Vehicle Classification using GPS Data

By Swadesh Vaibhav 2017A7PS0030P

Vipin Baswan 2017A7PS0429P

Saumil Agarwal 2018A7PS0268P

Thanks to-

Kamlesh Tiwari Sir

Priya Ma’am

MapMyIndia Personnel

The Objective of the Project

To classify vehicles into different classes based on their GPS footprints in time, at specific intervals.

Methodology Used

Machine Learning and Deep Learning Techniques

- Shallow Models like Decision Tree.
- Ensemble Models like Baggage Classifier.
- A Deep Learning Model, ANN.

Dataset

1. The dataset was provided by MapMyIndia. It consisted of the GPS histories of vehicles in 3168 files.

2. Each file had the following information about various vehicles:

    a. Vehicle ID

    b. Latitude and Longitude

    c. Time Stamps (Date and time format)

    d. Speed of vehicle

    e. Engine status and

    f. Vehicle label.

Preprocessing

Dividing the dataset into trainable format

1. All the files were processed and divided into files corresponding to each vehicle.

2. The vehicles were found to be distributed in the following classes-

    1. Car : 2549 (78%)
    2. Ambulance : 588 (18%)
    3. Motor Cycle : 18 (0.5%)
    4. Bus : 16 (0.49%)
    5. JCB : 1 (0.03%)
    6. Truck : 58 (1.7%)
    7. Tractor : 15 (0.46%)
    8. Mini Bus : 1 (0.03%)
    9. School Bus : 1 (0.03%)
    10. Not Available : 34 (1.04%, Removed from the training dataset)

Extraction of features

-

The Haversine Formula

Extraction of Features

1. The geo-locations of vehicles at various timestamps were used to calculate the distance travelled by each of them using the Haversine Formula.

2. Thereafter, we collected 6 features from the dataset:

    1. Average of the speed of the vehicle

    2. Average of the change in the position of the vehicle.

    3. Variance of the speed of the vehicle

    4. Variance of the change in the position of the vehicle.

    5. Average acceleration of the vehicle

    6. Variance of acceleration of the vehicle

PCA - Principal Component Analysis

To know about the contribution of each factor in the variance of the dataset, we did Principal Component Analysis (PCA). The results are as following:

1. 0.611 explained variance ratio using only 1st feature
2. 0.907 explained variance ratio using first two features
3. 0.960 explained variance ratio using first three features
4. 0.9998 explained variance ratio using first four features
5. 0.9999 explained variance ratio using first five features
6. 1.0 explained variance ratio using all features.

PCA

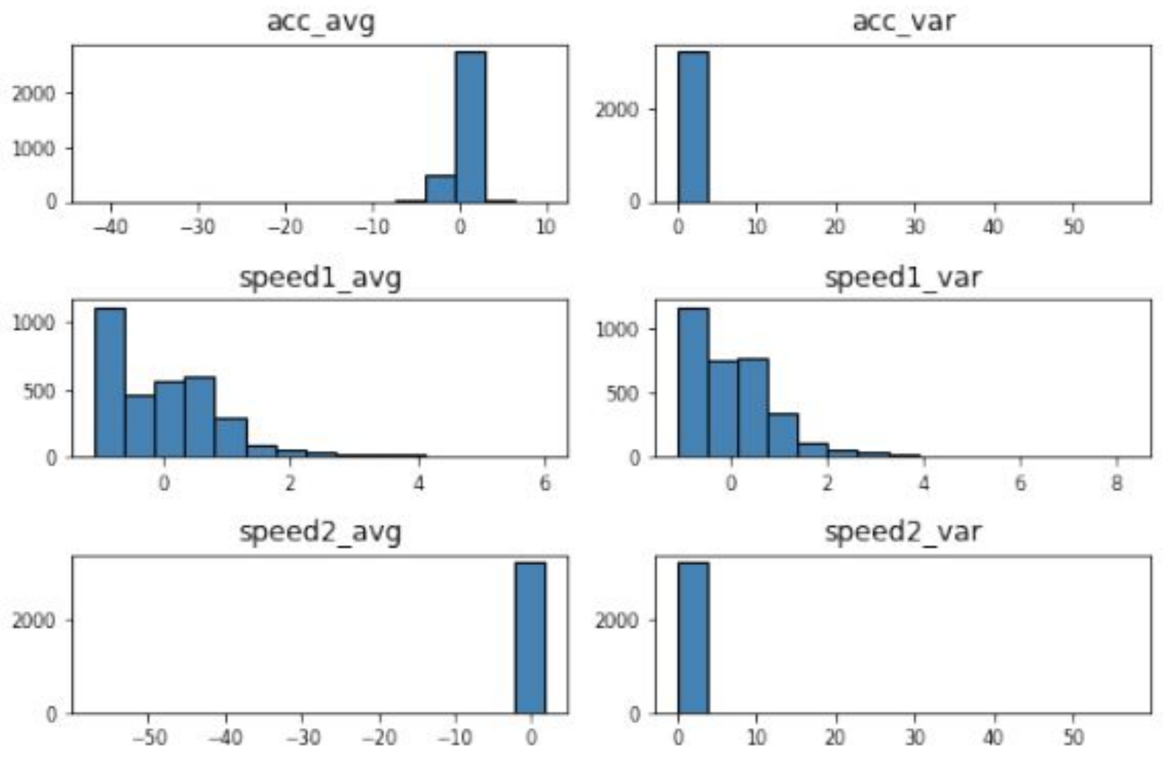
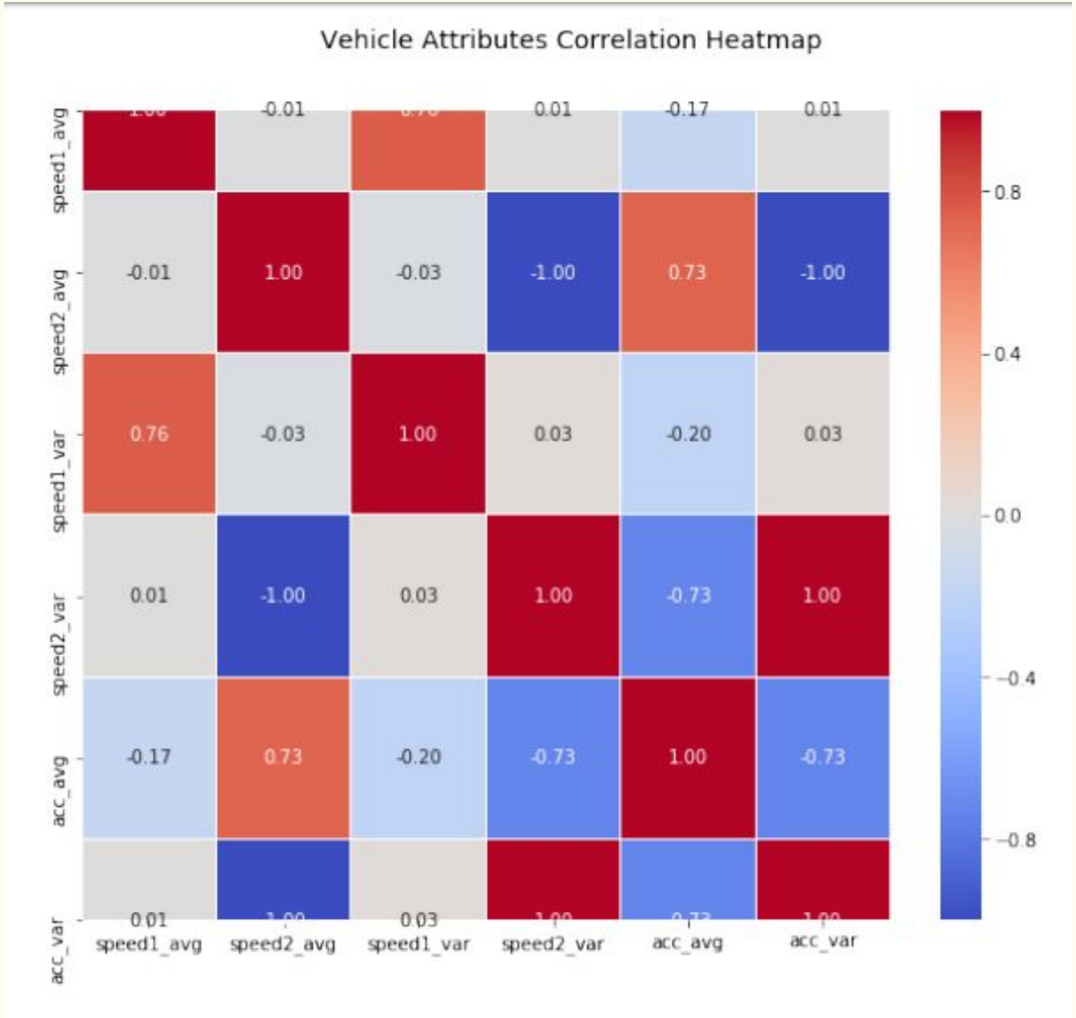


