Advanced Machine Learning Mini Project Report on: Automated Resume Screening and Skill Gap Analysis Using NLP

Amrutanshu Bhat

Anurag Kashyap

Ashwin Dhangar

Bammidi Pragati Rao

January 22, 2026

Abstract

The recruitment process often involves manual screening of a large number of resumes, making it time-consuming and prone to human bias. This project proposes an automated resume screening and skill gap analysis system using Natural Language Processing (NLP) techniques. Transformer-based models such as BERT and Sentence-BERT are employed to extract skills, generate semantic embeddings, and compute relevance scores between resumes and job descriptions. The system also identifies missing skills to provide actionable feedback to candidates. Experimental results demonstrate improved accuracy and efficiency compared to traditional keyword-based approaches, making the system suitable for real-world recruitment automation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Recruitment automation has become a critical requirement due to the exponential increase in job applications. Traditional resume screening methods rely on keyword matching, which fails to capture semantic relevance and contextual skill representation. This project addresses these limitations by leveraging NLP and deep learning techniques to automate resume screening and skill gap identification.

1.1 Problem Statement

Manual resume screening is inefficient, subjective, and not scalable. There is a need for an intelligent system that can automatically match resumes with job descriptions and identify missing skills.

1.2 Objectives

- Automate resume screening using NLP techniques

- Extract and normalize skills from resumes

- Compute semantic similarity between resumes and job descriptions

- Perform skill gap analysis

1.3 Scope

The system is applicable to IT and software engineering job roles and can be extended to other domains.

Chapter 2

State-of-the-Art Models

This chapter discusses existing state-of-the-art models used in resume screening and job matching, focusing on transformer-based architectures such as BERT and Sentence-BERT.

Chapter 3

Research Gaps and Motivation

This chapter identifies limitations in existing approaches and presents the motivation behind developing an automated semantic-based resume screening system.

Chapter 4

Dataset Analysis

This chapter describes the datasets used, their structure, preprocessing steps, and key statistical insights.

Chapter 5

Methodology and Approach

This chapter explains the overall system architecture, NLP pipeline, embedding genera-
tion, similarity computation, and skill gap analysis methodology.

Chapter 6

Experimental Results and Analysis

This chapter presents experimental results, performance evaluation, and comparative analysis with traditional keyword-based approaches.

Chapter 7

Conclusion

This chapter summarizes the contributions of the project and discusses future research directions.