

# HR-Analytics: A Study on the Functioning of the Employee Attrition

<sup>1</sup>Pratibha Desai, <sup>2</sup>Maheshwari Biradar

Department of Computer Science Engineering and Applications, D Y Patil International University, Pune-India

<sup>1</sup>[desaipratibha606@gmail.com](mailto:desaipratibha606@gmail.com)

<sup>2</sup>[biradar.maheshwari@gmail.com](mailto:biradar.maheshwari@gmail.com)

---

## Abstract

Every single organization has its own characteristics, such as productivity and strength, which rests on the feet of employees. Retaining a proper employee is a big challenge for all organizations in a competitive world. Attrition is a problem faced by many organizations where profitable and knowledgeable employees quit the organization periodically. Companies invest heavily in employee training by taking into consideration that the returns they would provide to the company in the future. If an employee quits the organization, it is a lack of golden chance cost to the company. A numerous organizations all over the world are trying to get rid of this serious matter. The major target is to create a model that can be used to predict whether employees will quit the company or not. Using this application, there is an identification of hidden reasons behind employee attrition and management can handle preventive solutions in regard to the attrition of all employees individually. These studies clarifies employee attrition rates through related attributes such as department, environmental satisfaction, job satisfaction, monthly rate, over time, working hours and many more using efficient algorithms such as Random Forest and Support Vector Machine.

*Keywords:* Attrition rate, machine learning, random forest, support vector machine, preprocessing, recruitment features, dependent variable analysis, exploratory data analysis, predictive models, exploratory data analysis

---

## 1. Introduction

Employees are a profitable asset to any organization. However, if they leave suddenly, it can be a huge cost for any organization. Attrition is the reduction in the number of employees in any company where employees may voluntarily quits the organization or may reach retirement age. Employee turnover is the number of current employees replaced by new employees over a period of time. Excessive employee attrition causes high employee turnover in any organization. This one after the other causes huge expenditure on human resources which contribute to recruitment of new employees, training and development of newly appointed employees as well as performance management. Again, attrition, which is voluntary, is inevitable. Thus, by improving employee morale and providing a favorable working environment can reduce this problem to a great extent. Attrition rate is defined as a company's hiring and termination criteria. An employee can quit a job for a numerous reasons. The business terminologies including 'turnover' and 'attrition' are always at odds with each other. There are several forms of "turnover" in an organization. A reduction in the number of employees is primarily considered a decrease. For the analysis of workforce data and other measurements necessary for workforce planning, these terminologies can be used interchangeably. The graphical user interface of the system is designed using Java JSP. The main technologies used were Java, JSP. Overall development was done in Eclipse Luna and I used the MYSQL GUI browser for the DB. A database that is primarily used to store user information such as

User names. The tool used for database functions is the MYSQL GUI Brower.

## 2. Literature review

Aniket Tambde, Dilip Motwani. (2019) "Employee Churn Rate Prediction and Performance Using Machine Learning," International Journal of Recent Technology and Engineering, (8) (September): 2277-3878. People who work for an organization are a key resource known as employees. If one of them instantly quits the company, it can have a great impact on the respective company and cause them a lot of losses. This model helps to predict employee turnover rates based on analytical data obtained and uses new machine learning algorithms to reduce prediction errors. Advantages have been found out that the system accurately predicts employee turnover and improves company performance. Disadvantages are applying different machine learning algorithms directly on the data set may not predict the expected accuracy and can be filled with overfitting or under fitting figures of the training data [1]. Amir Mohammad Esmaieeli Sikaroudi. (2015) "A Data Mining Approach to Employee Turnover Prediction," Journal of industrial and system engineering, (8)(November). Training and adaptation of employees require time and money. Employee workflow can be predicted from organizational and individual historical data to lower possible losses for the organization. Pros have been found out that In contrast, Naive Bayes is the most easy to understand model that has better classification performance. Disadvantages are PNN processing takes more time as the data volume increases [2]. Riyanto Jayadi, Hafizh M. Firmantyo, Muhammad T. J. Dzaka, Muhammad F. Suaidy, Alfitra M. Putra. (2019) "Employee Performance Prediction Using Naïve Bays," International Journal of Advanced Trends in Computer

