Lab07: Exploring Financial/Accounting Data Using Machine Learning Techniques

Introduction: In this lab, we aim to use machine learning techniques to explore financial and accounting data sourced from the dataset provided at 'https://github.com/masterfloss/data/raw/main/Exame500_2023.xlsx'. The dataset contains two sheets; the first contains the raw data, while the second provides translations for the column names.

Objective: This lab's main objective is to apply supervised and unsupervised learning algorithms to analyze financial and accounting data. Through this analysis, we seek insights into patterns, trends, and potential predictive relationships within the dataset.

Methodology:

1. **Data Acquisition:**
   o   We will retrieve the dataset from the provided GitHub link.
   o   Use the Pandas library in Python to load the dataset into our environment.
2. **Data Preprocessing:**
   o   Clean the dataset by handling missing values, outliers, and inconsistencies.
   o   Translate the column names using the information provided in the second sheet of the dataset.
3. **Exploratory Data Analysis (EDA):**
   o   Conduct exploratory data analysis to understand the distribution of variables, correlations, and potential relationships.
   o   Visualize critical aspects of the data, such as distributions, scatter plots, and correlation matrices.
4. **Unsupervised Learning:**
   o   Apply unsupervised learning techniques like clustering or dimensionality reduction to uncover hidden patterns within the data. Also use Cluster analysis.
   o   Visualize clusters and explore the characteristics of each cluster.
   o   Evaluate the effectiveness of the unsupervised learning approach.
5. **Supervised Learning:**
   o   Employ supervised learning algorithms such as regression or classification to predict relevant financial outcomes.
   o   Split the data into training and testing sets.
   o   Train various models and evaluate their performance using appropriate metrics.

Expected Outcomes:

- Identify variables influencing financial outcomes.
- Create prediction models with reasonable accuracy for relevant financial metrics.
- Discover hidden patterns and insights through unsupervised learning techniques.
- Based on the results obtained, obtain insights into potential areas for further analysis or intervention.

Conclusion: Through this lab, we aim to demonstrate the effectiveness of machine learning techniques in analyzing financial and accounting data. By applying supervised and unsupervised learning approaches, we anticipate gaining valuable insights to inform decision-making processes in financial management and accounting practices.