Machine Learning Data Analysis Project Report

1. Introduction

This project focuses on applying data analysis and machine learning techniques to a given dataset using Python. The main objective of the project is to understand the data, preprocess it, explore meaningful patterns, build a machine learning model, and evaluate its performance. The entire workflow has been implemented using a Jupyter Notebook (analysis (1).ipynb).

2. Objectives of the Project

The key objectives of this project are:

- Load and understand the dataset
- Clean and preprocess data
- Perform exploratory data analysis
- Select relevant features
- Build and train a machine learning model
- Evaluate model performance

3. Tools and Technologies Used

Python, Jupyter Notebook, NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn

4. Dataset Description

The dataset consists of multiple rows and columns with numerical and categorical features.

5. Data Preprocessing

Handling missing values, cleaning data, encoding categorical variables.

6. Exploratory Data Analysis

Used visualizations such as histograms, count plots, heatmaps, and scatter plots.

7. Feature Selection and Data Splitting

Independent variables (X) and dependent variable (y). Train-test split applied.

8. Model Building

A machine learning model was trained using Scikit-learn.

9. Model Prediction and Evaluation

Evaluated using accuracy score, confusion matrix, and classification report.

10. Results and Discussion

Analysis of model strengths and limitations.

11. Conclusion

Demonstrates a complete machine learning workflow.

12. Future Scope

Model improvement, automation, deployment.

13. References

Python, Scikit-learn, Pandas, NumPy documentation.