

Employee Attrition Prediction

Jayashree Prasad^{*1}, Prathamesh Shinde^{*2}, Manodnya Gaikwad^{*3}, Tanmay Mahindrakar^{*4}, Sohan Kakatkar^{*5}

Computer Engineering Department, Sinhgad College of Engineering, Pune

¹jrprasad.scoe@sinhgad.edu

²sprathmnda135@gmail.com

³manodnyagaikwad@gmail.com

⁴tanmay.mandy@gmail.com

⁵kakatkarsohan7@gmail.com

Abstract

Employees are the most valuable assets of an organization. To find, attract, develop and retain the right talent is a major part of management. Whenever a well-trained and well-adapted employee leaves the organization, it creates a vacuum. Therefore, the organization loses key skills, knowledge and business relationships. This study aims to identify attributes that contribute in employee attrition and numerical experiments are performed on these attributes using supervised learning methods like Support vector machine, Random Forest, Naive Bayes, Extreme gradient boosting. The performance of each of these supervised machine learning methods is analyzed through a robust and comprehensive evaluation process. This survey will help the human resource managers to identify the employees that are likely to leave the organization and predict the possible reasons for their decision, which will enable the HR managers to devise a retention plan or look for replacement.

Keywords - Employee Attrition, Machine Learning, Random Forest, Naive Bayes, Extreme Gradient Boosting, Decision Trees, Human Resource

I. INTRODUCTION

Employees leaving on short notice is a major concern for HR managers. It prevents them from coming up with a successful retention strategy. Finding a replacement cost is 1.5-2x times more than retention, including the hiring cost, training and onboarding costs. ML models would flag the employees who might be considering switching companies and possible reasons for doing so.

Attrition, in Human Resource terminology, refers to the phenomenon of the employees leaving the company. Attrition in a company is usually measured with a metric called attrition rate, which simply measures the no of employees moving out of the company (voluntary resigning or laid off by the company). The purpose of this study serves to predict the employee who is willing to leave the company and also the employees that could be dismissed with having the least repercussions in the company. It aims to provide insight into each process by gathering data and then using it to make relevant decisions about how to improve these processes.

The objective of this survey is to provide insight into each process by gathering data and then using it to make relevant decisions about how to improve these processes by training the model based on previous attrition data available and predicting it in future for better company HR management. This study uses documented attributes affecting employee attrition to predict and does not consider undocumented factors that may lead to attrition.

II. RELATED WORK

The study of Prediction of Employee Attrition with work-place related variables uses classification models for work-place variables rather than using demographic or behavioral variables [5]. The results suggest a model for Artificial Neural Network dominates with the maximum accuracy. The most influential variable turned out to be the attrition of the managers which somehow further triggers employee attrition. This approach suggests that work-place related policies are easy to formulate in an organization but demographic and behavioral aspects were yet to be studied in detail for future work.

Along with ANN, other research papers using various ML algorithms such as Extreme Gradient Boosting method in prediction of Employee turnover prediction problem were introduced for better performance [6]. The author points out that the Human resource data is not well maintained in real life scenarios and would therefore require a lot of preprocessing. The author leverages the ability of Xgboost method to generalize noise ridden data, to make a case for it to be used in case of the above mentioned scenario. This paper compares the performance of XGBoost with respect to other methods like Logistic regression, Naïve Bayes, Random Forest, K-nearest neighbor. Run time and maximum memory utilization has also been used for comparison.

Contributing to this research, a model for predicting Employee Attrition using Machine Learning based approach i.e. XGBoost is proposed which is highly robust [2]. XGBoost belongs to boosted tree algorithm and works on the principle of gradient boosting. As compared to others, practices a more regularized-model reinforcement to regulate overfitting and thus improvises performance. It is a fast method consisting of parallel tree construction and planned to be fault tolerant under the distributed setting. In the near future, the research would go in the direction to make this model a “Predictive Mode” and solve various issues, i.e. Advanced ones not predicting- “Who is going to Leave?” and “Why the Employees are doing turnover?”. The model will become more accurate, scalable and ready to implement as such in top IT organizations HR departments.

For identifying the key attributes which contribute most in predicting employee turnover, it uses two level dimensionality reduction. In the first level basic Sequential Backward Selection Algorithm (SBS) is used to remove less significant attributes and in the second level Chi-square and Random Forest Classifier is used to identify the significance of each of the attributes. There was considerable improvement when only significant attributes were used for prediction. The study also claims that using significant attributes not only makes the building model less complex but also prevents over fitting. The reason for high accuracy in this model is the dimensionality reduction which results in a narrow perspective in this domain.

