

Asian Journal of Research in Computer Science

Volume 17, Issue 12, Page 58-65, 2024; Article no.AJRCOS.127244 ISSN: 2581-8260

# Employee Attrition Prediction Based on Gradient Boosting Approach

# Sunil Bhutada <sup>a\*</sup>, K Rajya Lakshmi <sup>a</sup> and G Rajaramesh <sup>a</sup>

<sup>a</sup> Department of Information Technology, Sreenidhi Institute of Science and Technology (Autonomous), Hyderabad, India.

# Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/ajrcos/2024/v17i12529

**Open Peer Review History:** 

This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/127244

Original Research Article

Received: 23/09/2024 Accepted: 25/11/2024 Published: 06/12/2024

# ABSTRACT

In today's organizational landscape, predicting employee attrition has emerged as a critical concern. The departure of trained, technical, and pivotal staff members poses significant challenges, including financial setbacks incurred in their replacement. To address this, organizations harness current and historical employee data to discern prevalent attrition triggers. Employing established classification methodologies such as Decision Tree, Logistic Regression, Random Forest, Support vector machine, and Gradient boosting Algorithms are constructed using human resource data. Leveraging feature selection techniques, these models facilitate proactive measures to mitigate attrition risks. By accurately forecasting attrition, companies not only fortify their workforce stability but also enhance economic resilience through diminished human resource expenditures. This proactive approach not only aids in retaining valuable talent but also fosters sustainable growth by fostering an environment conducive to employee retention and organizational stability.

\*Corresponding author: E-mail: sunil.b@sreenidhi.edu.in;

*Cite as:* Bhutada, Sunil, K Rajya Lakshmi, and G Rajaramesh. 2024. "Employee Attrition Prediction Based on Gradient Boosting Approach". Asian Journal of Research in Computer Science 17 (12):58-65. https://doi.org/10.9734/ajrcos/2024/v17i12529. Keywords: Employing attrition; machine learning; gradient boosting; prediction; feature selection.

# **1. INTRODUCTION**

Employee retention is a multifaceted issue influenced by various factors such as workplace environment, gender and pay equity and personal circumstances like family relocation or health concerns (Kaur and Vijay, 2016). Attrition poses significant challenges for organizations, especially when skilled employees depart for better opportunities. This leads to financial losses and disrupts workflow. Analysis of present and historical employee data helps identify patterns and underlying causes of attrition, enabling proactive measures for retention. Understanding the complexities surrounding emplovee departures is crucial for implementing effective foster strategies to а supportive work environment and mitigate turnover, ultimately ensuring organizational stability and success (Shaik, 2019).

Identifying employee attrition aids in predicting and addressing retention issues. Utilizing this data allows organizations to mitigate attrition rates and implement strategies for employee retention.

In our analysis, we employ various data classification methodologies to effectively address employee attrition. Decision Tree, characterized by its tree structure comprising branches, root nodes, and leaf nodes, facilitates attribute testing to determine class labels. Naive Bayes, leveraging the Bayes Theorem, assumes independence among features, allowing for classification based on individual feature probabilities (Shiva and Subhani, 2019) For instance, in classifying fruits, features like color, shape, and size contribute independently to the probability of a fruit being identified as an apple. Logistic Regression, a statistical technique, evaluates datasets with independent variables to predict outcomes. By utilizing these methodologies, we can accurately identify patterns and factors contributing to employee attrition, enabling organizations to implement targeted retention strategies (Teja et al., 2019). comprehensive approach empowers This decision-makers to proactively mitigate attrition rates and foster a supportive work environment conducive to employee retention and organizational success.

System architecture, the backbone of any software solution, undergoes meticulous design phases crucial for translating requirements into tangible systems. The design phase initiates the transformation from the problem domain to the solution domain, ensuring alignment with system requisites (Butada and Shaik, 2020). lts significance lies in its profound impact on software quality, influencing subsequent stages like testing and maintenance. Central to the design phase is the creation of a blueprint document outlining the system's structure, akin to an architectural plan guiding implementation, testing, and maintenance endeavors. This phase typically bifurcates into both detailed and conceptual system design (Reddy et al., 2020).



