**PREDICTING SUCCESSFUL HIRES USING MACHINE LEARNING**

**KEERTHANA K**

III year, M.Sc., Data Science, Department of Computing

Coimbatore Institute of Technology, Coimbatore, India

keerthanak2511@gmail.com

**ABSTRACT**

**Recent trends have changed the candidate recruitment process. It’s a candidate driven market out here. Compared to a few years ago, candidates now have far more power during the job search. When there are more job seekers looking for work than jobs available it’s becoming common for candidates to reject a job offer. Candidates refusing job offers is rarely out of company’s hands completely. According to research study and every recruiter and HR professional everyday work experience, the current job market is 90% candidate driven. That means company don’t pick talent anymore, talent picks company.**

**Consequently, finding and hiring perfect job candidates, especially those with in-demand skills, has become an extremely hard, expensive and time-consuming. No one wants to waste time with rejected offers. In recent trend, candidates get multiple job offers from different companies and give the false promises to the recruiter leaving them to guess their joining status.**

**The aim of this paper is to predict which candidates are more likely to accept a job offer based on predictive modelling and machine learning (ML). The gathered dataset is about prior candidates who have and have not accepted such offers. By using this model recruiters can make more informed decisions about the likelihood of candidate success, thus bettering their overall rate of successful hires. Predictive models can be trained over time to respond to new data, delivering the results to HR needs. The predictive model is being built using classification algorithms like neural networks and random forests, which fall in the category of Supervised Machine Learning.**

**Keywords: Predictive Models, Machine Learning, Supervised Learning, Job Acceptance, Recruitment.**

1. **INTRODUCTION**

Human Resource (HR) departments have a tradition of collecting vast amounts of HR data. Unfortunately, this data often remains unused. As soon as organizations start to analyze their people problems by using this data, they are involved in HR analytics. By using people analytics, companies don’t have to rely on gut feeling anymore. Analytics enables HR professionals to make data-driven decisions. Furthermore, analytics helps to test the effectiveness of HR policies and different interventions. People analytics helps organizations to make intelligent, more strategic and more informed talent decisions. With people analytic, organizations can find better applicants, make smarter hiring decisions, and increase employee performance and retention.

It is a data-driven access toward Human Resources Management. Machine learning has been moderately but surely entering the HR domain, and multiple use cases such as attrition prediction, right hiring and human resource training have been organized. It is also believed that machine learning can gather the success of a potential candidate. Using machine learning it enables us to answer critical questions about the organization. Machine learning algorithms are usually classified as supervised or unsupervised algorithms. Supervised machine learning algorithms can affix what has been learned in the past to new data using labeled examples to predict future events. Starting from the investigation of a known training dataset, the learning algorithm generate an inferred function to make predictions about the output values. The system is able to provide targets for any new input after enough training. The learning algorithm can also examine its output with the correct, intended output and find errors in order to alter the model accordingly.

The focus of this paper is to build the supervised machine learning model to predict the successful hires in an organization. The factors affecting the hiring process will be analyzed. For the prediction, the model Random Forest Regression is used.

1. **REASON FOR UNSUCCESSFUL RECRUITMENT**

In the recent times, candidates are frequently turning down job offers, which may be a sign that something is going wrong in the hiring process in the companies. Workers want to be paid for what they’re worth, and if they don’t think the organization is offering them a fair value for their skills, they’ll move on. Unresponsive companies lose out candidates because candidates aren’t willing to wait forever to hear back and they may accept other jobs in the meantime. Candidates don’t always pick the best paying position they usually go for the best fit. It’s not always all about money, workplace culture and other factors like: Age, Marital Status, Candidate expectation etc., also do matter. They look for jobs closer to home because the commute isn’t feasible in the long term.

Reasons that candidate also cite for the offer decline include better offer from the competitor, current company retention plan or the offer is not lucrative enough. People decline job offers because they don’t believe there are opportunities to advance at the company. Job offers being declined by candidates also hampers the recruiters’ creditability as the turn-around time also increases. Job offer decline is one of most time consuming and expensive activity as one is at closure stage and the hiring agency needs to start all over again.

