BDA Mini Project Report

Project Title: Enrolment_Data_March_to_July

Name: Anumula Bhavani

University: Malla Reddy University

Department: Computer Science Engineering

Guide: Professor Jogeshwarrao

Introduction

This project focuses on analyzing large-scale enrollment data using PySpark. It demonstrates data preprocessing, transformation, and visualization techniques applied to big data in a distributed computing environment.

Objectives

- To implement big data analysis using Apache Spark. - To perform data cleaning, aggregation, and visualization. - To extract insights from enrollment datasets.

Methodology

The dataset was processed using PySpark DataFrames. After loading and cleaning, data transformations were applied to extract key metrics such as total enrollment trends. Visualizations were generated to display insights effectively.

Results

The analysis produced meaningful insights on enrollment patterns from March to July. Spark efficiently processed the large dataset, and multiple visual graphs were generated.

Conclusion

This project demonstrated how PySpark can handle and analyze big datasets efficiently. It provided valuable insights and established a foundation for further predictive analytics.

Future Scope

Future enhancements include integrating Spark MLlib for predictive modeling, automating dashboards, and deploying analytics pipelines on cloud platforms.