Machine Learning Approach for Forecasting Job Appeasement and Employee Corrosion

Swetha M S
Department of Information Science & Engineering. BMS Institute of Technology & Management.
Bengaluru, India
swethams_ise2014@bmsit.in

Mahalakshmi S
Department of Information Science & Engineering BMS Institute of Technology & Management.
Bengaluru, India
maha.shanmugam@bmsit.in

Pushpa S K
Department of Information Science & Engineering BMS Institute Of Technology & Management.
Bengaluru, India
pushpask@bmsit.in

Amrutha T Madihalli
Department of Information Science & Engineering. BMS Institute of Technology & Management
Bengaluru, India
amruthatm1105@gmail.com

Ananya
Department of Information Science & Engineering. BMS Institute of Technology & Management.
Bengaluru, India
ananyasingh0722@gmail.com

Anand Bhardwaj
Department of Information Science & Engineering. BMS Institute of Technology & Management.
Bengaluru, India
anand.bh231@gmail.com

Abstract—Employee turnover imposes costs on the organization. The quit may also cause significant and costly disruptions to the production process. The recent increase in the technological capacity to gather large magnitude of data and analyze it has changed how decision makers use them to decide on making the optimal decision. Employee attrition very similar to customer churn is an important and deciding factor affecting the revenue and success of the company. To avoid this problem, many companies now are taking guide via machine learning strategies to expect employee churn/attrition. In this paper, we are analyzing data from the past and present using different classifications like SVM, Random Forest, Decision tree, Logistic Regression, and an Ensemble model to come up with a better predictive model for the present dataset. Through this we are hoping to help the company predict employee churn and take effective measures to retain the employees and improve their economic loss due to the loss of valuable employees.

Index Terms—Support Vector Machine (SVM), Random Forest, Decision tree, Logistic Regression

I. Introduction

Organizations must consider a lot of factors to keep them as a leading Company in the competitive market of today [3]. Machine Learning and their techniques have given them a useful tool to get an edge in the market after analyzing data collected over years.

Attrition also called wastage rate or total turnover rate can be considered as a silent killer which destabilizes a company from within [2]. Employees may choose to leave the company for a lot of reasons like equal pay, lack of appreciation, long working hours and many more [5]. As employees are the central source of any company, employee attrition has a negative impact on the revenue of the company along with other various consequences like having to invest more in hiring and training new employees, more pressure on the present employees and a radical downslope in the expected performance of the company[1]. Hence Analytics done using Machine Learning and their tools helps us to understand the issue and source of it, as well as come up with effective solutions for it. Using the past and present data for predicting attrition helps in identifying the causes for the churn and stopping the increasing churn over rate.

In our methodology, we have used different classification methods like SVM, Random Forest and Decision tree along with a hybrid model to understand and analyse the performance of different predictive models and compare them using different classification metrics. SVM (Support Vector Machine) are kernel based algorithms used which serves as a tool to separate different classes. Kernel transforms the input data into higher dimensions where it can be solved using linear classifier by drawing a hyper plane. For example, facial expression recognition is of the uses of SVM where it filters out different expressions into their own class divided by hyper-plane. Decision Tree (DT) appears like a tree shaped algorithm to examine and determine a course of action or show statistical probability.

A company may deploy decision trees as a kind of decision support system. Let's consider booking a train to travel as example. First we look into our calendar to see if a train is available on that date .If available, we look at the time suitable to us. Then we consider the price is within our range etc. Like this at each step we make decisions and go further deep down the branch till an outcome has arrived that is the train being booked.

Random Forest builds a forest with a number of decision tree and is an ensembling method. Logistic Regression uses independent variables for coming to a conclusion. A last method used is Stacking, an ensembling method which combines the predictions from well performing algorithms and gives out a better performance. Retention is more important than hiring. The foremost successful organizations are successful because they look after their employees and that skills to retain them within the organization.

This motivated us to research the connection between work fulfilment and representative maintenance by developing a completely unique algorithm to use to the model in making better predications to assist the HR management retain their employees and put them within the right jobs consistent with the satisfaction levels.

