UNIVERSITI TEKNOLOGI MARA

CLASSIFICATION AND VISUALIZATION ON ELIGIBILITY RATE OF APPLICANT'S LINKEDIN ACCOUNT USING NAÏVE BAYES

NURUL ATIRAH BINTI AHMAD

BACHELOR OF COMPUTER SCIENCE (HONS.)

AUGUST 2023

ACKNOWLEDGEMENT

Alhamdulillah and all the praises to Allah S.W.T, I am extremely grateful for the blessings and opportunities bestowed upon me, allowing me to complete my thesis successfully. I recognize that it is through Allah's guidance and support that I have been able to reach this milestone. I would like to express my deepest gratitude to my supervisor, Ts. Dr. Khyrina Airin Fariza Binti Hj Abu Samah, whose invaluable guidance and unwavering support have been instrumental throughout this thesis journey. Her expertise and knowledge have greatly enriched my understanding of the subject matter. I am truly fortunate to have had such a dedicated and knowledgeable supervisor who consistently provided me with the necessary direction and encouragement.

I am also grateful to express my appreciation to Dr. Raihah Binti Aminuddin, my lecturer, for her kindness and support throughout the semester in teaching and assisting during the project. Furthermore, I would like to extend my gratitude to my examiner, Madam Nurazian Mior Dahalan, for examining my thesis and giving out her feedback and useful comments to improve my project's quality.

Lastly, I would like to express my heartfelt appreciation to my family for their unwavering love, encouragement, and patience throughout this journey. I would also like to thank my classmates and friends for their countless help in many ways.

ABSTRACT

The recruitment process is crucial in organizations, as it involves selecting qualified applicants from a diverse pool. In the digital era, social media platforms like LinkedIn have become famous for recruitment, and recruiters widely use them to find potential employees. Selecting suitable applicants is essential for an organization's success. However, recruiting unqualified applicants may affect the organizations. These can happen because the recruitment process requires much time, high costs, and possibly biased results when done manually. Thus, this project aims to classify and generate a list of potential job applicants by analyzing various attributes of their LinkedIn accounts, such as title, location, skills, education, language, certification, and experience. This project implements the Naive Bayes algorithm as the classification algorithm. The collected data from LinkedIn profiles then undergoes data preprocessing. It is set to label since it has no label class. The classification is set to two categories: Eligible or Ineligible. The dataset is split into training and testing sets to develop the Naive Bayes (NB) model. The NB model achieved a commendable accuracy of 89.8%, indicating good performance in classifying potential job applicants. The system's functionality based on the use case and usability tested by the Human Resources expert has been tested to evaluate its system requirements. The usability testing yields a score of 80% which indicates that the system is acceptable. The classification results are visualized, allowing users to identify eligible applicants efficiently. Thus, this project can help users find suitable applicants for the job. Future research could also focus on expanding the criteria considered for classification, such as cultural fit and behavioural attributes.

TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR APPROVAL	**i**
STUDENT DECLARATION	**ii**
ACKNOWLEDGEMENT	**iii**
ABSTRACT	**iv**
TABLE OF CONTENTS	**v**
LIST OF FIGURES	**ix**
LIST OF TABLES	**xiii**
LIST OF ABBREVIATIONS	**1**
CHAPTER ONE: INTRODUCTION	**2**
1.1 Background of Study	2
1.2 Problem Statement	3
1.3 Objectives	5
1.4 Scope	5
1.5 Significance of the Study	6
1.6 Summary	6
CHAPTER TWO: LITERATURE REVIEW	**7**
2.1 Overview of LinkedIn	8
2.2 Overview of Human Resources Recruitment	9
2.2.1 Fields of Job Offered	10
2.2.2 List Attributes in LinkedIn Profile	10
2.3 Web Scraping with Python	13
2.4 Keyword Extraction	14
2.4.1 Term Frequency and Inverse Document Frequency	14
2.4.2 Rapid Automatic Keyword Extraction Algorithm	15
2.4.3 Comparison of Word Retrieval Algorithms	15
2.5 Machine Learning Approaches	16
2.5.1 Supervised Learning	16
2.5.2 Unsupervised Learning	17
2.5.3 Comparison between Supervised Learning and Unsupervised Learning	17

2.6 Visualization Techniques 18
2.6.1 Line Chart 19
2.6.2 Bar Chart 19
2.6.3 Pie Chart 20
2.6.4 Scatter Plot 21
2.6.5 Heat Map 21
2.6.6 Word Clouds 22
2.6.7 Comparison Between the Types of Visualization Techniques 23
2.7 Development Approaches 24
2.7.1 Web Development Approach 24
2.7.2 Native Development Approach 25
2.7.3 Features Comparison Between the Web and Native Development Approaches 25
2.8 Related Work 26
2.8.1 Applicability of Naïve Bayes Model for Automatic Resume Classification 26
2.8.2 Resume Classification System using Natural Language Processing and Machine Learning 27
2.8.3 A Machine Learning Approach for Automation of Resume Recommendation System 28
2.8.4 Comparison between Three Related Works and This Project 29
2.9 Summary 31
CHAPTER THREE: METHODOLOGY 32
3.1 Introduction 32
3.2 Modified Waterfall Methodology 32
3.3 Requirement Phase 36
3.4 Design Phase 37
3.5 Implementation Phase 40
3.5.1 Software and Hardware Requirements 42
3.5.2 Data Visualization 43
3.6 Testing 43
3.6.1 Functionality Testing 43
3.6.2 Usability Testing 45
3.7 Deployment and Maintenance 46