iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VIII **Month of publication:** August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37796

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Prediction of Employee Attrition Using Machine Learning Approach

M. B. Shete[1], A. G. Patil[2]
[1]*M.Tech. (E&TC) student, Department of E&TC Engineering, PVPIT, Budhgaon, Sangli, India*
[2]*Associate Professor, Head of Department of E&TC Engineering, PVPIT, Budhgaon, Sangli, India.*

***Abstract:** In the world of technology, there are various zones through which different companies may adopt technologies which sustenance decision-making, Artificial Intelligence is the most creative advancement, generally used to help various companies and institutions in business approaches, authoritative aspects and individual's administration. As of late, consideration has progressively been paid to Human Resources (HR), since professional excellence and capabilities address a development factor and a genuine upper hand for organizations. Subsequent to having been acquainted with deals and showcasing offices, man-made brainpower is additionally beginning to direct representative related choices inside HR the board. The reason for existing is to help choices that are put together not with respect to emotional viewpoints but rather on target information investigation. The objective of this work is to break down how target factors impact representative weakening, to distinguish the fundamental driver that add to a specialist's choice to leave an organization, and to have the option to foresee whether a specific worker will leave the organization. After the testing, the proposed model of an algorithm for the prediction of workers in any industry, attrition is tested on actual dataset with almost 150 samples. With this algorithm best results are generated in terms of all experimental parameters. It uncovers the best review rate, since it estimates the capacity of a classifier to track down every one of the True positive rates and accomplishes a generally false positive rate. The introduced result will help us in distinguishing the conduct of representatives who can be attired throughout the following time. Trial results uncover that the strategic relapse approach can reach up to 86% exactness over another. There are the few algorithms that can be used for processing the data, K-Nearest Neighbour, logistic regression, decision Tree, random Forest, Support Vector Machine etc.*
***Keywords**: Employees Attrition, Machine Learning, Support vector machine (SVM), KNN (K-Nearest Neighbour)*

I. INTRODUCTION

Today attrition is one of the serious issues looked by industry across the world. It is the most consuming issue for the business, and high whittling down rates lead to many issues in the limit of the association like losing the skilled assistances and information, cost identified with preparing and organization. It is seen that many ascribe lead to the steady loss of a representative. Which incorporates working climate, work fulfillment, behaviour of seniors, work timing, and most significant is payor motivating forces. Additionally, the expectation model assumes a fundamental part in discovering the conduct of representatives. Ideal conveyance of any assistance or item is the essential objective of any association lately because of high rivalry in enterprises. On the off chance that a capable employee will leave company, the organization can't do the job at characterized times. It might turn into the justification the deficiency of that organization. Thusly, organizations are keen on knowing the worker's weakening. They can make an appropriate substitute or game plans prior.

This framework can anticipate which representative might leave an association with what reason, so they can make a few restorative moves to guarantee that workers stay in the association and can lessen the steady loss. A portion of the worker maintenance methodologies to control weakening are spurring representatives, open workers to fresher jobs, taking steady input from workers, and so on. Following are the couple of algorithms that can be utilized for preparing the data Support vector machine (SVM), k-Nearest Neighbour, Decision Tree, Random Forest, logistic regression etc. Based on the accuracy measurement well performed algorithm will be used for this dataset.

AI is most significant innovation towards information examination for quality expectation and assessment. There are different calculations in AI which are utilized to anticipate the proper class of new or inconspicuous information. In our framework we utilized distinctive AI calculations to discover the purposes behind worker wearing down. The AI calculations which are utilized in framework are depicted beneath:

A. Support Vector Machine (SVM)

The novel approach toward the attrition prediction is Support Vector Machine strategy. This approach utilized for grouping as well as relapse issues. It can tackle straight and non-direct issues. The calculation makes a stroke or a hyper plane what isolates the information into modules or classes [9]. At the point once, obscure information is assumed as information it detects which class it has a placed with. The edge amongst the hyper plane and the help vectors are pretty much as extensive as conceivable to decrease the blunder in grouping.