Science and Engineering, 8(6), (December): 2278-3091. Employee Performance makes the company successful in reaching its end. Predicting the future performance of employees is essential for the success of a company. In this article, we used the simple Bayesian classification method to predict the performance of the company's employees. Pros are in contrast, Naive Bayes is the most easy to understand model and has better classification performance. The disadvantages are that based on the confusion matrix, we found a small amount of false-positive results which means that the cost of using the simple and unpretentious method is low [3]. Rohit Punnoose and Pankaj Ajit. (2016) "Prediction of Employee Turnover in Organizations Using Machine Learning Algorithms," *International Journal of Advanced Research in Artificial Intelligence* (5) (October). Employee turnover is an important issue for organizations because it affects the workplace, leaves unfavorable, lasting growth and productivity strategy. This paper describes an Extreme Gradient Boosting Technique used to predict employee turnover. Advantages are XGBOOST is a more robust method that shows notably higher accuracy for predicting employee turnover. Disadvantages: The XGBoost classifier outperforms other classifiers in terms of accuracy and memory usage [4]. Ibrahim Onuralp Yigit; Hamed Shourabizadeh. (2017) "An approach for predicting employee churn by using data mining" *International Artificial Intelligence and Data Processing Symposium*, (8) (September). Predicting employee turnover, which is closely related to predicting customer attrition, is a crucial business challenge. Apply popular classification methods such as decision trees, logistic regression, SVM, KNN, random forest and Naive Bayes to HR data. Advantages have been found out that the system predicts employee turnover, thus reducing labor costs. Disadvantage is Model training is slow [5]. John M. Kirimi and Christopher Moturi. (2016) "Application of Data Mining Classification in Employee Performance Prediction" *International Journal of Computer Applications* 146(7):28-35 (July). This paper proposes a prediction model for predicting employee performance that enable talented professionals to refocus on human competency measures and thus enhance the human capital performance evaluation process. It has been found to have the advantage of high performance. Disadvantage is that you need more precision [6]. Sandeep Yadav; Aman Jain; Deepti Singh. (2018) "Early Prediction of Employee Attrition using Data Mining Techniques" *International Advance Computing Conference* (December). We present a framework for predicting employee turnover by analyzing their detailed behaviour and characteristics using classification techniques such as logistic regression, SVM, trees, Adaboost and neural networks. Advantage is Helps to emphasize the attributes that have the greatest impact on employee turnover. Disadvantage have been found out that It cannot enhance the cost of hiring and retaining quality employees [7]. S K Monisaa Tharani; S N Vivek Raj. (2020) "Predicting employee turnover intents in IT&ITS industry using machine learning algorithms" *International Conference on I- SMAC* (October). Predict employee turnover intents in the near future and identify key characteristics that influence employee turnover intents. The XG Boost algorithm is used to predict employee turnover. Advantages are the high prediction accuracy and disadvantage have been found out that XG Boost is slow in training [8]. Heng Zhang; Lexi Xu; Xinzhou Cheng. (2018) "Analysis and Prediction of Employee Turnover Characteristics based on

