Traffic Accident Prediction System

1. Introduction

Traffic accidents pose a significant challenge to urban safety and traffic management. This project aims to predict accident-prone areas (hotspots) based on historical accident data, environmental conditions, and traffic patterns. By leveraging machine learning techniques, particularly K-Means clustering, this system helps in identifying risk zones and enhancing road safety measures.

2. Objectives

- Identify accident-prone locations based on traffic and environmental factors.
- Cluster locations into risk categories using K-Means clustering.
- Provide an easy-to-use web interface for real-time prediction.

3. Technologies Used

- **Programming Language:** Python
- **Web Framework:** Django
- **Machine Learning:** Scikit-learn
- **Data Handling:** Pandas
- **Frontend:** HTML, CSS, JavaScript
- **Database:** MySQL / PostgreSQL

4. Database Design

The database follows a normalized relational model to store user data, accident records, clustering results, and predictions efficiently.

4.1 Entity-Relationship Model

- **Users**: Stores user authentication and access details.
- **Accidents**: Contains accident-related data, including location, traffic conditions, and severity.
- **Weather Conditions**: Reference table for predefined weather types.
- **Lighting Conditions**: Reference table for predefined lighting conditions.
- **Clusters**: Stores clustering results of accident-prone locations.
- **Model Metadata**: Stores trained model information.
- **Predictions**: Stores user-generated accident hotspot predictions.

4.2 Database Schema Overview

Table	Description
Users	Stores user login and role information
Accidents	Records accident location, traffic volume, and severity
Weather Conditions	Stores predefined weather conditions
Lighting Conditions	Stores predefined lighting conditions
Clusters	Stores risk zone clusters
Model Metadata	Stores trained ML model information
Predictions	Stores user input and predictions

5. Dataset

The dataset includes historical accident records, traffic density, weather conditions, and visibility data. The key features used for prediction include:

- Latitude & Longitude
- Traffic Volume
- Speed Limit
- Visibility
- Weather Condition
- Lighting Condition

6. Methodology

a. Data Preprocessing

- Encoding categorical variables (`Weather Condition`, `Lighting Condition`) using Label Encoding.
- Scaling numerical features (`Latitude`, `Longitude`, `Traffic Volume`, `Speed Limit`, `Visibility`) using a StandardScaler.

b. Model Development

- **Clustering Algorithm:** K-Means
- The trained model groups locations into accident risk clusters.

c. Prediction Workflow

1. User inputs accident-related parameters via a web form.
2. Categorical variables are encoded using pre-trained label encoders.
3. Numerical variables are scaled using a pre-trained StandardScaler.

4. The processed data is fed into the K-Means model for cluster prediction.
5. The output cluster, along with latitude and longitude, is displayed on a map.