بِسْمِ اللهِ الرَّحْمٰنِ الرَّحِيْمِ

আন্তর্জাতিক ইসলামী বিশ্ববিদ্যালয় চট্টগ্রাম

International Islamic University Chittagong

Department of Computer Science and Engineering

Project Report

Course Code: CSE-3636	**Course Title:** Artificial Intelligence

Project Name: Accidental Injury Prediction

Submitted By	
Name : Bibi Ayesha AKhter Omera	**ID No.** : C223204
Name : Zannatun Naim	**ID No.** : C223207
Name :Fahmida Yasmin	**ID No.** : C223210
Section : 6AF	**Semester:** 6th

Submitted To
Name : Sara Karim
Designation : Adjunct Lecturer, Dept of CSE,IIUC.

Marks :	**Signature:**

1. Introduction

Accidents are a major global issue, causing significant harm to individuals, families, and communities, as well as leading to substantial financial losses. These events are influenced by various factors, such as road conditions, driver behavior, weather, and vehicle types. The ability to predict the severity of accidents before they happen can drastically reduce the impact of these incidents. This project aims to develop a machine learning-based system that can predict accident severity—whether it will be minor, medium, or major—by analyzing historical accident data. By identifying patterns in factors like the driver's age, experience, and the conditions under which the accident occurred, the system can provide valuable insights to improve road safety.

The project leverages a dataset from Kaggle, containing detailed information about past accidents, including weather conditions, road types, vehicle movements, and accident causes. The machine learning models used in this project, including Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), and Naive Bayes, analyze this data to predict accident severity. By processing these features, the system can offer real-time predictions and potentially reduce the risk of severe accidents. The ultimate goal of the project is to integrate this predictive tool into traffic management systems, aiding in early intervention and promoting safer driving practices on the roads.

2. Language/Model Used & System/Environment

For this project, we used Python programming language, leveraging various machine learning libraries like scikit-learn for model building. The following models were used for classification tasks:

- Logistic Regression
- K-Nearest Neighbors (KNN)
- Support Vector Classifier (SVC)
- Naive Bayes
- AdaBoost

These models were implemented in a Jupyter notebook environment with Python 3.7+, and we used Kaggle datasets for training and testing the models.

3. Define Feature/Function

The system aims to predict the severity of accidents using a set of features/inputs collected from real-world accident data.

Description of Features/Inputs: Each input has several options

1. **Age**: several age's input are
   - 18-30
   - 30-50
   - over 51
   - under 18
   - unknown
2. **Gender**:
   - Male
   - Female
3. **Educational Level**:
   - Above high school
   - Junior high school
   - Unknown
   -Elementary school
   - High school
   - Illiterate
   - Writing & reading
4. **Vehicle Driver Relation**:
   - Employee
   -Unknown
   - Owner
   - Other
5. **Driving Experience**:
   - 1-2 years
   - Above 10 yr
   - 5-10 yr
   - 2 yr
   - Unknown
   - No license
   - Below 1 yr
6. **Lanes or Medians**:
   - Divided Lanes
   - Lanes Curb Medians:
7. **Types of Junction**:
   - No Junction
   - Y shape
   - Crossing
   - O shape
   - Other
   - T shape
   -X shape

8. **Road Surface Type**:
   - Asphalt roads
   - Gravel roads
   - Wet roads (due to rain)
   - Snowy roads
9. **Light Conditions**:
   - Daylight
   - Night (with street lighting)
   - Night (no street lighting)
   - Dusk/Dawn (low visibility)
10. **Weather Conditions**:
    - Clear weather
    - Rain
    - Fog
    - Snow
11. **Type of Collision**:
    - Collision with roadside-parked vehicles
    - Rear-end collision
    - Head-on collision
    - Side-impact collision

12. **Vehicle Movement**:
    - Going straight
    - Turning left/right
    - Reversing
    - U-turn
13. **Pedestrian Movement**:
    - Crossing the street
    - Walking on the sidewalk
    - Not a pedestrian
    - Running across the street

14. **Cause of Accident**:

- Moving backward
  - Speeding
  - Drunk driving
  - Distracted driving (e.g., texting while driving)

