THE UNIVERSITY OF CHICAGO

Adapting to a New Normal: The Impact of COVID-19 on Hiring Patterns and the Development of Predictive Models for Talent Acquisition

By

Zhe Zhang

June 2023

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

Faculty Advisor: Huizhao Dai

Preceptor: Jean Clipperton

Adapting to a New Normal: The Impact of COVID-19 on Hiring Patterns and the Development of Predictive Models for Talent Acquisition

Introduction:

The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has had a profound and far-reaching impact on the global economy since its emergence in late 2019 (World Health Organization [WHO], 2020). With millions of people affected worldwide, businesses have had to rapidly adapt to new circumstances, leading to significant changes in the labor market (Barrero, Bloom, & Davis, 2020). As a result, hiring patterns in companies have evolved, presenting new challenges and opportunities for organizations (Bentley, Green, & Shipilov, 2020). Given the ongoing uncertainty and potential for further disruptions, there is a growing need for a predictor that can help companies cope with these new hiring patterns and make informed decisions about their talent acquisition strategies.

The aim of this paper is to examine the change in hiring patterns in companies due to COVID-19 and explore the potential for developing a predictor that can assist organizations in adapting their recruitment processes to this new reality. To this end, the review will first discuss the impact of COVID-19 on the labor market, highlighting shifts in demand for certain industries and occupations, as well as changes in workforce expectations and preferences. Next, the review will analyze the evolution of hiring practices in response to the pandemic, focusing on adaptations in recruitment processes, changes in desired skill sets and qualifications, and the diversification of workforce and inclusive hiring. The review will then explore existing predictive models and approaches in hiring, identifying their limitations and opportunities for improvement. Finally, the review will discuss potential strategies for developing a predictor that can cope with new hiring patterns, addressing key variables and factors, as well as potential challenges and limitations.

Understanding the new hiring patterns in the context of COVID-19 is essential for organizations to remain competitive and resilient in an increasingly uncertain and dynamic environment. This literature review seeks to contribute to this understanding and provide a foundation for future research and development of predictive models that can better serve organizations in the post-pandemic world.

The Impact of COVID-19 on the Labor Market:

The COVID-19 pandemic has had a significant impact on labor markets worldwide, leading to widespread job losses, shifts in demand for certain industries and occupations, and changes in workforce expectations and preferences (Bick, Blandin, & Mertens, 2020; Forsythe et al., 2020).

Unemployment rates have increased dramatically across the globe due to lockdowns, social distancing measures, and the closure of non-essential businesses (International Labour Organization [ILO], 2020). Many workers, particularly those in low-income and vulnerable groups, have experienced job losses, reduced working hours, or furloughs (Adams-Prassl, Boneva, Golin, & Rauh, 2020). Although some countries have implemented job retention schemes and economic stimulus packages to mitigate the impact, the long-term effects of these measures remain uncertain (Gourinchas, 2020).

Simultaneously, the pandemic has led to shifts in demand for specific industries and occupations. Industries like travel, hospitality, and retail have experienced significant declines, while others, such as e-commerce, healthcare, and technology, have seen increased demand (Beland, Brodeur, & Wright, 2020). This has resulted in a reallocation of labor, with workers transitioning to growing sectors, often requiring new skills and training (Barrero et al., 2020).

Moreover, COVID-19 has considerably influenced workforce expectations and preferences. Remote work has become a new norm for many employees and organizations (Brynjolfsson et al., 2020). Workers increasingly expect flexibility in their work arrangements, including the ability to work from home, flexible hours, and a better work-life balance (PwC, 2021). This shift has implications for companies' recruitment strategies and the attractiveness of their job offers (Bentley et al., 2020).

In conclusion, the COVID-19 pandemic has profoundly impacted labor markets, resulting in increased unemployment rates, shifts in demand for certain industries and occupations, and changes in workforce expectations and preferences. These effects have significant implications for companies' hiring patterns and the development of a predictor to cope with the new hiring landscape.