To overcome the research gap, further studies attempt to identify the factors causing the departure of the employees from the organization and to take appropriate steps to minimize that. Algorithms such as linear support vector machine, C5.0 Decision Tree classifier, Random Forest, k-nearest neighbor and Naïve Bayes classifier have been implemented on the employee dataset [8]. The experimental results prove that Random Forest outperformed all other classifiers. Visual representation of data suggests that employees who tend to leave the organization were the ones who were not promoted in the last five years, had high working hours and were getting promoted but didn't have a reasonable salary. Even though the basic factors were identified using this approach, this study requires further exploration to minimize the prediction error rate.

Association rule mining using Apriori algorithm has been applied to form association rules using which the C5.0 decision tree has been trained to predict attrition of employees [7]. To further optimize the model Grey Wolf Optimizer (GWO) and Particle Swarm Optimization was implemented. The

results acquired from both optimization techniques were compared which proved that Grey Wolf Optimized C5.0 with association algorithm is more efficient in time and memory consumption as compared to other techniques with C5.0. The Grey Wolf Optimizer algorithm used in this approach has the potential to be superseded by some other nature based algorithm which creates an immense scope for extension of work in future endeavors.

Contributing to the retention strategies, the domain of this study focuses on recruitment and retention challenges that the IT/BPO industry currently faces and to examine ways to reduce high turnover rates among first year Employees in the leading Domestic Call Center based in Indore [3]. The findings imply that employees need manageable workloads, support and recognition from their co-workers and management, and opportunities for growth and innovation. Retirement is one major cause of employee attrition, and since people tend to retire around a specific age this is a factor that can be accounted and planned for. Other causes of employee attrition, such as personnel who quit due to prolonged illness, dissatisfaction with the company, or other reasons, can be more difficult to estimate. This study highlights how softer factors related to organizational culture, inconsistent shift timings and motivational factors are impacting employee attrition.

To add perspective, this research compares traditionally used classification algorithms with an ensemble learner and building it [9]. This model will be able to predict the employees turnover more precisely, based on the accuracy obtained from the individual classifications and the weights assigned to them. Based on the weighted average the final classification is done which gives an improved performance that is more superior to the results given by individual classifiers. However, its sensitivity of the data is slightly less than Random forest which can be studied more in detail in the future.

Integrating all machine learning algorithms studied by past researchers for better vision, this paper studies the performance of ten supervised machine learning algorithms like Decision tree, Naive Bayes, Random forest, Support Vector Machine three sizes of population via small, medium, large. Extensive experiments are performed to identify the class of algorithms that perform better in case of different population sizes [1]. The author points out that accuracy is not a correct and complete evaluation metric owing to bias in the values of target attribute, therefore the author uses recall, precision, F1 score and ROC curve alongside accuracy for evaluation of models. In conclusion this paper gives insights into choosing an algorithm and evaluation metric based on the size of the population.

To proactively combating voluntary attrition of employees this research deals with identifying top talents in the organization and their effects after their departure from the organization [4]. The most basic way mentioned for retention is salary increase. Although salary being an important aspect, other factors such as advancement opportunities, work environment as well as other factors dominate in the current world and hence gives room for more future work.

TABLE I.
LITERATURE REVIEW

Author	Publisher	Paper Title	Methodology	Pros	Cons
Yue Zhao, Maciej K. Hryniewicz, Francesca Cheng, Boyang Fu, Xiaoyu Zhu,	Springer Nature Switzerland AG 2019	Employee turnover prediction with machine learning : a reliable approach	ML algorithms are implemented on different types of dataset, their performances are compared	Extensive implementation of all algorithms on different types of dataset	Dataset used are small with less number of attribute
Dilip Singh Sisodia ; Somdutta Vishwakarma	IEEE, 2018	Evaluation of ML models for employee churn prediction	Comparative study of various ML algorithms	Finds out which ML algorithm is performing well	Doesn't have much detail about the methodologies used
Abdelrahim KasemAhmad, Assef Jafar, Kadan Aljoumaa	Springer International Publishing, 2019	Customer churn prediction in telecom using machine learning in big data platform	DT, RF and GBoost algorithms are implemented to predict customer churn	Ensemble Algorithms are used to predict customer churn	This paper is to predict customer churn
Moninder Singh, Kush R. Varshney, Jun Wang and Aleksandra Mojsilovic	2012 IEEE	An Analytics Approach for Proactively Combating Voluntary Attrition of Employees	An Analytics Approach for Proactively Combating Voluntary Attrition of Employees	Clustering techniques used Focus on retention strategies	Only algorithms are mentioned no result is mentioned

Vishnuprasad Nagadevara	Review of Business Research, 2012 - researchgate	Prediction of employee attrition using workplace related variables	Classification models namely Classification of regression trees, regression trees, Chi-square automatic interaction detection and ANN are used for prediction	Classification models such as Artificial Neural Network are easy to implement due to considering only work related variables	Behavioral aspects are yet to be studied in detail
Rachna Jain and Anand Nayyar	IEEE 2018	Predicting Employee Attrition using XGBoost Machine Learning Approach	XGBoost belongs to boosted tree algorithm and works on the principle of gradient boosting.	It is a fast method consisting of parallel tree construction and planned to be fault tolerant under the distributed setting.	The model is not yet implemented in real-world organizations.