B. K-Nearest Neighbours

K-Nearest Neighbour is viewed as a lazy learning calculation that characterizes informational collections dependent on their closeness with neighbours. It is perhaps the most crucial and straightforward characterization techniques and probably the most ideal decision for an order investigation of the information [7]. The instruction using KNN comprise determining contiguous material emphases and afterward selecting the class reliant on on the classes of the neighbours.

C. Decision Tree

All Decision tree approaches recursively isolated insights hooked on twigs to develop a tree to further develop the forecast precision. Decision tree approach is a customary calculation utilized for execution characterizations reliant on the selections completed in one phase. This gives decision tree prearranged depiction of the choice groups [10]

D. Random Forest

The most significant approach is Random Forest approach which is utilized for Classification and Regression subjects in Machine Learning. It rests on the knowledge of group understanding, which is a course of joining numerous classifiers to tackle an intricate issue and to improve the exhibition of the model. Rather than relying upon one decision tree, the Random Forest receipts the anticipation from respectively tree and forecast which have larger part of votes will be the last yield. As the quantity of trees builds the precision too increments and keeps it from the over fitting issue.

II. IMPLEMENTED ALGORITHM

The proposed framework comprises of various AI procedures. To assemble prototypical, we take representative dataset which includes all over a significant time span records of the workers, then, at that point we perform information reprocessing Data Pre-processing is the development in which the statistics gets transformed, or encoded, to carry it to such an express that the machine container undoubtedly investigates it. Dataset is divided into two categories, the basic one is train information and second one is test information. The mainstream of the evidence is utilized for making and more unsure portion of data is utilized for testing. The point of preparing is to select a forecast accurately as regularly as could really be expected. The test information is utilized to perceive how well the machine can anticipate new answers and to approve AI model conduct. A short time later, utilizing distinctive AI calculations we have constructed the model. In the wake of building model, client can give the new information to the framework. Moreover, client can pick calculation as per their decision and actually take a look at the outcome. Yield of the framework is in two structures - one is graphical portrayal and other is in polar structure that is 'Yes' or 'No' design. In the wake of assessing result the purpose for the whittling down is additionally given by the framework All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

A. Architecture of Algorithm

A calculation dependent on Machine Learning space is created as follows which is represented in Figure [1]. AI strategy is utilized to foresee the representative turnover and to anticipate the danger of a worker leaving. AI (ML) is the investigation of PC calculations that work on consequently through experience. It is viewed as a piece of man-made brainpower. AI calculations fabricate a model dependent on example information, known as "preparing information", to settle on forecasts or choices without being unequivocally customized to do as such.

1) *Data Collection:* The amount and nature of your information direct how precise our model is. The result of this progression is for the most part a portrayal of information which we will use for preparing. Utilizing pre-gathered information, via datasets from Kaggle, UCI, and so on, still squeezes into this progression.
2) *Data Preparation:* Wrangle data information and dataset it in the mood for getting ready. Clean what may require it Randomize data, which kills the effects of the particular solicitation wherein we accumulated or possibly regardless set up our data. Envision data to help with recognizing material associations between components or class unpredictable attributes or perform other exploratory examination. Split into getting ready and evaluation sets.

3) *Choose a Model*: There are numerous approaches for different errands; select the right one.
4) *Train the Model:* The goal of getting ready is to address a request or make a conjecture precisely as consistently as could truly be anticipated. Direct backslide model: estimation would need to learn values for m (or W) and b (x is input, y is yield). Each accentuation of connection is a readiness step.
5) *Evaluate the Model:* Uses some quantity or mixture of capacities to "portion" target implementation of perfect approach. Test the algorithm against beforehand inconspicuous information. This inobtrusive material is envisioned to be justly descriptive of perfect implementation in reality.