Simple solution for this issue could be to find out why the candidate is declining a job offer. If candidates are turning down jobs across the board, the organization might need to look at their own recruiting processes. Once they identify where the problems are, they need to make the changes.

1. **EXAMINING THE DATASET**

The dataset used here was generated using the help of an organization. The organization specified the key factors that the recruiter collects when a candidate attends an interview.

The data generated is for 500+ candidates. The main factors considered for the analysis are as follows:

* **Unique Id** - Each candidate attending an interview are identified using a ID
* **Gender** - Gender of the candidate - MALE / FEMALE
* **Age Band** - 20 TO 30, 30 T0 40, ABOVE 40
* **Marital Status** - Yes (Married) / No (Not Married/ Divorced)
* **Having Kids** - YES / NO
* **Spouse Working Status** - Yes (Working)/ No (Not Working)
* **Degree Level** - Bachelor(B), Masters(M), Pd (PhD)
* **Candidate Ambition Factor** - High, Medium, Low
* **Experience (In Years)** - Years of previous experience of the candidate
* **Distance To Office From Home** - nearby (<10 km), intermediate, far (>25 km)
* **Shift Type** - REGULAR, 2ND SHIFT, NIGHT
* **Percentage Hike Asked By The Candidate**
  + **Percentage Hike Offered By The Recruiter**
  + **Notice Period - For Joining**
  + **Work From Home –** Yes (Offered)/ No (Not Offered)
  + **Onsite Willingness Level -** High, Medium, Low
  + **Having Offers from Other Company -** Yes/ No
  + **Percentage Hike Offered By The Other Company**
  + **Offer Accepted By The Candidate Or Not** – Yes (Accepted)/ No (Not Accepted)

Here “**OFFER ACCEPTED BY THE CANDIDATE OR NOT**” attribute is considered as the dependent variable while the other attributes are independent variable. Dependent variable is something that relies on other factors. Here the other factors are the independent variables.

1. **MODELING THE DATA USING RANDOM FOREST REGRESSION**

Random Forest is an adoptable, easy to use machine learning algorithm. Random forest builds multiple decision trees and merge their predictions together to get a more accurate and stable prediction rather than relying on individual decision trees. Each tree in a random forest learns from a random sample of the training observations. For building a predictive model in python several packages/ libraries are used. These libraries contain built-in modules that provide access to system functionality such as file I/O. Modules written in Python provide standardized solutions for many problems that occur in programming.

The python libraries used in our paper are:

* **Pandas** - most popular python library that is used for data analysis.
* **NumPy** - general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with arrays.
* **Seaborn** - Python data visualization library based on matplotlib.
* **Matplotlib** - a python library used to create 2D graphs and plots by using python scripts.
* **Scikit learn** – a python library has a lot of effective tools for machine learning and statistical modeling including classification, regression.

The main aim is to predict the successful hires and it is important to see which variables are contributing the most in hiring. But before that we need to know if the variables are correlated if they are, we might want to avoid those in model building process. The variables which are not poorly correlated (i.e. correlation value tend towards 0), will be chosen and move forward with them and will leave the ones which are strongly correlated (i.e. correlation value tend towards be 1).

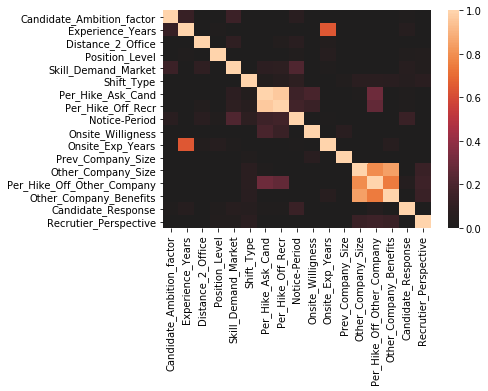


Figure 1: Correlation Heatmap of HR Data (Numerical variables)

From the above heatmap it clearly noticed that all the factors are poorly correlated to each other. Hence all the above factors (numerical variables) are considered for the model building.

