Predicting Employee Attrition using Random Forest

[1]M. Sirisha, [2]Md.Ruksana, [3]P.Saraswathi, [4]M.Rajitha

[1,2,3,4] UG Student, Department of Computer Science Engineering,

Dr K V Subba Reddy College Of Engineering For Women, Kurnool, Andhra Pradesh, India

Article Info

Page Number: 481-486

Publication Issue:

Vol. 71 No. 2 (2022)

Abstract

This paper aims to address the issue of employee retention in today's businesses. However, they are unable to identify the actual causes of their resignation. That could be due to a variety of factors, such as: financial, cultural, etc.). How a company treats its employees and ensures their happiness is unique to each business. However, the satisfaction rate is frequently ignored. As a result, employees frequently quit their jobs abruptly and without giving a reason. Researchers have become increasingly interested in machine learning (ML) methods over the past few decades. It is capable of suggesting solutions to numerous issues. Then, machine learning has the potential to anticipate employee attrition by making predictions. Using a real data set with a sample size of 1469, the authors of this paper compare cutting-edge options for the proposed machine learning algorithms. Managers could use the findings as a warning to alter their strategies or behavior. It could also be used to suggest new policies to managers in order to keep employees employed by the company. The goal of this study is to compare and contrast various machine learning approaches for predicting which employees will likely leave their company. With nearly 50 useful information units, the data set contains information about both current employees and employees who had already left their jobs. Numerous factors are combined in this last: cultural, professional, social, financial, and relational factors In this paper, six distinct ML algorithms were utilized. The Random Forest algorithm performed the best in terms of predicting employee attrition, as demonstrated by the experiments. The best prediction accuracy, 85.12, is regarded as satisfactory accuracy

Article History

Article Received: *25 December 2021*

Revised: *20 January 2022*

Accepted: *24 February 2022*

Publication: *28 March 2022*

1. Introduction

Human resource attrition, or the gradual retirement, designation, or other retirement of workers over time The financial burden that high employee turnover places on an organization is a major issue. In addition, regular employee turnover prevents your organization from expanding its collective knowledge base and experience over time. Some of the common expenses associated with losing and replacing employees include job postings, hiring procedures, paperwork, and training for new hires. In order to solve this issue, human resources analytics (HR analytics) should be implemented. Machine learning is a subset of artificial intelligence (AI). Instead of being explicitly programmed to do so, it focuses on teaching computers to learn from data and improve with experience. Algorithms in machine learning are taught to analyze large data sets for patterns and correlations in order to make the best decisions and predictions. Applications for machine learning get better with use and become more accurate with more data. Machine learning is used in everything from our homes to shopping carts to entertainment media to healthcare.

Concentric subsets of AI include both machine learning and its components, deep learning and neural networks. AI uses data processing to make decisions and forecasts. Without the need for additional programming, AI is able to process that data and use it to learn and become smarter thanks to machine learning algorithms. All of the subsets of machine learning that are below artificial intelligence are derived from it. Machine learning is included in the first subset; Deep learning is part of that, and neural networks are part of that. Neural networks find hidden patterns or internal structures in unsupervised learning data. It is used to get rid of datasets that don't have labeled responses in the input data.

A common method of unsupervised learning is clustering. It is utilized in exploratory data analysis to uncover data-hidden patterns and clusters. Market research, commodity identification, and gene sequence analysis are all examples of applications for cluster analysis