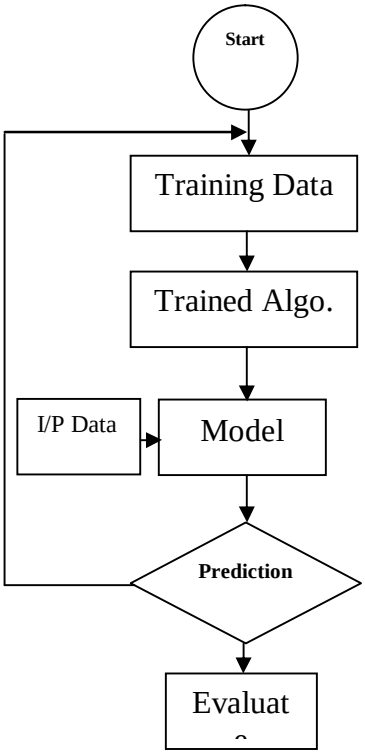


Figure [1] Block Diagram of Algorithm [1]

III.DATA ANALYSIS

Data gathering raises to the collection of relevant data from all available sources to perform analysis. The data used for this employee attrition analysis was obtained from GitHub Website [9]. This data set contains 1500 records and 50 attributes. The categorical values are converted to numeric values in order to make the classification algorithm more effectual. For example, categorical attribute 'Business Travel' contains three values such as Travel-Rarely, Travel Frequently, Non-Travel. Hence it is converted to 1, 2 and 3 respectively

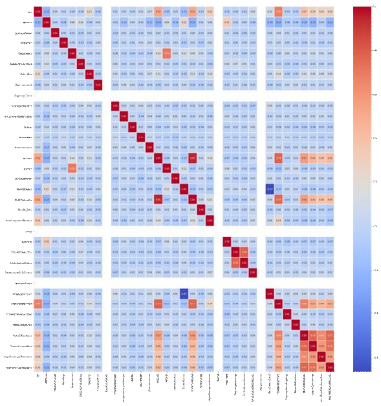


Fig. 1 Correlation Matrix

A. Some Graphs with Explanation

While concluding the result and analysis, the several graphs generated by the system with respect to attrition:

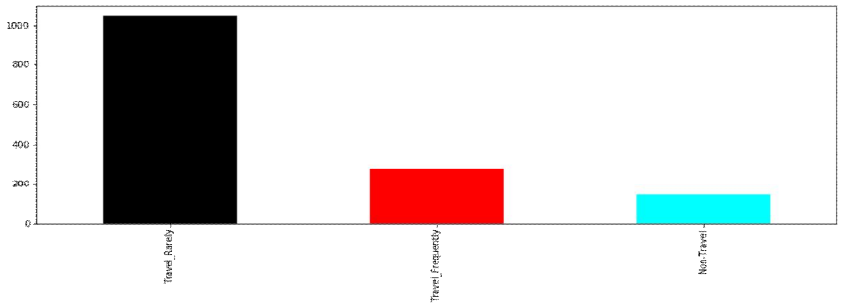


Fig. 3 Business Travel Vs Attrition

Fig. 3 depicts bar graph of evaluation between the Business Travel and respect to Attrition. Proportion of attrition with Travel Hardly is 16%, Travel Frequently is 25 % and Non-Travel is 10.00%. Attrition rate of Travel Frequently is more than other as there are 250 employees who are below the group of Travel Frequently and there remain 60 staffs who are leave-taking the organization. There are 1000 employees who Travel Rarely and out of which only 126 employees are send-off hence the attrition rate is low. For Non-Travel total 150 employees are there and 12 employees are leaving.

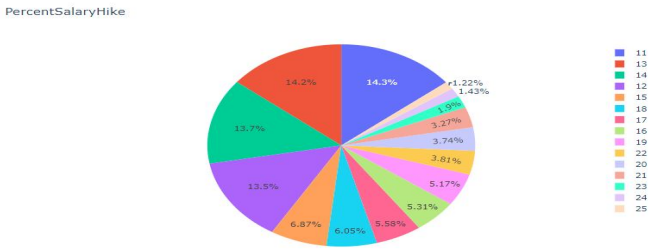


Fig. 4 Monthly Income

Fig. 4 signifies the bar graph of assessment among the monthly income in percentage salary with Attrition. Attrition rate of monthly income is increased.