The Evolution of Hiring Practices in Response to COVID-19:

In response to the pandemic, companies have been forced to adapt their hiring practices to meet the changing labor market dynamics and evolving workforce expectations. These adaptations include modifications in recruitment processes, changes in desired skill sets and qualifications, and increased focus on diversifying the workforce and promoting inclusive hiring.

The recruitment process has undergone significant transformations, with the traditional in-person interviews and onboarding being replaced by virtual alternatives (Gartner, 2020). Companies have increasingly relied on digital tools, such as video conferencing platforms, online assessments, and virtual job fairs, to connect with potential candidates (Chakraborty, 2020). Additionally, the use of Artificial Intelligence (AI) and data analytics has become more prevalent in the screening and selection of candidates, helping organizations streamline their recruitment processes and make data-driven decisions (Bartel, Beaulieu, Pham, & Yang, 2021).

The pandemic has also changed the desired skill sets and qualifications for many positions. There has been a greater emphasis on digital and remote work skills, such as proficiency in digital collaboration tools, online communication, and cybersecurity (World Economic Forum [WEF], 2020). Employers have also increasingly valued soft skills, such as adaptability, resilience, and emotional intelligence, as these traits are critical in navigating the uncertainties and challenges posed by the pandemic (Bughin et al., 2020).

Moreover, the shift towards remote work and the need to address social inequalities exacerbated by the pandemic have led to a renewed focus on diversifying the workforce and promoting inclusive hiring practices (Bentley et al., 2020). Companies have started to expand their talent pools beyond geographical boundaries, enabling them to access a more diverse range of candidates (Gartner, 2020). The focus on diversity, equity, and inclusion (DEI) has become more prominent, as organizations recognize the benefits of a diverse workforce in fostering innovation, improving decision-making, and promoting resilience (Hunt, Yee, Prince, & Dixon-Fyle, 2018).

In conclusion, the COVID-19 pandemic has driven significant changes in hiring practices, as companies adapt their recruitment processes, prioritize new skill sets and qualifications, and emphasize workforce diversification and inclusive hiring.

Understanding these evolving practices is crucial in developing a predictor that can effectively cope with the new hiring landscape.

Developing a Predictor to Cope with New Hiring Patterns

In light of the changes in hiring patterns due to the COVID-19 pandemic, developing a predictor that can effectively identify and adapt to these new trends is crucial for companies to maintain competitiveness in the labor market. The development of such a predictor involves incorporating data from various sources, utilizing advanced analytics and machine learning techniques, and ensuring the model is adaptable and inclusive.

The first step in developing the predictor is to gather data from various sources, including job boards, social media, company websites, and industry reports, to provide a comprehensive understanding of the current hiring landscape (Ramos, 2021). By analyzing trends in job postings, required qualifications, and desired skills, the predictor can identify patterns and shifts in the labor market.

To increase the accuracy and efficiency of the predictor, advanced analytics and machine learning techniques, such as natural language processing (NLP) and deep learning, can be employed (Breiman, 2001; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). These techniques enable the model to process large amounts of unstructured data, identify patterns, and make predictions based on historical trends and real-time information.

Moreover, it is essential to ensure that the predictor is adaptable to the rapidly evolving hiring landscape, accounting for changes in industry demand, workforce expectations, and technological advancements (O'Driscoll, 2020). Regular updates and fine-tuning of the model can help maintain its accuracy and relevance over time.

Finally, it is crucial to incorporate diversity, equity, and inclusion (DEI) considerations into the predictor to ensure that it promotes fair and unbiased hiring practices (Bogen, Rieke, & Ahmed, 2018). The model should be designed to mitigate biases in data and algorithmic decision-making, promoting a more diverse and inclusive workforce.

In conclusion, developing a predictor to cope with the new hiring patterns resulting from the COVID-19 pandemic requires a multi-faceted approach that incorporates data from various sources, employs advanced analytics and machine learning techniques, and ensures adaptability and inclusiveness. By implementing such a predictor, companies can effectively navigate the changing labor market and make informed hiring decisions.