The prevailing systems accuracy in predicating isn't much and organizations hesitate to include the model, so we are

motivated to develop a model which satisfies the HR management to use the system [4].

II. Literature Servey

Employment fulfilment is critical to high profitability, inspiration and low turnover rate. Managers particularly the HR group face the difficulties of securing techniques to boost position fulfilment, so their organizations stay serious. Associations who regularly neglect to upgrade work delight or fulfilment are at a danger of losing their most skilled individuals to the contenders. Bosses and chiefs who attempts to augment the potential, innovative capacities, and abilities of the entire workforce include a more prominent favorable position inside the challenge than those that don't. Representatives that are occupied with their work have a superior degree of employment fulfilment. Persuaded laborers give the protection organizations frantically required in these disordered occasions. Right now, the HR the board for representative maintenance they need utilized SVM and random forest as classifier calculations to anticipate the laborer fulfilment and steady loss. We propose to build up a substitution calculation that gives better exactness utilizing Linear Model Tree close by random forest.

[6] Usha P.M. et al., According to the author data mining serves as a method for identifying and analyzing hidden patterns within extensive datasets. The article's primary focus is on discerning the various factors that influence employee attrition within the human resource department of firms or organizations. To achieve this, the researchers utilized data mining approaches, specifically emphasizing the Weka platform. Weka, a data science tool for predictive analytics, employs algorithms like K-Nearest Neighbors (KNN) to cluster data and gain insights into the variables contributing to attrition. Additionally, Weka was employed to assess and analyze the performance and effectiveness of the various algorithms used in the study.

[7] Gunjan, V.K. et al., employed Apache Spark, an open-sourced general-purpose cluster framework, for Big Data Analytics. Their research utilized Multi-Layer Perceptron in Spark to predict attrition, and the output was comprehensively analyzed using graphs, which plotted each attribute and its association with attrition values. Additionally, they provided a user-friendly interface designed to convey results in easily understandable language."

[8] Aniket Tambde et al., focused on creating a professional model for examining and forecasting employee turnover rates. They employed the Random Forest and K-Nearest Neighbors (KNN) machine learning algorithms. Notably, when considering the confusion matrix, Random Forest outperformed KNN in predictive accuracy.

[9] Nesreen El-rayes et al., gathered and analyzed data from randomly selected resumes on Glassdoor, a job-search website, to forecast employee turnover resulting from job changes. Their study highlighted the effectiveness of tree-based models, particularly Random Forest (RF) and Gradient Boosting (GB), which exhibited strong predictive performance when compared to other binary classification techniques.

[10] Mehul Jhaver et al., utilized the Gradient Boosting technique with its regularization-based robustness to forecast employee turnover. They compared this approach with three other common algorithms: Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR). The results indicated that Gradient Boosting outperformed the other three algorithms, emphasizing its efficacy in predictive modeling for turnover scenarios."

III. Proposed System

To design a hybrid model to predict the employee turnover and to understand the reasons for it by going over the past data and enhancing the prevailing model, which may allow the corporate to predict the longer-term occurrence of the event.

By identifying and then comprehending the variables that contributed to attrition or turnover rates, businesses and people will be able to reduce employee turnover, boost productivity, and foster professional development. Managers can demand remedial actions to build and maintain their successful firm thanks to these important and predictive information.

The main goal is to find out the employee turnover using the provided dataset. We are applying many various ways of classification algorithm like Random Forest (RF), Decision tree (DT), Support vector machine (SVM), Logistic regression (LR). We then apply stacked model which is one of the ensemble methods to get a hybrid model. Decision tree is used as Meta-model. It is trained on the base model's prediction and on the test data. The inputs to the basic models may also be included in the training data for the meta-model. The meta-model will fit the training data, and a final prediction that is more accurate than the predictions provided by individual machine learning algorithms is obtained. This model is referred to as a "stacked hybrid model".

The main objectives are:

• To provide appropriate and obvious incentives for the company to prevent staff departures or, at the very least, to be prepared to predict and analyze what variables contributed most to turnover rate.