Fig. 1. System architecture

Detailed Design, is known as high-level design, delineates the system's modules, their specifications, and their interrelations to achieve desired outcomes. Conceptual System Design delves deeper into the internal logic of each module outlined in System Design. While System Design focuses on module identification, Detailed Design delves into logic planning for each module (Reddy et al., 2021).

The following paper provides the literature review in section 2. Section 3 presents the methodology. Section 4 states the results and analysis. The final section concludes the paper.

# 2. LITERATURE REVIEW

Various software development tactics focus on the entire life cycle of systems development (the development life cycle), which includes the full process of creating or redesigning systems (Mamatha et al., 2021). It involves systematic stages from initiation to maintenance, guiding engineers through planning, design, implementation, testing, and deployment. SDLC ensures structured, efficient development and optimization of systems (Reddy et al., 2021).

Predictive analysis in current mechanisms often employs dictated data mining techniques. Knowing that it is prohibitively costly and challenging to hire back qualified workers, employee attrition exerts a significant financial, operational, and resource burden on firms, underscoring the necessity for innovative retention strategies (Vijayalakshmi and Shaik, 2023).

Standard attrition identification advances by employing unstructured text data, which boosts scientific performance. using determining the presence of potential customers with greater dropout risk, this study aids those with marketing authority make intelligent decisions that will increase the sustainability of client tactics and profitability (Reddy et al., 2021).

Attrition among employees can be a major problem for businesses, especially when highly trained, technical, and critical employees leave for a better opportunity elsewhere. The increased interest in machine learning among company leaders and call centers necessitates that researchers investigate its application within businesses (Shaik et al., 2023). One of the most serious issues confronting business owners is the loss of talented employees. In a lot of

research papers, there are machine learning models developed using various algorithms like Machine. Support Vector Decision trees. XGBoost, K Nearest Neighbors, ANN, Random Forest, etc. Machine learning has been used to predict employee behavior in several types of research. To predict employee performance, the authors employed decision trees (ID3 C4.5) and the Naive Bayes classifier in their research. They discovered that task title was the most important factor, while age had no discernible effect. The authors used a dataset of 1585 records and 27 characteristics to test manifold data mining techniques for predicting staff churn. They employed Bayes algorithms, SVM, logistic regression, decision trees, and random forests as machine learning methods. For all of these findings, a support vector machine with an accuracy of 85.12 percent should be considered.

# 3. METHODOLOGY

# 3.1 Gradient Boosting Algorithm

Gradient Boosting is a formidable boosting handle that turns consecutive weak learners into strong learners. Each new model is trained using gradient descent to minimize the loss function, such as the average squared error or entropy cross of the prior model. According to the current ensemble's predictions, the algorithm calculates the gradient of the loss function in every loop. It then trains a new weak model to minimize this gradient. After adding the new model's predictions to the ensemble, the procedure goes through until a stopping specification is established (Vijayalakshmi and Shaik, 2023).

More research in the field of attrition may be found. Because the strategy to forecast employee attrition is quite similar to erosion, it enables us to predict alternative ways" (Shaik et al., 2023) In (Fernandes and Pinto, 2023), combining various training previous observations per employee from Training Data improves the predicted performance of retention models compared to using simply the most relevant data. Another issue is that instead of obtaining several samples from the whole term of the individuals, they limit it to a small piece of data, implying that many jobs are once again eliminated.

#### 4. RESULTS AND ANALYSIS

Employing established classification methodologies such as the Gradient boosting Algorithm is constructed using human resource data. Leveraging feature selection techniques, these models facilitate proactive measures to mitigate attrition risks. By accurately forecasting attrition, companies not only fortify their workforce stability but also enhance economic resilience through diminished human resource expenditures. This proactive approach not only aids in retaining valuable talent but also fosters sustainable growth by fostering an environment conducive to employee retention and organizational stability.

For implementation analysis, the data set is gathered from the Kaggle database, an

open-access repository. Then trained data set Machine Learning models using the k-fold validation methodology, using 75 percent 25 percent dataset splits. XGBoost is the machine learning algorithm employed in research to pick the most accurate model out of all of them and compare their accuracies".

Initially, the null values are checked in the dataset with particular functions. Then the outcome is depicted in the shape of a warmth map that showcases the null values in the dataset pictorially.