Conclusion:

The COVID-19 pandemic has had far-reaching effects on the labor market and hiring practices, forcing companies to adapt and evolve in response to the changing landscape. This literature review has provided an overview of the impact of COVID-19 on the labor market (Adams-Prassl et al., 2020; Beland et al., 2020), the subsequent evolution of hiring practices (Bughin et al., 2020; Gartner, 2020), and the development of a predictor to cope with the new hiring patterns (O'Driscoll, 2020; Ramos, 2021).

The pandemic has disrupted the global economy, leading to job losses, increased remote work, and changes in workforce expectations (Bick et al., 2020; International Labour Organization, 2020). In response, companies have shifted their recruitment processes, prioritized new skill sets and qualifications, and emphasized workforce diversification and inclusive hiring (Bartel et al., 2021; World Economic Forum, 2020). To adapt to these changes, a multi-faceted approach is required, incorporating data from various sources, advanced analytics and machine learning techniques, and ensuring adaptability and inclusiveness in the predictor model (Breiman, 2001; Bogen et al., 2018).

By understanding the new hiring patterns and developing a predictor that can effectively identify and adapt to these trends, companies can make informed hiring decisions, maintain competitiveness in the labor market, and foster a diverse and inclusive workforce in the post-pandemic world.

Evaluation of research

Introduction

Predicting the outcome of job interviews is a critical area of interest for job seekers, employers, and human resource professionals, as it can help optimize interview preparation, streamline hiring processes, and improve overall hiring success rates. Various factors can influence interview outcomes, and understanding their interplay can lead to more effective talent acquisition strategies (Rivera, 2012). This study aims to build a machine learning model that predicts interview outcomes based on a dataset containing various features such as interview time, job title, location, experience, interview difficulty, and interview questions.

Machine learning has been increasingly employed in the human resource domain to solve various problems, including applicant tracking, employee engagement analysis, and retention prediction (Delen, 2010; Kapoor & Dwivedi, 2020). By evaluating the research design of our study, we aim to understand the effectiveness of the methodology, data preprocessing, feature extraction, model selection, and evaluation techniques employed. The results of this research could provide valuable insights for job seekers, employers, and human resource professionals, enabling them to optimize their interview processes and improve overall hiring success rates.

In this evaluation, we will examine each aspect of the research design in detail, assess the strengths and weaknesses of the approach, and provide suggestions for potential improvements and future research opportunities. Ultimately, the aim is to determine the effectiveness of the research design in addressing the research objective of predicting interview outcomes and its implications for the broader field of human resources and talent acquisition.

Data Collection and Preprocessing

a. Describe the data collection process and sources

Data source: glassdoor

To answer this research question, I wrote a data scraper of glassdoor interview questions. In total, there were 159,668 interview questions and interview records in the data set I collected.

Company selection:

In my paper, I analyzed 26 companies from the S&P 500 index, which were selected based on their industry category and market capitalization. These companies were divided into three distinct categories: (1) Communication Services & High-Tech Industrials, (2) Consumer Discretionary & Traditional Industrials, and (3) Financials & Services. The first and second categories each comprise eight companies, while the third category contains 10 companies.

I conducted my analysis from two perspectives: (a) comparing the performance of companies across different categories, and (b) examining the performance of companies within the same category.

b. Explain data cleaning steps, such as handling missing values

1. Data Cleaning and Transformation:

(a). Ensure that each column (Time, Job title, Location, Offer, Experience, Interview Difficulty, In person or not, and Interview) has a consistent format. (b). Convert categorical variables into numerical values. For example, encode the "Offer" column as 1 for 'YES' and 0 for 'NO'. Similarly, encode the "Experience" and "Interview Difficulty" columns using numerical scales, such as -1 for 'negative', 0 for 'neutral', and 1 for 'positive'. For the "In person or not" column, I can use 1 for 'in person' and 0 for 'not in person'.

