WORKFORCE MEN DEBILITATION PREDICTION INCLUDING EXPOSED EMPLOYEES USING BIG DATA AND MACHINE LEARNING

Dr. Thota Siva Ratna Sai,[1] Thanush Kumar[2], Bindu Swetha[3], Lokesh Naidu[4], Moses Madhukar[5]
[1] Assoc. Professor, Department of Computer Science and Engineering
[2,3,4,5] Student, Department of Computer Science and Engineering
[1,2,3,4,5] Qis College of Engineering and Technology

Abstract- It takes time and effort to cultivate a great employee, yet it simply takes a moment to terminate their employment. The problem of employee turnover is one that affects even the most successful companies, and it has a direct bearing on the success of an organization because of the quality of the employees it is able to keep. There are monetary losses, holes in the company's ability to execute, re-recruiting expenses, and productivity losses whenever a high-performing individual quits. Keeping the right people on staff and lowering turnover rates are equally important to a company's success. According to NASSCOM, the worldwide staff churn rate in 2019 is between 18% and 20%, highlighting the need for statistical analysis to mitigate the risks posed by such high rates of employee departure. In this research, supervised classification algorithms are used to identify at-risk employees based on historical data on employee turnover rates (especially in the BPO industry).

Keywords—Churn rate, attrition rate, company, employee, risk, classification.

I. INTRODUCTION

When employees leave a firm for any reason, it is called attrition. There are consequences for the person and the business when employees leave. The monetary cost to the company is enormous, and it risks losing customers and clients if service quality declines as a consequence. The person, however, may face challenges in the form of less income, increased out-of-pocket expenses, and a diminished social support system. Attrition may be prevented if the underlying causes are addressed. According to Harvard Business Review, few employees in middle management or below have access to the executive suite. The inability to create trust between employees and management is a direct result of a lack of team building activities. Most people in the workforce are unmotivated to work because their jobs are mundane and need them to do the same tasks over and over again. When workers aren't forced to think outside the box, they become less productive on the job and are less likely to contribute anything of value to the company in the long run. With B.P.O.'s rapid expansion over the last decade, workers now have several opportunities to try something new in the workplace. According to the research, employees tend to go to organizations that provide better salaries compared to the market standard. The experience obtained is meaningless because of the low level of talent and inventiveness required for the work. Furthermore, advancement possibilities inside an organization are few.

The success rate of companies in mitigating these risks is low because of a lack of organized and powerful analytics. In order to assist them better manage these risks, we want to use technology to provide preemptive warnings in the following ways: • Anticipating probable attrition trends in the form of employee groups, locations, types of work, levels, etc. Benefiting from information on high-performing workers, we can determine the most effective profile for future hires. These analyses are not a one-and-done task, but rather an ongoing process that adapts to changing circumstances. Reducing the total number of workers is known as EMPLOYEE ATTRITION. Attrition occurs when a company loses workers over time. This indicates a greater rate of staff turnover than is being replaced. Any time an employee stops working for the firm, whether voluntarily or involuntarily, whether due to layoff, leave of absence, or death. The term "BIG DATA" refers to data sets that are very large and still increasing at an exponential rate. Big data is information that is too big or too complicated for the storage or processing capacities of today's standard data management systems.

II. RELATEDWORKS

When given the opportunity, workers leave one company to join another, since the global competitive landscape means that those with the necessary skills and abilities may choose from a sea of potential employers. The negative consequences of employee turnover on workplace productivity and the timely achievement of organizational goals have elevated employee turnover to the status of the most pressing problem facing all businesses today. To combat this issue, businesses are increasingly turning to machine learning strategies for help in foreseeing staff departures. Accurate forecasting allows businesses to take steps for staff retention and succession at the appropriate times.

They developed a new model for forecasting employee turnover in the existing system, called XGBoost, which is based on machine learning and has shown to be quite accurate in the past. For the purpose of verifying the efficacy of the suggested system for Employee Attrition, the data set is gathered through online database and fetched to the system, with the system displaying extremely impressive and precise findings with regards to Employee turnover behaviour. The topic of employee turnover has been the subject of several in-depth investigations, and many successful data mining issues have been demonstrated. This study focuses on voluntary attrition, however it acknowledges the existence of research on forced attrition.

According to Kransdorff A.'s research, the rate of voluntary attrition has been rising steadily over the last decade, and workers often leave their employers after six years in order to further their careers. The company will suffer major consequences if upper management is not notified as soon as possible that its top achievers are considering leaving [1]. Functional turnover (i.e. strong performers leave and weak performers remain) may aid minimise optimum organisational

performance, but too much of it can have a negative impact on productivity and output, as shown by the research of Julie T. Johnson, Rodger W. Griffeth, and Mitch Griffin. Loss of business and connections are the most common results, putting the achievement of the company's goals at risk [2]. [3]. Also, Abbasi, Sami M., and Kenneth W. Hollman have suggested a notion of dysfunctional turnover. When employees have a dysfunctional turnover (excellent performers remain and poor performers go), it may harm the company's organisational structure by slowing down the rollout of new initiatives, stifling creativity, and decreasing productivity [4, 3]. In today's highly competitive business environment, this might have serious consequences for a company's growth. According to research conducted by Dolatabadi, Sepideh Hassankhani, and Farshid Keynia, the higher the rate of staff turnover, the greater the financial loss to the company. According to his estimates [5,] even a turnover rate of 5% may cost an organisation around 1.5 times an employee's yearly salary.

In addition, Abbasi, Sami M., and Kenneth W. Hollman have identified and emphasised a select group of key factors that contribute to employee turnover of their own will. Toxic work environments and practises are also on the list, along with problematic recruiting practises, poor management, an insufficient emphasis on employees' contributions, and an absence of incentive programmes that can compete with the market. This research also found that managers' work styles, working conditions, and effective recruitment and hiring techniques all have a role in encouraging employee turnover [4]. Age, tenure, compensation, overall work happiness, and employees' views of justice are all factors in voluntary departure, according to a meta-analytic research conducted by Cotton, J.L. and Tuttle, J.M. [6]. Studies by Allen and Griffeth, as well as D. Liu, T. R. Mitchell, T. W. Lee, B. C. Holtom, and T. R. Hinkin, reveal similar results on pay, benefits, working conditions, satisfaction on the job, management support, career development opportunities, and more. [7] [8]. In order to reduce staff turnover, several BPO companies have used technology solutions. An information management firm, for instance, must actively work to reduce employee turnover if it is to maintain its position as a provider of "best-in-class" services at an affordable charge [9]. The "Early Warning System" (EWS) they implemented with RAG (Red, Amber, Green) indications has successfully reduced the turnover rate by a substantial margin. With the help of the IBM Watson team, M. Singh et al. have released a framework that uncovers the root causes of employee turnover and predicts who would leave the company in the future. The methodology is based on the difference between the EACB and EACA, or the expected cost of attrition before and after the retention period, respectively [10]. The primary goal of this project is to describe, demonstrate, and evaluate machine learning methods for detecting attrition at a well-known health-sector revenue cycle management firm, with the ultimate goal of providing a solid technical solution to the problem by anticipating the causes of attrition. With a focus on voluntary turnover, the HR team may utilise the findings to identify and address the most pressing causes of employee departure. The results of the literature review are summarised in TABLE I. above. Studies from various articles are included in the following

sections to highlight the inadequacy of some models and to promote models that show good outcomes.

