



FINAL YEAR PROJECT REPORT

**AUTO COMPLETING SENTENCES IN URDU
USING NATURAL LANGUAGE
PROCESSING**

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

By

**EISHA TAHIR
NAJEEBA ALI**

**72980 (BSCS)
72992 (BSCS)**

SUPERVISED

**BY
DR RAHEEL SIDDIQI**

BAHRIA UNIVERSITY (KARACHI CAMPUS)

FALL-2024

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: Eisha Tahir .

Name : Eisha Tahir

Reg No. : 72980

Signature: Najeeba Ali

Name : Najeeba Ali

Reg No. : 72992

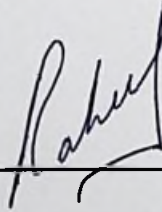
Date : 18-November-2024

APPROVAL FOR SUBMISSION

We certify that this project report entitled **“AUTO COMPLETING SENTENCES IN URDU USING NATURAL LANGUAGE PROCESSING”** was prepared by **Eisha Tahir and Najeeba Ali** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Science in Computer Science at Bahria University, Karachi.

Approved by,

Signature:



Supervisor: DR RAHEEL SIDDIQI

Date : 18- November-2024

The copyright of this report belongs to Bahria University according to the Intellectual Property Policy of Bahria University BUORIC-P15 amended on April 2019. Due acknowledgement shall always be made of the use of any material contained in, or derived from, this report.

© 2024 Bahria University. All right reserved.

Specially dedicated to
my beloved parents
(EISHA TAHIR)

Specially dedicated to
my beloved parents
(NAJEEBA ALI)

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 our project supervisor, DR RAHEEL SIDDIQI for his invaluable advice, guidance and his enormous patience throughout the development of the research.

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

AUTO COMPLETING SENTENCES IN URDU USING NATURAL LANGUAGE PROCESSING

ABSTRACT

This project addresses the lack of advanced text generation tools for Urdu, the national language of Pakistan, which remains underdeveloped compared to English due to limited resources. To ease tasks like chatting, emailing, and document typing in Urdu, the project developed a model for Urdu text prediction and generation. The process began by gathering data from Urdu newspapers and blogs to build a suitable dataset. Various models were evaluated, including RNN, LSTM, GRU, and N-Gram. The model accuracy results showed that the N-Gram model achieved 76%, the RNN combined with LSTM reached 83%, and the RNN combined with GRU reached 70%. We connected our front end with each model, allowing users to select any model according to their requirements and needs. The final deliverable is a user-friendly webpage where users can input incomplete Urdu sentences, and the system predicts and suggests possible completions, enhancing efficiency in Urdu writing tasks.

TABLE OF CONTENTS

DECLARATION	ii
APPROVAL FOR SUBMISSION	iii
ACKNOWLEDGEMENTS	vi
ABSTRACT	vii
TABLE OF CONTENTS	viii
LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF SYMBOLS / ABBREVIATIONS	xiii
LIST OF APPENDICES	xiv

CHAPTER

1	INTRODUCTION	1
	1.1 Background	1
	1.2 Problem Statement	1
	1.3 Aims and Objectives	2
	1.4 Scope of Project	2
2	LITERATURE REVIEW	4
	2.1 Background	4
	2.2 Related Work	4
	2.3 Comparison with Existing Studies	6
	2.4 Comparison with Our Project	8
	2.5 Chapter Summary	8

3	REQUIREMENT SPECIFICATIONS	9
	3.1 Existing System	9
	3.2 Proposed System	9
	3.3 Requirement Specification	10
	3.4 Use Case	11
4	DESIGN AND METHODOLOGY	13
	4.1 Proposed Methodology (Framework/Architecture)	13
	4.2 Process Model	14
	4.3 Modules Discussion	14
	4.4 Project Diagram	15
	4.4.1 Dataset	16
	4.4.2 Sequence	17
	4.4.3 Context	18
5	SYSTEM IMPLEMENTATION	19
	5.1 Language Model	19
	5.2 Implementation of N-gram	19
	5.2.1 Data Preprocessing	19
	5.2.2 Model Architecture	20
	5.2.3 Training	21
	5.2.4 Evaluation	21
	5.3 Implementation of RNN+LSTM	21
	5.3.1 Data Preprocessing	21
	5.3.2 Model Architecture	22
	5.3.3 Training	22
	5.3.4 Evaluation	22
	5.4 Implementation of RNN+GRU	22
	5.4.1 Data Preprocessing	22
	5.4.2 Model Architecture	23
	5.4.3 Training	23
	5.4.4 Evaluation	23

		x
5.5	Front-End Integration	23
	5.5.1 Objective	23
	5.5.2 Development	23
6	SYSTEM TESTING AND EVALUATION	28
6.1	N-Gram Model	28
6.2	RNN+LSTM Model	30
6.3	RNN+GRU Model	32
7	CONCLUSION	37
7.1	Summary	37
7.2	Recommendations for future work	38
	REFERENCES	39
	APPENDICES	41