Fig. 2. Heat Map



Fig. 3. Count of attritions

Bhutada et al.; Asian J. Res. Com. Sci., vol. 17, no. 12, pp. 58-65, 2024; Article no.AJRCOS.127244

Satisfaction_level	1	0.11	-0.14	-0.02	-0.1	0.059	-0.39	0.026	0	.8
Last_Evaluation	0.11	1	0.35	0.34	0.13	-0.0071	0.0066	-0.0087		
Number_of_projects	-0.14	0.35	1	0.42	0.2	-0.0047	0.024	-0.0061	0	.4
Average_monthly_hours	-0.02	0.34	0.42	1	0.13	-0.01	0.071	-0.0035		
Daily_wourking_hours	-0.1	0.13	0.2	0.13	1	0.0021	0.14	0.067	0	1.0
Work_accident	0.059	-0.0071	-0.0047	-0.01	0.0021	1	-0.15	0.039		0.4
Employee_left	-0.39	0.0066	0.024	0.071	0.14	-0.15	1	-0.062		
Promotion	0.026	-0.0087	-0.0061	-0.0035	0.067	0.039	-0.062	1	1	0.8
	Satisfaction_level	Last_Evaluation	Number_of_projects	Average_monthly_hours	Daily_wourking_hours	Work_accident	Employee_left	Promotion		

Fig. 4. Correlation matrix

@ Tone	casting Job Holder Abrasio-	× 🕀		the second second second second second	and the second se	-	
€ →	C @ 127.0.0.1.50	00/index			\$		1 2 3 9 4 9 3 4 1
H Apps	🔛 Menu Maker 🔮 J	L Pen by Captain A. 🛛 🕄	) ACT 🤠 Disa	Sie Past date ti 🚦 JAVA Web Applicati 🏦 regex - Re	egular Esp., 🤰 How can Tallow onL.	🛛 datatest/html/cht.	
		Forecast	ing Jol	Holder Abrasion Us	ing Datamini		ue
	10000						
	Team Leader	La Employee	20 mk	A Manager		- Keguter - Q Log	
				Login Form	×		
				User Type.			
				Manager			
				Username			
				manager			
				Password.			
				Loin			
					Close		
					_		
	o 🖿 🛒		4	W CONTRACTOR OF	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	The state	+ 1 (1 455 PM

Fig. 5. Login page



Fig. 6. Given credentials

S For	ecasting Job Holder Abrasio	× +		Concerning Street, or other				- 0 - ×
← → III Apps	C ① 127.0.0.1-50	IOD/Index A Pen by Captain A 🧟	ACT 🍓 Disa	able Past date ti 📑 JAVA Web Applicati	🚖 regex - Regular Exp 🏦 How can I allow	nh., 📀 🖬 😳	ni (ht	1 🔍 1
		Forecast	ing Jol	b Holder Abrasio	n Using Datami	ning Tec	hnique	
#	🔓 Team Leader	<b>11</b> Employee	20 HR	🚨 Manager		2+ Register	+) Login	
				Registra	tion Form 🗙			
				User Type	Name			
				Team Leader 0	Enter Name			
				Enter Email Id	Enter password			
				Mobile Number				
				Enter Mobile Number				
				Register				
				_				
					Close			
		1 2 6	-	w	a state	Section of the sectio	- 13 40	458 PM

Bhutada et al.; Asian J. Res. Com. Sci., vol. 17, no. 12, pp. 58-65, 2024; Article no.AJRCOS.127244

Fig. 7. Fill Registration form

Forecasting	Job Holder Abrasio 🗙 +		-				
	127.0.0.1:5000/hrvedetails	🔿 ACT 🌲 Disable Past date ti-	IAVA Web Applicati	🤌 renex - Repular Fan. 🧳 Ho	🖈 😊 🖬		e 🗣 🗛 🤪 i
			E sava neo appresan	Teges negota copt. 2 in		exercised energy	
	Forecas	ting Job Hold	ler Abrasic	on Using Dat	amining T	echnique	
🛠 Home	Employee Details	View Projects 💶 Assig	n Employee Details	Calculate Attrition	(+ Logout		
			Emple	waa Liet			
			Latipic	Syce List			
	EMPLOYEE NAME	MOBILE NUMBER	PROMOTION	DEPARTMENT	SALARY	WORKACCIDENTS	
	Shive	7978418005		7	50000	1	
	ginija	7451236544			1000		
<u> (</u>			1.000	and the second second	the second second	A CONTRACTOR	• ☎ €0 5:00 PM 09-Dec-19

