Team ID: PNT2022TMID31356

Project Name: Smart Lender - Applicant Credibility Prediction for Loan Approval

PRIOR KNOWLEDGE:

1.Python language

2.Machine learning

 Unsupervised learning

Unsupervised learning is the training of a machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data.

Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore the machine is restricted to find the hidden structure in unlabeled data by itself.

Unsupervised learning is classified into two categories of algorithms:

- **Clustering:** A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.
- **Association:** An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.
-

Types of Unsupervised Learning:-

Clustering

- Exclusive (partitioning)
- Agglomerative
- Overlapping
- Probabilistic

Clustering Types:-

- Hierarchical clustering
- K-means clustering
- Principal Component Analysis
- Singular Value Decomposition
- Independent Component Analysing
-

Supervised learning

Supervised learning, as the name indicates, has the presence of a supervisor as a teacher. Basically supervised learning is when we teach or train the machine using data that is well labelled. Which means some data is already tagged with the correct answer.

After that, the machine is provided with a new set of examples(data) so that the supervised learning algorithm analyses the training data(set of training examples) and produces a correct outcome from labelled data.

Supervised learning is classified into two categories of algorithms:

Classification: A classification problem is when the output variable is a category, such as “Red” or “blue” , “disease” or “no disease”.
Regression: A regression problem is when the output variable is a real value, such as “dollars” or “weight”.
Supervised learning deals with or learns with “labeled” data. This implies that some data is already tagged with the correct answer.

Types:-

- Regression
- Logistic Regression
- Classification
- Naive Bayes Classifiers
- K-NN (k nearest neighbors)
- Decision Trees
- Support Vector Machine

Advantages:-

- Supervised learning allows collecting data and produces data output from previous experiences.
- Helps to optimize performance criteria with the help of experience.
- Supervised machine learning helps to solve various types of real-world computation problems.

Disadvantages:-

- Classifying big data can be challenging.
- Training for supervised learning needs a lot of computation time. So, itrequires a lot of time.

3. Knowledge about classification and regression problems

4. Data preprocessing technique

5. Visualization techniques

6. ML Models
   - Decision tree
   - KNN model
   - Xgboost model
   - Random forest model

7. Evaluation metrics

8. HTML,CSS, flaskframework

9. Jupiter Notebook

10. MYSQL