• To design and construct a model that determines whether a specific employee will quit the company or not.

• To more accurately anticipate the turnover rate using a hybrid approach.

• To develop and enhance various retention tactics for selected personnel.

A company's ability to retain a successful business model and a positive company culture is a surefire indicator of its ability to succeed and expand in the future. Organizations and people can prevent this from happening and even boost employee productivity and improved growth by first recognizing and then comprehending the reasons that were associated with turnover rate or attrition. These helpful and predictive insights give managers the chance to demand corrective actions to build and maintain their profitable business.

IV. Design and Implementation

A version is intended to be used for prediction, analysis, or interpretation. While the interpretation is qualitative, the projection may be examined statistically. How well a model per-

forms on untrained data can be used to gauge its predicted accuracy. The use of techniques like validation may be evaluated. It is important to make intelligent choices when it comes to the algorithms that can be used, as well as the biases and reductions on the hypothesis area of potential fashions that could be developed for the problem. To create a hybrid model, we use the stacking model, one of the ensemble approaches. The meta-model utilized is a decision tree.

Here, Decision Tree Model, Logistic Regression Model, Random Forest Model, and SVM Model are employed as the basis models.

A. System Architecture

The whole system architecture is represented in the flowchart below. Gathering the data is the first stage in the process; in this case, we used the employee dataset. Next, we loaded the data and performed scrubbing, in which we normalized all the missing values and cleaned the dataset. 25% of the dataset will be used for testing, while the remaining 75% will be used for training. The training dataset will undergo data pre-processing, which aids in positioning the data correctly and getting it ready for machine learning algorithms.

Once data preparation is completed, we perform data analysis and visualization on different features of the dataset by plotting several graphs. This helps in giving a clear picture of the distribution of personal traits and helps choose the right qualities to use during prediction. Additionally, we use feature importance on the dataset that informs us of the significance of the features based on the ratings the feature importance model assigns to them. The three features that are the strongest estimators of the outcome variable, or employee turnover, for the employee dataset that we have used are "satisfaction," "year at company," and "evaluation."

Next step is modelling the 19, where we train four different ML algorithms: Decision tree (DT), Random Forest (RF), Support Vector Machine (SVM) and Logistic Regression (LR). Then we get the prediction of each algorithm.

Now, in order to obtain a hybrid model, we apply the stacking model, one of the ensemble approaches. The meta-model utilized is a decision tree. It is trained using the test data and the predictions from the basic model. The training data for the meta-model may also comprise the inputs to the fundamental models. The forecast generated by the meta-model will be more accurate than those made by separate machine learning algorithms since it will fit the training data. A "stacked hybrid model" is the name given to this model. This hybrid model combines multiple diverse skilled machine learning techniques while also reducing the errors in predictions made by the basic models.

B. Algorithms

1) Algorithm for Decision Tree
We are making use of ID3 algorithm.
Step1. Using Equations (1) and (2), determine the information gain for each and every attribute:

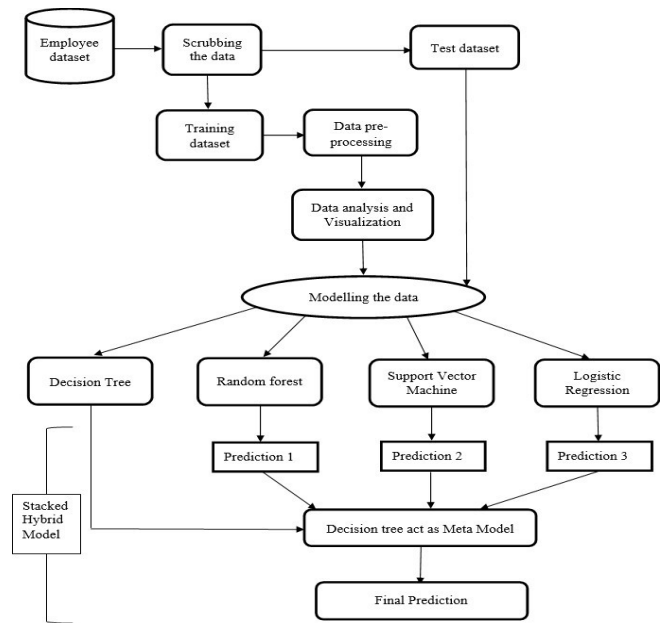$$\text{Entropy}(S) = -\sum_{i=1}^{c} p_i \log_2 p_i \qquad (1)$$

Fig. 1. System Design

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \quad (2)$$

Step2 Step 3 should be taken if two attributes yield the same gain value, otherwise step 4 should be used.

Step3. Choose a quality at random.

Step4. Pick an attribute with a high gain value right away.

Step5. Make the chosen attribute the decision tree's root node.

Step6. Make the value of the chosen attribute a child node.

Step7. If there are still unclassified examples, proceed to step 8, otherwise, move on to step 9.

Step8. Continue with the following characteristics. Repetition is required.

Step9. End.

2) Algorithm For Random Forest
Random forest can be created using 2 different stages:

a) Random forest creation:

Step1. Choose the features randomly such that number of selected features must be less than total features.

Step2. From the selected features, pick a root node 'r' using the best split point.

Step3. Make root node's value as child node using best split.

Step4. Repeat the step1 to step3 until single decision tree is formed.

Step5. Repeat step1 to step4 to create many numbers of trees and construct a forest with that.

b) To make prediction from the random forest created in the first stage:

Step1. Consider the test features and use the rules of every decision tree that has been generated at random to predict and store the outcome.

Step2. For every expected result, determine the votes.

Step3. Find the expected outcome with highest voted as the final prediction from algorithm of random forest.

3) Algorithm for Svm

Step1. Define an optimal hyper-plane which must be maximum margin.

Step2. Find the solution for non-linear data as well with the help of kernel method.

Step3. Project data to a high dimensional space where the classification with linear decision surfaces are easier.

a) Steps to represent an optimal hyper-plane:

Step1. Take the training data of n points:

(X1, y1), (x2, y2) …. (Xi, Yi)

Where xi- p-value vector for point 1

Yi- binary class value of 1 or -1

Thus, there are two classes 1 and -1

Step2. Assuming that the data is indeed linearly separable, the classifier hyper-plane is defined as set of points that satisfy the Eq. (3).

$$\vec{w} \cdot \vec{x} + b = 0 \qquad (3)$$

Step2. Calculate the hard margin which can be defined as Eq.(4) and Eq. (5):

$$x_i \cdot w + b - 1 \qquad (4)$$

$$x_i \cdot W + b = -1. \qquad (5)$$

Step3. Calculate the width of hard margin using. Eq. (6):

$$2/\|w\| \qquad (6)$$

Where 'w' is the width of the margin.

Step4. Finally, we can find 'w' (weight vector) for the features such that there is a widest margin between two classes.

4) Algorithm for Logistic Regression

Step1. Given a data (x, y), build a randomly initialized matrix for weight. Then, by features, we multiply it using Eq. (7).

$$a = w_0 + w_1 x_1 + w_2 x_2 + \cdots w_n x_n \qquad (7)$$

Where x- matrix of values y- vector.

Step2. Pass the output obtained in step1 to a link function given in Eq. (8).

$$\hat{y}i = 1/\left(1 + e^{-a}\right) \qquad (8)$$

Step3. Calculating the cost for this iteration whose formula is in Eq. (9):

$$\text{cost}(w) = \left(-1/m\right) \Sigma y_j \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \qquad (9)$$

Step4. Calculate the derivative of this cost:

And update the weights using Eq. (10) and Eq. (11):

$$d w_j = \sum_{i=1}^{i=n} \left(\hat{y} - y\right) x_j^i \qquad (10)$$

$$w_i = w_j - \left(\alpha * d w_j\right) \qquad (11)$$

5) Algorithm For Stacked Hybrid Model

There are 3 stages:

a) Build the ensemble

Step1. Determine the list of base models.

Step2. Determine the meta-model algorithm.

b) Train the ensemble:

Step1. All 'L' base models are made to learn on the training dataset.

Step2. Run k-fold cross validation on each base model, then compile the cross-validated predictions from each, denoted by p1, p2,, pL.

Step3. Combine N cross-validated predicted values from every base model to form new N*LN*L feature matrix using Eq. (12). The "level-one" data was named in accordance with the original response vector.

$$n\left\{\left[\begin{matrix}p_1\end{matrix}\right] \cdots \left[\begin{matrix}p_L\end{matrix}\right]\left[\begin{matrix}y\end{matrix}\right] \rightarrow n\left\{\overbrace{\left[\begin{matrix} Z \end{matrix}\right]}^{L}\left[\begin{matrix}y\end{matrix}\right]\right.\right. \qquad (12)$$

Step4. Train the meta-model on the "level-one" data using Eq.(13) .

$$Y = f(Z) \qquad (13)$$

c) Predict on new data:

Step1. Generated predictions from the base models will be taken.

Step2. Feed those predictions into the meta-model to generate the final result.

C. Implementation

In current system only limited numbers of techniques are used from the huge collection of data mining techniques for prediction. In above advanced system generated we have applied few algorithms like k-nearest neighbor (KNN), Support vector machine (SVM), Logistic regression (LR), Decision tree (DT) and ensemble model. Fundamentally our data set contain employee, satisfaction grade, assigned projects and work efficiency spent in firm.

In our system we scrub the data so that there are no null values in the data set and if any should remove the null values. We choose the best features for employee attrition, and they are in Table -1.

TABLE I. ATTRIBUTES

Features	Data type
Job Satisfaction level	Number [10]
Number of Projects	Number [10]
Average Monthly hours	Number [10]
Any Work Accident	Number [10]
Last Evaluation	Number [10]
Time Spent in Company	Number [10]
Department	Varchar [20]
Any Promotions	Number [10]
Salary	Varchar [20]
Turnover	Number [10]

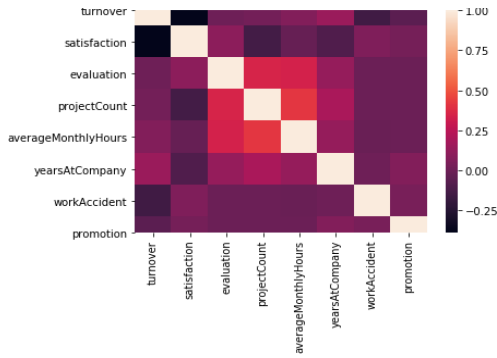We use the co-relation matrix and heat map to analysis the below:

Fig 2 Heat map

CASE 1: Positive (+ve) interdependence

Here project count (PC), average monthly hour (AMH), evaluation i.e., rank of employee is considered. The employee with the ability of covering maximum number of monthly hours and outputting the completion of multi assigned project with is comparatively high has been observed with highly ranked.

CASE 2: Negative (-ve) Relationships

Here turn over (TO), satisfaction is highly tie-up with each other. In other words, we state that employees having low satisfaction are directly proportional to employees leaving the firm. The heat map is shown below Figure 2. To inspect the scattering on the features.

The following are some essential views and thus features:

• Satisfaction - Employees were divided into two groups based on their level of satisfaction: low satisfaction and high satisfaction.

• Evaluation - Using a bimodal function, employee performance was classified as low (below 0.6) or high (greater than 0.8).

• Average Monthly Hours - Employees' working efficiency is an important attribute, but being able to fully utilize a high performance is also important, so work time is analyzed with respect to average monthly hour worked by each employee and is further classified based on less than 150 hours and more than 250 hours. This means that the higher the working efficiency, the higher the average monthly hours worked. These features discussed above are interviews.
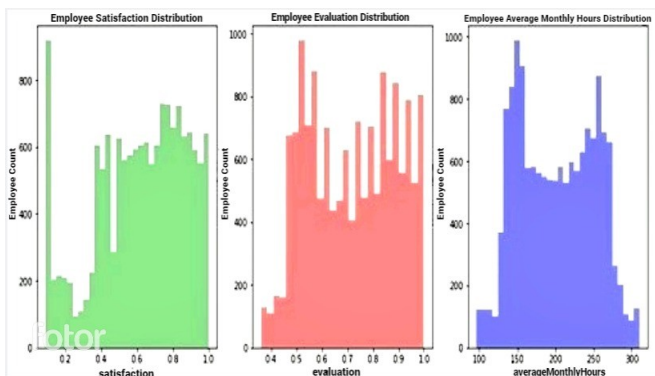

Fig. 3. Distribution Plot

The satisfaction v/s evaluation is the utmost gripping graph. The employees who left the firm were analyzed under 3 main clusters.

• *Cluster 1 (Hard-working and Sad Employee):*

It is important to treat the employees with working efficiency more than 0.75 so that they remain in the firm. Such employees improve the firm and there can someday lead departments in firm. When such employees are not treated well i.e. satisfaction less than 0.2. Such employees may leave the firm.

• *Cluster 2 (Bad and Sad Employee):*

In this case employees with satisfaction level 0.35-0.45 whose working efficiency will be below 0.58. Retaining such moderate employees are necessary as they can be directed by high evaluated to work in order reduce their work load. Hence if their satisfaction is low tendency to leave the firm is more.

• *Cluster 3 (Hard-working and Happy Employee):*

The employees with satisfaction level at between 0.7-1.0 and working efficiency greater than 0.8.Such employees should be treated well in firm in order to keep them associated to firm. They are well satisfied with their work and their performance is highly evaluated.
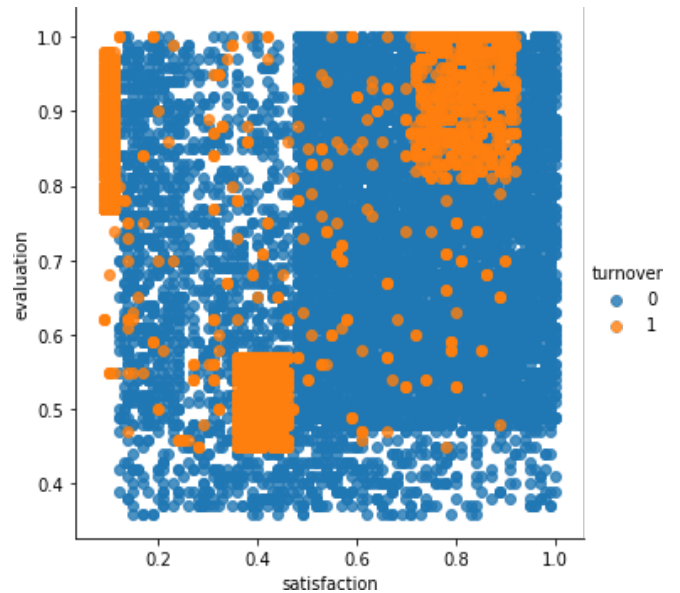

Fig.4. Satisfaction v/s Evaluation

V. RESULTS AND DISCUSSION

The best model performance out of the four (Decision Tree Model, SVM Model, Logistic Regression Model, Random Forest Model) was Random Forest. Hence Random Forest along with stacking classifier is the best model to predict employee attrition with an accuracy of 0.98. This is Concluded by using below Table-2.

When we compare the performance of all the algorithms which are ran on the dataset, we plot a graph i.e. the model performances graph by which we get to know that, the highest score is for the Hybrid Stacked Model which is shown below in the figure-5.

A. Final Interpretation

With the use of all this data, a final analysis is completed to determine the most likely reasons an employee departed a firm. These interpretations are given:

TABLE II.