Machine Learning" *International Symposium on Communications and Information Technologies (ISCIT)* (December) "Analysis and prediction of employee turnover characteristics based on machine learning". We use GBDT and LR algorithms to fit the attribute model that affects employee turnover. Advantages are the results of the forecasting model are more accurate. Disadvantages: The algorithm is a bit susceptible to noise and overfitting [9]. Francesca Fallucchi Marco Coladangelo, Romeo Giuliano and Ernesto William De Luca. (2020) "Predicting Employee Attrition Using Machine Learning Techniques" *Computers MDPI* 9(86) (November). Classification algorithms considered are Gaussian, Naive Bayes, Logistic Regression classifier, K-nearest neighbors, Decision tree classifier, Linear Support Vector Machine. Advantages have been found out that work seeks to provide answers to some of the usual questions of responsible HR management. Disadvantages are the System processes do not end with extracting knowledge from revealed models. Knowledge must be expressed and displayed in such a way that end users can accept it [10]. Madara Pratt, Mohcine Boudhane and Sarmacakula. (2021) "Employee Estimation Using Random Forest Algorithm." *Baltic J. Modern Computing*, Vol. 9 (November). It is an approach based on the random forest algorithm, which involves gathering a set of estimators constructed from bootstrap samples. Advantages: Random forests are usually well organized than simple decision trees. Disadvantages: Interpretation is hard [11]. I Setiawan, S Suprihanto, A C Nugraha, J Hutahaean. (2020) "HR Analytics: Employee attrition analysis using logistic regression." *International Conference on Innovation in Engineering and Vocational Education* 830 (November). Human resource management (HRM) aims to cover employee performance and commitment, analyze employee cooperation models, study employee engagement and turnover, and develop employee life expectancy estimates. They used R studio for data integration, exploratory data analysis, data preparation, logistic regression, model evaluation and visualization. Advantages: We establish 11 variables that have a remarkable effect on employee turnover. Disadvantages: To lower staff attrition, the company needs to improve its HR department [12]. Rajat Keshri, Dr. Srividya P (2020) HR analytics: "Prediction of Employee Turnover Using light GBM Algorithm" *Computers MDPI International Journal of Innovative Science and Research Technology* 5 (April) 2456-2165. Different factors on growth and the production of the company is affected. Companies and many multinational companies use machine learning techniques to predict worker turnover to find the solution of this issue. Advantages have been found out that the supervised learning involves the dataset given with the output each data. The capacity to collect data or generate data output from experience. Advantages are can't handle difficult tasks and can't categorize data based on unique features [13]. S. Mandane and D. Chitre. (2020) "A Survey of Employee and Customer Churn Prediction Methodologies" *Advances in Mathematics: Scientific Journal* vol.9 3955-3962. Employee and customer attrition is a major issue for many businesses today. One of the vital concerns of organization is reducing employee turnover due to the high costs of replacing former employees. Advantages have been found out that the random forest can be considered the best classification algorithm for churn prediction. Disadvantages have been found out that the Decision trees are less accurate than other data mining models, so they are not an ideal choice for churn prediction [14]. Eugene V. Juaneza (2021) "An Empirical Study on the PNP Maritime Group Attrition System" *International Journal of Academic and Industry Research*, Volume 2 25-44. Study Design and method: independent variables (internal factors such as age, gender, years of service) to establish the existence of sample-population relationships. Advantages: It assist researchers to prepare and conduct research in a fast and organized manner. Disadvantages: It takes a lot of time [15]. Dr. Gopinath.R (2020) "Employee Attrition Prediction in Industry using Machine Learning

Techniques” International Journal of Advanced Research in Engineering and Technology Vol.11(12) (December) 3329-3341. We look for ways to retain professional talent to save on hiring and training costs. Predicting whether a particular employee will leave and helps organizations make proactive decisions. Advantages have been found out that attribute selection strategies based on machine learning classifiers are very functional, increase classification accuracy, precision, true positive rate and lower error rate. Disadvantages are Human resource complications, unlike physical systems, cannot be defined by scientific formulas [16]. Praveen Ranjan Srivastava, Prajwal Eachempati.(2021) "Intelligent Employee Retention System for Attrition Rate Analysis and Churn Prediction. An ensemble machine learning and Multi-Criteria Decision-Making App" Journal of Global information management 29(6) (January) The purpose of this paper is to look into the factors determining employee turnover with the use of data sets of employee records. We also develop the predictive strength of deep learning to predict employee turnover for ensemble machine learning techniques such as random forests and gradient boosting on real-time employee data from average-sized rapid-moving consumer goods (FMCG) companies. It is also considered that the advantages are the discovery of the fuzzy AHP model show that these factors play an important role in the decision making of frustrated employees. Disadvantages are Using the machine learning techniques described above, but does not use a multifactor hybrid approach to confirm the causes for employee turnover [17]. Nesrine Ben Yahila 1, JihenHlell, and Ricardo Colomo-Palacios 2.(2021) "From Big Data to Deep Data to Support People Analytics for Employee Attrition Prediction." Institute of Electrical and Electronics Engineers 9 (April) 2169-3536. In the age of data science and big data analytics, people analytics can help to lower headcount by switching the way organizations and their human resources (HR) managers attract and retain talent. In this circumstance, workforce attrition poses significant challenges and risks to organizations, as it affects not only productivity but also planning continuity. In this context, the significant contributions of this study are: Advantages have been found out that the accuracy of the model is measured before and after the feature selection algorithm. That is, for the earliest we use the whole real data set with 16 features. Disadvantages are the model is evaluated using only 11 selected features after applying the feature selection process [18]. Ahmad Al Kuwaiti, Vinoth Raman, Arun Vijay Subbarayalu, Palanivel R.M., Sivasankar Prabakaran.(2016) "Predicting The Exit Time Of Employees In An Organization Using Statistical Model." International Journal of Scientific and Technology Research Volume 5, Issue 12 (December) 2277-8616. Employees are beneficial to any organization, and every organization offers a well and more flex-time environment to retain the best resourceful employees. By itself there are constant efforts to prevent or prolong the termination/exit of employees from the organization. Advantages are the results revealed that as the time between arrivals increased, the expected time before employees left the office also increased. Disadvantages are it takes time [19]. Ahmed Richard Eduvie, Johnwendy C. Nwaukwa, Felix Uloko, Eric Taniform. (2021) "Predicting employee attrition using decision tree algorithm." Global Scientific Journal 9, ISSUE 9 (September) 9186-2320. Organizational decision-making is essential for HR teams in terms of predicting employee turnover. Manually predicting which employee will leave when and for what reasons is

impossible, mainly on a large scale, because there are so many complex and related variables that affect how likely an employee is to leave, if at all. This is where machine learning comes into play. Machine learning finds statistical designs in wide quantity of historical data to reveal things that humans might miss. Pros are the system can be used by any organization and the forecasting features can be enlarged or altered based on how different HR teams access employees. Disadvantage: They only used two data types in their dataset: coefficients and integers [20]. Praphula Kumar Jain, Madhur Jain, Rajendra Pamula.(2020) "Explaining and predicting employees attrition: a machine learning approach." Springer Nature Journal (March). Employee burnout is a difficulty that several organizations face, and valuable and proficient employees quits the organization every day. Several companies around the world are trying to solve this serious problem. The major goal of this is to grow a model that helps to predict whether an employee will quit the organization or not. The major plan is to calculate the effectiveness of employee grading and satisfaction in your company, which helps reduce employee turnover. Advantages are the accuracy of DT and RF is notably improved, and it appears that these classifiers can be applied to predict whether an employee is probable to quit. Disadvantages: It does not accurately analyze characteristics and relationships with other variables, which can lead to inaccurate predictions [21]. Fahad Kamal Alsheref, Ibrahim Eldesouky Fattohan WaaleedM.Ead.(2022) "Automated Prediction of Employee Attrition Using Ensemble Model Based on Machine Learning Algorithms." Computational Intelligence and Neuroscience (June). Talented employees are a rare commodity for large organizations. The problem of retaining experienced and talented employees threatens business owners. The difficulty of employee burnout can be very expensive for employers because it fetch a lot to recompense for skill and organization. For this cause, this study presents an automated model that can predict employee turnover formed on unique predictive analysis techniques. Advantages: The results showed that low values indicate completeness of the model. Disadvantages: MLP training is slow [22]. Doohe Chung, Jinseop Yun, Jeha Lee, YeramJeon. (2023) "Predictive model of employee attrition based on stacking ensemble learning." Expert System with Application 215 (April). As human resources are the most important resource of an organization, the reduction of labor force is one of the main issues of companies. However, employee turnover arise for a variety of reasons, and it is hard for HR personnel and leaders of any department to recognize the hints in advance. In this study, prediction models were built based upon 30 variables and 8 prediction models, including logistic regression, Random Forest, XGBoost, SVM, ANN models, and ensemble models. Advantages have been found out that after model construction and analysis, the cumulative- based set model showed the best performance. Disadvantages are Single machine learning models such as RF, XGoost, logistic regression and ANN failed to carry out better than ensemble models [23]. Shawni Dutta, Samir Kumar Bandyopadhyay. (2020) Employee attrition prediction using neural network cross validation method. "International Journal of Commerce and Management 6(3) (June) 80-85. This white paper focuses on predicting the likelihood of attrition in advance by executing automated tools. The proposed system executes a feed-forward neural network along with a 10-fold cross-validation method under a particular stage to predict employee turnover. The proposed method is evaluated against six classifiers including support vector machines, k-nearest neighbor, simplest bays, decision trees, Ada boost and random forest classifiers. Advantages have been found out that need to improve classification accuracy. Several unwanted learners are merged into one learner using the logical technique of boosting the classification algorithm. Disadvantage is in this system, K-NN is intended as a lazy learner because it only stores training samples in the training phase [24]. Shikha Nkhera, Divya. (2019)