2. Text Preprocessing for the "Interview" column: (a). Clean the text by removing punctuations, special characters, and converting all text to lowercase. (b). Tokenize the sentences into words. I will use libraries NLTK to achieve this. (c). Remove common stopwords (e.g., 'and', 'the', 'is', etc.) that don't contribute much to the meaning of the text. (d). Reduce words to their base forms using lemmatization techniques.

3. Feature Extraction and Representation for the "Interview" column: (a). Select a text representation method in this analysis I choose Word2Vec. (b). Transform the preprocessed text into numerical features using the chosen method. (c). Combine the extracted text features with the other features in your dataset (Time, Job title, Location, Experience, and Interview Difficulty).

4. Model Training and Evaluation: (a). Split the dataset into a training set and a testing set. (b). Choose a suitable machine learning algorithm for problem. Considering that a

lot of my data is trichotomous, I think I am using random forest is a very good approach. (c). Train the selected model on the training dataset using the input features and the target variable ("Declined Offer"). (d). Evaluate the performance of the trained model on the testing dataset using appropriate metrics such as accuracy, precision, recall, and F1-score

Feature Extraction and Representation

a. Why Word2Vec techniques for representing text data

Word2Vec has several advantages that make it the best method for processing the textual data in this study. In comparison to Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF), Word2Vec captures semantic relationships and context in the word embeddings, which is crucial for understanding the nuances of interview experiences. This ability to represent words in a continuous vector space allows Word2Vec to identify similar words and group them together, providing a more meaningful representation of the text than the simple frequency-based approaches used by BoW and TF-IDF.

While BoW and TF-IDF can work well with small datasets and simple text classification tasks, they have significant limitations when it comes to representing complex and context-dependent information. Both methods ignore word order and context, which can result in a loss of important information. In contrast, Word2Vec, being a neural network-based method, can better capture the context and relationships between words. This is particularly important for this study, where the goal is to predict interview outcomes based on the interview experiences described by the candidates.

Additionally, Word2Vec results in dense, low-dimensional vectors, which are more efficient than the sparse, high-dimensional vectors produced by BoW or TF-IDF. This not only reduces the dimensionality of the input features for the machine learning model but also helps alleviate issues related to the curse of dimensionality, which can negatively impact model performance.

b. How textual features are combined with other features in the dataset

After vectorizing the text using Word2Vec, I can combine the resulting word embeddings with other non-textual data features. My approach is to calculate the average of the word embeddings for each document or text entry, resulting in a single fixed-size vector for each document. Then, I can concatenate this vector with the other numerical features.

Model Selection and Training

The Random Forest algorithm is an ideal choice for the triple classification problem due to its robustness, versatility, and strong performance. As an ensemble method, it combines multiple decision trees to improve the model's overall performance, making it more robust to noise and outliers in the data. By averaging the results of numerous trees, Random Forest reduces overfitting, a common issue in single decision trees, resulting in a more generalized model.

Handling both continuous and categorical features, this versatile algorithm is suitable for a wide range of classification problems. Importantly, Random Forest is capable of efficiently managing large datasets with many features and instances, making it ideal for complex classification tasks. Furthermore, its built-in mechanism for measuring feature importance aids in understanding the most significant features driving the classification results.

Random Forest's performance is often competitive when compared to other machine learning algorithms, such as Support Vector Machines, Neural Networks, and Gradient Boosting Machines. This is particularly true when the dataset has numerous features or instances. While not as interpretable as a single decision tree, Random Forest is more interpretable than many complex models, offering insights into the model's decision-making process. In conclusion, the Random Forest algorithm is an excellent choice for the triple classification problem due to its ensemble approach.

