MIDTERM PROPOSAL

Team Name:

FLAM – "Fusion Learning & AI Minds"

(drawn out of first letters of members first names)

Team Members:

Faiza Abdullah

Lufei Yu

Michael Joseph

Muhammadayan Syed

Professor:

Viswanatha Rao

ROAD ACCIDENTS DATASET:

PROPOSAL FOR MACHINE LEARNING COURSE PROJECT

Dataset Description: It is a comprehensive collection of information on road traffic accidents reported in the UK during 2022. It contains 61,352 rows and 20 columns, capturing details about accidents, vehicles, and casualties.

Project Proposal: We propose developing a predictive model to classify the severity of casualties in road accidents (e.g., fatal, serious, or slight) using supervised learning techniques. By leveraging features like casualty age, sex, vehicle type, pedestrian status, and location-based factors, we can train models such as logistic regression, decision trees or random forests to identify key predictors of severity.

The workflow would involve:

- Data preprocessing: Handling missing values, encoding categorical variables and splitting the dataset into training/testing sets (80/20 ratio).

- Feature engineering: Creating derived features, such as age groups or interaction terms between pedestrian movement and vehicle type, to enhance model performance.

- Model training and evaluation: Using cross-validation to tune hyperparameters, and metrics to assess performance, with a focus on imbalanced classes (e.g., fatal cases are rare).

- Interpretation: Employing techniques to explain feature importance, revealing insights such as the role of age or deprivation in severity.

This project aims to demonstrate ML's application in real-world safety analysis, potentially achieving 75-85% accuracy based on similar studies. Outcomes could inform policymakers on targeted interventions, like enhanced pedestrian safety in deprived areas. The analysis will highlight patterns in road safety, such as higher vulnerabilities among certain age groups or vehicle types.