"Predictive Modeling of Employee Turnover in Indian technical Industry Using Machine Learning Techniques". Journal Indexing and Metrics 23(1) (March). The objective of this paper is to grow a predictive model based on employee data to address the issue of employee turnover in the Indian IT sector. Advantages have been found out that this model does a better job of predicting who will quit the company than predicting who will not quit. Disadvantages are requires feature set refinement to develop more accurate models and lower misclassification rates [25]. Dr. GopinathR .(2020)" Employee Attrition Prediction in Industry using Machine Learning Techniques" International Journal of Advanced Research in Engineering and Technology Vol.11(12)(December) 3329-3341. Companies are consistently looking for ways to retain professional personnel to save training costs. Predicting whether or not a particular employee would leave will assist the organizations to make bold decisions. Advantages have been found out that the attribute selection strategies based on machine learning classifiers are very beneficial, grow classification accuracy, precision, true positive rate and lower error rate. Disadvantages: Human resource complications, unlike physical systems, cannot be defined by scientific formulas [26]. Sri Ranjitha Ponnurul, Gopi Krishna Merugumala2. Srinivasulu Padigala3, Ramya Vanga4, Bhaskar Kantapalli5. (2020) "Employee Attrition Prediction using Logistic Regression" International Journal for Research in Applied Science and Engineering Technology vol: 45.98(12) (May)2321-9653. This model can help to predict the voluntary attrition of the employees by machines learning algorithms including RF, decision tree, SVM, Naive-Bayes, KNN, Trees and PCA.

Advantages have been found out that we can get the accuracy of your model by comparing the predicted values with the actual test values. Disadvantage is the accuracy of the model cannot be obtained without comparing the predicted values with the actual experimental values [27]. Dr. Praveen Ranjan Srivastava and Prajwal Eachempati. (2021) "Intelligent Employee Retention System for Attrition Rate Analysis and Churn Prediction" Journal of Global Information Management 29(6) (January) 1-29. We focus to build the predictive power of deep learning to predict employee turnover on ensemble machine learning methods such as RF, GBT, DNN, and Fuzzy (Analytic Hierarchical Process). Advantages have been found out that this model helps to show that functional factors play an important role in the decision making of dissatisfied employees. Disadvantages are Functional factors could also be included in future studies to examine their impact on employee turnover [28]. Moninder Singh, Kush R. Varshney, Jun Wang and Aleksandra Mojsilovic. (2012) "An Analytics Approach for Proactively Combating Voluntary Attrition of Employees" International Conference on Data Mining Workshops. This approach uses predictive modeling, clustering, attrition, and proactive retention to execute proactive retention measures for thousands of employees over multiple geographies and business units at a large Fortune 500 company. Advantages have been found out that has current range, approximation or relative value while not distorting the actual image. Disadvantages are due to the sensitivity of the data, it does not come up with an accurate value for financial or attrition data [29]. Sudipta Bhatta, Isfaf UzZaman, Nuzhat Raisa, Shazzadul Islam Fahim, Sifat Momen.(2022) "Machine Learning Approach to predicting Attrition Among employees at work" Computer Science

Online Conference. Attrition is a big issue for any company. In addition to losing qualified employees, turnover can result in lost time and money for any company. We report Classifiers: Zero R, Decision Trees, K-NN, SVM and RF. Advantages are all classifiers got different degrees of outcomes, but voting classifiers got top results Disadvantages: SVM, RF, and ensemble learning were compared with performance measurement techniques, so the data sets are unbalanced [30].

### 3. Methodology

#### Algorithm 1/Pseudocode

The Random Forest algorithm is a supervised classification algorithm. A decision tree is a decision support tool. A decision tree uses a tree diagram to show the possible results. If you feed a training dataset with objectives and features into a decision tree, it will generate some set of rules. These acts can be used to make predictions. The Random Forest algorithm has 2 phases, one is generating the random forest and the other is prediction from the random forest classifier generated in the first phase.

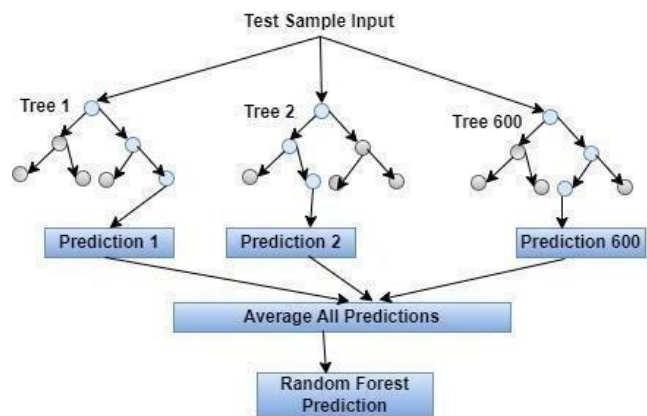


Fig.1

Random Forest

#### GENERATING RANDOM FOREST:

1. Randomly choose "K" features from the set of "m" features.
2. Compute the square meter of node 'd' using the best break points from 'K' features.
3. Divide the node into daughter nodes using the best interval.
4. Repeat the steps a-c until the number of nodes is "1".
5. Repeat the steps in a turn 'n' times to create a forest and create 'n' trees.

#### PREDICTION USING RANDOM FOREST CLASSIFIER:

1. Select test features, use the rules of each randomly generated decision tree to predict the result, and save the predicted result (target).
2. Computation of votes for each predicted goal. We selected the predicted target as the final prediction from the random forest algorithm.

#### Algorithm2/Pseudocode

SVM is a strong classifier that can differentiate between two classes. SVM analyse the test image into the class that has the greatest distance from its training neighbours. The SVM training algorithm set up a model that predicted whether the test image would fall into this class or another class. SVM requires huge training data to determine decision boundaries and uses single-mode detection (forehead), which is computationally expensive. SVM is a learning algorithm for classification that tries to find the best discriminant level that reduce errors on invisible patterns.



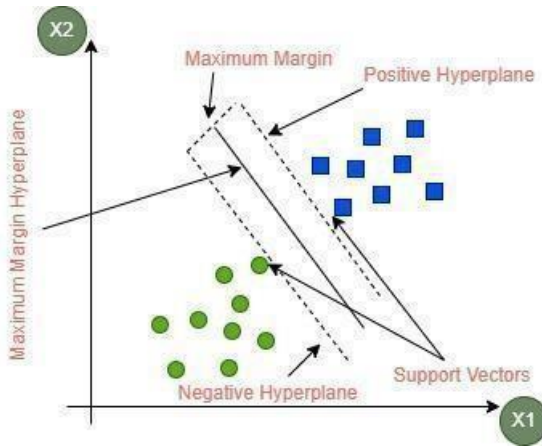


Fig.2. Distinguishing Hyper Plane to Minimize the Error

Data with indistinguishable arrivals are plotted into a high-dimensional feature space that can be parted by hyperplanes. This scheme is properly handled by the kernel.

For the result values -1, 1, equivalent to the set of training examples, the SVM finds the optimal separation hyperplane specified by the equation  $WT \cdot X + B$ , which uses the distance between the two classes, as shown in the figure above. Try to get.

#### 4. Results

##### Random Forest Algorithm

**Step 1:** This algorithm uploads a downloaded dataset (reads csv file) from Kaggle, i.e, the data set of employee turnover. From there, We have some features like:

1. Monthly income
2. Job level
3. Number of years with current manager
4. Age
5. Total working years

**Step 2:** We then perform a dependent variable analysis on the data set to remove redundant data which is useless for prediction

**Step 3:** We need to build a decision tree of n records from the dataset. To build a decision tree, we need to select the root node from among the best features using feature selection criteria using two methods:

1: Gini Index

Mathematical Formula:

$$\text{Gini} = 1 - \sum_{i=1}^n (p_i)^2$$

2: Information Gain

Mathematical Formula:

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

The selective measure attribute is repeated until a leaf node or end node cannot be split into sub-nodes.

**Step 4:** Each decision tree produces an output. To increase predictive power, decision trees are generated before averaging the number of trees making predictions and algorithms. The maximum number of random features in the forest determines the minimum number of leaves required to consider node splitting and internal node splitting.

**Step 5:** Once you have your training and test data, build a predictive model.

**Step 6:** Final results are considered based on majority or average. Classification and regression are each effortlessly visualized.

##### Support Vector Machine

Initially, we need to load the dataset we want to work with.

In the next step, we need to examine the data and determine the core of the data. There are basically four popular kernels. Then by examining the data we can learn what kind of data it is and how to partition it with hyperplanes.

##### A. Polynomial kernel

It represents the similarity of the vectors in the training data set in the feature space. A polynomial in the main variables used in the kernel.

$$F(X1, X2) = (a + X1^T \cdot X2)^b$$

##### B. Linear kernel

Almost all the kinds of classification problems are linearly divisible, so they are generally suggested for text classification.

$$F(x) = W^T \cdot X + B$$

##### C. Gaussian Radial Basis Function (RBF)

It is one of the high powered and widely used kernels in SVM. Usually the selection is non-linear data.

$$F(X1, X2) = \text{EXP}(-\gamma * \|X1 - X2\|^2)$$

##### D. Gamma

The gamma parameter determines how far the effect of a training example extends. This means that high gamma only examines points close to the acceptable super plane, while low gamma considers points further away.

##### E. Sigmoid

More preferred in neural networks than support vector machines, but sometimes specific use case

$$F(x, y) = \tan h(\alpha * X^T \cdot y + C)$$

1. Data pre-processing.
2. Divide the data into attributes and tags
3. Split the data into training and test sets
4. In the next step, we should train the support vector machine algorithm according to our needs.
5. Then you can make predictions from the data.

#### 5. Applications

The proposed algorithm has numeral parameters: Number of trees in the forest. Its default value is 500. Note that this parameter is not actually a calibration parameter in the sense that a bigger value of this parameter will consistently lead to more firm predictions than a smaller value of this parameter. The number of mtry variables selected to split each node. Its default value is  $mtry = \text{pd}$  in classification. This is probably the most important parameter to calibrate as it can considerably affect forest performance. The least number of node size observations under which a node is no longer break. The default value for this parameter is  $\text{nodesize}=1$  for classification. Generally, this setting is left at the default value. Numeral observations every year in every single bootstrap sample. By default, every single bootstrap sample contains = n observations drawn with renewal in the initial sample  $D_n$ . Several authors were attentive in the selection and influence of these parameters (Biau and Scornet, 2016; Genuer and Poggi, 2017). Generally, the default parameter values perform well. In

fact, few theoretical results are obtainable for Breiman random forests. However, we can cite an important result recently found, concentrated on the convergence of random forests in an additive model (Sconet et al., 2015).

Theoretical warranty were also obtained for clarified versions of the method (Wager, 2014). A summary of the main theoretical results is illustrated in the first algorithm before its update.

## 6. Conclusion

People are the backbone of any organization. The rate of growth and market penetration depends on our employees. Today, the growing population and skilled people make it a huge success for any company.

But the main problems that are commonly dealt with in any organization are only attrition. This is a difficult task and maintenance is also a key task. This system helps to implement employee workflow forecasts in the relevant organization. The analysis is done by taking into account certain characteristics such as monthly income, last year of promotion. Current position in the company, salary etc. Using machine learning to understand the key variables that influence workforce management. We use random forest algorithm and SVM to create a prediction model to identify different causes of employee turnover.

## 6. References

[1] S. Jahan, "Human Resources Information System (HRIS): A Theoretical Perspective", *Journal of Human Resource and Sustainability Studies*, Vol.2No.2, ArticleID:46129, 2014.

[2] C. Cortes and V. Vapnik, Support-vector networks. *Machine learning*, 20(3), 273-297, 1995.

[3] Rohit Punnoose ,PankajAjit, " Prediction of Employee Turnover in Organizations using Machine Learning Algorithms", (IJARAI) *International Journal of Advanced Research in Artificial Intelligence*, Vol. 5, No.9, 2016.

[4] "Guest Editorial: Special Issue on Early Prediction and Supporting of Learning Performance" *IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES*, VOL. 12, NO.2, APRIL- JUNE 2019.

[5] Renuka Agrawal, Jyoti Singh, and Zadgo ankar.S, "Formative Assessment For Performance Evaluation Of Faculty Using Data Mining", *International Journal Of Advances In Electronics And Computer Science*, ISSN: 2393-2835.

[6] Hossein Alizadeh, Buinzahra Branchand Islamic, 2016, "Introducing A Hybrid Data Mining Model To Evaluate Customer Loyalty", *Engineering, Technology Applied Science Research* Volume. 6, No.6, 1235-1240.

