Webology (ISSN: 1735-188X)
Volume 19, Number 2, 2022

Employee Retention And Attrition Analysis: A Novel Approach On Attrition Prediction Using Fuzzy Inference And Ensemble Machine Learning

M. K. Sharma[1] , Dhanpal Singh[2] , Manaswita Tyagi[3] , Ayushi Saini[4] , Nitesh Dhiman[5] , Riddhi Garg[6]

[1]Department of Mathematics, Chaudhary Charan Singh University, Meerut-250004, India.

[2]Department of Mathematics, Keshav Mahavidyalaya, University of Delhi, Delhi-110034, India.

[3]Department of Mathematics, Keshav Mahavidyalaya, University of Delhi, Delhi-110034, India.

[4]SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamil Nadu- -603 203, India.

[5]Department of Mathematics, Chaudhary Charan Singh University, Meerut-250004, India.

[6]Department of Mathematics, IFTM University, Mordabad, 244102-India.

Abstract

In Today's world, AI has become an essential tool for achieving and creating the unthinkable. It is helping in creating innovative solutions for almost every industry there is. In the wake of this ever-growing demand of computerized intelligence, the analysis of automation by HR (Human Resources) in the form of predictive form can be formulated. As the Human Resource departments is solely responsible for recruiting and bring valuable talent to the industry, it becomes essential that this task is done with maximum efficacy. Through this project, we intend to predict which employee would prefer a job change and which employee would stay in a company and hence, help constitutes as an active research domain is how AI based intelligence can be interpreted and utilized assess the input resources required to put in an employee. The present techniques are in sufficient to deal the current uncertainties caused due to confliction in their nature. In order to work on this, we propose using natural language processing, opinion mining, fuzzy logic and various widely used classifiers namely Random Forest (RF), Cat Boost Classifier, Support Vector Machine (SVM), and Naïve Bayes (NB). We also used Mamdani based fuzzy inference system using nine input and out output, for the attrition prediction in a company. In future, this model can be improved further by incorporating more data.

Keywords

Human Resource Management, Human Resource Analytics, Classification algorithms, Machine Learning, Model Validation.

Introduction

Human Resource Management (HRM) is a newer part of business administration which is responsible for the management of people. Some of the responsibilities are talent acquisition, employee attrition, employee benefits, employee compensation and more (Góes,& Oliveira, 2020). Among these tasks, acquisition presents a greater challenge as it is vital that the resources acquired by the management, bring innovation, and add more value to the company. Upon acquisition, the development and training phase begins, and it is provided at the cost, time and resource utilization of the company. At this point of time, it will benefit the management, to know, which employee will really want to work for the company upon finishing of the training. This will help the management make more optimized decisions when it comes to the efforts required to put in providing the training (Jiangjing, et. al., 2020). To be able to judge or classify the employees, many factors are to be considered related to their demographics, education, future aspirations, and experience. Classification of data must be done with maximum accuracy and efficacy, and to establish that, optimized classifiers must be used. Employee data pool serves as a major tool in HR Analytics. HR Analytics refers to a field of analytics which deal with the analysis of data of employees in a company to help in improving employee performance and increase employee retention (Ghorbani, & Ghousi, 2020). The data in such a case serves as the fuel for the process and drives it towards achieving important insights.

Now, what makes this process more efficient is the right set of algorithms that should be used starting from the data collection, cleaning to making a model for classification. In this paper, we focus on the efficacy of various algorithms such as CART, Naïve Bayes, kNN, Random Forest, SVM, XG gradient boost and Cat Boost (Zhang, & Liu, 2019). Companies are always trying to improve their employees' performance and retention thus, building a solution that classifies employees as likely to stay and not likely to stay in a company would help in better resource planning.(Zhang,at el.,2019)proposed an Efficient method for kNN Classification with Different Numbers of Nearest Neighbors. (Mienye, et al., 2020) gavean idea for detecting fraud in medical science using the concept of carboost. (Hancock, & Khoshgoftaar, 2020)gave a prediction technique by using the ensemble learning approach.(Fayaz, et al., 2020)classified the spam of products in ensemble machine learning models. (Sevakula, & Verma,2017)presented a concept for assessing the ability of vote point. (Li, et al., 2018)explained the efficiency of ensemble learning in solar cells. (Schapire, & Singer, 1999) improved the algorithm in boosting over confidence prediction. (Shone, et al., 2018) worked on deep learning in the network detection. (Shang, et al., 2020) developed a simple stochastic method by using variance for machine learning. (Ebenuwa, et al., 2019)gave a variance ranking method for the attribute selection procedure for imbalance data in the binary classification. (Rivera-Lopez,& Canul-Reich,2018) gave an approach for getting the optimality in parallel axis decision tree by using the evolution of differential.(Samuel,et.al.,2017) gave a prediction method by using the Fuzzy AHP decision support system for heart failure prediction method which laid a very effective tool in the machine learning. (Griffethet. al., 2020) developed an analysis for the metadata to find the correlation between the antecedent and consequences. The proposed method was very useful in research implications. (Ponnuru, et al., 2020) proposed a method based on logistic regression for the prediction of employee retention. (Usha, & Balaji, 2019) analyzed the machine learning implication on employee retention. (Rombaut,& Guerry,2018)predicted a seminal work on the human resource voluntary database analysis. (Nagadevara,2008) tried to introduce the use of software in the application of data mining technique in employee retention. (Alao,&Adeyemo, 2013) analyzed the employee retention based on the decision algorithm based on fault tree analysis. (Keramati, et al., 2014)also proposed improved prediction in telecommunication based on data mining technique. (Syam,& Sharma, 2018)gave a method for sales renaissance through machine learning and artificial intelligence. (Syam, & Sharma, 2018) analyzed various application of Human artificial intelligence. (Jarrahi, 2018; Yanqing, et al., 2019) evaluated the impact of artificial intelligence in decision making. (Kaur,& Vijay,2016)analyzed the major factors in fetching the results of attrition or retention in industry. (Al-Radaideh ,& Nagi,2012)proposed prediction techniques to evaluate the performance of employees using data mining techniques.

