Resume Classification Using Natural Language Processing and Machine Learning Algorithm

A Project presented to the National University in partial fulfillment of the requirement for the degree of Bachelor of Science (Hon's) in Computer Science & Engineering

Submitted By

Najib Ahmed

Registration no: 17502005012

Session: 2017-2018

Jannat Ara

Registration no: 17502005011

Session: 2017-2018

Supervised By
Md. Imran Hossain
Head
Department of CSE
Daffodil Institute Of IT

Department of Computer Science & Engineering

Daffodil Institute of IT Dhaka

Under National University Bangladesh

Date of Submission: 29 August, 2023

APPROVAL

The Project "**Resume Classification Using Natural Language Processing and Machine Learning Algorithm**" is submitted to the Department of Computer Science & Engineering, DIIT under the National University of Bangladesh in partial fulfillment of the requirements for the degree of Bachelor of Science (Hon's)in Computer Science and Engineering and approved as to its style and content.

---------- ----------

Examiner **Examiner**

Md. Imran Hossain

Head and Project Supervisor

Department of CSE

Daffodil Institute Of IT

DECLARATION

We affirm that the project work titled "**Resume Classification Using Natural Language Processing and Machine Learning Algorithm"** is being submitted in partial fulfillment for the degree of **B.Sc. (Hon's) in Computer Science & Engineering** is the original work carried out by me. It has not formed part of any other project work submitted for any degree or diploma, either in this or any other University.

Submitted By

Najib Ahmed

Registration no: 17502005012

Session: 2017-2018

Jannat Ara

Registration no: 17502005011

Session: 2017-2018

ACKNOWLEDGEMENT

Despite our efforts, the success of this project depends largely on the encouragement and guidanceof our mentors. We would like to take this opportunity to express our gratitude to the people whoare playing a vital role in the successful completion of this project.

Our sincere thanks to **Prof. Dr. Mohammad Shakhawat Hossain,** Principle DIIT who has allowed us towork on this project and showed encouragement.

Our cordial thanks to our Project Supervisor **Md. Imran Hossain,** Lecturer and Head of the Department of Computer Science & Engineering, DIIT for his valuable guidance and support to meet the successful completion of our project.

I express my gratitude to **Poly Bhoumik**, Senior Lecturer, DIIT, Dhaka, for having provided us with the facilities to do the project successfully.

I express my gratitude to **Saidur Rahman**, Senior Lecturer, DIIT, Dhaka, for having provided us with the facilities to do the project successfully.

I express my gratitude to **Safrun Nesa Saira**, Lecturer, DIIT, Dhaka, for having provided us with the facilities to do the project successfully.

I express my gratitude to **Nusrat Jahan Sarker**, Lecturer, DIIT, Dhaka, for having provided us with the facilities to do the project successfully.

I express my gratitude to **Moumita Akter**, Lecturer, DIIT, Dhaka, for having provided us with the facilities to do the project successfully.

I express my gratitude to **Mushfiqur Rahaman**, Lecturer, DIIT, Dhaka, for having provided us with the facilities to do the project successfully.

Finally, we would be grateful to the **National University, Bangladesh,** and the coordinators of the **Bachelor of Science in Computer Science and Engineering** degree program for giving usthis opportunity to apply the knowledge that we have gained through the study of the degree program.

ABSTRACT

Using automated methods that combine machine learning and natural language processing, resumes will be predicted and scored in accordance with their relevance to certain jobs. The system will extract important data from resumes to quickly and effectively identify the best-fit candidates from a large number of resumes for a job opening, lowering the costs associated with recruitment and ensuring that the recruitment process is in line with the organization's overall strategic vision.

