SRI LANKAN JOB SEEKER'S INTEREST AND JOB ADVERTISEMENT'S DESCRIPTION MATCHING USING MACHINE LEARNING TECHNIQUES

Flora Lakshinee Muthukumarasamy

A dissertation submitted in partial fulfilment of the requirement for Master of Science in Business Analytics

School of Computing

Informatics Institute of Technology, Sri Lanka

in collaboration with

Robert Gordon University, UK

2022

Abstract

Amidst the heavy competition in the employment market to secure a job that matches with one's qualifications, it is quite a task to seek one's dream job. Apart from the unavailability of vacancies, a jobseeker encounters certain other issues in finding his/her dream job. Technology-related problems occupy major space in present context since technological advancements have completely taken over the major elements of employment market: advertising and recruitment. Thus, if either the jobseeker is unaware of technological aspects involved in employment hunting or the technology itself causes glitches that will deprive individuals of the opportunity of securing their dream job.

This study attempts to devise a solution to a technology-based problem in seeking a better job. The problem exists in the job description of certain most-used job portals. It is a question that whether the description which is in image format would satisfy job-seeker's requirements. Thus there is an evident mismatch between these two aspects. The researcher aimed at minimizing this mismatch by developing a model using machine learning techniques. The model was designed such that it converts the image data to text and filters the first ten job descriptions that align with the interest of the jobseeker. Through Topic Modeling, researcher find the inherent grouping and the keyword list of the given job description. The algorithms NB, RFC, SGD, and LSVC provide greater accuracy than the cutoff values. So the job descriptions versus sector predictions are fed to the Stacking Ensemble Classifier to get the final prediction with 73% of accuracy. Use the Phase Matcher technique to filter the predicted sector jobs and match the keywords, which are from the topic modelling results. For a given JD, matching percentages of the keywords are returned eventually. This model is expected to assist jobseekers in finding a better job that matches with their qualifications that way they would not miss any opportunity of securing their dream job.

Keywords: Web Scraping, OCR, Text Preprocessing, Classification, Topic Modeling, BERT, Phase Matcher