Bahria University
Discovering Knowledge

FINAL YAER PROJECT REPORT

CV SORTER:
A TEXT MINING BASED APPROACH TO SELECT A CV

By

AAQIB SULTAN SIDDIQUE MOHAMMAD	**(41261)**
AHMED ALI GHOURI	**(41268)**
MUHAMMAD AZEEM	**(41314)**
MUHAMMAD ANIQUE IMAM	**(41302)**
MUHAMMAD DANISH	**(41316)**

SUPERVISED BY

(ASIA SAMREEN)

BAHRIA UNIVERSITY (KARACHI CAMPUS)

2019

ACKNOWLEDGEMENTS

We would like to thank everyone who had contributed to the successful completion of this project. We would like to express my gratitude to my research supervisor, Miss ASIA SAMREEN for her invaluable advice, guidance and his/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 me encouragement.

/

CV SORTER:
A TEXT MINING BASED APPROACH TO SELECT A CV

ABSTRACT

Human resource department of organizations receive infinite CVs on regular basis, making it more complex for the HR person to find the best job candidates. In fact, an employer receives an average of 144 CVs per job opening. Usually different people follow the different style therefore required information about the candidates is sometimes missed or takes time to search for. It is usually observed that many people add lot of unnecessary details in their CV. Upon visiting the market, it has been noticed that industries heavily depends on evaluation tools to select the most suitable applicant for the job. In fact, many companies are turning to systems that saves time and rapidly finds the best candidates. HR process of finding the best candidates is challenging because of both sequence of identifying, and quality of determining the candidates.
Purpose of this project is to provide an efficient CV extraction tool using text mining algorithm (K-Nearest Neighbours), and also using the features of Natural Language Processing (NLP).The main concept of this research based project is to create a best way to choose the appropriate CV accurately and in speedy process by extracting the best matching resumes. Which will allow HR department to choose the right candidate with minimal standard process.

TABLE OF CONTENTS

DECLARATION			**2**
APPROVAL FOR SUBMISSION			**3**
ACKNOWLEDGEMENTS			**6**
ABSTRACT			**7**
TABLE OF CONTENTS			**8**
LIST OF FIGURES			**11**
LIST OF TABLES			**12**
LIST OF APPENDICES			**13**
CHAPTER			
1	**INTRODUCTION**		**1**
	1.1	Unstructured and Structured data	1
	1.2	Text Mining and Natural Language Processing (NLP)	2
	1.3	Applications of Text Mining	2
	1.4	Text Mining Areas of Application	3
	1.5	Recruitment Procedures in Renown Companies	3
	1.6	Background	5
	1.7	Problem Statement	5
	1.8	Aims and Objectives	6
	1.9	Scope of Project	6
2	**LITERATURE REVIEW**		**8**
	2.1	E-Recruitment: A step towards paperless HR	8
		2.1.1 Techniques:	9

2.2 Candidate recruitment system using keyword based searching 9

2.2.1 Information extraction 10

2.2.2 Document Processing 10

2.2.3 Mathematical model 11

2.3 E-Recruitment Technical system through text mining processes 11

3 **DESIGN AND METHODOLOGY** 14

3.1 Process Model 14

3.2 Activity Diagram: 15

3.3 Use Case Diagram 16

3.4 K-Nearest Neighbour Algorithm 17

3.4.1 Sentence separation 17

3.4.2 Text Processing 17

3.4.3 Regular expression 18

3.4.4 Spacy Python Library 18

3.5 System Requirements 20

3.5.1 Software Requirements 20

3.5.2 User Requirement 20

3.6 Outcomes and Benefits of the Project 20

3.6.1 Outcomes 20

3.6.2 Benefits 21

4 **IMPLIMENTATION** 22

4.1 Multiple Files Upload 22

4.2 Word Extraction 22

4.3 K-Nearest Neighbours (KNN) 23

4.3.1 KNN Algorithm Working 23

4.4 Testing 32

4.4.1 Test Case 32

4.4.2 Test Case Template 32

4.4.2.1 Test Scenarios 33

5	**RESULTS AND DISCUSSIONS**		**34**
	5.1	Results	34
		5.1.1 Findings	34
		5.1.2 Comparison with prior studies	34
		5.1.3 Limitations of the resume sorting system	35
		5.1.4 Casual arguments	35
	5.2	Discussions	36
6	**CONCLUSION AND RECOMMENDATIONS**		**37**
	6.1	Conclusion	37
	6.2	Recommendations	38
REFERENCES			**39**
APPENDICES			**42**
	6.3	FRAMEWORK MODEL:	43