Uncertainty, vagueness and impression present in many real-life situations. To handle such conditions (Zadeh, 1965) introduced the concept of fuzzy logic. Fuzzy logic has been used in many applications such as; pattern recognition, air conditioners (AC), washing machines, transmission systems, helicopters system, knowledge-based systems for multi-objective decision making problem and medical field. Fuzzy inference system (Jingzheng, et al.,2021) is a mathematical tool that used to interpret the input values and based on some sets of fuzzy rules, assigns output value. There are two major inference systems have used in many real-life problems first is Mamdani approach based and second is Sugeno’s approach based. Mamdani fuzzy inference (Mandani, & Assilian, 1975) was first introduced as a method to develop a control system by synthesizing a set of linguistic rules provided by experienced knowledge. Sugeno fuzzy inference(Sugeno,1985) or Takagi-Sugeno-Kang fuzzy inference uses the output value in the form of constant or a linear function of the input values. In comparison to Mamdani inference system, the Sugeno system uses a weighted average of a data points rather than compute a defuzzification process. In this present work, we will use the Mamdani's inference system for the prediction of the attrition rate of the company. During this process, we will adopt nine input and one output-based inference system.

This research article is divided into seven sections including the present section. In the second section of this paper, we have discussed many methods based on different probabilistic methods such as regression analysis, Baye's theorem, and central tendencies. We also discussed some gradient based methods. In section third of our research paper, we proposed a new methodology. This section is divided into three parts: one is data preprocessing, validation of the model and architecture of the proposed model (as shown in Fig. 1&programmatic flow shown in Fig. 2), also accuracy comparison of the proposed method is given in table 1. In section four of the paper, we presented an algorithm and block diagram for our method. The results of our approach have been calculated in section five of this research article. In this section, we analyzed on XG Boost algorithm performance. For evaluating the performance, we developed an ensemble machine learning algorithm based on decision tree. In the sixth section, we have developed a nine input and one output based fuzzy inference system for the attrition prediction. In last segment of our article, we focused on the predictive performance based on the joint combination of prediction from multiple models.

Methods

The following supervised learning methods for classification are used in this paper as the data collected is already labeled and output classes would be likely to stay (1) and not likely to stay (0). Brief explanation about the algorithms:

1. CART: Classification and Regression Trees algorithm is used for both categorical and continuous target variables. This decision tree-based algorithm is based on a tree structure and asks a yes or no question at every node. The three main features of this algorithm are splitting rules, stopping rules, and the final prediction at the ending

node. For splitting at every node, it uses the Gini Impurity which is calculated at

Likelihood Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

Posterior Probability Predictor Prior Probability

$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

every node. We are able to achieve better model interpretability by using this algorithm.

2. Naïve Bayes: Naive Bayes algorithm based on Bayes' theorem assumes independence and equality among its features. These classifiers are very useful in real life applications such as document classification. This algorithm is able to estimate parameters on the basis of small training data. These classifiers are known to be faster compared to more sophisticated methods.

3. K Nearest Neighbour: k NN algorithm can be used for both, classification and regression [5]. Under classification, the output is a labeled class. A particular data point is classified on the basis of majority votes of its nearest k neighbors. This algorithm is particularly more advantageous when it comes to more noisy and larger data sets as it is able to handle them easily.

4. Random Forest: Random Forest uses central tendencies man and mode to classify a dataset. This algorithm can be used for classification as well as regression. Multiple decision trees are created in this algorithm at the time of training and for classification, the mode is the output and for regression the mean of the trees is the output. This algorithm usually results in the highest accuracy and is also able to handle big data efficiently. It often balances datasets automatically and has methods for missing data problem.

5. Support Vector Machine: Support Vector Machine results in a hyper plane clearly classifying the given data points into different groups of similar features. The new data points are then assigned to one of the already existing groups on the basis of similarity. This algorithm is easily able to handle high dimension data with more memory efficiency as it uses the training data as a subset for decision making.

6. XG Boost: XG boost is an implementation of gradient boosting decision trees for faster and better performance. When It comes to unstructured data such as image sor text, this algorithm is known to present the best results. It has evolved from decision

trees to random forest to gradient boosting and finally to XG Boost itself. It can be used to solve ranking, prediction, regression, and classifier problems. It supports all major languages such as Python, R, C++ and JAVA. The main feature of XG Boost is that it is an ensemble of learning algorithms that apply the principle of boosting weak learners using gradient boosting.

7. CatBoost: Cat Boost is an open-sourced machine learning library based on gradient boosting. It can be used in regression, multi-class classification, classification, and ranking [6].The best part about Cat Boost, developed by Yandex, is that it handles pre-processing of categorical features itself. The problem of over fitting is also avoided by the implementation of ordered boosting. Some of the other features of this algorithm are that it provides support for missing value and is also great for data visualization.

Proposed Methodology

The data used for this paper was taken from a Kaggle repository and it is imbalanced. The data has various labeled features such as gender, relevant experience, university enrolled in, education level, size of the company, discipline majored in, company type and last time the employee had a new job. There are a total of fourteen features in which four are numeric, four alpha numeric and there st categorical. The final column is the target column with targets 1 and 0.

1. Data Pre-processing: The data acquired is raw data, which comes from various different sources, is highly imbalanced, and contains several inconsistencies. The data set consists of categorical features, missing values and has several features that have a high correlation index. We use multiple data mining techniques combined with persuasive exploratory data analysis to pre-process the data.

i. Categorical Features - The categorical features in the data set are both nominal and Ordinal. Hence, it is necessary to encode the categorical features to improve the model. We use the Label Encoding technique to convert the data from labels into a machine-readable format.

ii. Handling Missing Values- In predictive analysis, missing data can prevent the predictions for some variables hence it is necessary to impute the values for such data. Although several techniques are available to handle the missing values, we use the mean of the non- missing observations to replace the null values.