Figure 1: Histogram for each feature


Vehicle Attributes Correlation Heatmap
speed1_avg
speed2_avg
speed1_var
speed2_var
acc_avg
acc_var
-0.01
0.01
-0.17
0.01
1.00
-0.03
-1.00
0.73
-1.00
0.76
0.03
-0.20
-0.73
-0.20
0.8
0.4
0.0
−0.4
−0.8

Training Classical ML Models

Training Paradigms used-

- For training, we used classical Machine Learning models. We got best performance (in terms of accuracy) from **Random Forest Classifier** and **eXtreme Gradient Boost (XGB) Classifier**, with accuracy of **92.44%** and **92.31%** respectively.
- The training and testing process used was 5-Fold Cross Validation.

Results of using various models on the dataset-

MODEL	Fold1	Fold2	Fold3	Fold4	Fold5	Avg
XGB Classifier	93.86	91.86	92.31	91.08	92.46	92.31
Random Forest	93.86	91.86	92.31	91.38	92.77	92.44
Bagging Classifier	92.93	90.17	91.23	91.08	92.31	91.54
Decision Tree Classifier	92.93	90.78	91.69	90.15	92.0	91.51
Gradient Boosting Classifier	92.63	90.63	91.85	89.69	91.38	91.24
Ada Boost Classifier	88.63	85.25	79.08	87.54	88.77	85.85
SVC	84.64	84.33	82.46	81.08	82.15	82.93
KNN	83.41	82.03	81.38	79.38	82.62	81.76
Bernaulli Naive Bayes	79.57	80.49	77.23	76.15	79.38	78.57
SGD Classifier	79.26	80.03	77.23	75.54	79.23	78.26

Table 1: Various Models used along with their fold and average accuracy (using 5-fold cross validation)

Visualising accuracy using Confusion Matrix

Class	Ambulance	Bus	Car	Motor Cycle	Tractor	Truck
Ambulance	107	0	16	0	0	0
Bus	1	0	4	0	0	0
Car	4	0	500	0	0	0
Motor Cycle	2	0	0	1	0	0
Tractor	3	0	0	0	0	0
Truck	8	0	5	0	0	0

Table 2: Confusion Matrix for XGB Classifier

Training the Deep Learning Model

Training Paradigm-

- We trained our model using Dense Neural Networks given by the Keras framework.
- It gave a test accuracy of 89.2%.
- The model architecture consisted of 6 fully connected Neural layers, with 8, 12, 16, 20, 15 and 10 neural nodes respectively.
- All the layers were activated using the ’relu’ function.
- The final output was obtained using the softmax function.

