

Analysis of Data Analyst Jobs

A system for predicting the best salary for the candidate

Afzal Mukhtar
Computer Science and Engineering
PES UNIVERSITY
Bengaluru, India
afzalmukhtar985@gmail.com

Aswin A Nair
Computer Science and Engineering
PES UNIVERSITY
Bengaluru, India
aswinanilnair@gmail.com

Hritika Rahul Mehta
Computer Science and Engineering
PES UNIVERSITY
Bengaluru, India
hritika2708@gmail.com

Abstract—The e-recruitment platform is growing and becoming an important part of our lives. Such platforms still have an issue on basing their choice of inputs for recruitment. But overall these systems are aimed at increasing accuracy and reducing the task of searching through a large number of candidates or jobs offered. Many have lost their jobs or unable to find jobs in the COVID-19 pandemic. Although there are numerous job recommender systems, here we focus on recommending jobs to candidates based on their skills. Usually, jobs are based on experience and skills are ignored, therefore we put a little focus on the skills for jobs as this pandemic can reduce the effective work experience a candidate can have.

Keywords—Salary Prediction, DataAnalyst Job Market, Machine Learning.

I. INTRODUCTION

Salary prediction provides an important insight to an aspiring job candidate about what the most in-demand skills are given a job title and even the most common, high-paying or most in-demand skills of the chosen job title. For an organization looking to hire fresh talent, a recommender system can provide a suggestion of the salary which can be offered to an individual given their set of skills. Thus a salary predictor serves a dual purpose.

One of the most important things in a job search is knowing where you are right now and where you want to go next. The mindset of a job seeker is very much like a product manager, trying to find the best product and market fit. Smart product managers know what their customers want. So they carefully find out customers' needs which are expressed through "top qualifications", "what you must have", in the job descriptions. After identifying what is commonly needed in the market, smart product managers will customize their products aka resumes, cover letters, online profiles, portfolios to demonstrate these characteristics and traits as much as possible.

Accurate recruitment of employees is a key element in the business strategy of every company due to its impact on companies' productivity and competitiveness. At present recruitment processes have evolved into complex tasks involving rigorous evaluations and interviews of candidates, intending to hire the best-suited professionals for each company's needs. With the advent of the Internet and the web, e-Recruitment has become an essential element of all hiring strategies. Many websites, such as CareerBuilder, Monster or Tecnoempleo, or even Social Networks, like LinkedIn, help companies and job seekers to find the best possible matches.

This paper documents our approach to building a recommender system to suggest appropriate jobs to individuals based on their skills. The dataset chosen for this

project is "[Data Analyst Jobs](#)" from Kaggle which has job listings from companies scraped from the Glassdoor website. We first go about a survey of various papers in the domain to get some knowledge on the various techniques and models used for the same. Further, we put into action these techniques to build a fully functional recommender system for the above-stated dataset.

II. LITERATURE REVIEW

A. Summary of Applying Data Mining Techniques in Job Recommender System for Considering Candidate Job Preferences [1]

By: Gupta, Anika and Dr. Garg, Deepak

Recommender systems are an important aspect of e-commerce. The recommendation system worked on in this paper is about recommending the best jobs based on salary and candidate's skills. Candidates can have similar looking profiles, but their choices may be different based on what position to choose, and what can they trade-off with – distance, salary, company rating. This recommender system takes Candidate information and applies classification rules along with their job preferences to produce a recommended jobs list.

The authors have assumed that all the text categorization is in place and discrete with well-labelled values, understandable by the system, and that the candidate or the system automatically updates the information dynamically. They have categorized the company into groups instead of using continuous numerical rating. Decision Trees are used for determining the criteria for attribute splitting. A threshold taken for the sample size determined if the value was selected or not, and that was used to fill the job category with 1 or 0, depending on if the rule exists or not.

