

NAME	MATRIC NO.
JOYADDAR MD JOBAYER	1731833

Answer 1: Supervised data mining is a type of machine learning learns from the labelled data.

The data which has a particular value, or a category assigned to it is a labelled data.

Answer 2: In data mining, data classifications is a typical strategy for organizing data sets that are both complex and huge. This method frequently employs algorithms that may be quickly modified to increase data quality. Therefore, data mining's classifications process is tightly linked to supervised learning. The goal of classifications is to link a variable of interest to variables that have been observed. The actual variable of interest is a "Qualitative" type variable.

Answer 3: The following are seven of the most common classification algorithms.

1. On the one hand, **structured data classification** is frequently applied to both structured and unstructured data. When data must be thoroughly classified into numerous classes, this approach is typically utilized.
2. A typical machine learning algorithm is **logistic regression**. This method is used to categories large datasets. Probabilities are considered throughout the implementation of this algorithm.
3. The next classification method would **be Naive Bayes**. It is based on the Bayes theorem, as its name suggests. It assumes of two distinct characteristics.
4. **Stochastic Gradient Descent** is an efficient and simple approach to suit the needs of linear data models.

5. **K-Nearest Neighbors** is regarded as a sluggish method of learning. Internal models aren't emphasized. For training reasons, it does, however, store several instances of data.
6. Classification algorithms remain incomplete without the **decision trees**. These are algorithms that focus on data attributes, which are brought together with classes.
7. Decision trees often expand to form random forests.

Example: Here's a simple example of a data mining classification method. Given a set of "n" qualities, which might be categorical or ordinal in nature. You'll also be given a set of "K" classes. A set of training cases is included in the collection, each of which has been meticulously labelled. The algorithms are expected to find a classification rule that can be used to forecast the class of each instance, according to the problem description. The values must be used in the forecast, as well as in each characteristic.

Answer 4:

To classify (Red, Domestic, SUV) we can use naïve bayes classifies.

Probability:

$$P(\text{Yes}) = 5/10 = 0.5$$

$$P(\text{No}) = 5/10 = 0.5$$

$$P(\text{Red} \mid \text{Yes}) = 3/5$$

$$P(\text{Red} \mid \text{No}) = 2/5$$

$$P(\text{Domestic} \mid \text{Yes}) = 3/5$$

$$P(\text{Domestic} \mid \text{No}) = 3/5$$

$$P(\text{SUV} \mid \text{Yes}) = 1/5$$

$$P(\text{SUV} \mid \text{No}) = 3/5$$

$$\Rightarrow P(\text{Yes} \mid \text{Red, Domestic, SUV}) = \propto P(\text{Red} \mid \text{stolen yes}) P(\text{Domestic} \mid \text{Yes}) P(\text{SUV} \mid \text{Yes}) P(\text{Yes})$$

$$= \propto 3/5 * 3/5 * 1/5 * 1/2$$

$$= \propto 9/250$$

$$P(\text{No} \mid \text{Red, Domestic, SUV}) = \propto P(\text{Red} \mid \text{No})$$

$$P(\text{SUV} \mid \text{No}) P(\text{Domestic} \mid \text{No}) P(\text{No}) = \infty * 2/5 * 3/5 * 3/5 * 1/2$$

$$= \infty * 18/250$$

18/250 (odds) → colour is safe and will not be stolen.

Answer 5:

Play = 1

Sleep = 0

Actual class Play	Actual class Sleep	Prediction
1	1	1
1	1	1
0	1	1
0	0	0
0	0	0
1	0	0
1	1	1
1	0	0
0	0	0
1	1	1



TP 4	FP 2
FN 1	TN 3

So, TP=4, TN=3, FN=1, FP=2

Accuracy: $TP + TN / TP + FP + FN + TN$

$$= 4 + 3 / 4 + 2 + 1 + 3$$

$$= 7/10$$

$$= 0.7$$

