Optimizing Matches between Job postings and Resumes

AZZIMANI Yassine
v24yasaz@du.se

***Abstract* - Nowadays, with a plethora of job search websites available, standing out is increasingly challenging, especially when vying for positions in large companies. In fact, some job postings attract tens of thousands of candidates, particularly those at GAFAM (Google, Amazon, Facebook, Apple, and Microsoft). Consequently, companies are increasingly relying on AI (Artificial Intelligence) and other algorithms to initially screen resumes. Analyzing job descriptions aids applicants in aligning with the desired profile, while scrutinizing resumes helps distinguish them from other candidates.**

***Keywords* – NLP, Sentimental Analysis, Human Resources**

INTRODUCTION

Currently, finding the best way to apply for job openings is not an easy task, especially for young graduates lacking work experience but armed with various skills acquired from university. Knowing which skills to highlight can mean the difference between matching with a job offer or missing out on it. In this scientific article, we aim to address how individuals can optimize their profiles to effectively align with the specific criteria and expectations outlined in job descriptions or research study requirements. To answer this question, we will utilize two datasets: one containing job offers and another containing resumes. Our study will focus on the engineering domain for precision. The data is sourced from Kaggle, an online platform where datasets are contributed by individuals, although some may be of poor quality. Therefore, a significant portion of our work will involve data cleaning, followed by the application of NLP (Natural Language Processing) algorithms such as sentiment analysis, data extraction, and data classification. In conclusion, we will provide key points to enhance the attractiveness of resumes and maximize their alignment with job offers in the engineering field.

LITERATURE REVIEW

After reading several articles about how artificial intelligence is shaping the field of Human Resources, particularly in the hiring process, one term that frequently appears is "Screening." This refers to the process of assessing and pre-selecting candidates based on specific criteria. Nowadays, screening is predominantly conducted by AI rather than humans, thanks to the effectiveness of Natural Language Processing (NLP). Many companies have developed screening tools utilizing NLP, such as Perfect.com, MyInterview.com, and talenteria.com, among others. The advantages of using these tools are significant, including time savings, shorter hiring cycles, the ability to select qualified candidates, and verifying information by cross-referencing data available on the internet.

This initial understanding reveals that many companies are adopting AI screening, and if a

company hasn't yet, it's likely only a matter of time before they do, given the numerous benefits of automated processes. Now, after some technical review, we can observe how algorithms identify words or similar words to a specific one. Let's take the example of 'Artificial Intelligence.' Using a classifier like BERT (Bidirectional Encoder Representations from Transformers), we generate a word cloud containing words like 'Python,' 'NLP,' or 'Machine Learning,' among others. Each word is assigned a score based on its semantic meaning compared to the original word.

To leverage this information in our study, we need to compile a list of skills required in the engineering domain and identify the closest semantic matches within our competency pool to include on our resumes.

METHODOLOGY

Based on employment objectives highlighted in this study, the target area is how to make the resume attractive as possible according to the many needs in the engineering domain. As a result, it will make it easier for applicants to make a resume based on the current job market.

Data description and cleaning

In our study, we will utilize a dataset sourced from the internet, available on Kaggle. The Job Offers dataset primarily focuses on four main domains: Artificial Intelligence, Machine Learning, Data Science, and Big Data Jobs. This dataset comprises 3200 rows in CSV format. Similarly, the Resume dataset covers the same domains as the Job Offers dataset, specifically IT Resumes, with 224 Microsoft word files. Both datasets are obtained from real data scraped from the internet, providing valuable insights for our study.

Feature selection

For this study, we will take the entire dataset of job offerings and clean it as needed. For instance, it's common for some rows to have missing information, such as salary or experience level. This is often because companies may not disclose salary details before an interview with an applicant, especially in Europe. After cleaning the data and handling missing values, we have created a correlation matrix between the columns.

Data mining

In our case, we conducted a small part of data mining, focusing on data cleaning and filling in missing values. However, the dataset was of particularly good quality and did not necessarily require extensive mining.

RESULT AND ANALYSIS

Job offers dataset analysis

Geographical influence

To begin the analysis, we will examine the distribution of the job offers dataset and identify the regions of the world it covers. This will ensure our study is as precise as possible.