It is assumed that all the text categorization is already in place and we have discrete and well-labelled values that are easily understandable by the system. It is assumed that either the candidate or the system itself updates the age, experience, education and skills field. It uses TF-IDF value for feature extraction and information gain with a threshold value for feature addition. Location preference matrix is not considered as it is assumed that the group has applied equally in all 4 regions and hence adding these will not result in any new information.

The way the data was categorized and the algorithms used will be useful in our project. This paper recommended jobs and categorize continuous variables into discrete categories which can reduce the number of values by a large extent and thus improve model performance in speed and accuracy.

The use of Decision Trees can be used as a classifier on certain aspects of the problem we are trying to solve. The categorization of companies based on their requirement can help us in categorizing the prediction of salary.

B. A survey of job recommender systems. [2]

By: Shaha T. Al-Otaibi and Mourad Ykhlef

This paper talks about how recommendation systems influence the field of job recruitment. The main focus is on how to implement recommendation systems for identifying suitable employees for a particular job description. The author presents an overview of the various recommendation systems that are in use, in the current day such as Collaborative Filtering approach, Content-based filtering approach and Knowledge-based approach, he also goes on to list various challenges associated with them such as performance decrement due to increase in data, limited scalability, a requirement for prior domain knowledge etc.

The paper also has attached a case study to recommend candidates for a specific job, which was implemented using a content-based filtering approach as it was assumed to be the best approach. However, this led to the matching of candidates that had specific terms related to the job in their CV but did not take into account their level of expertise in those areas.

The Hybrid approach as when used for job recommendation case study doesn't seem to improve the quality of job recommendation.

Content-based filtering method doesn't seem to distinguish properly among jobs that have already been recommended to the user and requires a sufficient amount of learning period before understanding the user's preferences

With our topic being the analysis of Data Analyst jobs and the paper's prime discussion is on the various methods used for job recommender systems, this paper has given us an insight into how we can use the similar approach for our project.

C. Salary Prediction in the IT Job Market with Few High-Dimensional Samples: A Spanish Case Study [3]

By: Ignacio Martín , Andrea Mauriello , Roberto Battiti , Jose' Alberto Hernandez'

This paper focuses on predicting the salaries of jobs offered on an online portal called "Tecnoempleo" that specialises in

IT jobs for people residing in Spain. E-recruitment of candidates is a hot topic of research and there are many different approaches like an automatic evaluation of CV, a ranking of the skills of candidates which are used by companies to aid in the choice of a candidate.

Salary prediction by ranges is the main focus of this paper, several models like – linear models, logistic regression, K-nearest neighbours, multi-layer perceptrons, support vector machines, random forest, adaptive boosting with decision trees along with ensembles of the models stated before are considered to formulate a solution.

In this work data is collected from an e-Recruitment website with geographical limits to Spain, the posts for jobs not based in Spain were removed. Here the missing values were removed because in most cases there was no default value with which they could be substituted. The information which is unstructured like the title of the job post, description of the offered post and requested profile were also removed. These comprise some of the assumptions of the paper considered for this report.

We need to employ similar feature engineering to dataset concerning the Salary estimate and Size columns by converting them into a numeric form. It is very clearly concluded about how salary classification into a range makes job profiling easier and hence a website easier to use for a person hence we could also attempt a classification approach to our problem.

III. REFERENCES

- [1] A. Gupta and D. Dr. Garg, "Applying Data Mining Anika," 2014. [Online]. Available: https://gdeepak.com/pubs/Applying_Data_Mining_Anika.pdf.
- [2] S. T. Al-Otaibi and M. Ykhlef, "A survey of job recommender systems," 26 July 2012. [Online]. Available: <https://academicjournals.org/journal/IJPS/article-full-text-pdf/B19DCA416592.pdf>.
- [3] I. Martín, A. Mariello, R. Battiti and J. A. Hernández, "Atlantis Press: International Journal of Computational Intelligence Systems," 2018. [Online]. Available: <https://www.atlantis-press.com/journals/ijcis/25899235/view>.