[7] Amir Mohammad Esmaieeli Sikaroudi ,Rouzbehghousi and Ali Esmaieeli sikaroudi, 2015 "A Data Mining Approach To Employee Turnover Prediction"(Case Study: Arak Automotive Parts Manufacturing), *Journal Of Industrial And Systems Engineering* Volume.8, No.4.

[8] Anjali A. Dudheand Sachin Sakhare. R, January 2018, "Teacher Ranking System To Rank Of Teacher As Per Specific Domain", *Journal On Soft Computing ICTACT*, Volume: 08, Issue:02, Issn: 2229-6956.

[9] Hamidah Jantan, Abdul Razak Hamdan, And Zulaiha Ali Othman, 2009, "Knowledge Discovery Techniques For Talent Forecasting In Human Resource Application ", *International Journal Of Industrial And Manufacturing Engineering* Volume:3, No:2.

[10] Umadevi, Dr. B and Dhanalakshmi, R. April 2017. "A Comprehensive Survey of Students Performance Using Various Data Mining Techniques ". *International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064*. Volume 6 Issue 4.

[11] Dilip Singh Sisodia, Somdutta Vishwa karma, Abinash Pujahari "Evaluation of Machine Learning Models for Employee Churn Prediction", *Proceedings of the International Conference on Inventive Computing and Informatics (ICICI2017) IEEEXPlore Compliant -Part Number: CFP17L34-ART, ISBN:978-1-5386-4031-9*

[12] -Part Number: CFP17L34- ART , ISBN:978-1-5386-4031-9

[13] Rajat Keshri, Dr .Srividya P "Prediction of Employee Turnover Using Light GBM Algorithm" *International Journal of Innovative Science and Research Technology* ISSNNo: 24562165 Volume5, Issue4, April2020 Department of Electronics and Communication, R V College of Engineering.

[14] S. MADANE AND D. CHITRE1 "A SURVEY OF EMPLOYEE AND CUSTOMER CHURN PREDICTION METHODOLOGIES " *Spec Issue on ICAML 2020 Advance sin .*

[15] Eugene V. Juaneza MS degree holder. P/Major, Philippine National Police. "An Empirical Study on the PNP Maritime Group Attrition System" Juaneza, E. V. (2021). *An Empirical Study on the PNP Maritime Group Attrition System. International Journal of Academe and Industry Research*, Volume 2, Issue2, pp. 25-44.

[16] Dr. M. Subhashini and Dr. R. Gopinath "EMPLOYEE ATTRITION PREDICTION ININDUSTRY USING MACHINE LEARNING TECHNIQUES" *International Journal of Advanced Research in Engineering and Technology (IJARET)* Volume 11, Issue 12, December2020, pp. 3329-3341.

[17] NESRINE BEN YAHIA 1, JIHEN HLEL1, AND RICARDO COLOMO-PALACIOS 2, (Senior Member, IEEE). "From Big Data to Deep Data to Support People Analytics for Employee Attrition Prediction" Received March 29, 2021, accepted April 13, 2021, date of publication April 20, 2021, date of current version April 27, 2021 D.E. Thompson et al., *Interpersonal distance preferences, J. Non-verbal Behav.* 4 (2) (1979) 113–118.

[18] Ahmed Al Kuwaiti, Vinoth Raman. "Predicting The Exit Time Of Employees In An Organization Using Statistical Model." *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH* VOLUME 5, ISSUE 12 DECEMBER 2016 ISSN2277-8616.

[19] Richard Eduvie, Jonwendy C.Nwaukwa, Felix Uloko, Eric Taniform "Predicting employee attrition using decision tree algorithm." *GSJ: Volume 9, Issue 9, September 2021 ISSN2320-9186GSJ: Volume 9, Issue 9, September 2021, Online: ISSN 2320-9186 VERITAS UNIVERSITY, ABUJA.*

[20] Praphula Kumar Jain, Madhur Jain, Rajendra Pamula. "Explaining and predicting employees attrition: a machine learning approach" Received: 19 August 2019 / Accepted: 13 March 2020 / Published online: 25 March 2020 © Springer Nature Switzerland AG2020.

[21] Fahad Kamal Alsheref, Ibrahim Eldesouky FattohanWaaheed M. Ead(26,June2022). "Automated Prediction of Employee Attrition Using Ensemble Model Based on Machine Learning Algorithms" Received 7 March 2022; Revised 23 May 2022; Accepted1 June2022; Published 26 June2022.