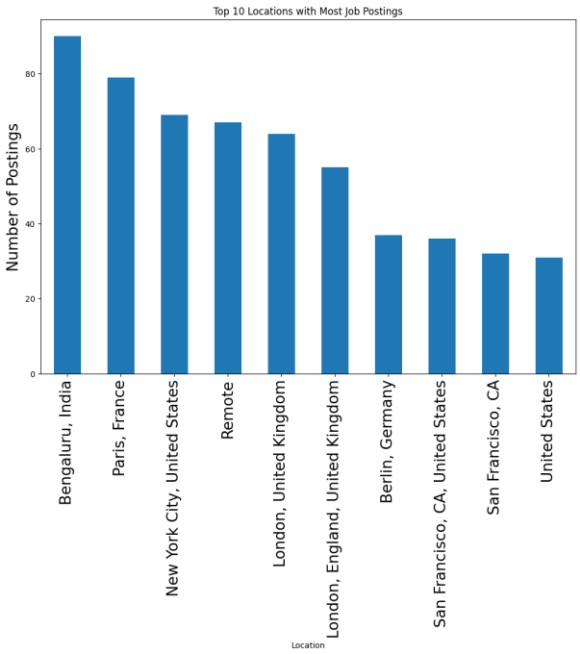

Figure 1:Top 10 Locations with Most Job Postings

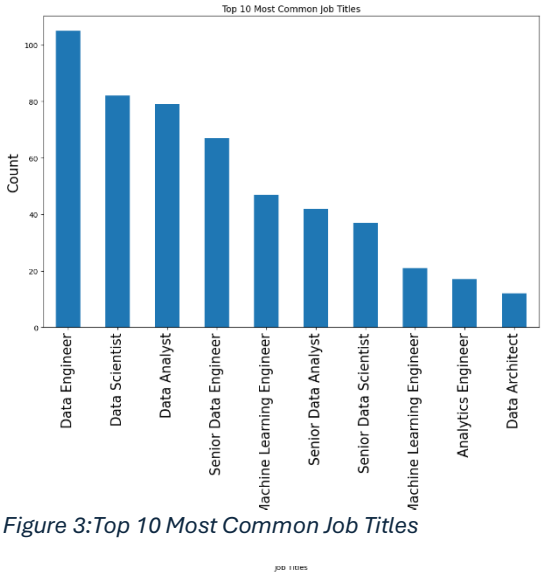

Figure 3:Top 10 Most Common Job Titles

Here, with this plot, we can see how well the dataset represents the global job market. Unlike many other datasets I have tried before, which often focus on a specific country (most commonly India or the USA), this dataset has a more balanced distribution across different regions. Additionally, the inclusion of 'Remote' jobs highlights the significant impact of the COVID-19 pandemic on the work world. Before the pandemic, remote jobs made up 15% of the market; today, they account for 45%. This change helps us understand the current job market and where to apply for a better chance of being recruited. Now, let's focus on the types of jobs that are most prevalent in our dataset.

Trendy job title

Today, almost every device has an internet connection and generates a large amount of data. With the significant advancements in artificial intelligence, we can now analyze this vast amount of data and use it beneficially. This is why data-related jobs are trending, as illustrated by the word cloud.

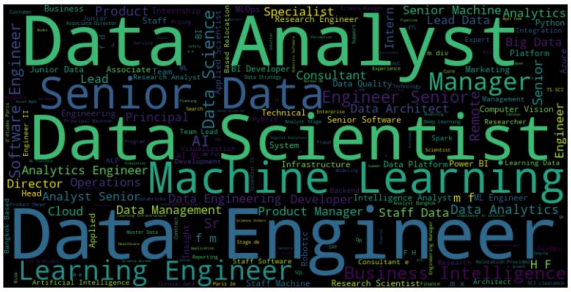

Figure 2: Word Cloud of Job Titles

Here in the figure, we can see the three major job titles in our database:

- Data Analyst
- Data Scientist
- Data Engineer

When we look deeper into the ranking of common jobs, this is what we observe.

This ranking clearly shows the significant presence of data-related jobs in the world of work, with these roles appearing prominently in the top rankings. These types of jobs are currently in high demand and are expected to see constant growth due to upcoming innovations, such as:

- Autonomous vehicles
- Assisted medical diagnostics.

Furthermore, when we study the correlation between salary and job title, we find the following plot.

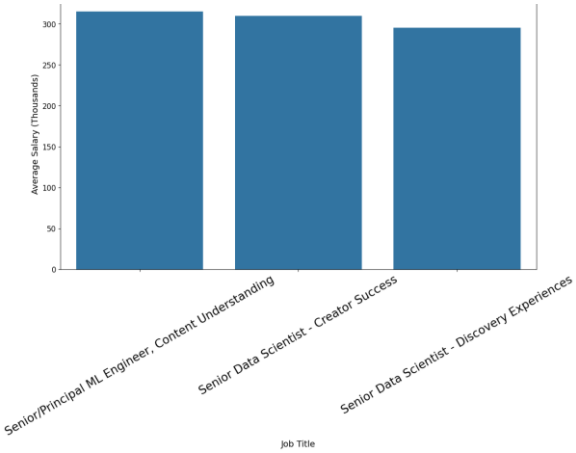

Figure 4:Top 3 Highest Paying Job Titles and Their Salaries

This shows that the demand for jobs in the data field is significant. Studying data-related fields, from Data Science to Business Intelligence, can be very beneficial for finding a job and earning a good salary. Now that we understand which domain is dominant in the job market, we need to identify the tools we need to master to have a chance to enter these types of jobs.

