Concept	Meaning	Remarks
CHAID	Outdated, Handles only categorical variables	
C4.5	Outdated, Handles both categorical and numerical variables	entropy
CART	Most popular and widely used model	gini
Ensemble	Use of multiple learning algo followed by voting for better performance	
Bagging	Bootstrap Aggregation: Random Sampling with Replacement, Shuffling by rows and cols	
OOB	Rows / Columns that are not included in a certain random sample. The model has no visibility to such records in that particular decision tree being created as they are not exposed to that data. Usually needed when accuracy is low	
n_estimators	The number of trees we want to build before taking the maximum voting or averages of predictions	
Linear Reg / Regression Eqn	y = mx + c [m = Slope, c = Bias]	predicted price = y = (w1 * sq-ft) + c
RSS	Residual Sum of Square: $\Sigma(i=1 - n)$ [Actual - Predicted]^2	
Pred Line	Line with the least RSS Value	
Eucledian Dist	d = $\sqrt{}$[(x2-x1)^2 + (y2-y1)^2 + (z2-z1)^2 + ...] --> Multi Dimensional	Default for Clustering Distance
Chebyshev Dist	d = Max(|a1-b1|, |a2-b2|, …. . |an-bn|)	Survey, Ordinal Data
Manhattan Dist	d = |a1-b1| + |a2-b2| + …. + |an-bn|	a.k.a. City Block Dist. Google Maps
WSS Plot	Elbow Curve to find out optimal number of clusters	First Elbow in the WSS Plot
Error Rate Graph	Graph to find out optimal value of K. Must be lowest Stable Result	
Linear Classification	Opposite classes divided by straight line	
Non Linear Classification	Opposite classes divided by curved line.	
Kernel Methods in SVM	Transform data to higher dimensions. Z = X^2 + Y ^{2} Polynomial: Non Linear Data Radial: Non-Linear Data Linear: Linear Data	
Parsimony Principle	Use the easiest Algorithm. If something can be solved by DT, do not go for RF or SVM	
C Hyperparameter	Controls width of Boundary in SVM	
Gamma Hyperparameter	Logistic Regression: Sigmoid Func+F21:H25	
Sigmoid Func	f(x) = 1 / (1 + e^-x)	
Probablistic Model	Prediction is based on probability	

Odds Ratio	Power of a variable to make the target variable 1 (weight calculation)	
R^2	**Coefficient of Determination: The square of Correlation Coefficient. $R^2 = r*r$** The amount of variation in the outcome (dependent variable) that is explained by the predictor (independent variable) Total amount of variation explained by the independent variables which is understood by the target variable Since $-1 < r < 1$ Hence $0 < R^2 < 1$	
Adjusted R^2	- When the predictor variables are increased, R^2 value increases, implying that increasing the count of variables leads to a good model, which is not true - Hence when the number of variables are increased, a better measure to evaluate the model's performance is Adjusted R^2 - Here the value of R^2 is adjusted for the no of predictor variables - Now when the variable count increases, the value of Adjusted R^2 will increase only when the added variables really adds to the model's explaination power	
Feature Engineering	Design features in such a way that the correlated variables do not harm the model. This can be done either by dropping one of the correlated variables or create a new variable that combines the features of 2 or more correlated variables	
Dimension Reduction Techniques:	1) Missing Values 2) Low Variance Filter 3) High Correlation Filter (Multicollinearity) 4) Random Forest (Limits max_features) 5) PCA	