CSE523 - Machine Learning

Project 11: Identify abnormal driving behavior using spatio-temporal analysis

Weekly Report-8

Group: Titans

**Team Members**

Enrollment Number	Name
AU2140043	Bhavya Khakhar
AU2140154	Harsh Loriya
AU2140170	Krishna Patel
AU2140187	Priyam Shah

Working on the Revised Approach

In our previous weekly report we explained how we shifted from our initial approach to the new approach and the challenges and problems faced by us to implement it . So we are working on the new approach and currently doing classification by trying out different models to improve the accuracy. We mainly tried three models for classification, one is Logistic Regression with the accuracy of 58%, the other one is Random Forest with the accuracy of 71% and the last one is KNN with the accuracy of 65%.

At first in the training, we are storing the data into an array and corresponding labels of the files and are fed in the model as a whole array for training. For testing the accuracy of the prediction, we used one file at a time and predicted a label. Then calculate the accuracy of the model.

Confusion matrix:
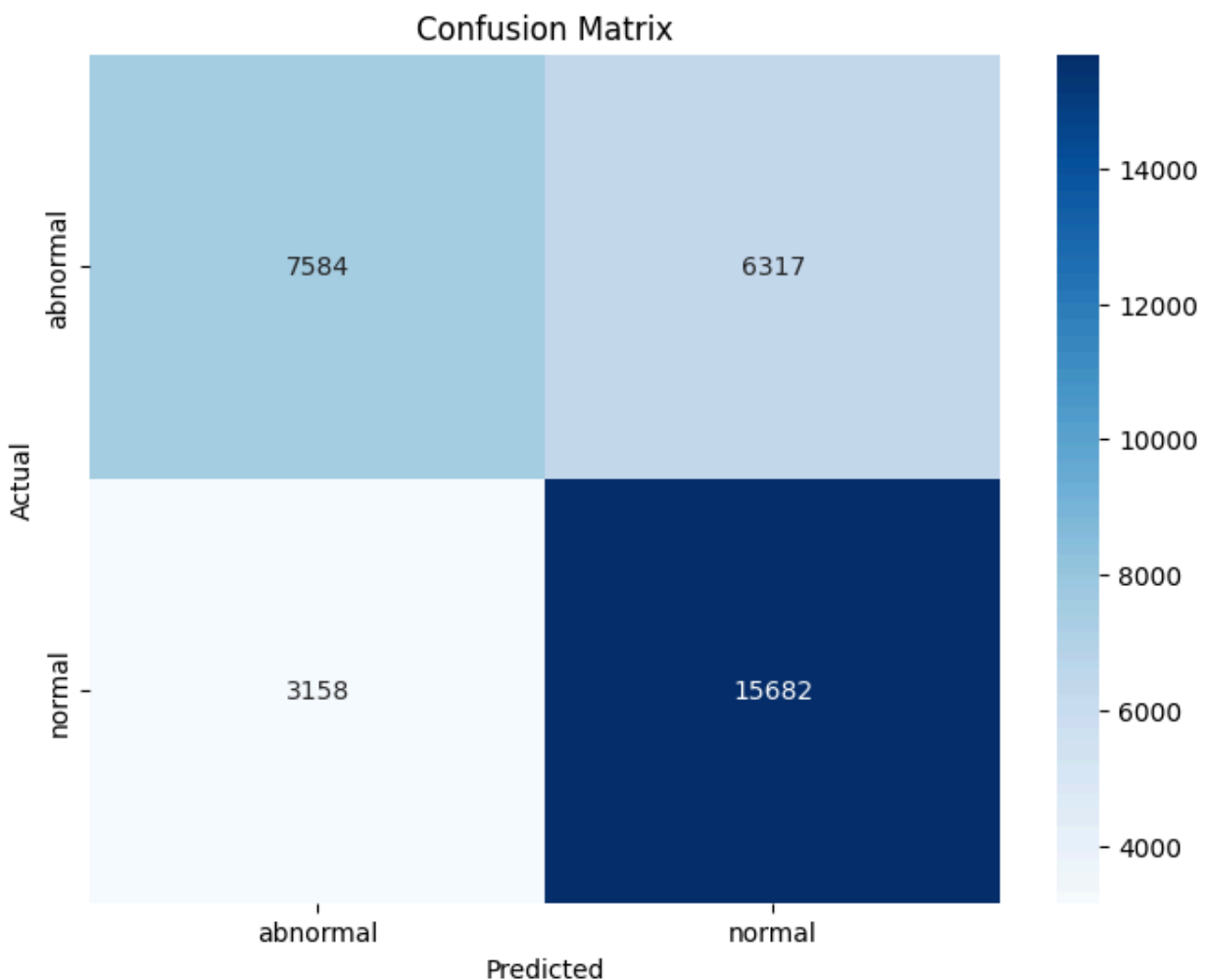
Justification of Random forest:

- Random Forest is robust to irrelevant features and noise in data, effectively handling complexity by selecting informative features and ignoring noisy ones for better generalization and higher accuracy.

- It effectively captures non-linear patterns in spatio-temporal data through its ensemble nature and decision trees, partitioning feature space and making predictions based on the average or majority vote of multiple trees.

- Random Forest is an ensemble learning method that combines multiple decision trees to make predictions, reducing overfitting and improving generalization performance, achieving higher accuracy than individual decision trees or simpler models.

- Also it effectively handles class imbalance in spatio-temporal datasets by averaging multiple trees' predictions, reducing bias towards the majority class and improving predictive performance on the minority class.

- Although Random Forest is a highly flexible model, it is less prone to overfitting compared to other complex models like K-Nearest Neighbors (KNN), especially when the number of trees in the ensemble is adequately chosen.

References:

- GfG. (2024b, February 22). *Random Forest algorithm in machine learning*. GeeksforGeeks. https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/

- *What is Random Forest? | IBM*. (n.d.). https://www.ibm.com/topics/random-forest