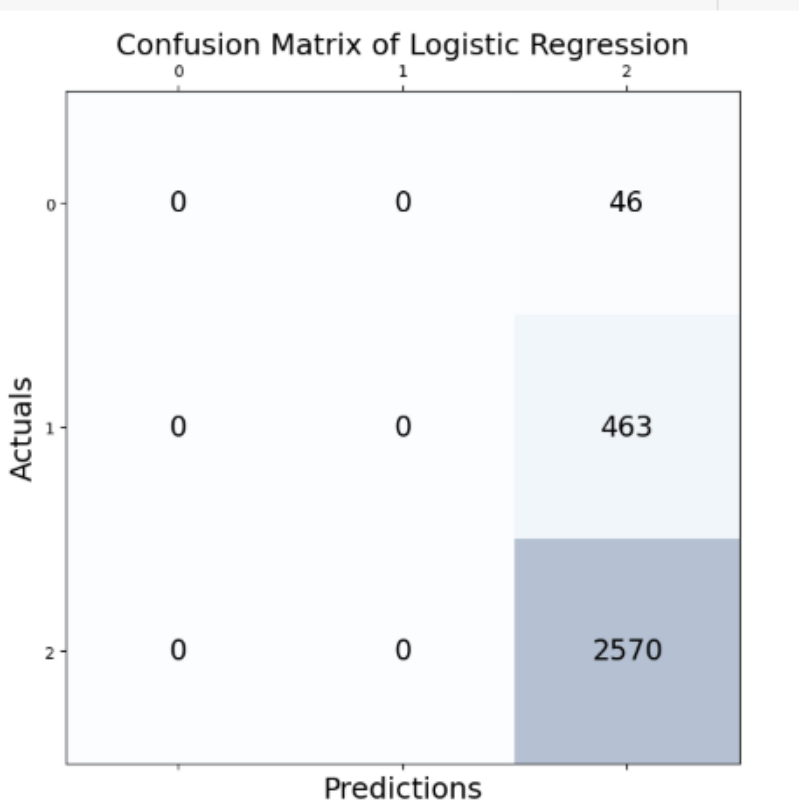
4. Methodology

The system works by processing historical accident data, identifying key patterns, and predicting the severity of future accidents. It uses supervised learning, where the model is trained on labeled data and tested on unseen data to ensure accuracy.

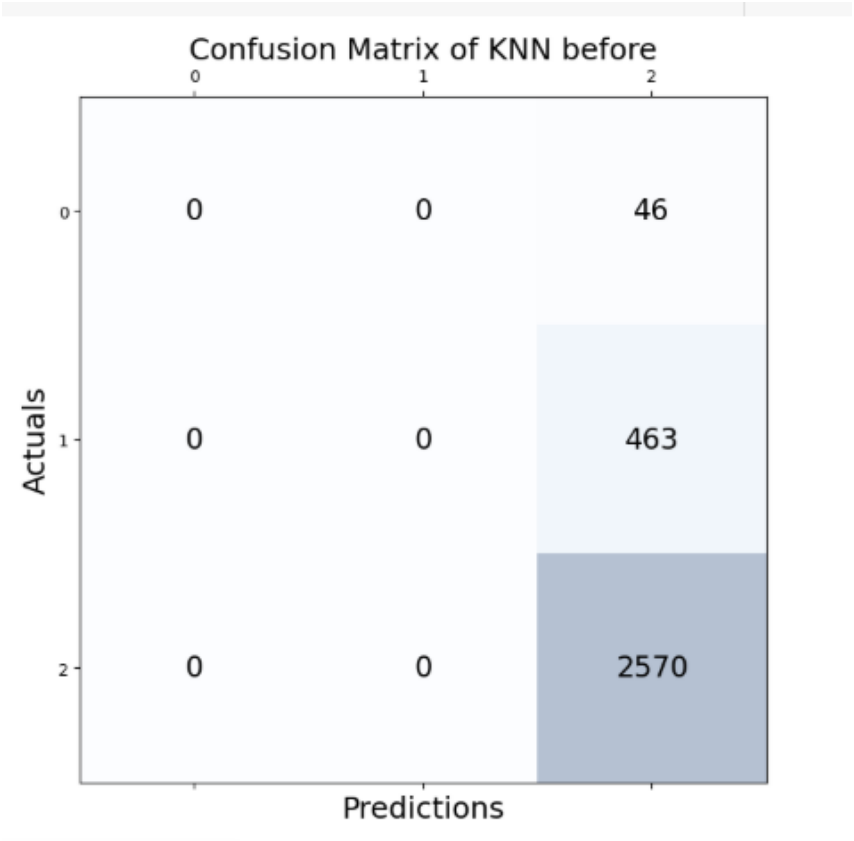
Process Flow:

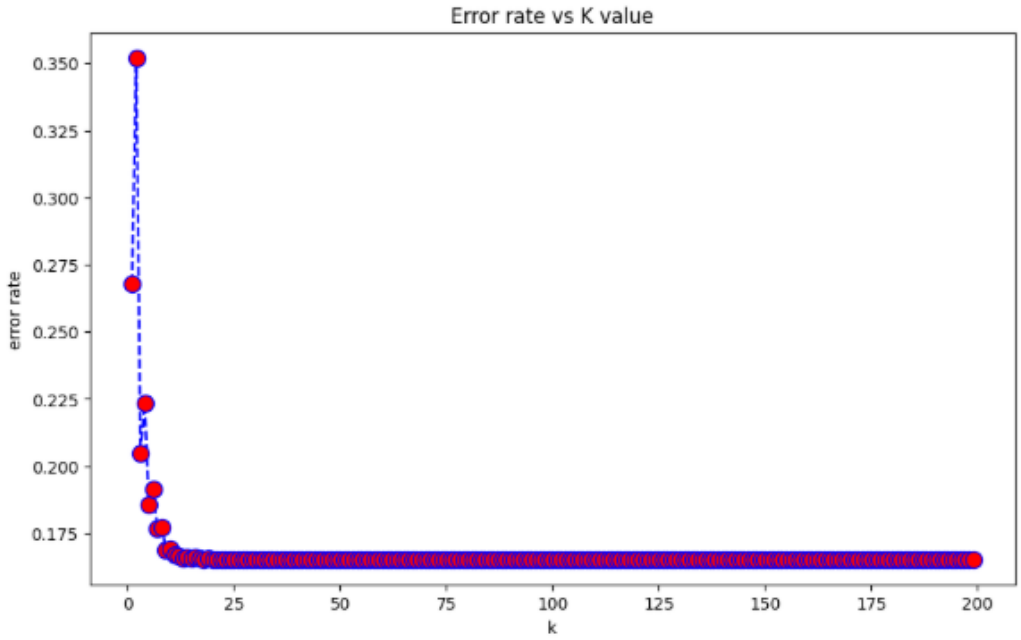
1. Data Collection: We used a dataset from Kaggle that includes various attributes related to past accidents.
2. Data Preprocessing: Data was cleaned, missing values were handled, and the features were normalized.
3. Model Training: We used the following models for training: Logistic Regression, KNN, SVC, Naive Bayes, and AdaBoost.
4. Evaluation: The models were evaluated based on their accuracy, precision, recall, and F1-score.

Confusion Matrix:
Logistic regression:

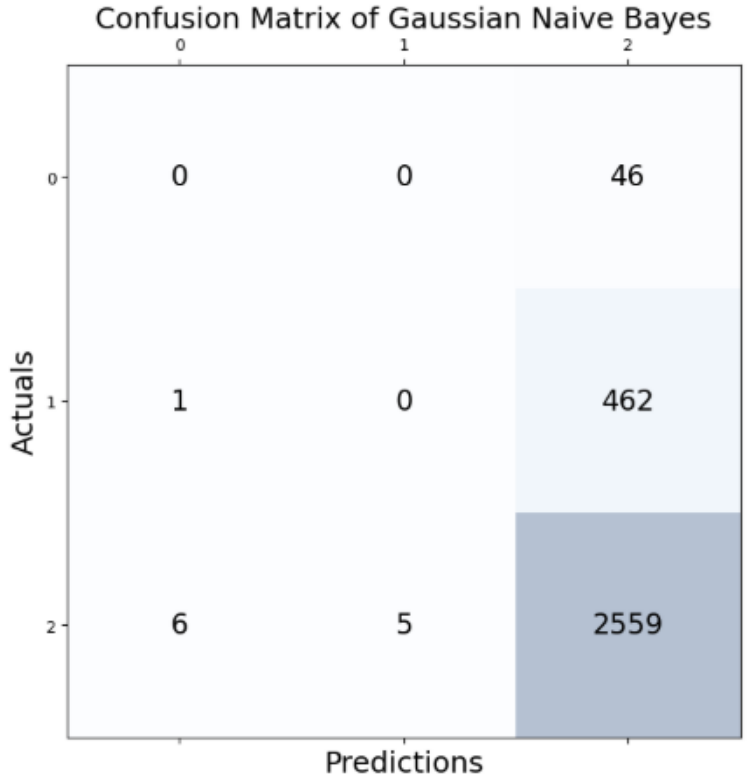


KNN:

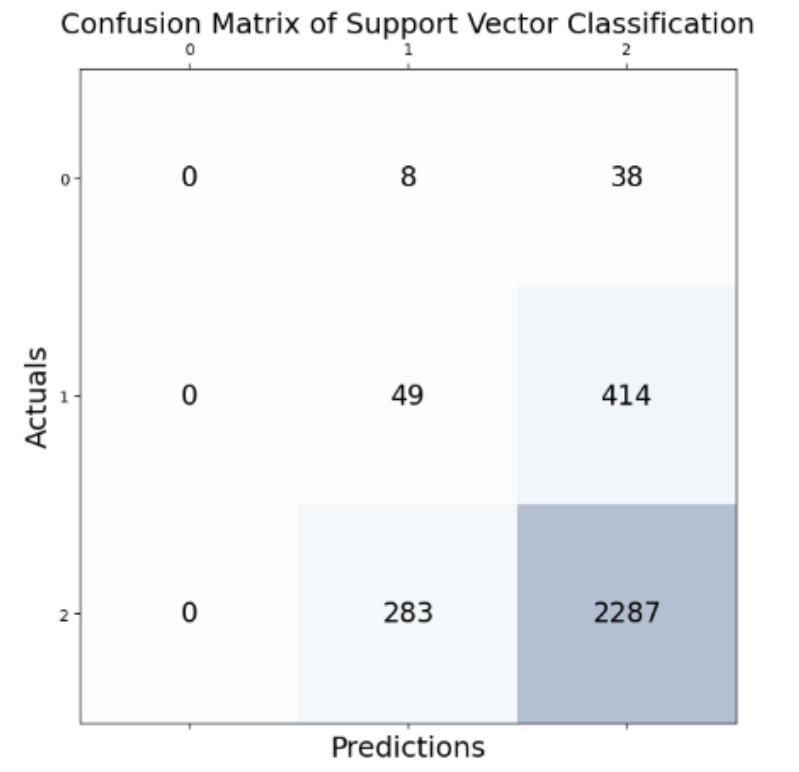

Confusion Matrix of KNN before
0
1
2
0
0
46
0
0
463
0
0
2570
Actuals
Predictions


Error rate vs K value
0.350
0.325
0.300
0.275
0.250
0.225
0.200
0.175
error rate
0
25
50
75
100
125
150
175
200
k