The model gave accuracy score of 0.75. The random forest works quite well even with the default parameters. That’s one of reason for the usage of Random Forest Classifier for this problem.

One of the best feature Random forest regression model is that it provides the importance of variables in the data model. For this HR Analytics problem, we are interested in knowing which factor contributed the most in the process of successful hiring and the factor the was the cause of unsuccessful hiring.

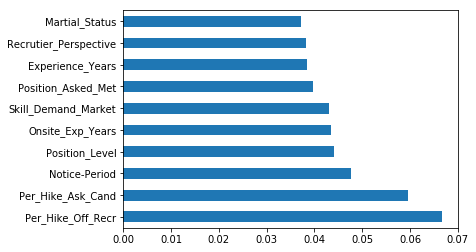


Figure 2: Horizontal Bar chart for factors (TOP 10) affecting hiring.

The above bar chart we can see the top 10 factors that are responsible for successful hiring in the organization. The major factor accountable is the “Percentage Hike Offered by the Recruiter”. A candidate offers a job only if he/she is satisfied with the Hike offered by the organization. The next factor is Position level. A candidate based on his present position level can take on a better decision whether to accept the job offer or not.

Years of experience of the candidate also matters. The more experienced the candidate is, the more freedom the candidate has in taking the decision. Lastly marital status, depending on the personal and family issues of the candidate the candidate accepts the job offer.

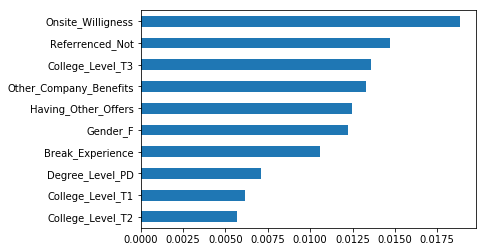


Figure 3: Horizontal Bar chart for factors (LEAST 10) affecting hiring.

From the factors in the above figure we can conclude that these are the key reasons for a candidate to turn down the job offer. It is clearly seen that a candidate’s college level and his degree level play an important role. Highly educated, having degree level as PhD has made these candidates to reject the offer.

Even gender plays a major role in career choices. Most of the candidates who have turned down their offer happens to be Females. Due to certain restrictions and other family issues females tend to decline their job offers.

1. **CONCLUSION**

Identifying and recruiting top talent is complicate enough, but successfully hiring that talent presents even more challenges. Candidates control the market, so they can be able to be selective. In this paper there may be few factors reasoned out for successful hiring, but there are few more factors that influence this process. The candidate might have turned down an offer because the offered salary could have been not up to the expectation. If the candidate is talented then the organization can try offering more salary. The organization should never try to force the candidate to accept the job.

Using machine learning and predictive analysis such kind of problems can be easily ended. Machine Learning is the one, where you can consider the learning is guided by a teacher. We have a dataset which acts as a teacher and its role is to train the model or the machine. Once the model gets trained it can start making a prediction or decision when new data is given to it.

1. **REFERENCES**
2. Alvin C. Rencher Methods of Multivariate Analysis, 2nd Edition, Wiley Inter-science, 2002 [Para 1,Para V]
3. Bishop, C. Pattern Recognition and Machine Learning. Berlin: Springer-Verlag, 2006.
4. Introduction to Machine Learning, Second Edition, Ethem Alpaydın, The MIT Press Cambridge, Massachusetts London, England 2010.
5. Machine Learning, Tom Mitchell, McGraw, 1997, 0-07-042807-7
6. R.P. Hooda, “Statistics for Business and Economics”, 5th Edition, Vikas Publishing House Pvt. Ltd., Noida, 2013.
7. S.C. Gupta, V.K. Kappoor, "Fundamentals of Mathematical Statistics", Sultan Chand & Sons, 10th Revised Edition, New Delhi, 2002.