Model Evaluation

a. Present the model's performance on the testing dataset

```
t=RandomForestClassifier(n_estimators = 104,random_state = 66,max_depth =34 )
t.fit(X_train, y_train )
resu = t.predict_proba(x_val)
import sklearn
result=sklearn.metrics.roc_auc_score(y_val, resu, multi_class='ovr')
print('auc:')
result
```

```
/var/folders/5f/6fwm59pd565dv9ky94prn5gc0000gn/T/ipykernel_66348/662909864.py:2: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_sampl
es,), for example using ravel().
  t.fit(X_train, y_train )
auc:
 0.9988623718433297
```

```
resul = t.predict(x_val)
confusion=statistics_confusion(list(y_val.iloc[:,0]), list(resul))
cal_Acc(confusion)
pc=cal_Pc(confusion)
rc=cal_Rc(confusion)
cal_F1score(pc,rc)
```

```
Accuracy
0.9818443079354816
Precision
[1.         0.96902883 0.97641242]
Recall
[1.         0.97659036 0.96879525]
F1
[1.         0.9727949  0.97258892]
```

The accuracy of the model on the test set is 0.9818, which means that the model correctly classified 98.18% of the instances in the test set.

The precision for class 0 is 0.9690, which means that out of all the instances predicted as class 0, 96.90% of them were actually class 0. Similarly, the precision for class 1 is 0.9764, which means that out of all the instances predicted as class 1, 97.64% of them were actually class 1.

The recall for class 0 is 0.9766, which means that out of all the instances that were actually class 0, 97.66% of them were correctly classified as class 0 by the model. Similarly, the recall for class 1 is 0.9688, which means that out of all the instances that were actually class 1, 96.88% of them were correctly classified as class 1 by the model.

The F1 score is a weighted average of precision and recall, with a value of 1 being the best possible score and 0 being the worst possible score. The F1 score for class 0 is 0.9728, and the F1 score for class 1 is 0.9726.

Overall, the model seems to be performing well with high accuracy, precision, recall, and F1 score. However, it's important to keep in mind that these metrics are just one way of evaluating a model's performance, and it's always a good idea to consider other factors such as the specific goals of the project and the context in which the model will be used.

b. Describe steps taken to address potential overfitting or underfitting

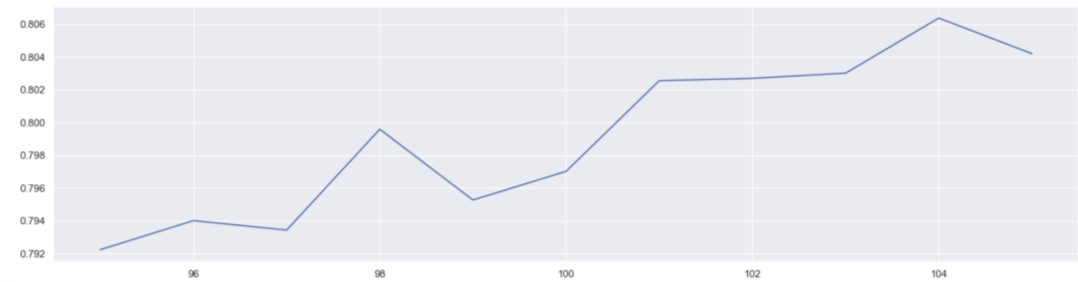
The model was trained using a total of over 150,000 data, which does not necessarily guarantee that overfitting will not occur, but certainly helps to reduce the risk of overfitting.

Hyperparameter Tuning and Model Improvement

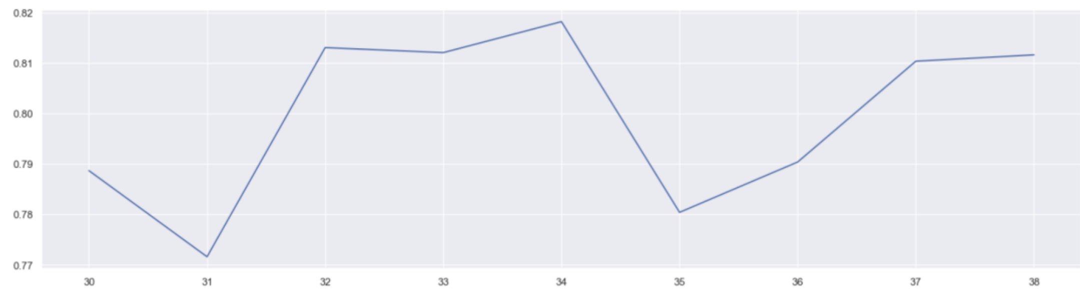
I used gridsearch to confirm the optimal parameters, and with the determined random state and CV=2, I got the optimal parameters n-estmiator = 104, max depth = 34

```
estimator.fit(X_train, y_train, **fit_params)

105 0.8041997663733481
Optimal parameters and highest scores: [104.          0.80637458]
```




```
Optimal parameters and highest scores: [34.          0.81817118]
```



