Q1. What the meaning of prediction and classification?

Definition:

- Classification and prediction are two forms of data analysis those can be used to extract models describing important data classes or to predict future data trends.

- Such analysis can help to provide us with a better understanding of the data at large.

- Classification predicts categorical (discrete, unordered) labels, prediction models continuous valued functions.

Classification:

- The goal of data classification is to organize and categorize data in distinct classes.

- A model is first created based on the data distribution.

- The model is then used to classify new data.

- Given the model, a class can be predicted for new data.

- In general way of saying classification is for discrete and nominal values.

Prediction:

- The goal of prediction is to forecast or deduce the value of an attribute based on values of other attributes.

- A model is first created based on the data distribution.

- The model is then used to predict future or unknown values.

Q2. Define Regression with an example (Real-Life example)?

Linear regression is one of the most commonly used techniques in statistics. It is used to quantify the relationship between one or more predictor variables and a response variable.

Businesses often use linear regression to understand the relationship between advertising spending and revenue.

For example, they might fit a simple linear regression model using advertising spending as the predictor variable and revenue as the response variable. The regression model would take the following form:

$$revenue = \beta_0 + \beta_1(ad\ spending)$$

The coefficient β_0 would represent total expected revenue when ad spending is zero.

The coefficient β_1 would represent the average change in total revenue when ad spending is increased by one unit (e.g. one dollar).

If β_1 is negative, it would mean that more ad spending is associated with less revenue.

If β_1 is close to zero, it would mean that ad spending has little effect on revenue.

And if β_1 is positive, it would mean more ad spending is associated with more revenue.

Depending on the value of β_1, a company may decide to either decrease or increase their ad spending.

Q3. Explain Decision Tree and its applications?

A decision tree is a support tool with a tree-like structure that models probable outcomes, cost of resources, utilities, and possible consequences. Decision trees provide a way to present algorithms with conditional control statements. They include branches that represent decision-making steps that can lead to a favorable result.

The flowchart structure includes internal nodes that represent tests or attributes at each stage. Every branch stands for an outcome for the attributes, while the path from the leaf to the root represents rules for classification.

Decision trees are one of the best forms of learning algorithms based on various learning methods. They boost predictive models with accuracy, ease in interpretation, and stability. The tools are also effective in fitting non-linear relationships since they are capable of solving data-fitting challenges, such as regression and classifications.

Types of Decisions

There are two main types of decision trees that are based on the target variable, i.e., categorical variable decision trees and continuous variable decision trees.

1. Categorical variable decision tree

A categorical variable decision tree includes categorical target variables that are divided into categories. For example, the categories can be yes or no. The categories mean that every stage of the decision process falls into one of the categories, and there are no in-betweens.

2. Continuous variable decision tree

A continuous variable decision tree is a decision tree with a continuous target variable. For example, the income of an individual whose income is unknown can be predicted based on available information such as their occupation, age, and other continuous variables.

Applications of Decision Trees

1. Assessing prospective growth opportunities

One of the applications of decision trees involves evaluating prospective growth opportunities for businesses based on historical data. Historical data on sales can be used in decision trees that may lead to making radical changes in the strategy of a business to help aid expansion and growth.

2. Using demographic data to find prospective clients

Another application of decision trees is in the use of demographic data to find prospective clients. They can help in streamlining a marketing budget and in making informed decisions on the target market that the business is focused on. In the absence of decision trees, the business may spend its marketing market without a specific demographic in mind, which will affect its overall revenues.

3. Serving as a support tool in several fields

Lenders also use decision trees to predict the probability of a customer defaulting on a loan, by applying predictive model generation using the client's past data. The use of a decision tree support tool can help lenders in evaluating the creditworthiness of a customer to prevent losses.

Decision trees can also be used in operations research in planning logistics and strategic management. They can help in determining appropriate strategies that will help a company achieve its intended goals. Other fields where decision trees can be applied include engineering, education, law, business, healthcare, and finance.

Advantages of Decision Trees

1. Easy to read and interpret

One of the advantages of decision trees is that their outputs are easy to read and interpret, without even requiring statistical knowledge. For example, when using decision trees to present demographic information on customers, the marketing department staff can read and interpret the graphical representation of the data without requiring statistical knowledge.

The data can also be used to generate important insights on the probabilities, costs, and alternatives to various strategies formulated by the marketing department.

2. Easy to prepare

Compared to other decision techniques, decision trees take less effort for data preparation. Users, however, need to have ready information in order to create new variables with the power to predict the target variable. They can also create classifications of data without having to compute complex calculations. For complex situations, users can combine decision trees with other methods.

3. Less data cleaning required

Another advantage of decision trees is that, once the variables have been created, there is less data cleaning required. Cases of missing values and outliers have less significance on the decision tree's data.

Disadvantages of Decision Trees

1. Unstable nature

One of the limitations of decision trees is that they are largely unstable compared to other decision predictors. A small change in the data can result in a major change in the structure of the decision tree, which can convey a different result from what users will get in a normal event. The resulting change in the outcome can be managed by machine learning algorithms, such as boosting and bagging.

2. Less effective in predicting the outcome of a continuous variable

In addition, decision trees are less effective in making predictions when the main goal is to predict the outcome of a continuous variable. This is because decision trees tend to lose information when categorizing variables into multiple categories.