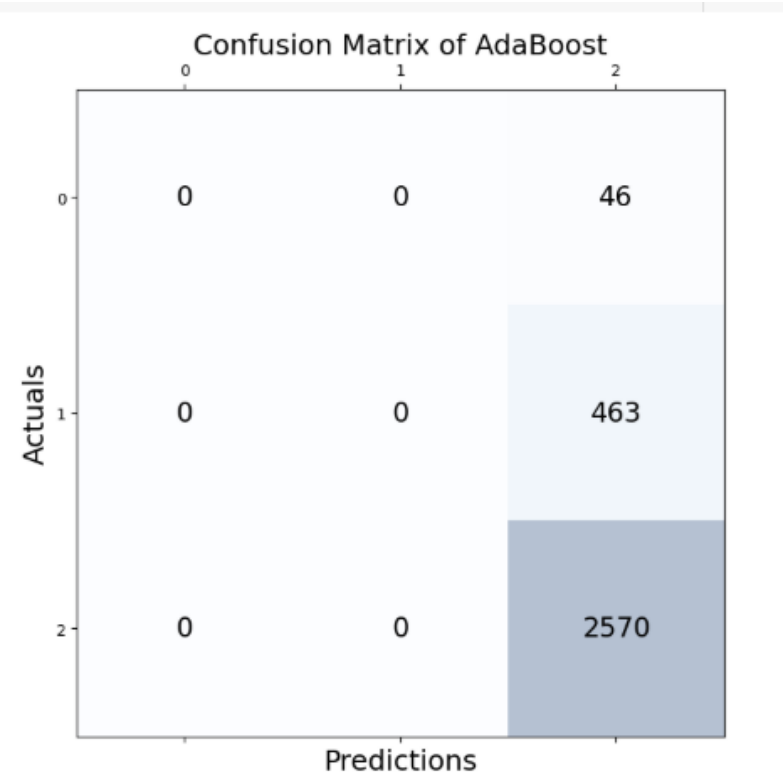
Gaussian Naïve Bayes:



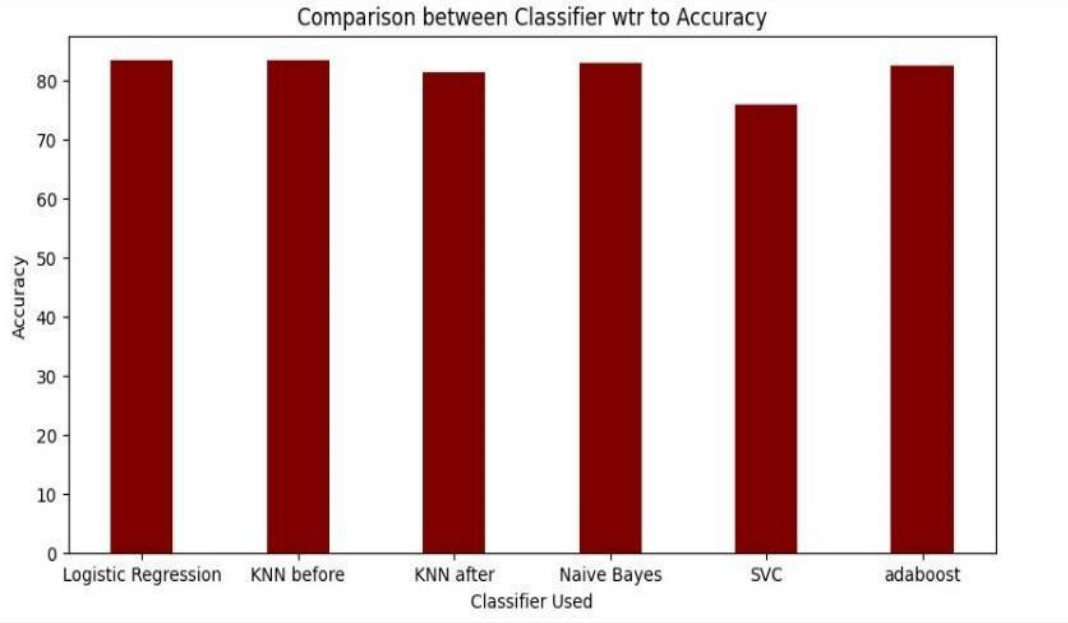
Support Vector Classification:



AdaBoost::



The image presents a bar chart comparing the accuracy of different classifiers used in a machine learning model.