S.NO	AUTHOR(S)	METHOD USED	PREDICTED VALUE	YEAR
1.	V.Saradhi & Girish Keshav Palshikar	Support Vector Machine	74.12	2011
2.	Rohit Punnoose & Pankaj Ajit	Logistic Regression	77.69	2016
3.	Dr. M. S. Loganathan & S. Ashwini	Naive Bayes	79.53	2017
4	L. Alaskar, M. crane, Mai A. Alduailij	Random Forest	80.33	2019
5	M. Hoffman & Steven Tadelis	XGBoost	82.04	2021

III. PROPOSED SYSTEM ARCHITECTURE

The proposed system contains the following modules and it is represented in Fig.1.

- ✓ Collecting Data
- ✓ Importing Libraries
- ✓ Pre-processing the data
- ✓ Model building
- ✓ Predicting value
- ✓ Validating Output

Libraries Used

- ✓ numpy
- ✓ pandas
- ✓ matplotlib
- ✓ seaborn
- ✓ sklearn

Used Algorithms

1.NAÏVE-BAYES

An application of Bayes' theorem, the Naive Bayes algorithm is a supervised learning technique used to resolve classification issues.

2.RANDOM FOREST

A Forest Chosen at ensemble learning, the act of merging several classifiers to solve a complicated issue and increase the performance of the model, forms the basis of this approach.

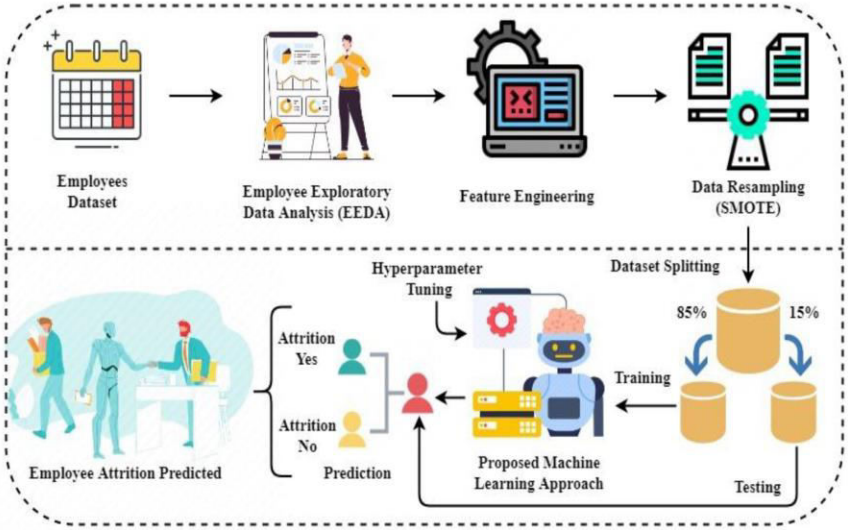


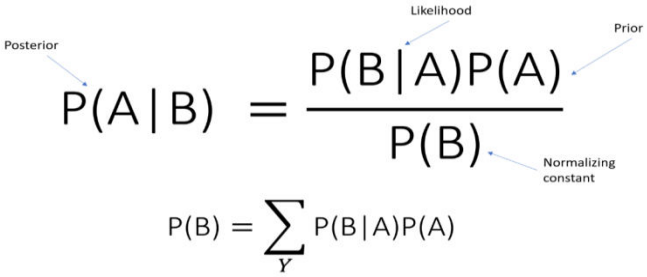
Fig.1Proposed system architecture

IV. RESULTS AND DISCUSSION

The output screens obtained after running and executing the system are shown from Fig.2 to Fig 4.