Correlation index-In order to evaluate the strength of the relationship within the various variables in our dataset we use correlation analysis. The heat map in sea born is used to visualize the correlation matrix. A correlation index can be used to check multi co linearity in our dataset.

2. Model Validation

Table 1 The accuracy comparison of proposed model with other classifiers

Classifier	Accuracy	AUC
CART	0.71	0.594774138304502
Naïve Bayes	0.76	0.6618851852680049
K Nearest Neighbour	0.74	0.5839958719658915
Random Forest	0.73	0.6022621292623996
Support Vector Machine	0.75	0.5020430109712738
XG Boost	0.78	0.6596879038165342
CatBoost	0.77	0.6475116036021141

The target variable is discrete and skewed according to our dataset as it is a binary classification problem. Hence, we use Area under the ROC Curve (AUC). AUC-ROC curve is used to measure the performance of the given model for classification problems at various threshold settings. ROC i.e., the receiver operating curve is used to measure how accurate the model is in distinguishing various classes. In our and case, the possibility of an employee leaving or not. The sci-kit library is used to split the data into training and test data sets. The test size is set as 0.25 and attributes with a low correlation index are selected in the training test datasets.

3. Architecture Diagram: The EDA is done using various packages from pyplot and sns. For data pre-processing and cleaning pre- processing module from sci-kit learn is used. The classification algorithms like XG Boost and Cat Boost are imported and other classifiers are used from sci-kit learn package in Python 3.4.

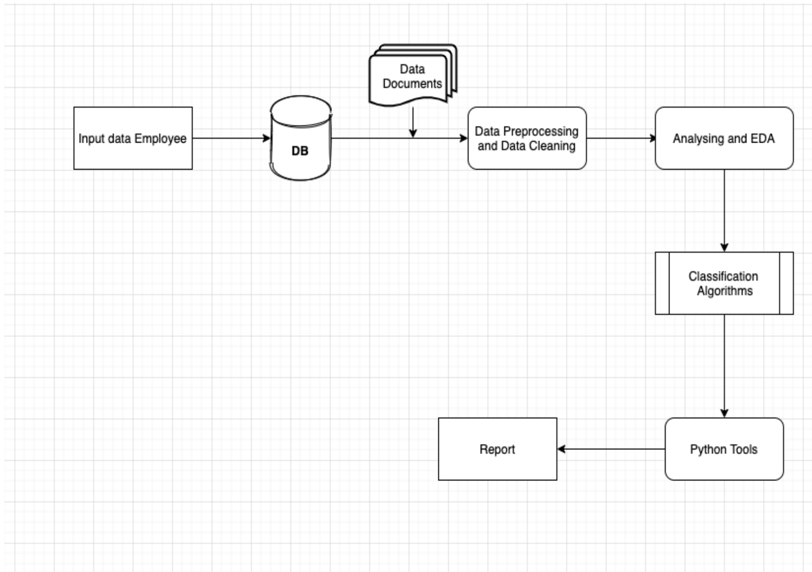


Fig. 1 Architecture Diagram for the data flow

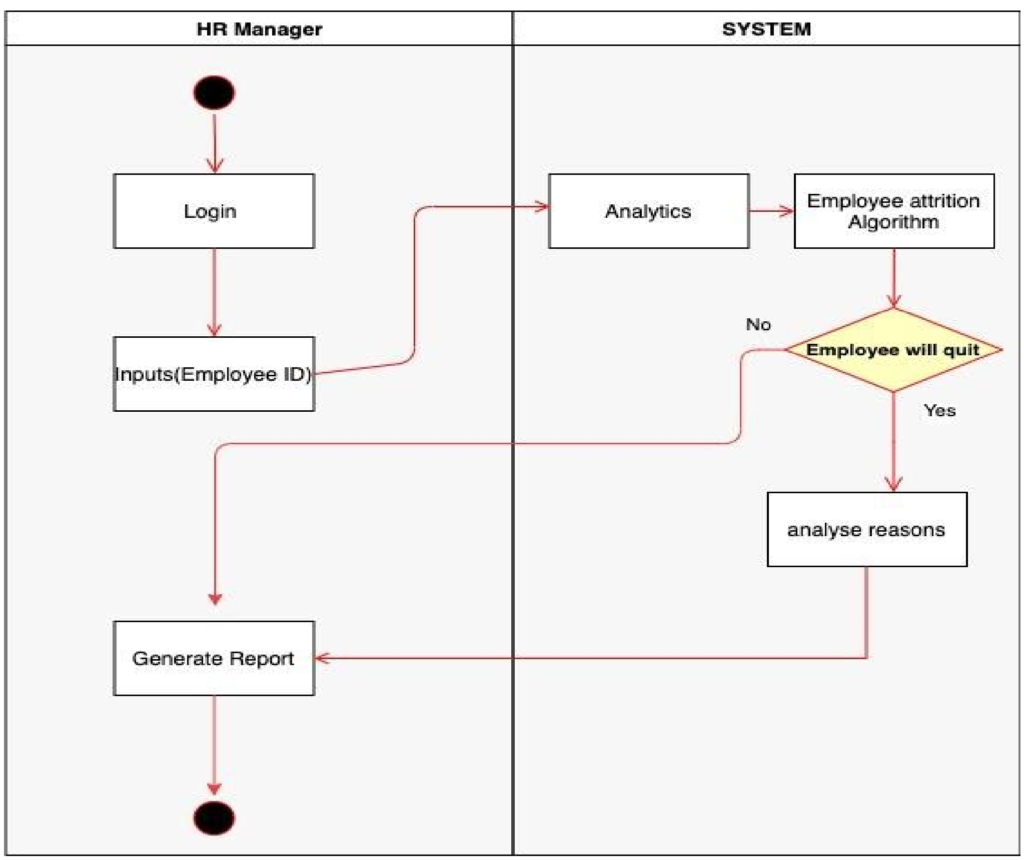


Fig. 2 Activity Diagram implementation for understanding the programmatic flow

Proposed Algorithm

In this part of the research paper, the steps of the proposed algorithm are described. The present section contains certain steps to develop the proposed algorithm to find a solution. The spans of increment or decrement impact of the present set of rules in the form of the mediative fuzzy correlation coefficient are as follows;

Step 1. Write the factors for which the employee attrition analysis is being done.