B. Result and Analysis

For desired results, there are numerous characteristics like department, gender, overtime, business travel, etc. Grounded on these standards, algorithm is developed with the support of dissimilar machine learning algorithms. This algorithm will predict the behaviour of employees who would like to will leave the organization or not. The foretold standards are equated through test values to estimate the correctness of each algorithm. The graph given below in the fig. 5 describes various factors with the help of ROC curve final output, so we can effortlessly achieve which procedure is best for our model. After the graph, we conclude that this algorithm Random Forest gives uppermost accuracy on the HR Employee Attrition dataset whereas Logistic Regression rigidities the lowermost accurateness aimed at the similar dataset.

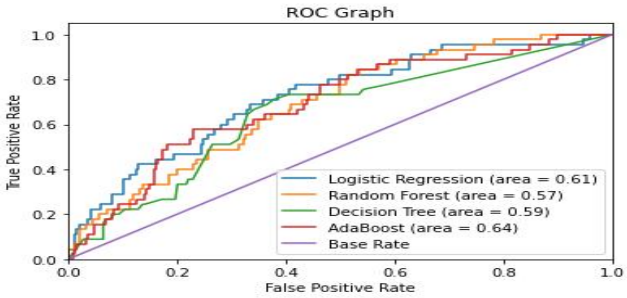


Fig. 5 ROC curve final output

1) Logistic Regression

```
tab1 = confusion_matrix(pred_value,att1_y_test)
tab1
```

```
array([[176,  90],
       [ 76,  95]], dtype=int64)
```

```
accu = tab1.diagonal().sum() *100 / tab1.sum()
accu
```

```
62.013729977116704
```

2) Decision Tree

```
array([[212,   0],
       [ 40, 185]], dtype=int64)
```

```
tab2.diagonal().sum()*100/tab2.sum()
```

```
90.8466819221968
```

3) Random Forest

```
tab_rf
```

```
array([[231,   0],
       [ 21, 185]], dtype=int64)
```

```
rfc.feature_importances_
```

```
array([0.09116461, 0.01972206, 0.08322913, 0.01709955, 0.07459788,
       0.03287762, 0.03323994, 0.        , 0.0839118 , 0.04255988,
       0.01470773, 0.07439083, 0.03561796, 0.04536333, 0.04501948,
       0.0358541 , 0.0507122 , 0.13928889, 0.08064302])
```

```
accu1 = tab_rf.diagonal().sum() *100 / tab_rf.sum()
accu1
```