Naive-bayes : Accuracy:80.95

Random Forest: Accuracy:86.41



$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(B) = \sum_{Y} P(B|A)P(A)$$

Fig.2 Naïve bayes

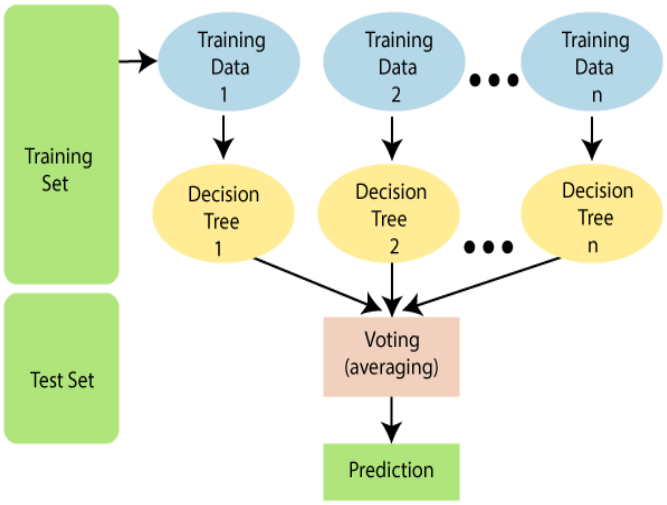


Fig.3 Predicting values

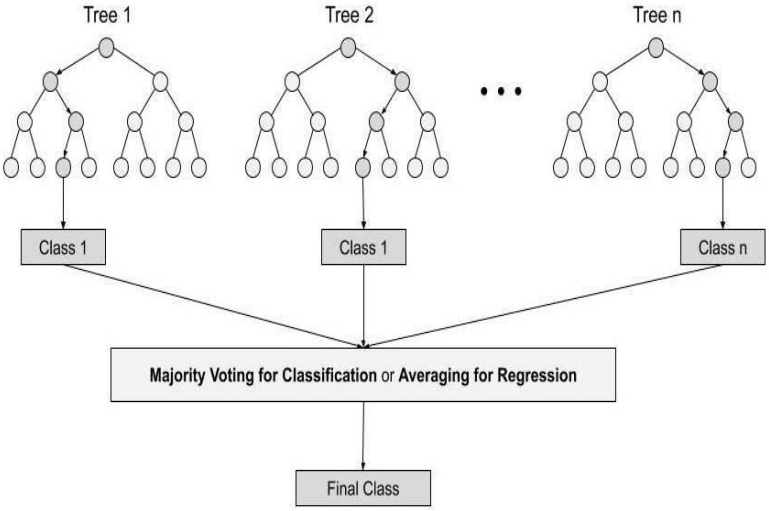


Fig.4 validating Output

V. FUTURE SCOPE AND CONCLUSION

Different machine learning methods, including Nave Bayes and random forest, are contrasted and analysed side by side in this study. Predicting Employee Turnover and Performance with 79.16% Accuracy Using Random Forest with Feature Selection. The loss of skilled workers may have far-reaching effects on a business. Most modern businesses realise they must do something to stem the tide of employee departures, and they do it by mining the data they collect in search of trends. However, even in very mature companies, success cannot be guaranteed using these approaches. Nonetheless, reducing the effects of turnover is one area where more developed companies may work together. Assuming a constant and stable attrition rate, businesses might

investigate its direct and indirect effects in order to find ways to save costs. Thereafter, procedures may be put in place to amplify these effects of talent dilution, guarantee that organisational transitions go through without a hitch, and reduce reliance on key employees.