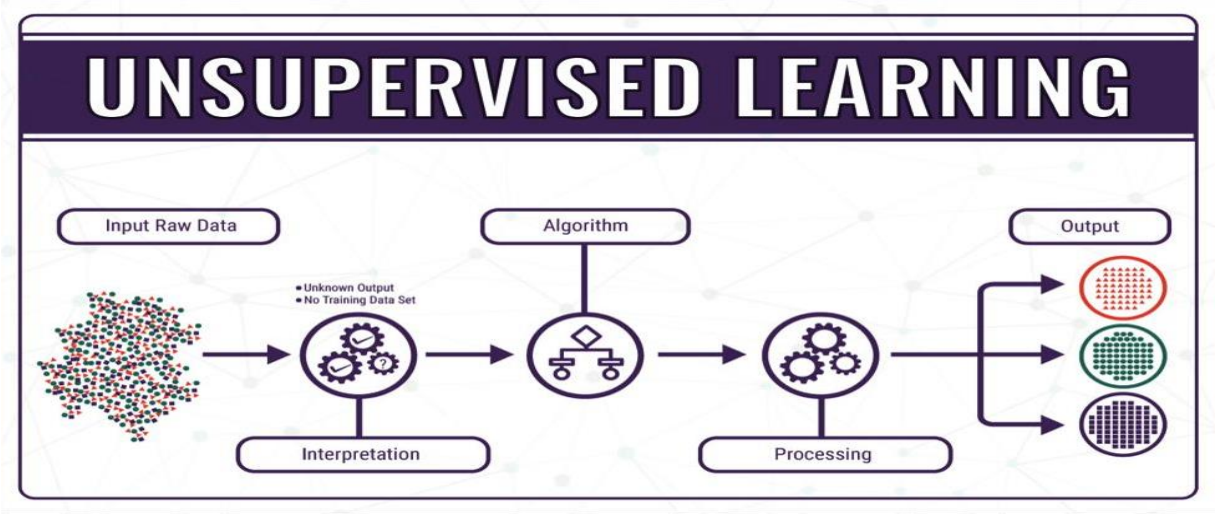

Fig.1 Un-Supervised Machine Learning

2. Literature Review

A class value can be explained or predicted by classification algorithms. Many AI applications rely on classification, but e-Commerce applications particularly benefit from it. Classification algorithms, for instance, can assist in predicting whether a customer will purchase a product. In this instance, "yes" and "no" are the two categories. Items can be categorized into a large number of categories with the help of classification algorithms, which are not limited to just two classes.

Because clustering techniques aim to group or group observations with similar characteristics, we fall into untrained ML with them. The output information is not used for training in clustering methods; rather, the algorithm determines the output. We can only see how good the solution is using visualization in clustering methods.

K-Means, where K is the number of clusters chosen by the user, is the most widely used clustering method. It is important to keep in mind that there are a number of methods for choosing the value of K, such as the elbow method.) assigns each data point closest to the center, which is generated at randoms.

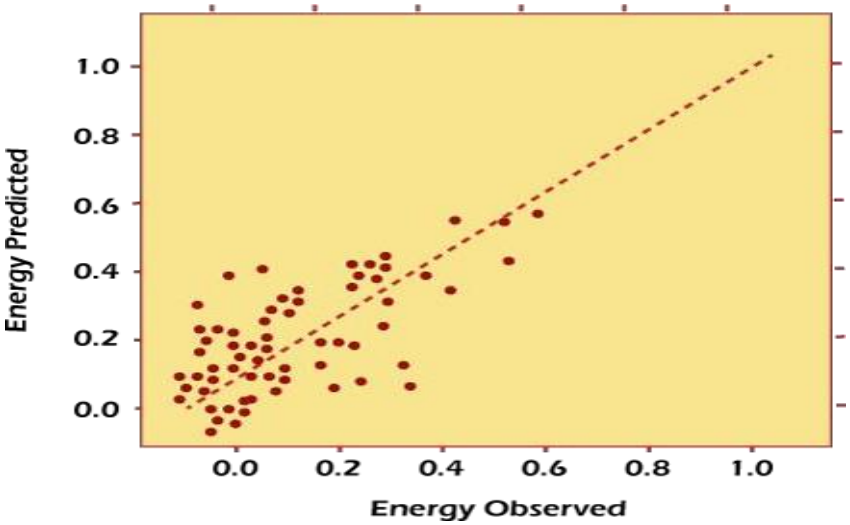

Fig.2 Regression Model

3. Proposed System

We use dimensionality reduction to get rid of the data set's least important information (sometimes columns that don't matter). For instance, images may have thousands of pixels that don't matter to your analysis. Or, during the manufacturing process, thousands of measurements and tests may be applied to each microchip, many of which yield redundant data. To make the data set manageable in these situations, you need a dimensionality reduction algorithm.

Imagine that you've made the decision to build your own bicycle because you're dissatisfied with the options you've found in stores and online. The bike you build with these great components will last longer than any other option. The concept of combining multiple predictive models—also known as supervised machine learning—is used by each model to produce superior-quality predictions.

The Random Forest algorithm, for instance, is an ensemble method that combines multiple decision trees that have been trained using various data set samples. Consequently, the quality of predictions made by a random forest is higher than that made by a single decision tree. Data structures and the most useful business insights. Here are some methods to use when working on projects that require machine learning. The prepared data's high quality does not always produce the expected outcomes.

Preparing data frequently necessitates domain expertise. Since data science engineers create products in a variety of domains, it is unlikely that they are well-versed in all of them. Given this, this data is incorrect, regardless of the high quality of the collected data. When processing incorrect data, an ML model will not benefit business optimization.

