FINAL YEAR PROJECT REPORT

URDU NEWS CLASSIFICATION

In fulfillment of the requirement
For degree of
BS (COMPUTER SCIENCES)

By

KHURRAM ALI TABISH	48502 BSCS
ALI ASGHAR	48469 BSCS
MUHAMMAD USMAN BHATTI	48443 BSCS

SUPERVISED

BY

MISS KOMAL FATIMA

BAHRIA UNIVERSITY (KARACHI CAMPUS)
FALL-2020

DECLARATION

We hereby declare that this project report is based on our original work except for citations and quotations which have been duly acknowledged. We also declare that it has not been previously and concurrently submitted for any other degree or award at Bahria University or other institutions.

Signature :

Name : KHURRAM ALI TABISH

Reg No. : 48502

Signature :

Name : MUHAMMAD USMAN BHATTI

Reg No. : 48443

Signature :

Name : ALI ASGHAR

Reg No. : 48469

Date :

APPROVAL FOR SUBMISSION

We certify that this project report entitled **"URDU NEWS CLASSIFICATION"** was prepared by **KHURRAM ALI TABISH, MUHAMMAD USMAN BHATTI, ALI ASGHAR** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Computer Science at Bahria University.

Approved by,

Signature : Komal fatima.

Supervisor : Miss KOMAL FATIMA

Date : 11/1/2021

ACKNOWLEDGEMENTS

We would like to thank everyone who had contributed to the successful completion of this project. We would like to express our gratitude to my research supervisor, Miss KOMAL FATIMA for her invaluable advice, guidance and her enormous patience throughout the development of the research.

In addition, we would also like to express my gratitude to our loving parent and friends who had helped and given us encouragement.

URDU NEWS CLASSIFICATION

ABSTRACT

In this age of information news is an important aspect of our daily lives. The need to stay up to date with everyday events is becoming greater day by day. However different types of people are interested in different types of news. As such a system is required that can classify news according to category to make it easier for users to find news that is relevant to them. There are existing systems for English language however that is not the case in Urdu and there is very limited work in regards to Urdu text classification as classifying text in Urdu can be a very challenging task. In this project, we are using pre compiled Urdu news datasets. Our datasets contains news related to six categories with Health, Science, Politics, Entertainment, Business and Sports. In order for the machine learning algorithms to work on the data we needed to apply pre-processing techniques like stop words removal and a feature extraction first. Feature extraction was performed by using TF-IDF and count vectorization techniques. We will use LSTM model for News classification targeting 80% accuracy.

TABLE OF CONTENTS

CHAPTER