TABLE OF CONTENTS

Approval ii
Declaration iii
Acknowledgment iv
Abstract v
List of Figures x
List of Tables x

Chapter 1: Introduction 2-4
1.1 Introduction 2
1.2 Project Aims 2
1.3 Objective 3
1.4 Features of Our System 3
1.5 Why We Select This Project 4
1.6 Limitations of Our Project 4

Chapter 2: Background Study 6-8
2.1 Background of this Project 6
2.1.1 First Generation Hiring System 6
2.1.2 Second Generation Hiring System 6
2.1.3 Third Generation Hiring System 6
2.2 Problems with the Current System 7
2.3 The Solution of the Problem 7
2.4 Methodology 7-8

Chapter 3: System Specification 10-20
3.1 Requirements (Hardware) 10
3.2 Requirements (Software) 10
3.3 Platform 10
3.4 Front End Language 10-12
3.4.1 HTML 11
3.4.2 CSS 11-12

3.4.3 JavaScript 12
3.5 Database 12-13
3.6 Back-End Language 13-14
3.6.1 Python 13
3.7 Integrated Development Environment 14-15
3.7.1 PyCharm 15
3.8 Framework 15
3.8.1 Django 15
3.9 Natural Language Tool Kit (Python Package) 16
3.10 Scikit-Learn (Python Package) 16
3.11 Word Embedding 16-17
3.11.1 Frequency-based embedding categories. 18
3.12 Machine Learning Algorithm 16
3.12.1 Supervised Learning 17
3.12.2 Unsupervised Learning 17
3.12.3 Reinforcement Learning 18
3.13 K-Nearest Neighbors (k-NN) 18
3.13.1 k-NN Algorithm 19
3.13.2 Key Aspect of k-NN Algorithm 20
3.14 Random Forest 20

Chapter 4 System Analysis 22-27

4.1 Analysis Model 22
4.1.1 Software Development Life Cycle (SDLC) 22
4.1.2 Necessity of SDLC 23
4.1.3 Software Development Life Cycle for Our Project 23
4.2 Agile Method 23-25
4.2.1 Advantages of Agile Method 24
4.2.2 Disadvantages of Agile Method 25
4.2.3 When to use Agile Method 25
4.2.4 Why have We Chosen Agile 25
4.3 NLP Methodology 26-30
4.3.1 Key Components of NLP Methodology 26
4.3.2 Term Frequency Inverse Document Frequency(TF_IDF) 26-27

Chapter 5: Feasibility Report 29-30
5.1 Feasibility Study 29
5.2 Technical Feasibility Study 29
5.2.1 k-NN Technical Feasibility 29
5.3 Operational Feasibility Study 29-30
5.4 Economically Feasibility 30
5.5 Performance Analysis 30

Chapter 6: Design of Proposed System 32-45
6.1 Introduction 32
6.2 Flowchart 32
6.2.1 Elements of Flowchart 33
6.3 System Architecture 34
6.4 Data Flow Diagram (DFD) 34- 40
6.4.1 External Entity 35
6.4.2 Process 35
6.4.3 Data Store 35
6.4.4 Data Flow 35
6.4.5 DFD Level 0 36
6.4.6 DFD Level 1 36
6.4.7 DFD Level 2 37
6.5 Use Case Diagram 38
6.6 Activity Diagram 39-40
6.7 Sequence Diagram 41-43
6.8 Entity Relationship Diagram 42-43
6.8.1 Terms of E-R Diagram 43-44
6.8.2 Relationship Cardinality 43
6.8 Ranking Model Design 44

Chapter 7: Implementation 46-50
7.1 Admin Page 46
7.2 Recruiter Page 46
7.3 Candidate Page 46
7.4 Home Page 47

7.5 Register Page 47
7.6 Login Page 48
7.7 Post Job 48
7.8 Job Listing 49
7.9 Job Description 49
7.10 Apply for Job 50
7.11 Shortlisted Candidates 50

Chapter 8: Conclusion 51-53

8.1 Future Enhancement 52
8.1 Conclusion 52
References 53

List of Figures

Figure 4.2: Agile method software model 24
Figure 6.2: Flowchart 32
Figure 6.2.1: Elements of Flowchart 33
Figure 6.3: System Architecture.....34
Figure 6.4.5: DFD Level 0 36
Figure 6.4.6: DFD Level 1 36
Figure 6.4.7: DFD Level 2 37
Figure 6.5: Use Case Diagram 38
Figure 6.6.1: User Activity Diagram 39
Figure 6.6.2: Data Extraction Activity Diagram.....40
Figure 6.6.3: Job Description Conversion Activity Diagram.....40
Figure 6.7: Sequence Diagram 41
Figure 6.8: Entity-Relationship Diagram 42
Figure 6.8.1: Components of ER Diagram.....43
Figure 6.9: Ranking Model 44

List of Tables

Table 5.5: Performance Evaluation.....30