Step 2. In step second we do the data processing on the collected data and do the exploratory data analysis.

Step 3. In this step we explore the requirement of encoding. If the encoding is required then we will move

- (i) We do the categorical encoding and explore the features of engineering.
- (ii) We predict the features which will lead to
  Attrition. In this we do the implementation of XG Boost for prediction.
- (iii) After (ii) we can share the features,

Step 4. If encoding is not required then we will move in the following manner:

- (i) We analyze the principal component of our model.
- (ii) We implement the XG Boost for Attrition Prediction on the data obtained in (i). By doing that we can predict the leading features for attrition.
- (iii) We can calculate the AUC and save the model obtained by the implementation of XG Boost prediction in (ii).

1. The architecture of proposed algorithm: The architecture of the proposed algorithm can be shown in the form of chart 1 as follows;

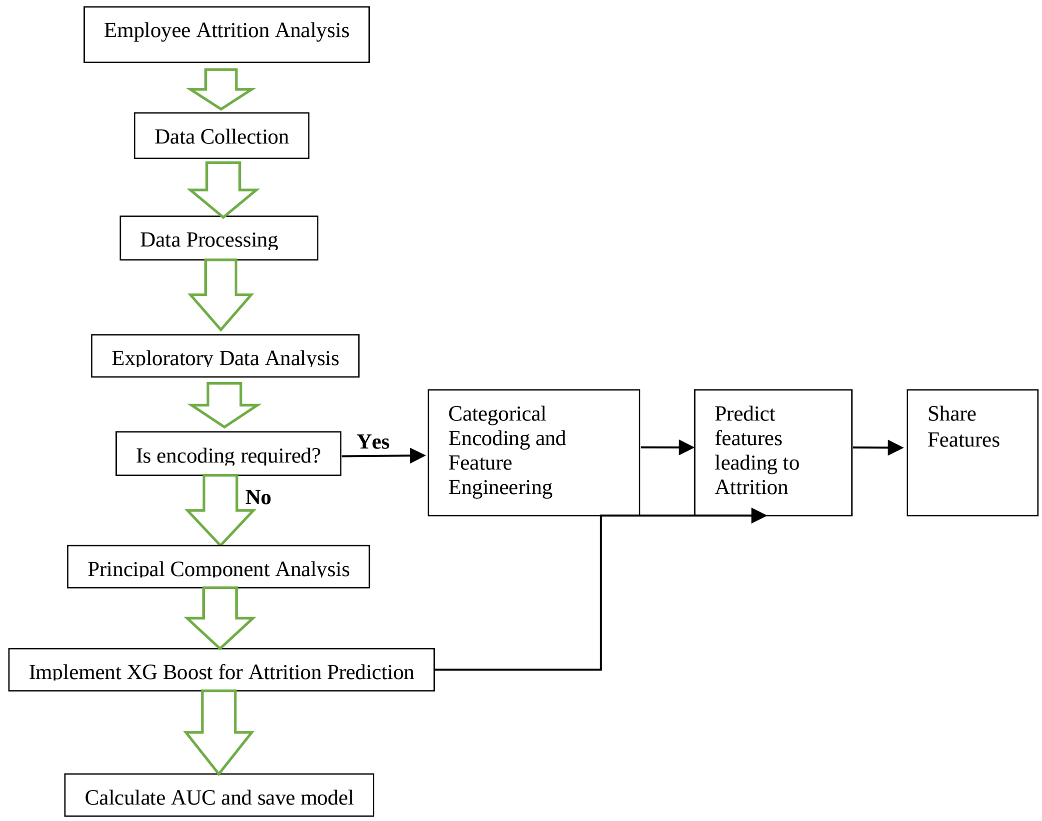


Chart 1 Architecture of the proposed Algorithm

Attrition Prediction Using Fuzzy Inference System

In this section, we consider the factors involve in a company, which affecting the attrition rate of the company, which have divided into several linguistic categories describe in table 2 as;

Table 2 Including factors, linguistic categories & their ranges

Including factors	Linguistic categories & their ranges		
CDI	Low [0, 0.3]	Medium [0.28, 0.62]	High [0.6, 1]
RE	Low [0, 0.3]	Medium [0.28,	High [0.6, 1]

		0.62]	
EU	Non-reputed [0, 0.52]		Reputed [0.5, 1]
EL	Low [0, 0.3]	Medium [0.28, 0.62]	High [0.6, 1]
MD	Non-disciplined [0, 0.52]		Disciplined [0.5, 1]
CS	Micro Entities [0, 0.3]	Medium Sized [0.28, 0.62]	Large Sized [0.6, 1]
CT	Limited company [0, 0.52]		Partnership based [0.5, 1]
CH	Low [0, 6]	Medium [5.8, 8.2]	High [8, 1]
T	Low [0, 0.3]	Medium [0.28, 0.62]	High [0.6, 1]
CA	Low [0, 0.3]	Medium [0.28, 0.62]	High [0.6, 1]

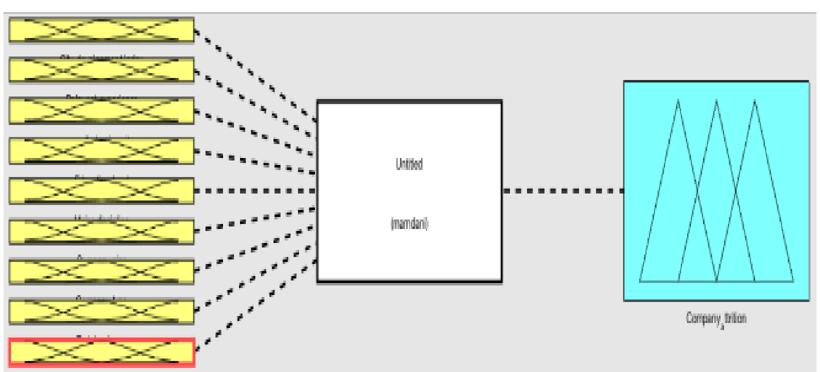


Fig. 3 Proposed fuzzy inference system

We have taken nine input and one output based Mamdani's fuzzy inference system (as shown in Fig. 3) the output factor has been divided into three linguistic categories i.e., low, medium and high which shows the company attrition prediction levels. In this work, we have taken only such rules, which gives major variation in the output of the proposed fuzzy inference system. The fired fuzzy rule of the system is described in table 3.

Table 3 Fuzzy fired rules for the inference system

	IF									THEN
Rule	CDI	RE	EU	EL	MD	CS	CT	CH	T	CA
1	L	L	NR	L	ND	ME	PB	H	H	H
2	M	M	R	M	D	MS	LM	M	M	M
3	H	H	R	H	D	LS	LM	L	L	L
4	M	M	NR	L	ND	ME	PB	H	H	L
5	L	M	NR	L	ND	ME	PB	H	H	H
6	M	M	NR	L	ND	ME	PB	H	H	H
7	L	L	R	H	D	LS	LM	H	H	H
8	H	H	R	H	D	LS	PB	M	L	M
9	H	L	R	M	ND	LS	LM	L	L	L
10	H	L	R	M	ND	LS	LM	L	L	L

CDI: City Development Index
RE: Relevant Experience
EU: Enrolled University
MD: Major Discipline
CS: Company Size
CT: Company Type
CH: Company Hours
T: Target
CA: Company attrition
L: Low
M: Medium
H: High
NR: Non-Reputed
R: Reputed
ND: Non-Disciplined

D: Disciplined
PB: Partnership Based
LM: Limited Company
LS: Large Sized
ME: Micro Entities
MS: Medium Sized

The output results of the fuzzy inference system are described in the table 4-6. This gives low, medium and high-risk categories for the prediction of attrition.

Table 4 Lower Company attrition rate

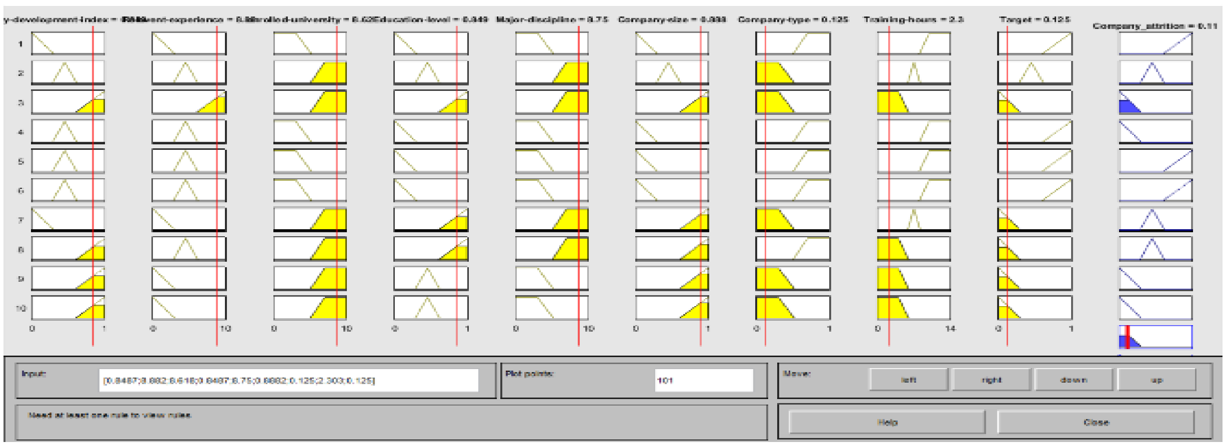


Table 5 Medium company attrition output

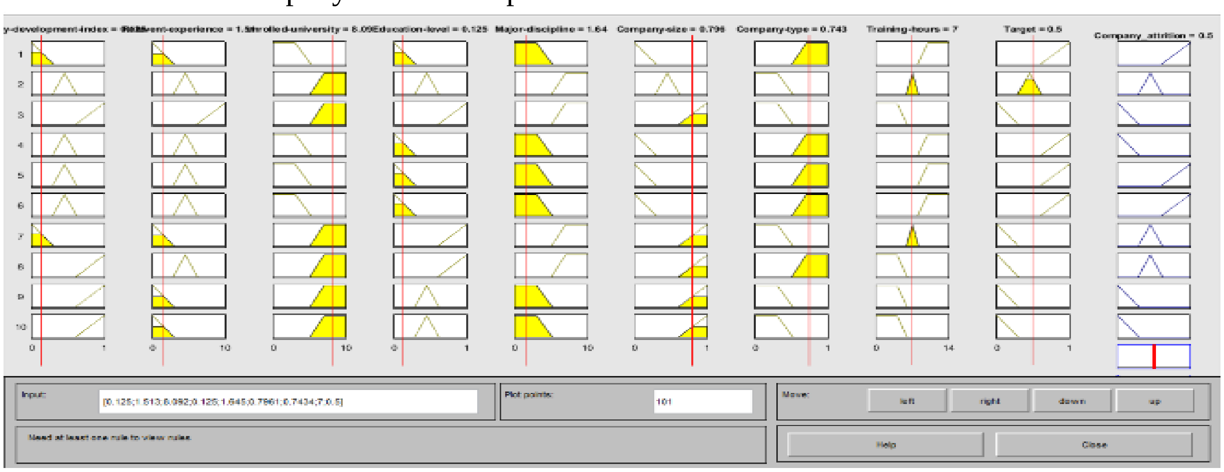
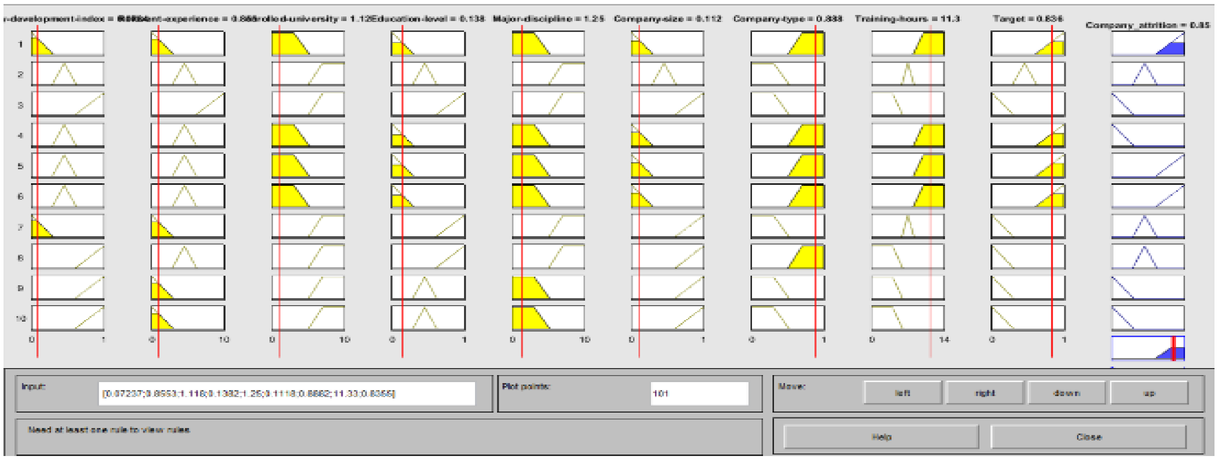


Table 6 Higher company attrition rate



Results

From the given table it can be noted that XG Boost Algorithm performs better than the other classification algorithms. XG Boost is a decision-tree-based ensemble Machine Learning Algorithm. It is a highly optimized gradient boosting algorithm that handles missing values and regularization to reduce over fitting and bias.

```
from sklearn.naive_bayes import Gaussian NB
from sklearn.ensemble import Random Forest
Classifier from sklearn. neighbors import
KNeighbors Classifier from sklearn.svm
import SVC
from sklearn.tree import Decision Tree
Classifier from xg boost import XGB
Classifier
from sklearn.ensemble import Gradient
    Boosting Classifier
from cat boost import Cat Boost Classifier
models = [] models. Append ((’Naive Bayes’,
Gaussian NB())) models. Append ((’KNN’, K
Neighbors Classifier ())) models. Append
((’SVM’, SVC(gamma=’auto’))) models.
append((’XG Boost’, Gradient Boosting
Classifier
    ()))
models. append (("Cat Boost", Cat Boost
Classifier())) models.append ((’CART’,
```

```
Decision Tree Classifier()))#
    evaluate each model
in turn results = []
names = []for name, model in
         models: model.fit
         (X_train, y_train)
         y_pred =
         model.predict(X_val)
         accuracy = accuracy_score(y_val,
         y_pred) print("{} : {}".format(name,
         accuracy))
```

Here we imported the six classifying algorithms to create a model with the best efficacy within optimised time. We then print the results with accuracy score to find out the best performing algorithm.

```
clf = XGB Classifier()# A parameter grid for XG Boost
params = {
         'min_child_weight': [1, 5, 10],
         'gamma': [0.5, 1, 1.5, 2,5],
         'subsample': [0.6, 0.8,1.0],
         'colsample_bytree': [0.6, 0.8, 1.0],
         'max_depth': [3, 4, 5]
         }random_cv=Randomized Search CV (estimator=clf,
             param_distributions=params,
```

We then print the results with accuracy score to find out the best performing algorithm. We then create a parameter grid for XG Boost only to apply randomized search and cross check the results. We also use the random_cv.fit() function to train and fit our data in the predictive classifier model.