Skills required

For the skills section, we will highlight which skills appear most frequently in the job offers. In the next plot, we present the 20 most found skills in the job listings.

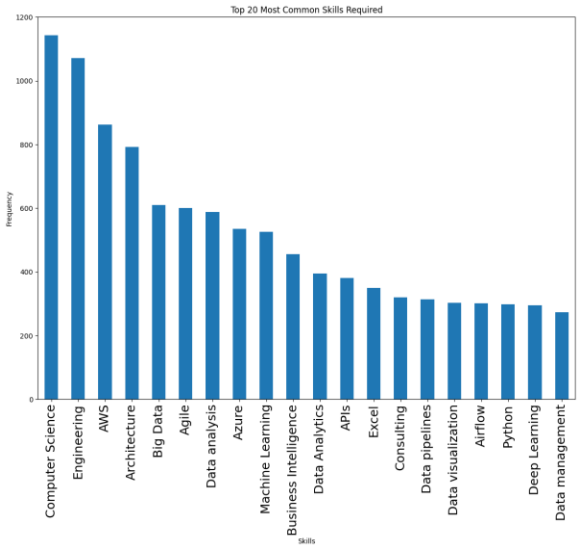

Figure 5: Top 20 Most Common Skills Required

In the plot, one thing is clear: in a dataset focused exclusively on IT jobs, only one programming language, Python, prominently appears. It is well known that Python is the most widely used language in companies due to its versatility across various domains, from web development to artificial intelligence. Therefore, listing numerous programming languages may not be the most effective strategy for finding a job.

As expected, data-related tools are trending in our rankings, and we also see the presence of some soft skills. This implies that having a varied skill set on our resume is essential.

Experience required

In this section, we will focus on a key aspect that we cannot enhance as quickly as we would like, yet it remains a crucial factor in the resume selection process. As the following chart shows, the number of jobs requiring either no experience or a significant level of experience is extremely large.

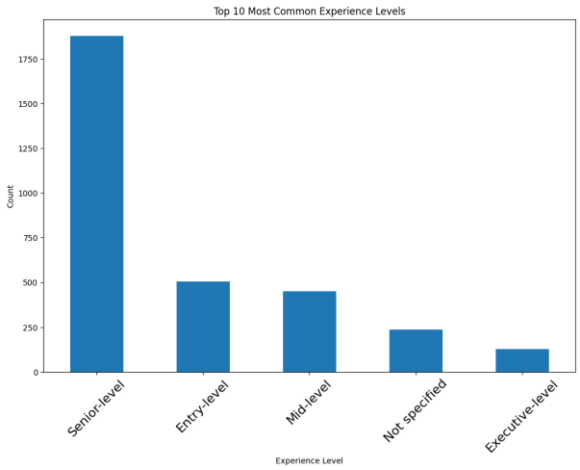

Figure 6:Top 10 Most Common Experience Levels

The combined total of entry-level and mid-level positions is insufficient to meet the demand for senior roles. This issue is well-documented in numerous articles discussing the challenges of recruiting exclusively for senior positions.

Every company has a different excuse:

- **In small startups** - "we are a very small team and we don't have time to mentor juniors, we need engineers who will be very productive from day 1"
- **In medium-sized companies** - "we are going to grow very fast, we need engine who can handle scale and faced such challenges before"
- **In big companies** - "our infrastructure is super complex, it'll take juniors too lo to ramp up".

Figure 7: Extract from https://zaidesanton.substack.com/p/hiring-only-seniors-is-worst-policy

This issue can cause considerable anxiety, especially among young workers who have recently graduated and naturally lack experience. The only upside to this drawback is that we can expect to find jobs more easily as we gain experience and, consequently, earn more money. Indeed, the following chart illustrates how salaries correlate with the expected level of experience.

