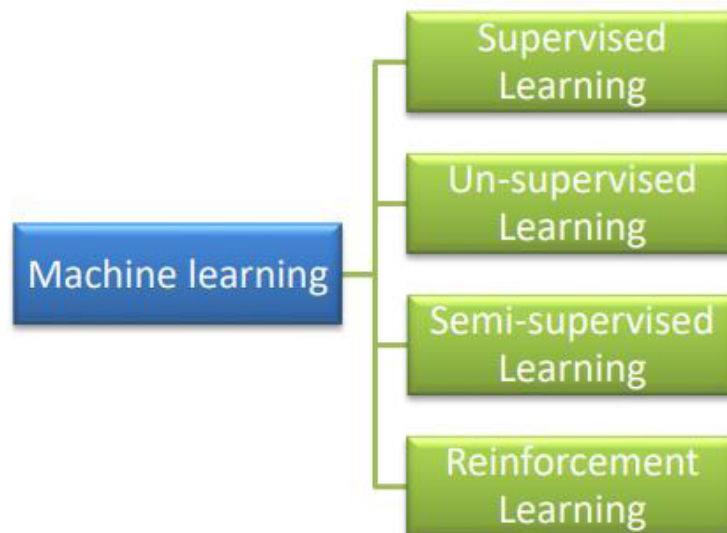


PRIOR KNOWLEDGE

Team ID	PNT2022TMID53349
Project Name	Car Resale value Prediction

1. MACHINE LEARNING

ML is the field of science which deals with programming computers that take information known as data and learn from that data. ML is a subset of Artificial Intelligence, which imitates human intelligence into machines, these machines think in the way of humans and mimic their behavior. Any device's attributes that match the human brain can be applied with the term ML, such as critical thinking, problem-solving, learning, decision-making, etc. In recent years, many sectors like government, health care, E-commerce, marketing, and sales have been using ML. According to the labeled data, methods in ML can be classified into three categories: supervised learning method, semi-supervised learning method and unsupervised learning method.



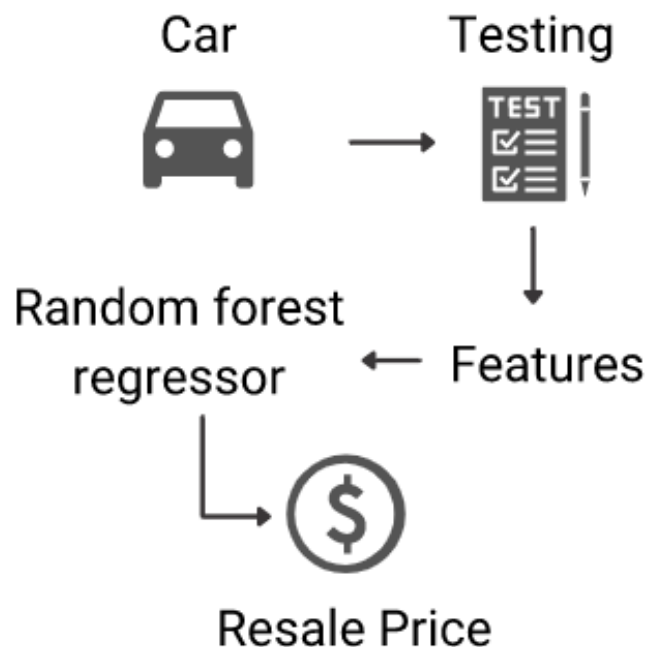
Machine Learning Categories

2. DECISION TREE

Decision Tree is a supervised learning ML algorithm used for both classification and regression problems. But they are commonly used to solve classification problems. DT is based on a tree-structure classifier in which the root node represents the entire samples or population, and branches represent the rules, internal nodes. The results represent the features of the dataset are defined by each leaf node.

3. RANDOM FOREST REGRESSOR

Random Forest is the most widely used algorithm that comes under the supervised learning category. This algorithm is based upon the concept of ensemble learning, further classified into multiple classifiers that are combined for efficient predictions. The combination of various classifiers is used to solve complex problems by increasing the model's performance. This algorithm is used for both regression and classification problems. RF consists of many DTs in the form of subsets of the provided dataset and takes an average of the subsets to improve the accuracy of the dataset.



Random Forest Regressor Model

4. XGBOOST REGRESSOR

XGBoost is an algorithm that has recently been dominating applied machine learning and Kaggle competitions for structured or tabular data. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance.



Past data are used to train the model, and then this trained model is used to test new data and then for prediction. The trained ML model's performance is evaluated using some portion of available past data (which is not present during training). This is usually referred to as the validation process. In this process, the ML model is evaluated for its performance measure, such as accuracy. Accuracy describes the ML model's performance over unseen data in terms of the ratio of the number of correctly predicted features and total available features to be predicted. Since the Regressor algorithm gives better performance and reduces the execution time of the model and is fitted on XGBoost Regression and Random Forest Regression algorithms.