**KONGU ENGINEERING COLLEGE**

**-Perundurai**

LATHIKA K

22CSR107

**ASSIGNMENT-1**

* **FEATURE:** Individual measurable properties (e.g., Make Year, Brand ,Variant etc) used as inputs to the model.
* **LABEL:** The output variable that the model aims to predict (e.g., Resale Price).
* **PREDICTION:** Project a portable dataset that relates back to the original data.(eg: In variant value 520d, the d represents that the car run in diesel which relates the information fuel).
* **OUTLIER:** A data point that deviates significantly from the rest of the data (e.g., if there was a car with resale price at 12000000in this dataset).
* **TEST DATA:** After trained a model test it on unseen data. It is a valid data.(e.g.,: Mileage gives use the accurate valid data).
* **TRAINING DATA:** Model goes through training dataset and validation dataset. It extract features.(e.g.,:Brand name of the car BMW).
* **MODEL:** It is a relationship between features and label. (eg:Relationship between brand and resale price).
* **VALIDATION DATA:**The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while turning model hyperparameters.(e.g.,: Resale Price).
* **HYPERPARAMETERS:** They are computational variables. It can be inferred by labels and features.(e.g.,:Brand).
* **EPOCH:** When all training data is used at once and is defined as total number of iterations of all training data in one cycle for training the ML model.
* **LOSS FUNCTION:** Quantifies the gap or the error margin of the car price predicted by the network to the actual price.
* **LEARNING RATE:** A hyperparameter used to govern the pace at which an algorithm updates or learns the values of a parameter estimate.
* **OVERFITTING:** When model cannot generalize and fits too closely to the training dataset instead.
* **UNDERFITTING:** When model has not learned the patterns in the training data well and is unable to generalize well on the new data.
* **REGULARIZATION:** Set of methods for reducing overfitting in models. It trades a marginal decrease in training accuracy for an increase in generalizability.
* **CROSS-VALIDATION:** To access the performance of a model and to prevent overfitting. It involves dividing the dataset into multiple subsets.
* **FEATURE ENGINEERING:** Leverages data to create new variables that aren’t in the training set. Produce new features.
* **DIMENSION REDUCTION:** A method representing a given dataset using a lower number of features.
* **BIAS:** It is a sort of mistake in which some aspects of a dataset are more weight.
* **VARIANCE:** The amount by which the performance of a predictive model changes when it is trained on different subsets of the training data.