

Resume Screening Using LSTM

Sakshi Navale¹, Samiksha Doke², Divya Mule³

1-3 Student, Department of Computer Engineering Jaihind College of Engineering, Kuran, Pune

Abstract -Resume Screening is process of determining whether a candidate is qualified for a role based his or her education, experience, and other information captured on their resume. It is a form of pattern matching and classification. Long Short-Term Memory (LSTM) networks are neural networks capable of learning long-term dependencies. Resume Screening using LSTM will take PDF file as input. Data will be extracted from the file. The data extracted will be saved in CSV file. The data from the csv file will then be cleaned and LSTM model will be used to make predictions. The LSTM will be trained and the saved to use in the web application. Applicants GitHub profile will also be retrieved. Web scraping will be used to find details of the GitHub profile. The analysis of the GitHub profile will affect result. User Details will be stored in the database. Data set will be cleaned and then vectorize. The vectorized dataset will then be used for training purpose. The model of bidirectional LSTM will be trained and weights will be saved for further use.

Key Words: Machine Learning, Supervised learning, Deep Learning, LSTM

1. INTRODUCTION

Resume screening is that the process of reviewing a resume to see if the candidate is qualified for the position. The choice of whether the candidate is qualified or not is predicted on education, experience, skills, and the other relevant information that will appear on the resume. Resume screening remains the foremost time consuming a part of recruiting screening resumes is estimated to require up to 23 hours for only one hire. When employment opening receives 250 resumes on the average and 75% to 88% of them are unqualified, it's no wonder the toughest a part of recruitment is screening the proper candidates from an outsized applicant pool.

Compounding the matter, a recent survey of talent acquisition leaders found that 56% will increase their hiring volume next year, but 66% of recruiting teams will either stay the identical size or shrink. In this paper, we purpose a system that will help to the recruiters as it will make sure to find best candidate in minimum time

2.LITERATURE SURVEY

Author of [1] discusses Web application for resume screening. Use of NLP pipeline. Text Extraction is done using sections-based segmentation. Semi supervised learning to train machine learning model. The system has some drawbacks. The web application shows results at the recruiter side in the form of ranking. One of the features can be described as

Resume is only matched to those job openings which they are interested in and have applied for.

The author of [2] makes use of OCR (Optical Character Recognition) Feature Extraction of Principal Component Analysis. Decision tree classification algorithm is used. But the system works only on Urdu Text. Paper [3] mentions that Conventional SVM can be optimized for text classification. The conventional SVM is optimized by selecting features using the entropy. The author of [4] compares three classification algorithms. Results show that support vector classifiers with the TFIDF feature shows more accuracy than naive bayes and KNN.

The author of [5] presents the web document classification based on fuzzy k-NN network, in the process of classification, TF/IDF is adopted for selecting features of document. The results show that classification performance is better than k-NN and SVM, but the speed of classification is bit slow than KNN.

3. PROPOSED SYSTEM

Resume Screening using LSTM will take PDF file as input. Data will be extracted from the file. The data extracted will be saved in CSV file. The data from the csv file will then be cleaned and LSTM model will be used to make predictions. The LSTM will be trained and the saved to use in the web application. Applicants GitHub profile will also be retrieved. Web scraping will be used to find details of the GitHub profile. The analysis of the GitHub profile will affect result. User Details will be stored in the database. Data set will be cleaned and then vectorize. The vectorized dataset will then be used for training purpose.

The model of bidirectional LSTM will be trained and weights will be saved for further use. The system will be implemented using VScode and Django framework. Python Programming language will be used for implementation. The system architecture of proposed system is shown in Fig-1. In proposed system there are two types of users.



- [1] Sujit Amin, Nikita Jayakar, Sonia Sunny, Pheba Babu, M. Kiruthika, Ambarish Gurjar, Web Application for Screening Resume, 2019 International Conference on Nascent Technologies in Engineering (ICNTE 2019), 978-1-5386-9166-3/19/\$31.00 ©2019 IEEE.
- [2] K. Khan, R. Ullah Khan, Ali Alkhalifah, N. Ahmad, Urdu Text Classification using Decision Tree, 978-1-4673-9268-6/15/\$31.00 ©2015 IEEE.
- [3] Zi-qiang wang, xia sun, de-xian zhang, xin li, an optimal svm-based text classification algorithm, Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian, 13-16 August 2006 1-4244-0060-0/06/\$20.00 ©2006 IEEE
- [4] Fang Miao, Pu Zhang, Libiao Jin*, Hongda Wu, Chinese News Text Classification Based on Machine learning algorithm, 2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetic.
- [5] Huabe iNie, Yi Niu, Juan Zhang, Web Document Classification Based on Fuzzy KNN Algorithm, 2009 International Conference on computational intelligence and Security 978-0-7695-3931-7/09 \$26.00 © 2009 IEEE