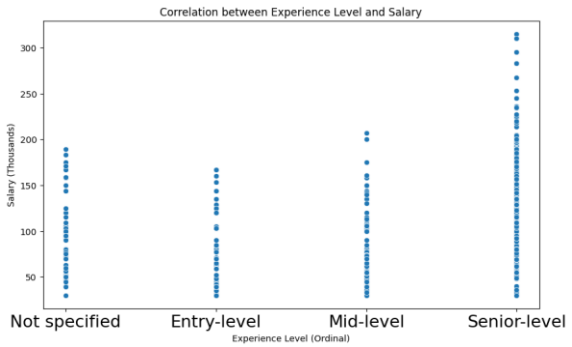

Figure 8: Correlation between Experience Level and Salary

As the chart illustrates, there is a strong relationship between the required experience and salary. The next chart will further demonstrate the positive correlation between these two factors.

Figure 9:Correlation between Experience (Years) and Salary

In this chart, I have omitted entries labeled as 'not specified' for the experience requirement to avoid potential biases that could skew the analysis.

Resume dataset analysis

In the second part of our analysis, we will examine a dataset of resumes. Studying resumes from the same field as our search can help us create a distinctive resume, thereby increasing our chances of securing a job. This section will be shorter than the first, as it deals with information that is less critical than that derived from the initial dataset. First, however, we need to convert an entire folder containing 224

resumes from Word to PDF format using a straightforward algorithm.

Education level

The analysis could begin with a quick assessment of the average educational level.

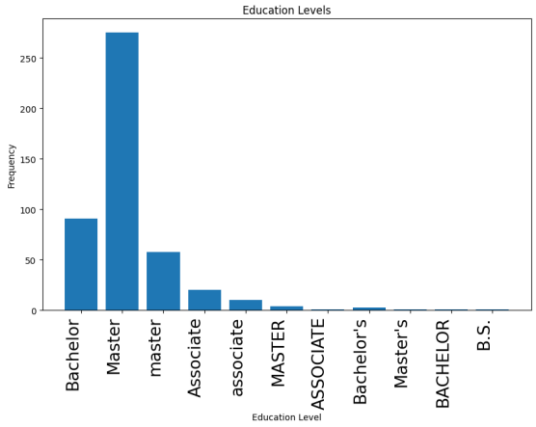

Figure 10:Education Levels

From this initial analysis, we can easily see that having a master's degree places us among a large group of candidates. Possessing a PhD, however, could significantly increase our chances of standing out.

Stand out through skills

Now, let’s examine which skills people typically highlight on their resumes.

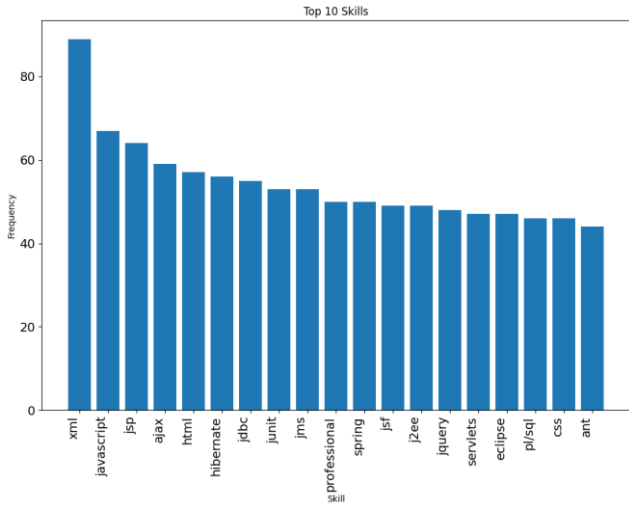

Figure 11: Top 20 Skills

In this chart, we can clearly see that the skills are predominantly oriented towards web development rather than engineering. This indicates a potential lack of resumes showcasing a broader range of skills. However, it is evident that adding soft skills could help match more job offers and increase the chances of being noticed.

CONCLUSION

To conclude, to create a contemporary CV that aligns with current job market demands, we should focus on emphasizing data-related skills when possible. Rather than solely highlighting coding abilities, it's better to showcase analytical and soft skills. For recent graduates, perseverance is key if a job isn't found quickly. With time and experience, the situation is likely to improve.

REFERENCES

https://vervoe.com/ai-in-resume-screening/

https://ideal.com/ebook-resume-screening-using-artificial-intelligence/

https://www.kaggle.com/datasets/joyshil0599/data-science-jobs-comprehensive-dataset

https://www.kaggle.com/datasets/palaksood97/resume-dataset/data

https://www.statista.com/statistics/1122987/change-in-remote-work-trends-after-covid-in-usa/

https://medium.com/@lunadoan/data-job-market-2024-insights-you-need-to-boost-your-career-d05c7e18a5c1

https://zaidesanton.substack.com/p/hiring-only-seniors-is-worst-policy