AN ARTIFICIAL INTELLIGENCE PROJECT PROPOSAL

on

SMS SPAM SHIELD: MULTI-CATEGORY XAI SPAM DETECTOR

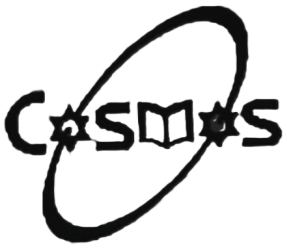
Submitted By

Alok Kumar Jha (230302)
Bibek Kumar Jha (230310)
Kushal Prasad Joshi (230345)

Submitted To

Er. Ranjan Raj Aryal

in partial fulfilment of requirement for the practicals of
Artificial Intelligence (CMP 346) course.



Cosmos College of Management & Technology
(Affiliated to Pokhara University)
Sitapaila, Kathmandu, Nepal

January 18, 2026

Cosmos College of Management & Technology
(Affiliated to Pokhara University)
Sitapaila, Kathmandu, Nepal

APPROVAL

This is to certify that the project proposal titled:

SMS Spam Shield:
Multi-Category XAI Spam Detector

has been reviewed and approved by the project assigner Er. Ranjan Raj Aryal for the further working on project in partial fulfilment of requirement for the practicals of Artificial Intelligence (CMP 346) course.

Project group memebers of Bacholer of Engineering in Computer Engineering named as Alok Kumar Jha (230302), Bibek Kumar Jha (230310) and Kushal Prasad Joshi (230345) can work on the project titled SMS Spam Shield: Multi-Category XAI Spam Detector and submit the final report to fulfill the requirement for the practicals of Artificial Intelligence (CMP 346) course by Pokhara University.

_____

Er. Ranjan Raj Aryal
Course Lecturer

Date of approval: _____

ABSTRACT

This project proposes **SMS Spam Shield: Multi-Category XAI Spam Detector**, an intelligent and explainable system for classifying SMS messages into multiple actionable categories such as *spam, phishing, promotional, transactional, and legitimate* messages. Unlike conventional binary spam filters, the proposed system aims to provide fine-grained classification while offering transparent, human-interpretable explanations for each prediction.

The system is designed to combine classical machine learning models, including Logistic Regression, Naive Bayes, and SVM, with a deep learning-based recurrent neural network (RNN/LSTM). An ensemble-based aggregation strategy is employed to improve robustness and generalization across diverse SMS patterns. To address the black-box nature of automated text classifiers, explainable artificial intelligence techniques such as LIME and SHAP are incorporated to generate token-level explanations and confidence measures for classification decisions.

The project focuses on English-language SMS messages and utilizes offline-trained models evaluated using standard multi-class performance metrics, including precision, recall, F1-score, and confusion matrices. By integrating ensemble learning with explainable AI (XAI), the proposed system aims to enhance both the accuracy and transparency of SMS spam detection, benefiting end users, system administrators, and researchers seeking interpretable and trustworthy text classification solutions.

Keywords: SMS spam detection, multi-category classification, explainable AI (XAI), ensemble learning, LSTM, LIME, SHAP.

PREFACE

This document is submitted in partial fulfillment of the requirements for the Bachelor of Engineering degree in Department of Information Communication and Technology (ICT). The proposed project, *SMS Spam Shield: Multi-Category XAI Spam Detector*, aims to address the increasing variety and sophistication of unsolicited SMS messages by developing an accurate and interpretable SMS classification system. The motivation for this work arises from the growing societal and economic impact of SMS-based spam, phishing, and fraudulent communication, as well as the increasing demand for transparency in automated decision-making systems used in security and communication domains.

Through this project, we seek to explore practical applications of artificial intelligence in cybersecurity, design a robust and user-friendly SMS spam detection framework, and contribute towards improving message safety and user trust through explainable classification mechanisms.

This project is intended to be carried out under the supervision of Er. Ranjan Raj Aryal, whose expertise and guidance are expected to be invaluable throughout the development process. With this proposal, we formally seek approval to proceed with the proposed work and look forward to the opportunity to contribute to academic learning and applied research in artificial intelligence.

We, Alok Kumar Jha (230302), Bibek Kumar Jha (230310) and Kushal Prasad Joshi (230345), hope that this proposal clearly communicates the objectives and planned approach of the proposed SMS Spam Shield: Multi-Category XAI Spam Detector, and serves as a strong foundation for its successful execution under the guidance of Er. Ranjan Raj Aryal.

ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to our respected supervisor, Er. Ranjan Raj Aryal, for his continuous support, encouragement, and expert guidance throughout the process of preparing this project proposal. His valuable feedback and insights have been instrumental in shaping the direction of our work.

We are also thankful to the Department of Information Communication and Technology (ICT) and all the faculty members of Cosmos College of Management & Technology (Affilated to Pokhara University), Sitapaila, Kathamandu, Nepal for their continuous support and for providing us with the opportunity and resources to carry out this proposed project.

We would also like to express our kind regards to the people around us who have directly or indirectly contributed to the successful completion of this proposal. Also we will thank our college friends who gave us valuable suggestions and feedback during the preparation of this project proposal.

Parts of this proposal were drafted and refined with the assistance of AI-powered language models, including ChatGPT [**?**]. The AI tools were used solely to help with structuring, phrasing, and clarity of the text. All research, analysis, design, and conclusions presented in this proposal are entirely the author's own work.

Finally, we extend our sincere thanks to our family and friends for their unwavering support and encouragement during this endeavour.

We are grateful to all of you.

Alok Kumar Jha (230302), Bibek Kumar Jha (230310) and Kushal Prasad Joshi (230345)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
LIME	Local Interpretable Model-Agnostic Explanations
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
SHAP	SHapley Additive exPlanations
SMS	Short Message Service
SVM	Support Vector Machine
XAI	Explainable Artificial Intelligence