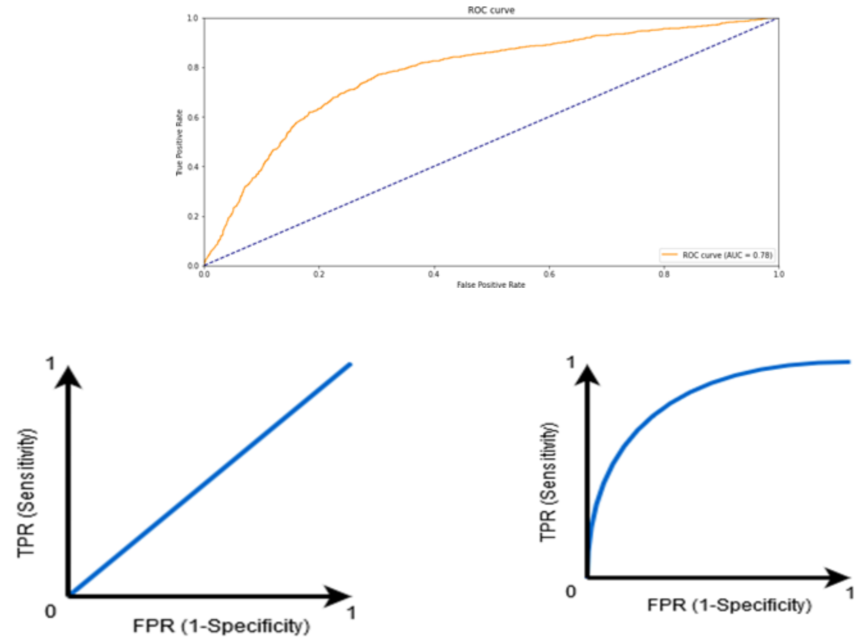


Fig. 4 The AUC for XGB Classifier

Then AUC for XGB Classifier given in Fig. 3.The target variable is discrete and skewed according to our dataset as it is a binary classification problem. Hence, we use Area under the ROC Curve (AUC). AUC-ROC curve is used to measure the performance of the given model for classification problems are at various threshold settings. ROC i.e., the receiver operating curve is used to measure how accurate the model is in distinguishing various classes. In our case, the possibility of an employee leaving or not is given as:

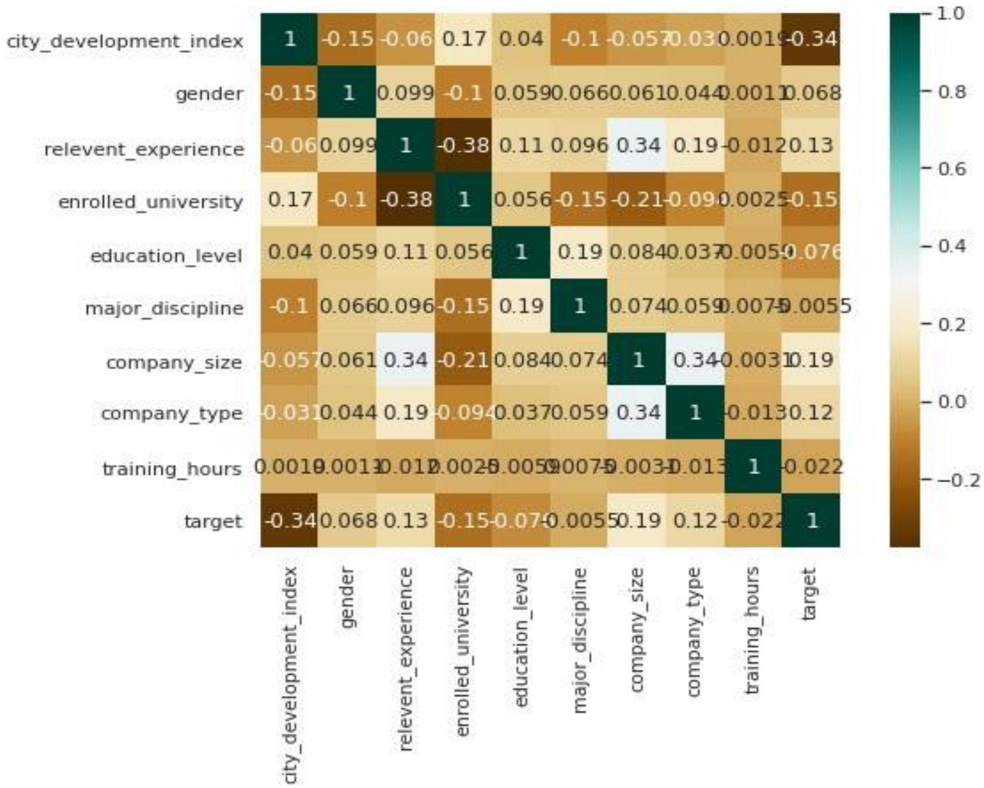


Fig. 5 Heat Map for Correlation Analysis

In order to evaluate the strength of the relationship within the various variables in our dataset we use correlation analysis. The heat map in seaborne is used to visualize the correlation matrix. A correlation index can be used to check multi co linearity in our dataset. A heat map (heat map for correlation analysis is given in Fig. 4) is a two-dimensional data visualization technique that depicts the extent of a phenomenon as color. The difference in colour can be by hue or intensity, providing visible visual clues to the reader on how the trend is grouped or varying over distance.

Then we also plot the area under the curve graph to find out the robustness of our classifier. Here we obtain a value = 0.93 which indicates that the classifier is remarkably robust and should be trusted with its results.

Results Interpretation and Conclusion

Thus, it can be inferred that XG Boost is a powerful algorithm that provides the best accuracy in our use case. The reason for that being the use of ensemble techniques is of its power and beauty. Employees are the backbone of the organization and it is very crucial for any company to successfully retain its employee. From our study, we came to the conclusion that various factors like city development index, gender, and education level are crucial in deciding the attrition rate for a company. In order to visualize the performance of this binary classifier which is built using XG Boost, we use the AUC-ROC metric. It summarizes the trade-off between the true positive and false positive sets. Since we have already balanced the target class using SMOTE, we can predict the probabilities using ROC (Receiver Operating Characteristic curve).In comparison to Mamdani inference system, the Sugeno system uses a weighted average of a data points rather than compute a defuzzification process. In this present work, we used the Mamdani's inference system for the prediction of the attrition rate of the company. During this process, we adopted nine input and one output-based inference system. Using the fuzzy inference system, we can deal with the current uncertainty that will help in the prediction of use the ensemble machine learning in employee retention and attrition.

Acknowledgement

The work has been carried out under the Uttar Pradesh Government Higher Education Department order number -91/2021/2095/sattar-4-2021-04(17)/2021.

Declaration of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this work.

References

Góes ASDO, Oliveira RCLD. (2020). A Process for Human Resource Performance Evaluation Using Computational Intelligence: An Approach Using a Combination of Rule-Based Classifiers and Supervised Learning Algorithms. IEEE Access. 8, 39403-39419.