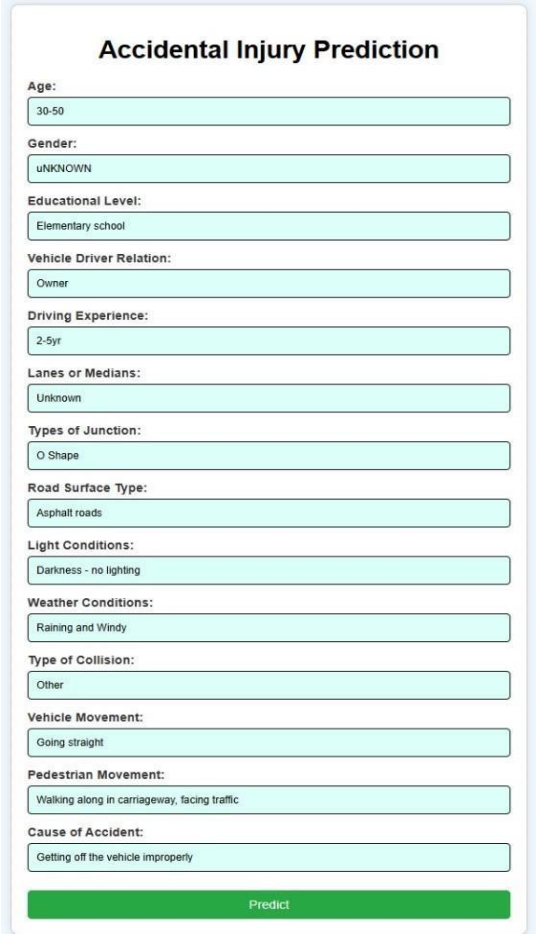
The image shows a bar chart comparing the accuracy of six machine learning classifiers: Logistic Regression, KNN (before and after tuning), Naive Bayes, SVC, and AdaBoost. Most classifiers achieve around 80% accuracy, with KNN showing similar performance before and after tuning.

Model Setup:
The models were implemented using scikit-learn, and each model was tuned using cross-validation to optimize hyperparameters. A dataset split of 70% training, 15% validation, and 15% testing was used for model training and evaluation.

Tool Functions:
The scikit-learn library provided various tools to implement machine learning algorithms like Logistic Regression, KNN, and SVC. Tools like GridSearchCV were used to find the best hyperparameters, and metrics like confusion matrix and classification report were used for evaluation.

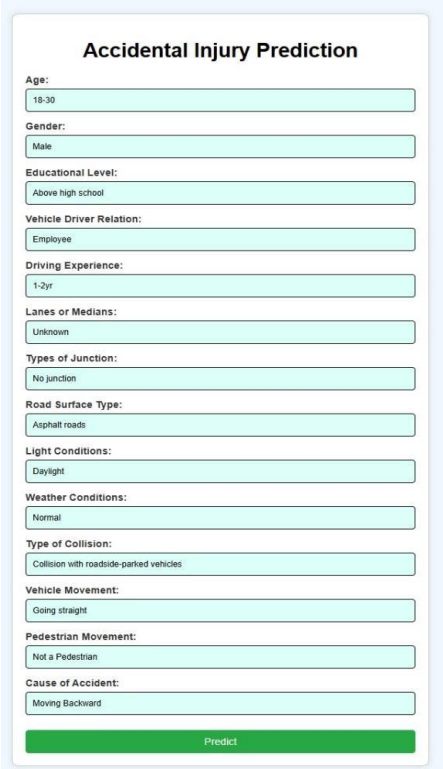
Result:

In this project,
Input: We aim to predict the severity of accidents using machine learning techniques by input of several data in form.
Output: Result shows prediction of accident severity as Maximum, Medium, or Minor. This is given bellow.



Picture ; The form collects key details such as the driver’s age, gender, educational level, driving experience, road surface type, weather conditions, vehicle movement, and cause of the accident. These factors are crucial for the machine learning model to predict accident severity accurately.

Picture ; The system predicts the accident’s severity, such as "Minor Injury," and provides a safety message advising the driver to drive slowly and think about their family’s safety. This helps raise awareness and encourages safer driving practices.



Picture ; Collects data

Picture ; Predicts the accident’s severity, such as "Medium Injury,"

1. Conclusion

In conclusion, this project demonstrates the power of machine learning in predicting accident severity. By analyzing data such as driver characteristics, road conditions, and weather, we can create a system that anticipates accidents and helps to improve road safety. This model serves as an effective tool to forecast potential accidents and take preventive actions.