

- **Streamlit Deployment:** The final model is deployed through a Streamlit web application, providing a user-friendly interface that allows real-time interaction and displays predictive results.
- **Legal and Ethical Compliance:** Adherence to the GDPR for data protection and the EU AI Act for high-risk AI systems is paramount. We implement rigorous data security measures, continuous system monitoring, and clear communication about how the system works and its limitations to ensure transparency and trust.

Continuous Monitoring and Feedback Integration

Post-deployment, the system is continuously monitored for performance accuracy and system reliability. User feedback is actively requested and integrated into ongoing system enhancements to ensure the model remains effective under evolving real-world conditions and user expectations.

Risk Level and Legal Obligations

Our AI-enabled road safety prediction system, integrated into Google Maps, has been classified as "High Risk" under the EU AI Act due to its profound impact on public safety and significant use of personal and environmental data. This classification necessitates stringent adherence to legal standards, including the GDPR and AI-specific regulations. To manage this, we ensure data accuracy and integrity as mandated by Article 10 of the EU AI Act, maintain rigorous documentation and record-keeping per Article 12, and uphold transparency and information provision in accordance with Article 13. Our compliance strategy includes regular system reviews to verify accuracy, ongoing legal compliance checks, and clear user guidelines to educate users about the system's functionalities and limitations. By focusing on these crucial aspects, we aim to enhance the reliability of our predictions while ensuring the system operates within legal and ethical boundaries, thereby safeguarding user data and trust.

Promoting Transparency in AI System Operations

To ensure users have a clear understanding of the AI system integrated within the Google Maps extension, we prioritize transparent communication through various channels. While we do not provide direct operational mechanics within the app's interface, we offer comprehensive resources and support documentation available on the associated website and user guides. These documents detail how the AI processes data to predict road safety risks, explaining the types of data analysed, including traffic patterns, weather conditions, and historical incident data.

We also describe the function of the algorithms used, such as neural networks and decision trees, in a user-friendly manner on our support pages. Here, users can learn about the role these technologies play in assessing risks and how they contribute to the overall predictions made by the system. By making this information accessible outside the app, we accommodate users who wish to delve deeper into the technology without cluttering the application interface.

Additionally, we maintain a FAQ section that addresses common queries about the AI system’s predictions, aiming to clarify any misunderstandings and set realistic expectations about the technology's capabilities and limitations. For transparency on data privacy, we provide clear information on how user data is collected, used, and protected, adhering to GDPR and other relevant legislation. This includes details on user rights and how to exercise them, ensuring users feel informed and secure.

Regular updates about changes to the system or data policies would be communicated through email newsletters and official social media channels, keeping users informed about developments that may affect their experience or data privacy.

Project Timeline

Week 4: Focus on Data Preprocessing and Initial Modelling

Tasks:

Data Preprocessing:

- **Preparing the Data:**
    - Clean and organize raw datasets.
    - Handle missing values, normalization, and standardization.
    - Address outliers and inconsistencies.
    - Ensure data is ready for analysis and modelling.

Initial Data Modelling:

- Develop baseline predictive/descriptive models using pre-processed data.
- Perform exploratory data analysis (EDA) to understand data patterns and distributions.
- Document initial findings and potential challenges.

Team Contributions:

- Lead data cleaning and normalization efforts.
- Handle standardization and outlier detection.
- Conduct exploratory data analysis (EDA) and document insights.
- Assist with data preprocessing and initial model development.
- Coordinate tasks and ensure timely completion of the preprocessing phase.

Week 5: Focus on Data Modelling

Tasks:

Developing the Machine Learning Model:

- Select appropriate algorithms for predictive modelling (e.g., regression, classification).
- Train models using the pre-processed data.
- Optimize model parameters to improve accuracy.
- Begin evaluating the models based on initial performance metrics.

Team Contributions:

- Develop and optimize predictive models.
- Evaluate model performance and perform hyperparameter tuning.
- Document performance metrics and compare different models.
- Assist with model training and optimization.
- Oversee the modelling process and ensure comprehensive documentation.

Week 6: Focus on Evaluation and Iteration

Tasks:

Evaluating the Model:

- Assess model performance using predefined metrics (e.g., accuracy, precision, recall, F1-score).
- Validate models against a holdout dataset or through cross-validation.
- Document evaluation results and insights for further improvements.

Fine-Tuning the Hyperparameters:

- Perform hyperparameter tuning to enhance model performance.
- Implement iterative improvements based on evaluation feedback.
- Refine and test models to achieve optimal results.

Team Contributions:

- Lead the evaluation process and document findings.
- Perform hyperparameter tuning and validate models.
- Document changes and their impact on performance.
- Support the iteration process with additional data analysis.
- Ensure all iterations are well-documented and communicate progress to the team.

Week 7: Focus on Deployment and Final Presentation

Tasks:

Final Model Deployment:

- Finalize the best-performing model for deployment.
- Prepare deployment scripts and ensure compatibility with production environments.
- Conduct final testing to validate model performance in real-world scenarios.

Final Presentation:

- Compile comprehensive documentation of the model development process.
- Prepare a deployment guide and user manual.
- Create presentation materials summarizing the project, methodology, and results.
- Present the final results and insights to stakeholders.

Team Contributions:

- Finalize deployment scripts and ensure model readiness.
- Conduct final testing and validation.
- Compile and organize documentation.
- Prepare deployment guide and user manual.
- Facilitate the handoff meeting and coordinate with the deployment team.

Reference

World Health Organization. (n.d.). *Road safety*. World Health Organization. https://www.who.int/health-topics/road-safety#tab=tab_1

Article 10: Data and data governance. EU Artificial Intelligence Act. (n.d.-a). https://artificialintelligenceact.eu/article/10/

Article 12: Record-keeping. EU Artificial Intelligence Act. (n.d.-b). https://artificialintelligenceact.eu/article/12/

Article 13: Transparency and provision of information to deployers. EU Artificial Intelligence Act. (n.d.-b). https://artificialintelligenceact.eu/article/13/ The AI act explorer. EU Artificial Intelligence Act. (n.d.-g). https://artificialintelligenceact.eu/ai-act-explorer/