# Fig. 8. Employee list



Fig. 9. Adding Employee details to the form



Bhutada et al.; Asian J. Res. Com. Sci., vol. 17, no. 12, pp. 58-65, 2024; Article no.AJRCOS.127244

Fig. 10. Employee attrition



Fig. 11. Suggestion to HR

# **5. CONCLUSION**

Employee attrition poses significant challenges for organizations, leading to financial, temporal, and effort-related losses. The departure of trained and experienced personnel creates gaps that are hard to fill swiftly and economically. impact, organizations Recognizing its increasingly rely on analyzing past and current employee data to predict future attrition and understand turnover drivers. Through data extraction algorithms, accurate predictive models for attrition emerge, aiding in proactive retention strategies. This process extends beyond merely distinguishing potential leavers from stayers. Utilizing advanced data analysis techniques enables organizations to assign attrition probabilities to individual employees, facilitating targeted retention efforts and fostering a more stable workforce.

Future enhancements for the system entail continual adaptation to evolving user needs and technological advancements. Upgrades will ensure adaptability to changing environments and emerging technologies. Enhanced security measures, such as implementing single sign-on solutions, will address future security concerns effectively. Regular updates and advancements will maintain the system's relevance and functionality in an ever-changing landscape, ensurina optimal performance and user satisfaction.

#### DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative Al technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of this manuscript.

# **COMPETING INTERESTS**

Authors have declared that no competing interests exist.

# REFERENCES

- Butada, S., & Shaik, S. (2020). IPL match prediction using machine learning. *IJAST*, *29*(5).
- Fernandes, S., & Pinto, A. (2023). Predicting employee turnover with advanced machine learning: A gradient boosting perspective. *Journal of Organizational Behavior*, *44*(5), 789-805.
- K, V., & Shaik, S. (2023). Predicting employee attrition methodologies of K-fold technique. *I. J. Mathematical Sciences and Computing*, 1, 23-36.
- Kaur, S., & Vijay, R. (2016). Job satisfaction A major factor behind attrition or retention in retail industry. *Imperial Journal of Interdisciplinary Research*, 2(8).
- Keertan J, S., & Shaik, S. (2019). Machine learning algorithms for oil price prediction. International Journal of Innovative Technology and Exploring Engineering, 8(8).
- Mamatha, M., Datta, S., & Shaik, S. (2021). Fake profile identification using machine learning algorithms. *International Journal of Engineering Research and Applications* (IJERA), 11(2).
- Reddy, R. V., Shaik, S., & Rao, B. S. (2021). Machine learning based outlier detection

for medical data. *Indonesian Journal of Electrical Engineering and Computer Science*, 24(1).

- Reddy, R. V., Shaik, S., Chandra, G. R., & Rao, B. S. (2020). Breast cancer prediction using classification techniques. International Journal of Emerging Trends in Engineering Research, 8(9).
- Reddy, S., Sri, R., & Shaik, S. (2021). Sentimental analysis using logistic regression. International Journal of Engineering Research and Applications (IJERA), 11(2).
- Shaik, S. (2019). Analysis of photo plethysmography signals with artificial networks neural using curvelet transform. Journal of Advanced Research in Dvnamical & Control Systems. 11(1).
- Shaik, S., Kumar, P. S., Reddy, S. V., Reddy, K. S., & Bhutada, S. (2023). Machine learning based employee attrition predicting. Asian Journal of Research in Computer Science, 15(3), 34-39.
- Teja, K. P., Reddy, V., & Shaik, S. (2019). Flight delay prediction using machine learning algorithm XGBoost. *Journal of Advanced Research in Dynamical & Control Systems*, 11(5).
- Zhang, Y., & Li, J. (2023). Utilizing gradient boosting for employee attrition prediction: An empirical study. *Data Science and Management Review*, *5*(3), 200-210.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the publisher and/or the editor(s). This publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© Copyright (2024): Author(s). The licensee is the journal publisher. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history: The peer review history for this paper can be accessed here: https://www.sdiarticle5.com/review-history/127244