CAPSTONE PROJECT PLAN

1. Project Title

Predicting the severity of road accidents using Machine Learning

2. Student Information

Name: Lucy Tumaini Yaa
Registration Number:
Program: Foundations of Data Science and Artificial Intelligence
Institution: Ngao Labs
Supervisor: Diana Auma
Date: 08 November, 2025

3. Project Overview / Introduction

This project focuses on predicting the severity of road accidents through the application of machine learning techniques. It utilizes historical accident data to identify key factors influencing accident outcomes. By developing a predictive model, the project aims to enhance understanding of accident patterns and provide data-driven insights that can inform policy decisions and improve road safety measures. Through data cleaning, exploratory analysis and machine learning modeling, the project will demonstrate the integration of AI tools in solving complex problems using data-driven approaches.

4. Problem Statement

Road accidents have been a major cause of injuries, deaths and property damage worldwide and understanding their severity is crucial for improving road safety. In many regions, including Kenya, it has been difficult to predict how severe an accident might be based on different contributing factors due to limited data analysis capacity. This project aims to develop a machine learning model that predicts how severe an accident might be (e.g. slight, serious or fatal) based on different contributing factors such as weather conditions, road surface, driver experience and vehicle type. The goal is to identify and understand what factors influence accident outcomes and provide data driven insights and strategies that can help improve road safety.

5. Project Objectives

General Objective:

To understand the severity of accidents using historical accident data. It includes analyzing the contributing factors of road accidents and assessing their contribution to the severity of road accidents.

Specific Objectives:

- To investigate the severity of road accidents per vehicle type.
- To investigate severity of road accidents based on driver experience.
- To develop a machine model to predict the severity of accidents based on various factors such as road surface, weather conditions, and time of the day.

6. Methodology

The project will follow the data science lifecycle, which includes:

i. Data Collection – Obtain data from reliable sources such as Kaggle.
ii. Data Cleaning and Preparation – Handle missing values, remove duplicates, and correct inconsistencies. Relevant variables will be selected and converted into suitable formats for analysis.
iii. Exploratory Data Analysis (EDA) – Use descriptive statistics and visualizations to identify trends, correlations and patterns related to accident severity.
iv. Feature Engineering – Create or transform relevant variables that improve model accuracy, performance and interpretability.
v. Model Building – Apply AI algorithms such as Logistic Regression, Random Forest and XGBoost using Python to predict the severity of road accidents.
vi. Model Evaluation – Assess performance using metrics such as Accuracy, Recall, Precision and F1 score.
vii. Interpretation and Reporting – Summarize insights and provide recommendations for improving road safety.

Tools and Technologies:

Programming Language: Python

Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn.

Software: Jupyter Notebook

Version Control: GitHub

7. Project Timeline

Week Activity

Week 1 Topic selection and background research

Week 2 Dataset sourcing, data cleaning and preparation

Week 3 Exploratory data analysis (EDA), Feature engineering, Model building and Interpretation and reporting.

8. Expected Deliverables

A cleaned and well-documented dataset.

Python notebooks for analysis and modeling.

A trained and tested predictive model.

A written project report and presentation slides.

Recommendations based on findings.

9. Expected Outcomes

Demonstration of how data science and AI can be applied to solve real-world problems.

A predictive model capable of accurately estimating the severity of road accidents.

A comprehensive report and visualizations that summarizes the analysis and model performance.

10. Ethical Considerations

All data used will be obtained from open and ethical sources.

Personal data, if any, will be anonymized to protect privacy.

The project will adhere to institutional research and AI ethics guidelines

11. References

Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly Media.

Khosravi, A., et al. (2019). Machine learning applications in road accident analysis: A review. Accident Analysis & Prevention, 123, 61–74.

Sharma, B., Katiyar, V. K., & Kumar, K. (2016). Traffic accident prediction model using support vector machines with Gaussian kernel. In Proceedings of Fifth International Conference on Soft Computing for Problem Solving: SocProS 2015, Volume 2. Springer Singapore.

UK Road Safety: Traffic Accidents and Vehicles

World Health Organization. (2020). Global status report on road safety 2018. Geneva: WHO.