ACKNOWLEDGEMENT

I am greatly indebted to our project guide **Dr. A. M. Rajurkar** for her able guidance throughout this work. It has been an altogether different experience to work with her and I would like to thank her for help, suggestions and numerous discussions

I gladly take this opportunity to thank **Dr. A. M. Rajurkar (Head of Computer Science & Engineering, MGM's College of Engineering, Nanded).** I am heartily thankful to **Dr. Lathkar G. S. (Director, MGM's College of Engineering, Nanded)** for providing facility during the progress of the project and for her kind help, guidance, and inspiration.

Last but not least I am also thankful to all those who help directly or indirectly to develop this project and complete it successfully.

Ambarish Udgirkar
[B.Tech CSE-A]

ABSTRACT

In today's digital communication landscape, email remains a dominant medium, yet users are often overwhelmed by the sheer volume and variety of messages they receive. This project presents a deep learning-based email classification system designed to automatically categorize emails into four major classes: Spam, Promotions, Social, and Updates. The proposed solution is built upon a hybrid CNN-BiLSTM architecture that combines the local feature detection strength of Convolutional Neural Networks (CNN) with the contextual memory and sequence processing capabilities of Bidirectional Long Short-Term Memory (BiLSTM) networks. This design enables the model to capture both keyword patterns and deeper semantic relationships across email content.

Real-world data was extracted directly from Gmail using the Gmail API, and further enhanced using Optical Character Recognition (OCR) to extract text from image-based emails. A comprehensive preprocessing pipeline was implemented to clean, normalize, and tokenize the email data, ensuring linguistic consistency and reducing input noise. With rigorous training, regularization, and hyperparameter tuning, the model achieved a test accuracy of 95.3%, outperforming earlier versions based on LSTM and CNN-LSTM architectures. The system was deployed via a Flask-based API, supporting real-time classification and seamless integration with other applications. Overall, the project demonstrates a reliable, scalable, and production-ready approach for intelligent email sorting, addressing both academic challenges and real-world user needs.comprehension.

TABLE OF CONTENTS

LIST OF FIGURES