

Name: Manisha Apchunde

Roll no: 281076

PRN: 22420244

Batch: A3

Assignment 03

Aim:

Visualize the data using R/Python by plotting the graphs for assignment no. 1 and 2. Consider suitable data set. Use Scatter plot, Bar plot, Box plot, Pie chart, Line Chart.

Objective:

The objective of this assignment is to visualize datasets using various types of plots such as scatter plot, bar plot, box plot, pie chart, and line chart. These plots help us better understand the distribution, relationships, and trends within the data.

Resources Used :

Programming Language: Python

Libraries Used:

pandas - for data handling

matplotlib - for plotting graphs

seaborn - for advanced visualizations

Introduction to Visualization:

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools make it easier to understand patterns, trends, and outliers in data.

Support Vector Machine (SVM):

- **Supervised Learning Algorithm:** SVM is a supervised learning algorithm, meaning it requires labelled data for training. It learns to classify data into different categories based on the features provided.
- **Classification and Regression:** SVM can be used for both classification and regression tasks. In classification, it separates data points into different classes, while in regression, it predicts a continuous outcome.
- **Hyperplane Separation:** The primary goal of SVM is to find the hyperplane that best separates different classes in the feature space. This hyperplane has the maximum margin, which is the distance between the hyperplane and the nearest data points of each class.

Real Life Applications:

1. Business: Sales and marketing data trends over months
2. Healthcare: Patient record visualizations for diagnosis
3. Finance: Stock market data visualizations over time
4. Education: Student performance analysis using bar and line charts
5. Social Media: Visualization of engagement metrics and trends

Limitations of SVM:

- SVM might not perform well with large datasets as it can be computationally expensive during the training phase.
- It may not handle noisy data or data with overlapping classes effectively.
- SVM does not directly provide probability estimates, which might be a limitation in some applications where probabilistic outputs are required.

Working of SVM:

- SVM maps the input data into a high-dimensional feature space using a kernel function.
- It then finds the hyperplane that best separates the data points of different classes.
- The hyperplane is chosen to maximize the margin, which is the distance between the hyperplane and the nearest data points (support vectors) of each class.

- During prediction, SVM classifies new data points based on which side of the hyperplane they fall on.

Conclusion:

Visualization plays a crucial role in data analysis. It helps in making sense of the data and uncovering hidden insights. By using different types of plots, we can draw meaningful interpretations that aid in better decision-making. In this assignment, we practiced using multiple chart types to represent various aspects of the dataset, enriching our data understanding and analytical skills.