Utilizing unsupervised machine learning methods is a good way to avoid misconceptions. The idea is to let the ML algorithm label the data for you instead of labeling it yourself. It is possible to identify not only the well-known data categories but also the hidden ones by employing algorithms for dimensionality reduction, data clustering, anomaly detection, and

association mining

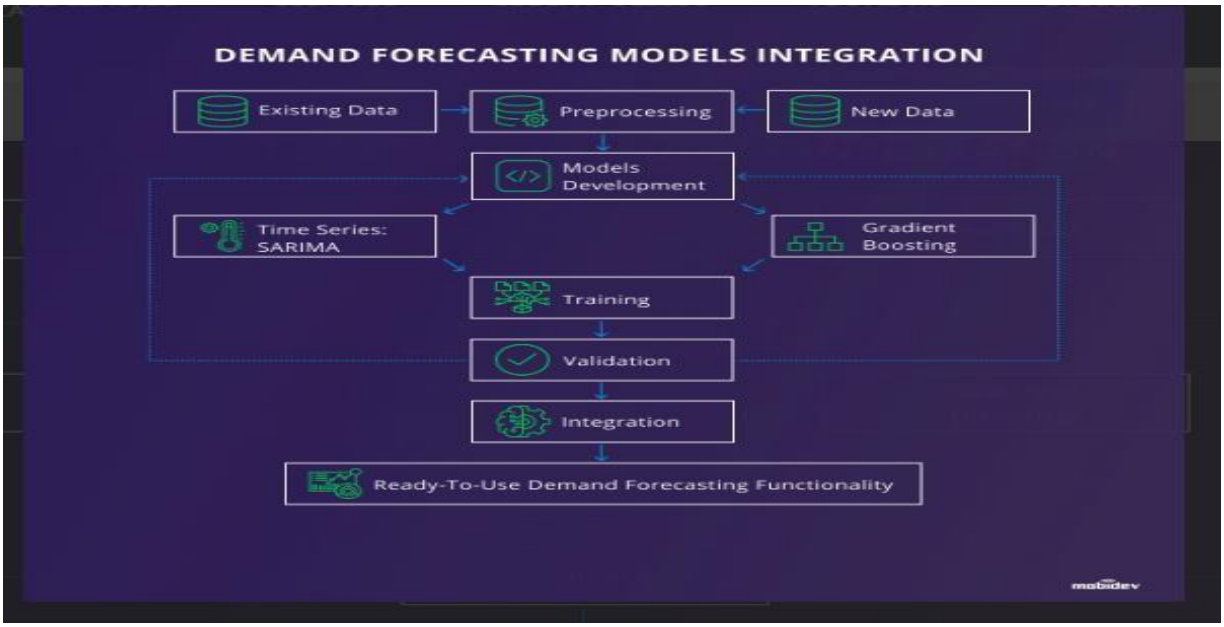

Fig.3 Proposed Neural network

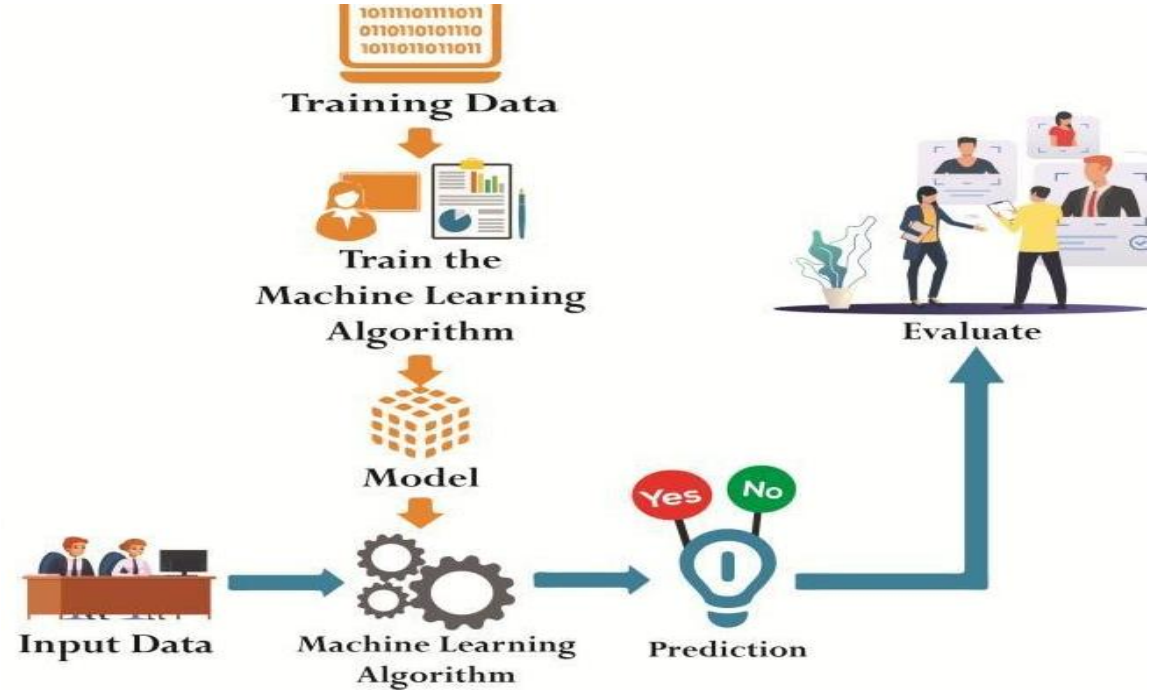

Fig.4 Proposed architecture

The goal of training is to consistently make accurate predictions. The test data are used to test the behavior of machine learning models and see how well the machine can predict new answers. The model was then built using a variety of machine learning algorithms. The user can provide the system with new input data after building the model. In addition, the user can examine the outcome using their preferred algorithm. The system produces two types of output: one is a graphical representation and the other is in polar form in the "Yes" or "No" format. The system also explains the reason for the attrition after evaluating the results

4. Conclusion

It is possible to draw the conclusion that employee turnover is not a natural phenomenon from the perspective of the employees. There are a number of factors that contribute to employee turnover. Employee turnover is observed to be caused by both internal and external factors. Location and work-life balance are important among the opportunities for growth and promotion outside of the workplace. Compensation, work schedules and shifts, working conditions, relationships with supervisors and managers, opportunities to utilize skills, and workload are all important internal factors. It is possible to draw the conclusion that the primary factors that contribute to an increase in employee turnover are internal to the organization from the perspective of managers. Although external factors also have an impact, the company's management can concentrate on altering internal factors to increase employee retention because it does not have any control over external factors

References

