# A DDS to determine the best driving strategy by using XGBOOST algorithm

**CHAITANYA PAMPANA***

**Assistant Professor, CSE Department,**

**SRINIVASA INSTITUTE OF ENGINEERING AND TECHNOLOGY, (AUTONOMOUS)**

**NH-216, Cheyyeru(V), Katrenikona(M), Amalapuram, E.G.Dist-533216.**

**ABSTRACT:**

In this paper author is describing concept for driving decision strategy by observing vehicle internal data such as steering and RPM level to predict various classes such as speed (steering), changing lane etc. All existing technique were concentrate on external data such as road condition and pedestrians etc but not concentrate on internal values. So to take efficient determination of steering condition and changing lane author is analysing internal data. All internal data will be collected from sensor and then store on cloud and then application will read data from cloud and then apply machine learning algorithms to determine or predict steering condition or changing lane. To implement this project author has introduce and algorithm called DDS (Driving Decision Strategy) algorithm which is based on genetic algorithm to choose optimal gene values which helps in taking better decision or prediction. DDS algorithm obtained input from sensor and then pass to genetic algorithm to choose optimal value which helps in faster and efficient prediction. In propose DDS algorithm we have used Random Forest with Genetic Algorithm and we are getting prediction accuracy up to 75% and to further enhance accuracy we are upgrading propose DDS with Genetic and XGBOOST algorithm and this extension XGBOOST algorithm giving much better accuracy compare to propose algorithm

**KEYWORDS:** DDS, Machine Learning, Vehicle

## 1] INTRODUCTION:

Currently, global companies are developing technologies for advanced self-driving cars, which is in the 4th stage. Selfdriving cars are being developed based on various ICT technologies, and the principle of operation can be classified into three levels of recognition, judgment and control. The recognition step is to recognize and collect information about surrounding situations by utilizing various sensors in vehicles such as GPS, camera, and radar. The judgment step determines the driving strategy based on the recognized information. Then, this step identifies and analyzes the conditions in which the vehicle is placed, and determines the driving plans appropriate to the driving environment and the objectives. The control step determines the speed, direction, etc. about the driving and the vehicle starts driving on its own. An autonomous driving vehicle performs various actions to arrive at its destination, repeating the steps of recognition, judgment and control on its own

## 2] LITERATURE SURVEY:

### 2.1] Ning Ye, Yingya Zhang *et al*

In Intelligent Transportation Systems (ITS), logistics distribution and mobile e-commerce, the real-time, accurate and reliable vehicle trajectory prediction has significant application value. Vehicle trajectory prediction can not only provide accurate location-based services, but also can monitor and predict traffic situation in advance, and then further recommend the optimal route for users. In this paper, firstly, we mine the double layers of hidden states of vehicle historical trajectories, and then determine the parameters of HMM (hidden Markov model) by historical data. Secondly, we adopt Viterbi algorithm to seek the double layers hidden states sequences corresponding to the just driven trajectory. Finally, we propose a new algorithm (DHMTP) for vehicle trajectory prediction based on the hidden Markov model of double layers hidden states, and predict the nearest neighbor unit of location information of the next k stages.

### 2.2] Li-Jie Zhao, Tian-You Chai *et al*

Real-time and reliable measurements of the effluent quality are essential to improve operating efficiency and reduce energy consumption for the wastewater treatment process. Due to the low accuracy and unstable performance of the traditional effluent quality measurements, we propose a selective ensemble extreme learning machine modeling method to enhance the effluent quality predictions. Extreme learning machine algorithm is inserted into a selective ensemble frame as the component model since it runs much faster and provides better generalization performance than other popular learning algorithms. Ensemble extreme learning machine models overcome variations in different trials of simulations for single model. Selective ensemble based on genetic algorithm is used to further exclude some bad components from all the available ensembles in order to reduce the computation complexity and improve the generalization performance. The proposed method is verified with the data from an industrial wastewater treatment plant, located in Shenyang, China.

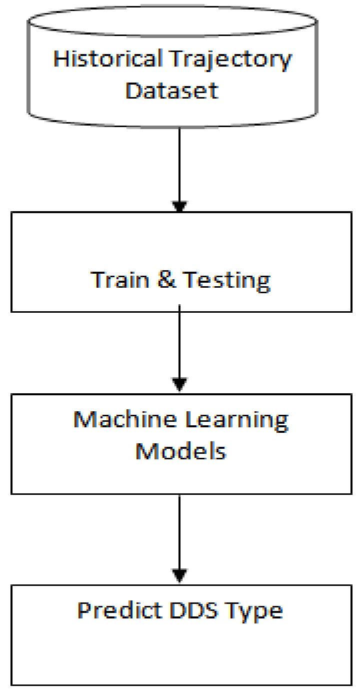
## 3] PROBLEM DEFINTION:

Self-driving cars use in-vehicle computers to compute data collected by sensors. As the amount of the computed data increases, it can affect the speed of judgment and control because of overload. These problems can threaten the stability of the vehicle. To prevent the overload, some studies have developed hardware that can perform deep running operations inside the vehicle, while others use the cloud to compute the vehicle's sensor data.

## 4] PROPOSED APPROACH:

To implement this project author has introduce and algorithm called DDS (Driving Decision Strategy) algorithm which is based on genetic algorithm to choose optimal gene values which helps in taking better decision or prediction. DDS algorithm obtained input from sensor and then pass to genetic algorithm to choose optimal value which helps in faster and efficient prediction.

## 5] SYSTEM ARCHITECTURE:




## 6] PROPOSED METHODOLOGY:

### 6] Data collection

To implement this project we are using historical vehicle trajectory dataset. In dataset if user is slowing down vehicle then it has some sensor value with class label as ‘lane changing’. Similarly based on values we have different classes in dataset. Machine learning algorithm will be trained on such dataset and then when we apply test data on trained model then algorithm will predict class for that test data.

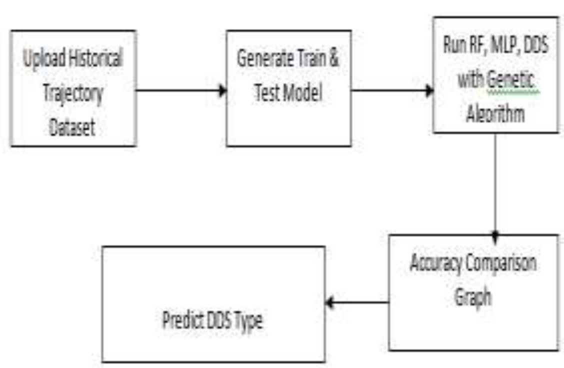
**Upload historical trajectory Dataset**: using this module we will upload dataset to application and then find out total number records.

**Generate train & test model:** This module is read dataset and to split dataset into train and test part to generate machine learning train model

**Run Random Forest:** Using this module we will split dataset into train and test and then build Random Forest trained model. Trained model will be applied on test data to calculate and test prediction accuracy

**Run MLP Algorithm**: Using this module we will split dataset into train and test and then build MLP trained model. Trained model will be applied on test data to calculate and test prediction accuracy

## 7] ALGORITHM:




### Random forest algorithm:

This model has three random concepts, randomly choosing training data when making trees, selecting some subsets of features when splitting nodes and considering only a subset of all features for splitting each node in each simple decision tree. During training data in a random forest, each tree learns from a random sample of the data points.

### MLP:

An MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training.

### Genetic algorithm

The genetic algorithm is a method for solving both constrained and unconstrained optimization problems that is based on natural selection, the process that drives biological evolution. The genetic algorithm repeatedly modifies a population of individual solutions. The genetic algorithm can address problems of mixed integer programming, where some components are restricted to be integer-valued.

### XGBOOST:

XGBoost (Extreme Gradient Boosting) belongs to a family of boosting algorithms and uses the gradient boosting (GBM) framework at its core. It is an optimized distributed gradient boosting library.

## 8] Evaluation Metrics:

F1-Score, Accuracy and Receiver Operating Characteristics-Area Under the Curve (ROC-AUC) metrics are employed to evaluate the performance of our models. For Computing F1-score and Accuracy, Precision and Recall must be evaluated by

- FPR=False Positive Rate
- TPR=True Positive Rate
- Accuracy
- Precision
- Recall
- F1-score

For this, the calculation of values is measured based on:

• True positive (TP) = No. of events, correctly determined.

• False negative (FN) = No. of events, inaccurately anticipated and not required.

• False-positive (FP) = No. of events, incorrectly predicted.

• True negative (TN) = No. of events, correctly anticipated and not required.

**False Positive Rate (FPR):** Itis a metric that can be used to assess machine learning accuracy. It is defined as:

**FPR=FP/(FP+TN)**

**True Positive Rate (TPR):** Itis a synonym for recall and is therefore defined as

**TPR=FP/(FP+TN)**

**Accuracy:** It is the most important performance measure and it is easily done by a ratio of correctly predicted observations to the total observations.

**Accuracy=(TN+TP)/(TP+FP+TN+FN)**

**Recall:** It is the ratio which correctly predicts positive observations among all observations in original data.

**Recall= TP/(TP+FN)**

**Precision:** It is used to calculate the correctly identified values. This means to calculate the total number of software's which are correctly predicted as positive from the total number of software's predicted positive. It is defined as

**Precision = TP/ (TP + FP)**

**F1-score:** The F-score is a way of combining the precision and recall of the model, and it is defined as the mean of the model's precision and recall. It is also called as F-score. It is defined as

**F1 Score = 2(Precision**
**Recall/Precision + Recall)**

ROC-AUC is another powerful metric for classification problems, and is calculated based on the area under ROC-AUC curve from prediction scores.

## 9] RESULTS:

Random Forest Prediction Results
Random Forest Precision : 67.59627129540341
Random Forest Recall : 67.61446886446886
Random Forest FMeasure : 66.64871828546056
Random Forest Accuracy : 67.3469387755102

Multilayer Perceptron Classifier (MLP) Prediction Results
Multilayer Perceptron Precision : 48.202468030054234
Multilayer Perceptron Recall : 45.21672771672772
Multilayer Perceptron FMeasure : 42.57765830346475
Multilayer Perceptron Accuracy : 48.97959183673469

DDS Prediction Results
DDS Precision : 74.93150684931506
DDS Recall : 75.28235653235652
DDS FMeasure : 75.0576134682095
DDS Accuracy : 75.51020408163265

Extension DDS Prediction Results
Extension DDS with XGBoost Precision : 73.76488095238095

Random Forest we got 67% accuracy and with Multilayer perceptron we got accuracy as 48% and with propose DDS we got accuracy as 75 and in below screen we can see accuracy of extension XGBOOST

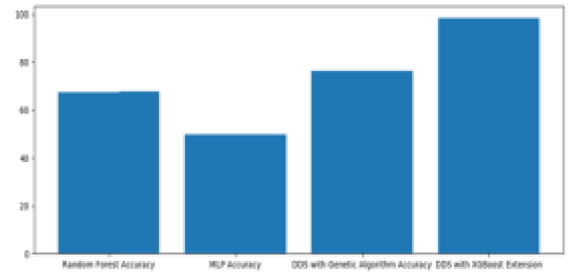
Random Forest Accuracy : 67.3469387755102

Multilayer Perceptron Classifier (MLP) Prediction Results
Multilayer Perceptron Precision : 48.202468030054234
Multilayer Perceptron Recall : 45.21672771672772
Multilayer Perceptron FMeasure : 42.57765830346475
Multilayer Perceptron Accuracy : 48.97959183673469

DDS Prediction Results
DDS Precision : 74.93150684931506
DDS Recall : 75.28235653235652
DDS FMeasure : 75.0576134682095
DDS Accuracy : 75.51020408163265

Extension DDS Prediction Results
Extension DDS with XGBoost Precision : 73.76488095238095
Extension DDS with XGBoost Recall : 75.0
Extension DDS with XGBoost FMeasure : 74.37452326468345
Extension DDS with XGBoost Accuracy : 98.46938775510205

In above screen in blue colour text we can see with extension XGBOOST we got 98% accuracy




In above graph x-axis represents algorithm names and y-axis represents accuracy and in all algorithms extension XGBOOST has got high accuracy.

## 10] CONCLUSION:

This paper proposed a Driving Decision Strategy. It executes the genetic algorithm based on accumulated data to determine the vehicle's optimal driving strategy according to the slope and curvature of the road in which the vehicle is driving and visualizes the driving and consumables conditions of an autonomous vehicle to provide drivers.

## 10] EXTENSION WORK:

We have used Random Forest with Genetic Algorithm and we are getting prediction accuracy up to 75% and to further enhance accuracy we are upgrading propose DDS with Genetic and XGBOOST algorithm and this extension XGBOOST algorithm giving much better accuracy compare to propose algorithm

**11] REFERENCES:**

[1] Y.N. Jeong, S.R.Son, E.H. Jeong and B.K. Lee, "An Integrated SelfDiagnosis System for an Autonomous Vehicle Based on an IoT Gateway and Deep Learning, " Applied Sciences, vol. 8, no. 7, july 2018.

[2] Yukiko Kenmochi, Lilian Buzer, Akihiro Sugimoto, Ikuko Shimizu, "Discrete plane segmentation and estimation from a point cloud using local geometric patterns, " International Journal of Automation and Computing, Vol. 5, No. 3, pp.246-256, 2008.

[3] Ning Ye, Yingya Zhang, Ruchuan Wang, Reza Malekian, "Vehicle trajectory prediction based on Hidden Markov Model, " The KSII Transactions on Internet and Information Systems, Vol. 10, No. 7, 2017.

[4] Li-Jie Zhao, Tian-You Chai, De-Cheng Yuan, "Selective ensemble extreme learning machine modeling of effluent quality in wastewater treatment plants, " International Journal of Automation and Computing, Vol.9, No.6, 2012