Model	Accuracy_score	Recall_score	Precision	f1_score	Area_Under_curve	Kappa_metric
Logistic Regression	0.7898	0.3657	0.5957	0.4532	0.644	0.3322
Decision Tress	0.9382	0.8825	0.8616	0.8719	0.9191	0.8312
Random Forest	0.9831	0.9478	0.9807	0.9639	0.971	0.9529
SVM	0.7658	0.7836	0.5054	0.6145	0.7719	0.4573
Stacked Model	0.9862	0.959	0.9828	0.9707	0.9769	0.9617

1. When underworked (less than 150 hours per month or 6 hours per day), employees frequently leave their jobs.
2. When workers are overworked (more than 250 hours per month or 10 hours per day), they frequently abandon their jobs.
3. Workers with extraordinarily high or bad evaluations ought to be considered for a high turnover rate.
4. The majority of turnover among employees is caused by those making low to medium salary.
5. Employees having a project count of 2, 6, or 7 faced the risk of being let go by the company.
6. The best indicator of staff retention is employee happiness.
7. Workers with four and five years with the company should be considered for a high turnover rate.
8. Employee satisfaction, year's At Company, and evaluation were the three key factors in deciding turnover.

B. Future Enhancement

This research could incorporate predictive analytics techniques to forecast employee attrition trends in a more dynamic and forward-looking manner. By considering economic indicators, industry-specific factors, and market conditions, organizations could gain valuable insights for proactive workforce planning. Additionally, implementing a real-time feedback mechanism that collects continuous input from employees, leverages sentiment analysis and natural language processing, and integrates this data into the attrition prediction model can enhance its accuracy. Moreover, the development of a recommendation engine to provide customized retention strategies for individual employees based on their predictive factors could empower HR departments to take targeted actions to reduce attrition and improve overall workforce satisfaction. Finally, exploring the use of blockchain technology for data security could enhance the trust and transparency of attrition prediction and management, ensuring the confidentiality and integrity of employee information while complying with data privacy regulations.
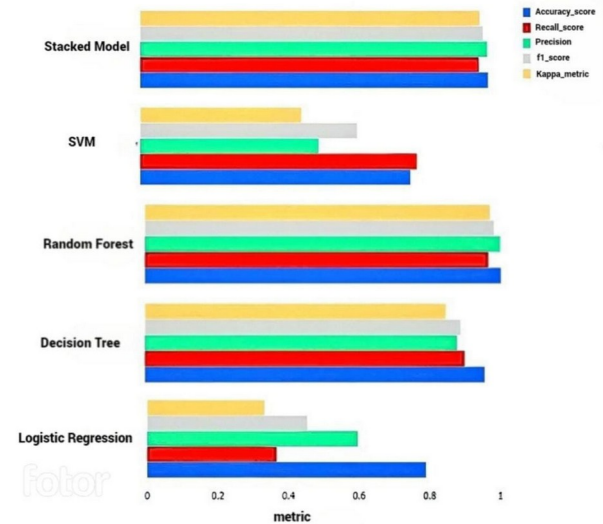

Fig. 5. Model Performances.

VI. Conclusion

As we progress with our research, it's obvious that **Random Forest along with stacking classifier and Decision tree as a meta-model** is the best model to predict employee attrition with an accuracy of **0.98**. Clearly we can say this is a best suited approach. By this we can also conclude by saying, it's important to choose the efficient base models in order to make stacking method more accurate otherwise stacking technique is not recommended.

We trained the computer with limited data (14000 odd records divided into 75percent training data and 25 percent test data), but it can also work effectively on a broad dataset. From the reference above it is very clear how the estimation of attrition plays an important role in the businesses

Given that large turnover percentages are caused by employees, whether they have high or negative ratings, they should be considered. When overworked (more than 250 hours per month or 10 hours per day) or underworked (less than 150 hours per month or 6 hours per day), employees frequently leave their jobs. Employees with low to medium salaries are most likely to leave the organization. Employees who have worked on 2, 6, or 7 projects had a higher departure rate. The best indicator of employee turnover is employee satisfaction. Consideration should be given to employees who have worked for the company for four and five years. The main determinants of turnover were years At Company, assessment, and employee satisfaction.