1. SaiyadainMirza S., TriptiPande Desai,(Emerging Issues in Human Resource Management ,An Indian perspective , Chapter 1, Excel Books , New Delhi 2007 p 5). https://www.educba.com/employee- attrition/The author describes that lack of clarity of the policies, discrimination, and lack of effective management are the major factors of employee attrition.
2. Dessler Garry (A framework of Human Resource Management Third edition, Pearson Education, New Delhi, 2008, P 642-643). https://www.slideshare.net/MohdSahil1/predicting-employee-attrition.The author describes that job stress is the main reason for employee attrition it can cause many problem to employees as well as the organizations. It may be due to internal problems, external problems or personal problems. It is also a reason of employee attrition in organizations
3. . W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Vaesens,"New insights into a churn prediction in the telecommunication sector. An profit driven datamining approach," European © 2020 JETIR February 2020, Volume 7, Issue 2 www.jetir.org (ISSN-2349-5162) JETIR2002420 Journal of Emerging Technologies and Innovative Research (JETIR) www.jetir.org 104 journal of operational research, vol. 218, no. 1, pp. 211-229, 2012.
4. K.Coussement and D. VandenPoel, "Integrating the voice of customers through call center emails into a decision support system for attrition prediction," Information & Management, vol. 45, no. 3, pp. 164– 174, 2008.
5. C.-P. Wei and I.-T. Chiu, "Turning telecommunications call details to attrition prediction: a data mining approach," Expert systems with applications, vol. 23, no. 2, pp. 103– 112, 2002. [4]. K. Coussement and
6. D. Van den Poel, "Attrition prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," Expert systems with applications, vol. 34, no. 1, pp. 313–327, 2008.
7. J. Burez and D. Van den Poel, "Handling class imbalance in customer attrition prediction," Expert Systems with Applications, vol. 36,no. 3, pp. 4626– 4636, 2009.

8. C.-F. Tsai and M. Y. Chen, “Variable selection by association rules for customer attrition prediction of multimedia on demand,” Expert Systems with Applications, vol. 37, no. 3, pp. 2006–2015, 2010.
9. K. Coussement, D. F. Benoit, and D. Van den Poel, “Improved marketing decision making in a customer attrition prediction context using generalized additive models,” Expert Systems with Applications, vol. 37, no. 3, pp. 2132–2143, 2010