VISVESVARAYA TECHNOLOGICAL UNIVERSITY,

Jnana Sangama, Belagavi-590018.

A Mini-Project Report

On

HAND-WRITTEN DIGIT RECOGNITION

Submitted for the partial fulfillment of the requirement of the **VIth** *Semester*

MACHINE LEARNING (21AI63)

Submitted in partial fulfillment of the requirement for the award of the degree of

BACHELOR OF ENGINEERING

In

Artificial Intelligence and Machine Learning

Submitted by

SUMITH SIGTIA, 1OX21AI039

Under the Guidance of

Ms. PRIYANKA S V

Assistant Professor

Department of Artificial Intelligence and Machine Learning

THE OXFORD COLLEGE OF ENGINEERING**,**

Bommanahalli, Bangalore 560068

Department of Artificial Intelligence and Machine Learning

THE OXFORD COLLEGE OF ENGINEERING

Bommanahalli, Bangalore 560068

2024-2025

THE OXFORD COLLEGE OF ENGINEERING

Bommanahalli, Bangalore 560068

CERTIFICATE

This is to certify that the Mini-Project entitled **"HAND-WRITTEN DIGIT RECOGNITION"** carried out by **Mr. SUMITH SIGTIA [1OX21AI039]** of VIth Semester students of The Oxford College of Engineering, in partial fulfillment for the award of Bachelor of Engineering in Artificial Intelligence and Machine Learning of **Visvesvaraya Technological University, Belagavi** during the academic year 2023-2024. The Mini-Project report has been approved as it satisfies the academic requirements in respect of DBMS Laboratory with Mini-Project work prescribed forthe said Degree.

Signature of the Guide
Ms. Priyanka S V
Asst. Professor, Dept. of AIML
TOCE, Bangalore

Signature of the HOD
Dr. P. Bindhu Madhavi
Professor & HOD, Dept. of AIML
TOCE, Bangalore

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without complementing those who made it possible, whose guidance and encouragement made our efforts successful.

My sincere thanks to highly esteemed institution **THE OXFORD COLLEGE OF ENGINEERING** for grooming up me in to be AIML engineer.

I express our sincere gratitude to **Dr S. N. V. L. NARASIMHA RAJU** Chairman**, The Oxford Educational Institutions**, Bengaluru for providing the required facility.

I express our sincere gratitude to **Dr. H.N. Ramesh**, Principal, **TOCE**, Bengaluru for providing the required facility.

I am extremely thankful to **Dr. P. BINDHU MADHAVI**, **Professor & HOD of AIML**, TOCE for providing support and encouragement.

I am grateful to **Mrs. Priyanka S V**, Asst. Professor, Dept. of **AIML**, **TOCE** who helped me to complete this project successfully by providing guidance, encouragement and valuable suggestion during entire period of the project. I thank all my AIML staff and others who helped directly or indirectly to meet my project work with grand success.

Finally, I am grateful to my parents and friends for their invaluable support guidance and encouragement.

SUMITH SIGTIA,1OX21AI039

ABSTRACT

This project focuses on the application of various machine learning techniques to the MNIST dataset, a large collection of handwritten digits. The goal is to build and evaluate models that can accurately classify the digits from 0 to 9. The MNIST dataset serves as a standard benchmark for testing machine learning algorithms and includes 70,000 grayscale images of handwritten digits, each of size 28x28 pixels.

Key Objectives

1. **Data Preprocessing and Visualization**:
    - Load and explore the MNIST dataset.
    - Visualize sample digits to understand the data distribution and characteristics.
2. **Binary Classification**:
    - Implement a binary classifier to identify whether a digit is '5' or not.
    - Train a Stochastic Gradient Descent (SGD) classifier and evaluate its performance using cross-validation, confusion matrix, precision, recall, and F1 score.
3. **Precision-Recall Trade-off**:
    - Explore the precision-recall trade-off by plotting precision and recall as functions of decision thresholds.
    - Determine the optimal threshold for maximizing precision and recall.
4. **Receiver Operating Characteristic (ROC) Curve**:
    - Plot the ROC curve and calculate the Area Under the Curve (AUC) to evaluate the classifier's performance.
5. **Random Forest Classifier**:
    - Train a Random Forest classifier and compare its performance with the SGD classifier using ROC curves and AUC scores.
6. **Multi-class Classification**:
    - Implement multi-class classification using Support Vector Machine (SVM) and One-vs-Rest (OvR) strategies.
    - Evaluate the models using cross-validation and confusion matrices.
7. **Normalization and Confusion Matrix Analysis**:
    - Normalize the dataset and plot confusion matrices to visualize classification errors.
    - Analyze misclassifications and propose strategies for improvement.

8. **Multi-label Classification**:
    - Implement multi-label classification using K-Nearest Neighbors (KNN) to classify digits based on multiple labels (e.g., large/small digit, odd/even digit).
    - Evaluate the model using F1 score.
9. **Noise Reduction**:
    - Add noise to the images and use KNN to clean the noisy images.
    - Evaluate the effectiveness of noise reduction techniques.

CONTENT