Jiangjing, L., Jun, Z., Chunliang, F., Linhua, G. (2020). Performance Evaluation Model of Human Resource Management in Public Institutions Based on Improved Decision Tree. International Conference on Wireless Communications and Smart Grid (ICWCSG). 310-315.

Ghorbani, R., Ghousi, R. (2020). Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques. in IEEE Access. 8, 67899-67911.

Zhang, J., Liu, H. (2019). On Human Resource Management and Big Data Analysis. 2019 IEEE International Conference on Smart Internet of Things (Smart IoT).345-349.

Zhang, S., Li, X., Zong, M., Zhu, X., Wang, R. (2018). Efficient kNN Classification with Different Numbers of Nearest Neighbors.in IEEE Transactions on Neural Networks and Learning Systems. 29, 1774-1785.

Mienye, I.D., Sun, Y.,Zenghui, W. (2020). An improved ensemble learning approach for the prediction of heart disease risk. Informatics in Medicine Unlocked. 20, 1002402.

Hancock, J., Khoshgoftaar, T.M. (2020). Medicare Fraud Detection using Cat Boost. 2020IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI).97-103.

Fayaz, M., Khan, A., Rahman, J.U., Alharbi, A., Uddin, M.I., Alouf, B. (2020). Ensemble Machine Learning Model for Classification of Spam Product Reviews. Hindawi Complexity Volume 2020. Article ID 8857570, 10 pages.

Sevakula, R.K., Verma, N.K. (2017). Assessing generalization ability of majority vote point classifiers. IEEE Transactions on Neural Networks and Learning Systems.28, 2985–2997.

Li, H., Cui, Y., Liu, Y., Li, W., Shi, Y., Fang, C., Li, H., Gao, T., Hu, L., Lu, Y. (2018). Ensemble learning for overall power conversion efficiency of the all-organic dye-sensitized solar cells. IEEE Access. 6, 34118–34126.

Schapire, R.E., Singer, Y. (1999). Improved boosting algorithms using confidence-rated predictions. Mach Learn. 37, 297–336.

Shone, N., Ngoc, T.N., Phai, V.D., Shi, Q. (2018). A deep learning approach to network intrusion detection. IEEE Trans. Emerg. Topics Comput. Intell. 2, 41–50.

Shang, F., Zhou, K., Liu, H., Cheng, J., Tsang, I.W., Zhang, L., Tao, D., Jiao, L. (2020). VR-SGD: a simple stochastic variance reduction method for machine learning. IEEE Trans Knowl Data Eng. 32, 188–202.

Ebenuwa, S.H., Sharif, M.S., Alazab, M., Al-Nemrat, A. (2019). Variance ranking attributes selection techniques for binary classification problem in imbalance data. IEEE Access. 7, 4649–66.

Rivera-Lopez, R., Canul-Reich, J. (2018). Construction of near-optimal axis-parallel decision trees using a differential-evolution-based approach. IEEE Access. 6, 5548–5563.

Samuel, O.W., Asogbon, G.M., Sangaiah, A.K., Fang, P., Li, G. (2017). An integrated decision support system based on ANN and Fuzzy AHP for heart failure risk prediction. Expert Syst Appl. 6, 8163–8172.

Griffeth RW, Hom, P.W., Gaertner, S. (2000) A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. J. Manag. 26, 463–488.

Ponnuru, S., Merugumala, G., Padigala, S., Vanga, R., Kantapalli, B. (2020). Employee Attrition Prediction using Logistic Regression. Int. J. Res. Appl. Sci. Eng. Technol. 8, 2871–2875.

Usha, P., Balaji, N. (2019). Analysing Employee attrition using machine learning. Karpagam J. Comput. Sci. 13, 277–282.

Rombaut, E., Guerry, M.A. (2018). Predicting voluntary turnover through Human Resources database analysis. Manag. Res. Rev. 41, 96–112.

Nagadevara, V. (2008). Early Prediction of Employee Attrition in Software Companies-Application of Data Mining Techniques. Res. Pract. Hum. Resour. Manag.2020–2032.

Alao, D., Adeyemo, A. (2013). Analyzing employee attrition using decision tree algorithms. Comput. Inf. Syst. Dev. Inf. Allied Res. J. 4, 17–28.

Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M., Abbasi, U. (2014). Improved churn prediction in telecommunication industry using data mining techniques. Appl. Soft Comput. 24, 994–1012.

Syam, N., Sharma, A. (2018). Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. Ind. Mark. Manag. 69, 135–146.

Jarrahi, M. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. Bus. Horiz. 61, 577–586.

Yanqing, D., Edwards, J., Dwivedi, Y. (2019). Artificial intelligence for decision making in the era of Big Data. Int. J. Inf. Manag. 48, 63–71.

Kaur, S., Vijay, R. (2016). Job Satisfaction – A Major Factor Behind Attrition or Retention in Retail Industry. Imperial Journal of Interdisciplinary Research. 2(8).

Al-Radaideh, Q.A., Nagi, E.A. (2012). Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance. International Journal of Advanced Computer Science and Applications. 3(2), 144-151.

Zadeh, L.A. (1965). Fuzzy sets. Information and Control. 8(3), 338-353.

Jingzheng, R., Yi, M., Ruojue, L., Yue, L. (2021). Multicriteria decision making for the selection of the best renewable energy scenario based on fuzzy inference system, Renewable-Energy-Driven Future. Technologies, Modelling, Applications, Sustainability and Policies. 491-507.

Mamdani, E.H., Assilian, S. (1975). An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller'. International Journal of Man-Machine Studies. 7(1), 1–13.

Sugeno, M. (1985). Industrial Applications of Fuzzy Control. Amsterdam; New York: New York, N.Y., U.S.A: North-Holland; Sole distributors for the U.S.A. and Canada, Elsevier Science Pub. Co.