ML ASSIGNMENT 5

Q1) data validation, data insertion, data preprocessing, feature extraction, training model, testing model, deployment. Data preprocessing includes variety of tasks like removing outliers and unnecessary data, normalisation, feature extraction, dimensionality reduction, removing null values etc.

Q2) quantitative data includes numerical values and numerical operations and comparisons can be performed on it. Qualitative data includes non numerical aspects like strings and texts, categories etc.

Q3) int(32,64), float, string, category etc.

Q4) removing outliers and unnecessary data, normalisation, feature extraction, dimensionality reduction, removing null values etc address these issues.

Q5) converting to numerical values (integers), label encoding, one hot encoding etc.

Q6) learning will be skewed and learning could be overfitting/underfitting. Missing values can be dropped or replaced with mean, median, mode etc.

Q7) Missing values can be dropped or replaced with mean, median, mode etc.

Q8) Data preprocessing includes variety of tasks like removing outliers and unnecessary data, normalisation, feature extraction, dimensionality reduction, removing null values etc. dimensionality reduction reduces dimensions of features while feature selection extracts features using a variety of means like dropping unnecessary features.

Q9) i) interquartile range is range between 25 and 75 percentiles.

ii) 25, 50 and 75 percentile ranges. When outliers are more in the lower region, lower whisker length surpasses upper whisker. Points outside the boundary lines can be removed.

Q10) data comes in batches at regular intervals. The gap between 25 and 75 percentiles is interquartile range. Cross tab puts one variable on rows and another variable on columns to allow bivariable comparison.

Q11) nominal data has no order or rank while ordinal data does. Histogram gives probabilities for bins which are created by dividing the data into equal size intervals and placing these bins in order from minimum to maximum. Boxplot gives interquartile range and boundaries that allow us to identify and remove outliers.