Predicting Severity of Car Accidents in Virginia Using Machine Learning

1. Project Team

- **Team Members**: Nasim Aalemi (Individual Project)
- **Title**: Predicting Severity of Car Accidents in Virginia Using Machine Learning

2. Business Problem or Question

- **Primary Question**:
  What factors contribute to the severity of car accidents in Virginia, and can we predict severity using machine learning?

- **Secondary Questions**:
  1. What are the most common causes of severe accidents in Virginia?
  2. How do environmental and road conditions impact accident severity?
  3. Which machine learning models perform best in predicting accident severity?

3. Data Sources

- **Primary Dataset:**
  https://www.virginiaroads.org/maps/1a96a2f31b4f4d77991471b6cabb38ba/about
  Virginia car accident data from the Virginia Open Data Portal or Virginia Department of Transportation (if needed).

4. Modeling Approach

- **Type of Modeling:**
  Classification (predicting accident severity as a categorical variable, e.g., low, medium, high).

- **Algorithms to Use:**
  1. Logistic Regression (baseline).
  2. Random Forest.
  3. XGBoost.
  4. Support Vector Machines (SVM).

5. Project Approach

- **Methodology:**
  CRISP-DM (Cross-Industry Standard Process for Data Mining).
  1. **Business Understanding:** Define the problem and objectives.
  2. **Data Understanding:** Explore and visualize the dataset.
  3. **Data Preparation:** Clean, preprocess, and engineer features.
  4. **Modeling:** Build and evaluate classification models.
  5. **Evaluation:** Compare model performance and interpret results.
  6. **Deployment:** Summarize findings and provide actionable insights.

6. Significant Assumptions or Constraints

- Assumptions:
    1. The dataset is representative of all car accidents in Virginia.
    2. The features provided (e.g., weather, road conditions) are accurate and complete.
- Constraints:
    1. Limited access to real-time or high-frequency data.
    2. Potential missing or incomplete data entries.
    3. Computational limitations for training complex models.

7. Adaptation of Existing Work

- Existing Work:
  This project is inspired by similar studies on car accident severity prediction, such as analyses of the Maryland Vehicle Crash Dataset.
- Expansion:
    1. Focus on Virginia-specific data to identify regional trends and factors.
    2. Compare multiple machine learning algorithms to determine the best-performing model for this dataset.

8. Requested Deviations from Project Requirements

- None: This project adheres to all stated requirements.

9. Project Timeline

1. Week 1-2: Data collection, cleaning, and preprocessing.
2. Week 3-4: Exploratory Data Analysis (EDA) and feature engineering.
3. Week 5-6: Model building and evaluation.
4. Week 7: Interpretation of results and final outcome preparation.
5. Week 8: Presentation and submission.

10. Expected Outcomes

1. A predictive model that accurately classifies accident severity.
2. Identification of key factors contributing to severe accidents in Virginia.
3. Actionable insights for policymakers and transportation authorities to improve road safety.

11. Conclusion

This project aims to leverage machine learning to predict and analyze the severity of car accidents in Virginia. Through identifying key contributing factors and building an accurate predictive model, the findings can inform targeted interventions to reduce accident severity and improve road safety. The use of the CRISP-DM methodology ensures a structured and comprehensive approach to solving this business problem.