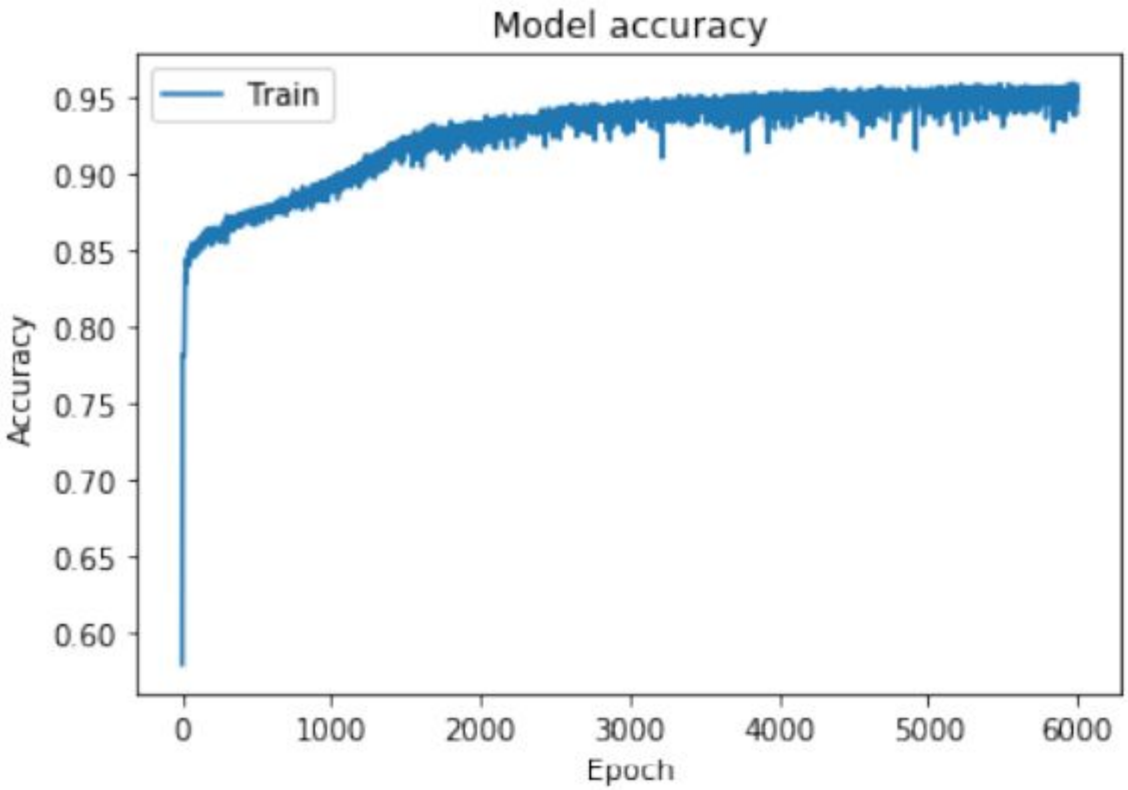


Figure 3: Model Accuracy of Neural Nets as a function of the number of Epochs

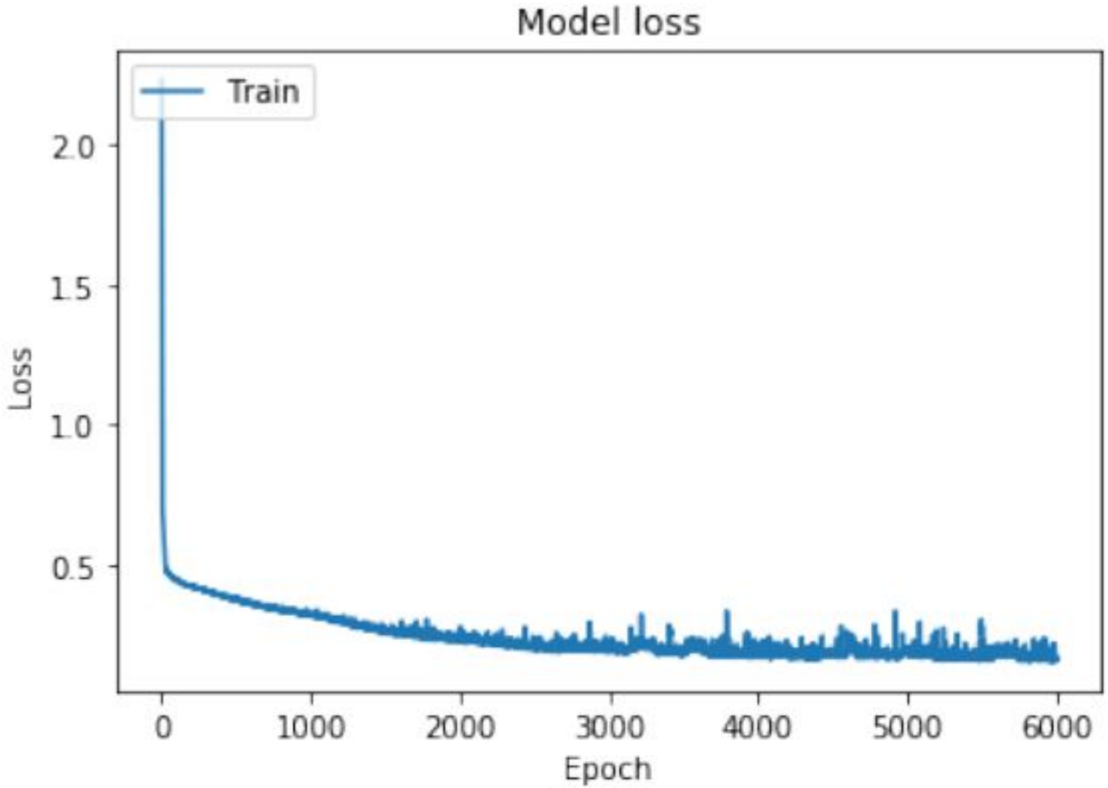


Figure 4: Model Loss of Neural Nets as a function of the number of Epochs

Conclusions

The following conclusions can be drawn from the results of the project-

- GPS Data can be used to predict the type of vehicles with an **accuracy greater than 90%**.
- Features that are crucial for this prediction are: **speed, mean speed, acceleration, variance in acceleration, mean acceleration**.
- **ML ensemble models** are very useful towards this classification.
- According to our inferences, **eXtreme Gradient Boost classifier and Random Forest Classifier** will give the best accuracy

THANK YOU!

- Questions?
- Doubts?
- Queries?
- Follow-up on our work?