- Efficiency of the stacking model depends on the base models.
- Base models must be as different as possible to get the better result.
- Stacking model makes the result difficult to interpret.

References

[1] Qin Zhou., (2017) "The Impact of Job Satisfaction effect on Turnover Intention: An Empirical Study based on the Circumstances of China".
[2] Phillips, J. D., (1990) "The price tag on turnover" Personnel Journal, vol. 12, pp. 58-61.

[3] Ms. Ankur Jain, (2018) "Impact of TQM on employees' job satisfaction in Indian software industry".

[4] Bosele, P. and Wiele, T.V.D. (2019) "Employee perceptions of HRM and TQM and its effects on satisfaction and Intention to leave", Managing Service Quality, 2019.

[5] Zhu Xiaoyan, Li Yanping, (2018) "A Study on Psychological Contract, Job Satisfaction and Turnover Intention in Banking Industry".

[6] Usha. P. M. and Dr. NV Balaji. ANALYSING EMPLOYEE ATTRITION USING MACHINE LEARNING. IOP Conference Series Materials Science and Engineering, 1085(1), 012029, 2021. DOI:10.1088/1757-899X/1085/1/012029

[7] Gunjan, V. K., Garcia Diaz, V., Cardona, M., Solanki, V. K., & Sunitha, K. V. N. (Eds.). Prediction of Employee Attrition and Analyzing Reasons, ICICCT-2019

[8] Aniket Tambde, Dilip Motwani. Employee Churn Rate Prediction and Performance Using Machine Learning. Published By: Blue Eyes Intelligence Engineering & Sciences Publication, 2019, DOI: 10.35940/ijrte.B1134.0982S1119

[9] Nesreen El-rayes, Michael and Stephen Taylor. An Explicative and Predictive Study of Employee Attrition using Tree-based Models. Conference: Hawaii International Conference on System Sciences. DOI:10.24251/HICSS.2020.175

[10] Mehul Jhaver; Yogesh Gupta; Amit Kumar Mishra. Employee Turnover Prediction System. Conference: 2019 4th International Conference on Information Systems and Computer Networks (ISCON) 2019 . DOI:10.1109/ISCON47742.2019.9036180

[11] Ganesan Santhanam, Raja Jayaraman, Dr.V. Badrinath,"Influence of Perceived Job Satisfaction and Its impacts on Employee Retention in Gulf Cooperation Countries" (2019).

[12] Richard, F. G., Joseph, M.L., Billy, B., (2016) "Job satisfaction, Life satisfaction and Turnover Intent: Among Food-Service Managers," Cornell Hotel and Restaurant Administration Quarterly, vol. 42(2).

[13] Sukhadiya J, Kapadia H, Prof. D'silva M, 2017 "Employee Attrition Prediction using Data Mining Techniques" International Journal of Management, Technology And Engineering ISSN NO : 2249-7455 Volume 8, Issue X, OCTOBER/2018.

[14] Sikaroudi1 A, Ghousi1 R, 2015, "A data mining approach to employee turnover prediction". Journal of Industrial and Systems Engineering Vol. 8, No. 4, pp106-121 Autumn 2015

[15] Rygielski C, Jyun-Cheng Wang b, Yen D, 2018 "Data mining techniques for customer relationship management" Technology in Society, Novemeber 2022

[16] Jain, R., & Nayyar, A. (Year). Predicting Employee Attrition using XGBoost Machine Learning Approach. Conference: 2018 International Conference on System Modeling & Advancement in Research Trends (SMART). DOI:10.1109/SYSMART.2018.8746940.

[17] Brockett, N., Clarke, C., Berlingerio, M., & Dutta, S. (Year). A System for Analysis and Remediation of Attrition. Conference: 2019 IEEE International Conference on Big Data (Big Data). DOI:10.1109/BigData47090.2019.9006333.

[18] Bindra, H., Sehgal, K., & Jain, R. (Year). Optimization of C5.0 Using Association Rules and Prediction of Employee Attrition. DOI: 10.29228/Joh.63712