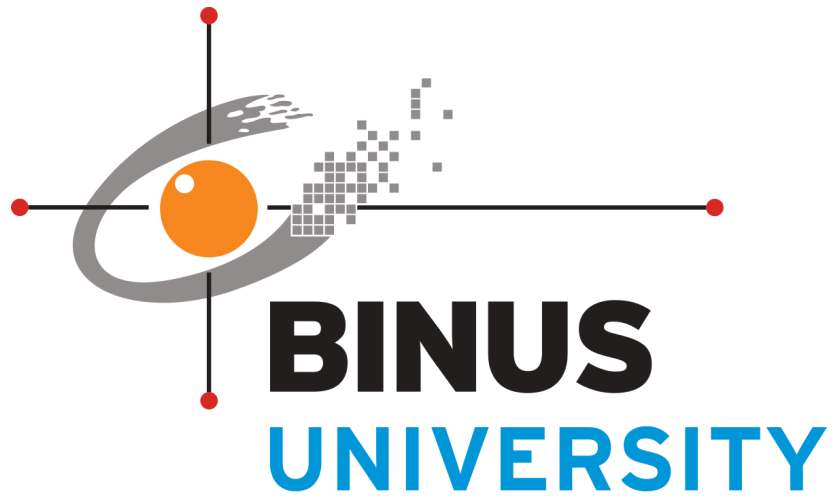


JOB-DESCRIPTION TO CANDIDATE SKILL MATCHER: ENHANCING RECRUITMENT EFFICIENCY THROUGH TEXT ANALYSIS

NATURAL LANGUAGE PROCESSING PROPOSAL

LB02



MADE BY:

GROUP 9

MEMBERS:

KEVIN ANTHONY LIEM 2702254096

STEVEN JONATAN 2702278240

WILBERT 2702231691

TABLE OF CONTENTS

TABLE OF CONTENTS.....	2
CHAPTER I: INTRODUCTION.....	3
CHAPTER II: METHODOLOGY.....	4
2.1 Methodology.....	4
2.2 Evaluation Metric.....	4
REFERENCES.....	5

CHAPTER I: INTRODUCTION

In today's labor market, both job seekers and employers struggle to effectively match job descriptions (JDs) to candidate profiles (resumes or CVs). One of the main challenges is the identification and mapping of skills. Job descriptions list required or preferred skills, while candidates often list their skills in varied formats. With the help of technologies such as natural language programming (NLP), the process can be automated and makes hiring faster, fairer, and help job seekers to better match their skills to job opening.

Previous studies have looked into different aspects of this problem, each addressing specific parts of the job matching process. One study introduced a model that uses document embeddings to extract implicit or hidden skills for job recommendations — skills that are not directly mentioned but can be inferred from the job descriptions [1]. Another study provided an overview of deep learning techniques for analyzing the job market, focusing on how these methods can be used to extract and classify skills from large collections of job postings [2]. In a more recent study, researchers have explored the potential of large language models (LLMs) to improve skill extraction in the job market domain, showing how these models can better understand complex relationships between job requirements and candidate skills [3].

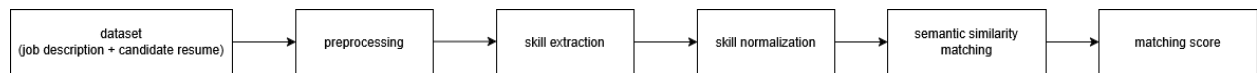
However, many of these studies mainly focus on extracting skills from job descriptions or resumes, instead of offering building a complete system that connects skills to match candidates and jobs. This project will introduce a system that extracts and normalizes skills from job descriptions and candidate profiles, computes a semantic similarity or matching score between a job description and candidate profile, and evaluates the matching performance. This approach aims to apply these techniques to a realistic recruitment scenario using transformer-based embeddings and localized datasets and contribute to automating job matching.

CHAPTER II: METHODOLOGY

Data:

We will collect a dataset of job postings and a dataset of candidate resumes from kaggle. For each job description we will extract the job title, job description text (responsibilities, requirements), and the listed skills. From each candidate profile/resume we will extract their listed skills, education, experience, and ideally a textual summary or job objective. The dataset will contain two excel files, where one will contain the job description and the other will be the candidate's resume.

Algorithm / Model:



1. **Preprocessing:** Clean and normalize text by tokenizing, lowercasing, removing stop words, and applying lemmatization or stemming.
2. **Skill Extraction and Normalization:** Use Named Entity Recognition (NER) or sequence classification models to identify and label skills in text. Extracted skills will then be mapped to a standardized taxonomy such as ESCO using embedding-based similarity.
3. **Candidate–Job Matching:** Compute similarity between candidate and job description using both explicit skill overlap and semantic similarity through sentence embeddings. A logistic regression or neural network may combine these features for final scoring.
4. **Evaluation Metrics:** For classification, we will use precision, recall, and F1-score. If modeled as a ranking problem, we will use Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (NDCG).

By combining explicit skill extraction with semantic similarity measures, we aim to go beyond simple keyword matching (which suffers from vocabulary mismatch) and capture deeper alignment between candidate and job description. Using a normalised skills vocabulary improves

interoperability and comparability. Using ranking metrics enables us to evaluate how well the system orders candidates for a job, which is closer to realistic recruitment settings.

REFERENCES

- [1] A. Gugnani and H. Misra, “Implicit Skills Extraction Using Document Embedding and Its Use in Job Recommendation,” *Proc. Thirty-Second Innovative Applications of Artificial Intelligence Conf. (IAAI-20)*, 2020.
- [2] E. Senger, M. Zhang, R. van der Goot, and B. Plank, “Deep Learning-based Computational Job Market Analysis: A Survey on Skill Extraction and Classification from Job Postings,” *Proc. 1st Workshop on NLP for Human Resources (NLP4HR 2024)*, pp. 1–15, 2024.
- [3] K. Nguyen, M. Zhang, S. Montariol, and A. Bosselut, “Rethinking Skill Extraction in the Job Market Domain using Large Language Models,” *Proc. 1st Workshop on NLP for Human Resources (NLP4HR 2024)*, pp. 27–42, 2024.
- [4] M. Zhang, K. N. Jensen, R. van der Goot, and B. Plank, “Skill Extraction from Job Postings using Weak Supervision,” *RecSys in HR '22: 2nd Workshop on Recommender Systems for Human Resources*, ACM Conf. on Recommender Systems, Seattle, USA, 2022.
- [5] L. Vásquez-Rodríguez, B. Audrin, S. Michel, S. Galli, J. Rogenhofer, and L. van der Plas, “Hardware-Effective Approaches for Skill Extraction in Job Offers and Resumes,” *RecSys in HR 2024*, 2024.