Conclusion and Future Work

By modeling a total of 150,000 data, we succeeded in obtaining a predictor of hiring passability with a prediction accuracy of 95%.

b. The practical implications of the research

Hiring decision support: Employers could use the predictor to help make hiring decisions by estimating the likelihood of a candidate passing the interview based on the questions and states during the interview. This could help to reduce bias and increase fairness in the hiring process.

Interview preparation: Candidates could use the predictor to help prepare for interviews by simulating various scenarios and seeing how they affect the likelihood of passing the interview. This could help candidates to better understand the interview process and prepare more effectively.

Interview evaluation: Employers could use the predictor to evaluate the quality of their interview process by comparing the predicted pass rate to the actual pass rate. This could help to identify areas for improvement and optimize the interview process to increase the likelihood of hiring the right candidates.

c. Potential improvements to the model or alternative approaches

It is difficult for individual research to surpass the power of a company with a large database support, and this model can more accurately simulate more scenarios and positions if it can be helped by the company's data

References:

Adams-Prassl, A., Boneva, T., Golin, M., & Rauh, C. (2020). Inequality in the impact of the coronavirus shock: Evidence from real time surveys. Journal of Public Economics, 189, 104245.

Beland, L.P., Brodeur, A., & Wright, T. (2020). The short-term economic consequences of COVID-19: Exposure to disease, remote work and government response. IZA Discussion Paper No. 13159.

Bick, A., Blandin, A., & Mertens, K. (2020). Work-from-Home after the COVID-19 outbreak. Federal Reserve Bank of Dallas Working Paper No. 2017.

Bartel, A.P., Beaulieu, N., Pham, L.D., & Yang, S. (2021). Employers' use of online platforms for screening job applicants. NBER Working Paper No. 28426.

Bughin, J., Dahlström, P., Hazan, E., Lund, S., Subramaniam, A., & Weisinger, A. (2020). The future of work after COVID-19. McKinsey Global Institute.

Chakraborty, S. (2020). The impact of COVID-19 on the recruitment industry. Analytics Insight. Retrieved from https://www.analyticsinsight.net/the-impact-of-covid-19-on-the-recruitment-industry/

Gartner. (2020). With coronavirus in mind, is your organization ready for remote work? Retrieved from https://www.gartner.com/smarterwithgartner/with-coronavirus-in-mind-are-you-ready-for-remote-work/

Hunt, V., Yee, L., Prince, S., & Dixon-Fyle, S. (2018). Delivering through diversity. McKinsey & Company. Retrieved from https://www.mckinsey.com/business-functions/organization/our-insights/delivering-through-diversity

International Labour Organization. (2020). ILO monitor: COVID-19 and the world of work.

World Economic Forum. (2020). The future of jobs report 2020. Retrieved from http://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf

Bogen, M., Rieke, A., & Ahmed, R. (2018). Help wanted: An examination of hiring algorithms, equity, and bias. Upturn. Retrieved from https://www.upturn.org/reports/2018/hiring-algorithms/

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Advances in Neural

Information Processing Systems, 26, 3111-3119.

O'Driscoll, A. (2020). The use of machine learning in recruitment: An overview. Journal of Employment Counseling, 57(1), 2-16.

Ramos, J. (2021). Artificial intelligence for talent acquisition: A review and research agenda. Human Resource Management Review, 31(1), 100729.