Author	Publisher	Paper Title	Methodology	Pros	Cons
Krishna Sehgal, Harlieen Bindra, Anish Batra and Rachna Jain	International Journal of Advanced Research in Computer Science and Management, 2013	An Analysis of attrition: retention strategy for IT/BPO Industry	In order to meet the objectives of the study, primary data is collected using the questionnaire method.	It examines ways to reduce high turnover rates among first year Employees in the leading Domestic Call Center	Scope of this study is very limited

Krishna Sehgal, Harlieen Bindra, Anish Batra and Rachna Jain	Springer Nature Singapore, 2019	Prediction of Employee Attrition Using GWO and PSO Optimised Models of C5.0 Used with Association Rules and Analysis of Optimisers	GWO and PSO optimised models of C5.0 are used with association rules and analysis of optimisers	Time to predict employee attrition and consumption of RAM have been optimised with GWO	Better nature based algorithms can be used
--	---------------------------------	--	---	--	--

III. CONCLUSION AND FUTURE WORK

Attrition being a major element contributing to the growth of an organization, researchers are still studying ways and methods to identify any possible attrition. Majority of the researches in this survey suggest that the future dominating algorithm in this sector is most likely to be Tree based algorithm. Recent studies have demonstrated that accuracy is not the only factor while evaluating models, but also other performance measures such as recall, precision, F1 score and ROC curve must be considered. Hence, increasing the overall perspective, this survey will contribute to design a robust and reliable model for employee attrition prediction and their respective retention strategies.

ACKNOWLEDGEMENTS

We feel great pleasure in expressing our deepest sense of gratitude and sincere thanks to all the Staff Members of our Department for their guidance during the project work, without which it would have been a very difficult task. We have no words to express our sincere thanks for the extreme assistance and cooperation extended by all the teaching, non- teaching staff members of our department, our parents and colleagues who helped us directly or indirectly for completing this project successfully.

REFERENCES

- [1]. Yue Zhao, Maciej K. Hryniewicki, Francesca Cheng, Boyang Fu, Xiaoyu Zhu, "Employee turnover prediction with machine learning : a reliable approach", Springer Nature Switzerland AG, vol. 869, 2019.
- [2]. Rachna Jain and Anand Nayyar, "Predicting Employee Attrition using XGBoost Machine Learning Approach", International Conference on System Modeling & Advancement in Research Trends, IEEE, 2018.
- [3]. Vibha Gupta, "An Analysis of Attrition: Retention strategy for IT/BPO Industry", International Journal of Advance Research in Computer Science and Management Studies, 2013.
- [4]. Moninder Singh, Kush R. Varshney, Jun Wang and Aleksandra Mojsilovic, "An Analytics Approach for Proactively Combating Voluntary Attrition of Employees", IEEE, 2019.

- [5]. Vishnuprasad Nagadevara, "Prediction of employee attrition using workplace related variables", Review of Business Research, 2012 , Research Gate.
- [6]. Rohit Punnoose, Pankaj Ajit, "Prediction of Employee Turnover in Organizations using Machine Learning Algorithms A case for Extreme Gradient Boosting", International Journal of Advanced Research in Artificial Intelligence, 2016.
- [7]. Krishna Sehgal, Harlieen Bindra, Anish Batra and Rachna Jain, "Prediction of Employee Attrition Using GWO and PSO Optimised Models of C5.0 Used with Association Rules and Analysis of Optimisers", Springer Singapore, vol.74, June 2019.
- [8]. Dilip Singh Sisodia ; Somdutta Vishwakarma, "Evaluation of ML models for employee churn prediction", IEEE, 2018.
- [9]. Shubham Karande and L. Shyamala, "Prediction of Employee Turnover Using Ensemble Learning", Springer Nature Singapore, vol 904, March 2019.
- [10]. Mirza Mohtashim Alam(&), Karishma Mohiuddin, Md. Kabirul Islam, Mehedi Hassan, Md. Arshad-Ul Hoque, and Shaikh Muhammad Allayear, "A Machine Learning Approach to Analyze and Reduce Features to a Significant Number for Employee's Turn Over Prediction Model", Springer, Cham, vol 857, 2018.