REFERENCES

1. Žliobaite, “Learning under concept drift: an overview,” Computing Research Repository by Cornell library, pp. –1–1, 2010.
2. Singh, Jay P. "Predictive validity performance indicators in violence risk assessment: A methodological primer." Behavioral Sciences & the Law 31.1 (2013): 8-22.
3. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794, ACM (2016).
4. “How do I select SVM kernels?,” Dr. Sebastian Raschka, May-2020. [Online]. Available:https://sebastianraschka.com/faq/docs/select_svm_kernel. html. [Accessed: 14-May-2020].
5. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. 20, 273– 297 (1995)
6. Zhao, Yue, et al. "Employee turnover prediction with machine learning: A reliable approach." *Proceedings of SAI intelligent systems conference*. Springer, Cham, 2018.
7. Géron, A.: Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. O’Reilly Media (2017)
8. Efron, B.S., Hastie, T.: Computer Age Statistical Inference. Cambridge University Press, Cambridge (2016)
9. Morgan, J.N., Sonquist, J.A.: Problems in the analysis of survey data, and a proposal. J. Am. Stat. Assoc. 58, 415–434 (1963)
10. Raschka, S. (2015). *Python machine learning*. Packt Publishing Ltd.
11. Cox, David R. "The regression analysis of binary sequences." *Journal of the Royal Statistical Society: Series B (Methodological)* 20.2 (1958): 215-232.
12. T. Fawcett, “An introduction to ROC analysis”, Pattern Recognition Letters 27 (8), 861–874, 2006.
13. Lessmann, Stefan, and Stefan Voß. "A reference model for customer- centric data mining with support vector machines." *European Journal of Operational Research* 199.2 (2009): 520-530.
14. Metz, Charles E. "Basic principles of ROC analysis." *Seminars in nuclear medicine*. Vol. 8. No. 3. WB Saunders, 1978.
15. Mauricio A. Valle & Gonzalo A. Ruz (2015) Turnover Prediction in a Call Center: Behavioral Evidence of Loss Aversion using Random Forest and Naïve Bayes Algorithms, Applied Artificial Intelligence, 29:9, 923-942, *DOI: 10.1080/08839513.2015.1082282.*
16. Gao, Ying. "using decision tree to analyze the turnover of employees." (2017).

17. W. C. Hong, S. Y. Wei, and Y. F. Chen, “A comparative test of two employee turnover prediction models”, International Journal of Management, 24(4), 808, 2007.
18. Saradhi, V. V., & Palshikar, G. K. (2011). Employee churn prediction.
19. *Expert Systems with Applications*, *38*(3), 1999-2006.
20. Yedida, R., Reddy, R., Vahi, R., Jana, R., GV, A., & Kulkarni, D. (2018). Employee Attrition Prediction. *arXiv preprint arXiv:1806.10480*.
21. Singh, Moninder, et al. "An analytics approach for proactively combating voluntary attrition of employees." *2012 IEEE 12th International Conference on Data Mining Workshops*. IEEE, 2012.
22. Sachdeva, Sujata. *Times of India Online,* 27 Nov. 2007, www.timesofindia.indiatimes.com. Accessed 6 May 2020.
23. D. Liu, T. R. Mitchell, T. W. Lee, B. C. Holtom, and T. R. Hinkin, “When employees are out of step with coworkers: How job satisfaction trajectory and dispersion influence individual-and unit-level voluntary turnover”, Academy of Management Journal, 55(6), 1360-1380, 2012.
24. D. G. Allen and R. W. Griffeth, “Test of a mediated performance – Turnover relationship highlighting the moderating roles of visibility and reward contingency”, Journal of Applied Psychology, 86(5), 1014-1021, 2001.
25. Cotton, J.L. and Tuttle, J.M., 1986. “Employee turnover: A metaanalysis and review with implications for research” Academy of management review, pp.55-70.
26. Dolatabadi, Sepideh Hassankhani, and Farshid Keynia. "Designing of customer and employee churn prediction model based on data mining method and neural predictor." *2017 2nd International Conference on Computer and Communication Systems (ICCCS)*. IEEE, 2017.
27. Abbasi, Sami M., and Kenneth W. Hollman. "Turnover: The real bottom line." *Public personnel management* 29.3 (2000): 333-342.
28. M. Stoval and N. Bontis, “Voluntary turnover: Knowledge management – Friend or foe?”, Journal of Intellectual Capital, 3(3), 303-322, 2002.
29. Johnson, Julie T., Rodger W. Griffeth, and Mitch Griffin. "Factors discriminating functional and dysfunctional salesforce turnover." *Journal of business & industrial marketing* (2000).
30. Kransdorff, Arnold. "Succession planning in a fast-changing world." *Management Decision* (1996).