50. MACHINE LEARNING :-

-> Series – One coloumn of data or holds only 1d data. Each element has an index. Dataframe – Table of rows and coloumns. Holds 2d data. Have both row and coloumn index.

->

-> Loc – Use to access a group of rows and coloumns by labels. Both start and end labels are included. Iloc - Use to access a group of rows and coloumns by integer positions. End position is not included.

-> Supervised – Training or shaping a model on the present dataset. Use to map datas from input to the desired output. Egs. Linear and logistic regression, svm, decision tree, etc. Unsupervised – No specific data is available to train the model for the labled responses. The model tries to learn the underlying structure of the data. Egs. K-means clustering, pca, etc.

->

-> Precision - Measures the accuracy of the positive predictions made by the model. P = TP / (TP + FP). Recall - Also known as sensitivity or true positive rate, measures the ability of the model to identify all relevant instances. R = TP / (TP + FN).

Accuracy = (TP + TN) / (TP + TN + FP + FN). Precision focuses on quality of positive predictions. Recall focuses on ability to capture all quality instances. Accuracy focuses on the overall correctness of the model.

-> Overfitting - occurs when a machine learning model learns the noise and outliers. This results in a model that performs exceptionally well on the training data but poorly on new, unseen data. Prevention – Cross validation, regularisation, pruning for decision tress, ensemble methods, etc.

-> Cross validation – For accessing ml models. Dividing the datasets into multiple folds. The models is trained on some percent on data, tested for other part and made to work for the third part.

-> Classification – Predict discrete labels or categories. Categorical data as output. Regression – Predict continuous data. Continuous value as output.

->