Final Project Report - Handwritten Digit Recognition

1. Project Overview

The goal of this project is to classify handwritten digits (0-9) using pixel-based features extracted from images. The model uses machine learning techniques trained on a dataset like MNIST or similar structured numerical datasets.

2. Objectives

- To build a machine learning model for digit classification.
- To deploy a user-friendly Streamlit web app for CSV-based predictions.
- To visualize prediction results in Power BI.

3. Dataset Details

- Source: MNIST / digit feature dataset.
- Shape: 784 feature columns (28x28 pixel values) and 1 label (target).
- File Used: sample_input.csv, digit_predictions.csv

4. Tools & Technologies Used

- Python
- Pandas
- Scikit-learn
- Joblib
- Streamlit
- Power BI

5. Model Used

- Algorithm: Logistic Regression / Random Forest / SVM
- Input: Feature matrix of 784 pixel values
- Output: Predicted digit (0-9)

6. Streamlit Web App

Features:
- Upload CSV files containing pixel data
- Model predicts digit for each row
- Displays:
  - Data Preview

- Prediction Table
- Bar Chart of digit counts
- Downloadable prediction CSV

7. Power BI Visualization

- Imported digit_predictions.csv
- Used Bar Chart to show digit frequency (0-9)
- Optional: Pie chart, Filters, Slicers for analysis

8. Project Files

- app.py: Streamlit application
- digit_classifier_model.pkl: Trained ML model
- sample_input.csv: Input sample file
- digit_predictions.csv: Output with predictions
- digit_prediction_powerbi.pbix: Power BI Report file

9. Conclusion

The handwritten digit recognition system was successfully implemented using machine learning and deployed with Streamlit. It allows users to easily upload input data and receive predictions with visualization.

10. Future Scope

- Add image upload and CNN model for image-based recognition
- Host on cloud (Heroku / Streamlit Cloud)
- Add language toggle for Marathi/English
- Integrate with SQL for storing predictions