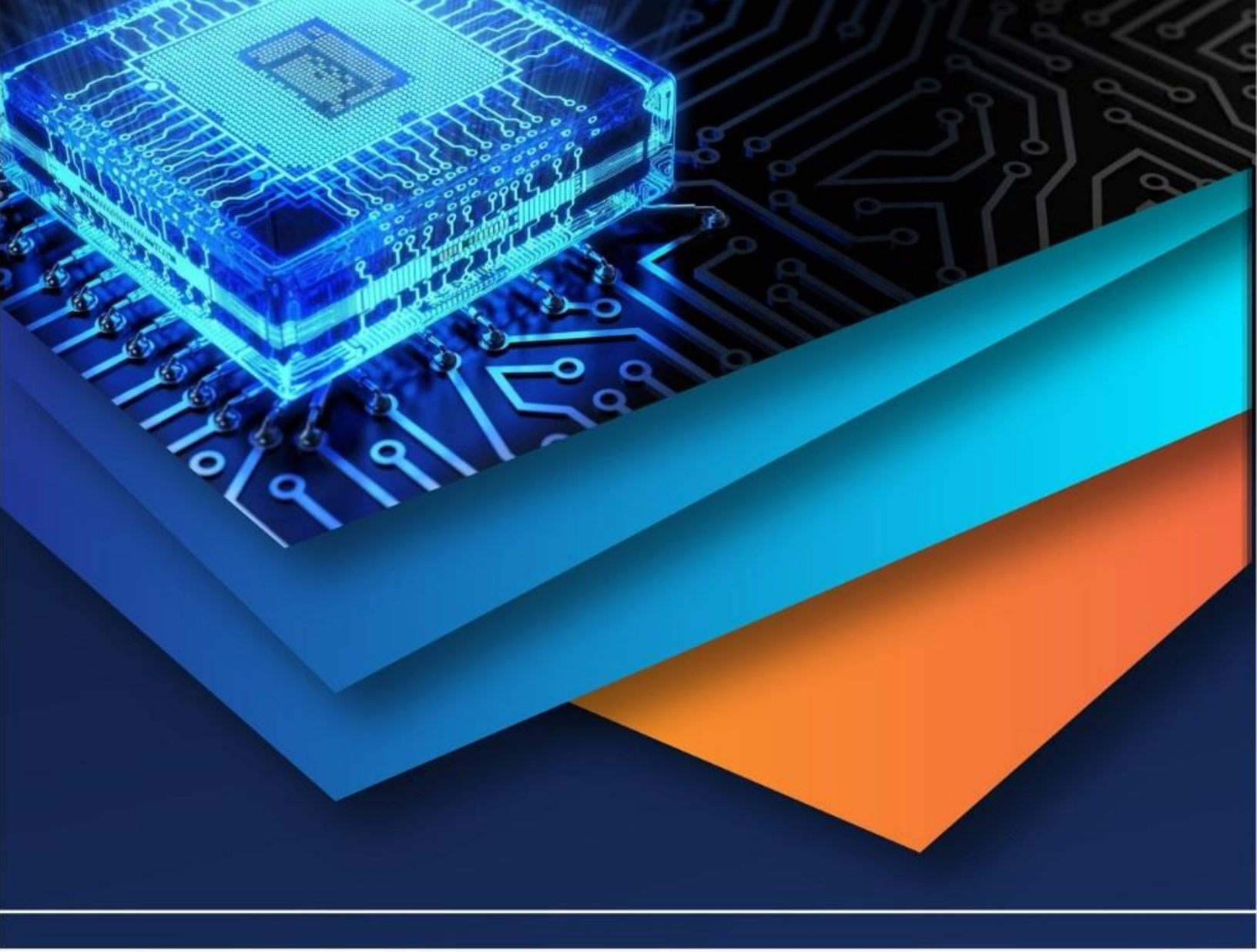
```
95.19450800915332
```

IV. CONCLUSIONS

On the basis of this result paper, algorithm for machine learning is performed precisely with desired results which predict or forecast the employees or workers behaviour who are interested to close the particular company or organization. Result showed that selected algorithm for predicting attrition performs accurately and more closely to desired outputs than any other algorithm. It is detected that; the reason of worker attrition is due to both external and internal factors. This study give strength help association for expressing the factors of employee attrition and can yield suitable stages to diminish the attrition rate.

REFERENCES

[1] Lopes, S. C., Guerra-Arias, M., Buchan, J., Pozo-Martin, F., & Nove, A. (2017). A rapid review of the rate of attrition from the health workforce. Human Resources for Health, 15(1). doi:10.1186/s12960-017-0195-2
[2] Thammasiri, D., Delen, D., Meesad, P., & Kasap, N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. Expert Systems with Applications, 41(2), 321-330. doi: 10.1016/j.eswa.2013.07.046
[3] Cotton J. and tuttle, J.M.,1986." Employee turnover: A metanalysis and review with implications for research" Academy of management review, pp.55-70.
[4] Liu, D., Mitchell, T.R., Lee, T.W., Holtom, B.C. and Hinkin, T.R., 2012." When employees are out of step with co-workers; How job satisfaction trajectory and dispersion influence individual-and unit-level voluntary turnover'. Academy of management journal, PP.1360-1380.
[5] heckert, T.M. and Farabee, A. M., 2006." Turn over intentions of the faculty at a teaching-focused university'. Psychological reports, pp.39-45.
[6] Rish, Irina. 'An empirical study of the naïve bays classifier', IJCAI Workshop on EMPIRICAL METHODS in AI.
[7] David a. Freedman, Statistical Models: Theory and practice, Cambridge university press p.128
[8] Rosenblatt, Frank, Principles of Neurodynamic: Perceptions and the Theory of Brain Mechanisms.
[9] Fawcett, T., 2006.An introduction to ROC analysis. Pattern recognition letters, pp.861-874



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY