

Employee Attrition System: Using Machine Learning to Evaluate Performance of the Staff

¹Asiya Anjum

¹University College of Engineering; UCE(A).

¹Department of CSE, ¹Osmania University, Hyderabad, India.

Abstract: Employee's attrition prediction has recently become a major issue in organizations. Employee turnover is a notable problem for organizations, particularly when highly qualified, technical, and key employees leave for better opportunities. This results in a loss of income because a trained employee must be replaced. As a result, we evaluate the common reasons for employee attrition using recent and historical employee data. Methods for supervising machine learning are described, demonstrated, and implemented. Evaluated for predicting employee turnover within an organization. In this study, numerical experiments for real and simulated human resources datasets representing organizations with small, medium, and large employee populations are performed on the human resource data using Decision Tree, Logistic Regression, SVM, KNN, Random Forest, and Naive Bayes methods. To prevent employee attrition, we apply the feature selection method to the data and analyze the results. This helps companies predict employee attrition and helps their economic growth by lowering human resource costs.

IndexTerms: Employee, Opportunities, Predicting, Decision Tree, Logistic Regression, SVM, KNN, Random Forest, and Naive Bayes methods, Machine Learning, Cost, Data mining, Data base, Data Science.

I. INTRODUCTION

An interdisciplinary field, Data Science deals with processes and systems that are used to extract knowledge or insights from large amounts of data. Data extracted can be either structured or unstructured. Data science is a continuation of data analysis fields like data mining, statistics, predictive analysis. A vast field, data science uses a lot of theories and techniques that are a part of other fields like information science, mathematics, statics, chemo metrics and computer science. Some of the methods used in data science includes probability models, machine learning, signal processing, data mining, statistical learning, database, data engineering, visualization, pattern recognition and learning, uncertainty modelling, computer programming among others. What is clear is that data science is solving problems. Data is everywhere, and the uses we are making out of it (science) are increasing and impacting society more and more. The Data Science Trends are largely a continuation of some of the biggest trends of the recent years including Big Data, Artificial Intelligence (AI), Machine Learning (ML), along with some newer technologies like Block chain, Edge Computing, Server-less Computing, Digital Twins, and others that employ various practices and techniques within the Data Science industry. When there is a loss of employees in any organization, there are a lot of problems which are caused, starting from an empty position in the organization. Filling these positions is a lengthy process of interviewing candidates, training them and integrating them into teams. This makes retention of valuable talent essential to the smooth functioning of the organization. HR is constantly looking out for ways to predict which employee is unhappy with the current job in order to try and convince the employee to stay or to cushion the blow of loss of talent by looking for replacements. Accurate employee attrition prediction has tremendous monetary and productivity benefits for the organization and so machine learning is used to train classifier models using the dataset.

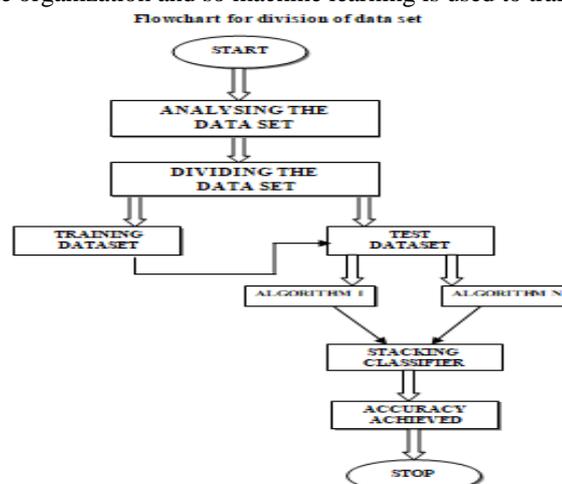


Fig.1.1. Flowchart for Stacking and Meta classifier

The dataset is called IBM HR Analytics Employee Attrition & Performance, taken from Kaggle. The dataset consists of 35 variables such as Age, Daily Rate, Hourly Rate, Job Satisfaction, Overtime and Monthly Income being some of the important factors that contribute to attrition. The target variable is Attrition which has two values - Yes and No. Supervised learning is a process of training the model to map the function between labeled input variables and target output variable. The target in this dataset is a binary variable and so classification is used to solve this problem. Classification is a supervised learning technique which uses a model to

predict categorical values from the input data. Stacked Classifier combines multiple classifying algorithms which are trained on the training set and is used with a meta-classifier which predicts the output class using the output values from the Stack Classifier. The data is used to train and ensemble Stacked Classifier consisting of three classifying algorithms: Support Vector Machine, Decision Tree Classifier and Adaptive Boosting with the meta classifier algorithm used being Logistic Regression for obtaining better accuracy. The dataset goes through cleaning, pre-processing, feature engineering, training and modeling to obtain an accuracy of 90% or more using Stacked Classifier.

II. PROBLEM STATEMENT:

Evaluating the staff performance using the dataset consists of variables such as Age, Daily Rate, Hourly Rate, Job Satisfaction, Overtime and Monthly Income (being some of the important factors that contribute to attrition) for all the employees of an organization to predict which employee is unhappy with the current job so that the HR can immediately start looking for their replacements.

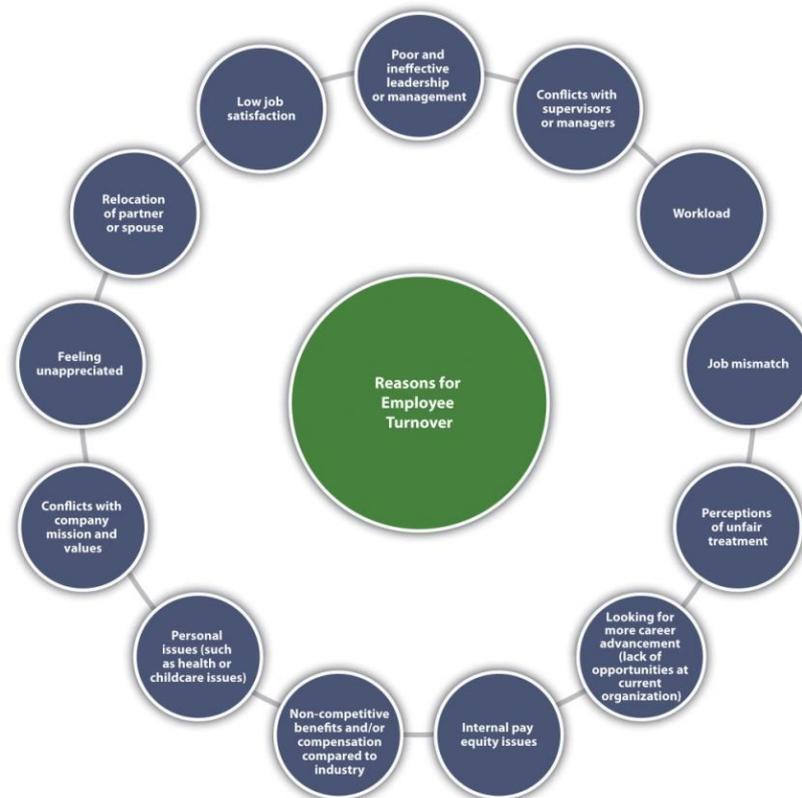


Fig.1.2. Reasons for Employee Attrition

III. TYPE OF EMPLOYEES TURNOVER:

1. **Internal vs. external turnover:** Like recruitment, turnover can be classed as 'internal' or external. Internal turnover involves employees leaving their current position, and taking a new position with the same organization. Both positive (such as increased morale from the change of task and supervisor) and negative (such as project/relational disruption,) effects of internal turnover exist, and thus this form of turnover may be as important to monitor as its external counterpart. Internal turnover might be moderated and controlled by typical HR mechanisms, such as an internal recruitment policy or formal succession planning.

2. **Skilled vs. unskilled employees:** Unskilled positions often have high turnover, and employees can generally be replaced without the organization or business incurring any loss of performance. The ease of replacing these employees provides little incentive to employers to offer generous employment contracts; conversely, contracts may strongly favor the employer and lead to increased turnover as employees seek, and eventually find, more favorable employment. However, high turnover rates of skilled professionals can pose as a risk to the business or organization, due to the human capital (such as skills, training, and knowledge) lost. Notably, given the natural specialization of skilled professionals, these employees are likely to be re14 employed within the same industry by a competitor. Therefore, turnover of these individuals incurs both replacement costs to the organization, as well as resulting in a competitive disadvantage to the business.

3. Voluntary vs. involuntary turnover:

1. Involuntary: - In this case, the employee ceases to work for the company due to being laid off or terminated. It could be because the company is trying to cut costs, or the employee has violated company policy.

2. Voluntary: - Voluntary turnover is when an employee terminates employment on their own accord. There are several possible causes:-

- 2.1 Relocation of family
- 2.2 Going back to school
- 2.3 Starting a family
- 2.4 Taking care of an elderly relative
- 2.5 General job dissatisfaction such as low pay, lack of benefits, or poor management

IV. CAUSES OF EMPLOYEE TURNOVER:

There are several factors that cause high turnover within companies. This report will focus on voluntary turnover, because voluntary turnover is something that companies are more able to control.

Employees voluntarily quit for several reasons, specifically:

1. Pay is too low
2. Lack of benefits
3. Tasks are too repetitive
4. Circumstances listed above such as family, school, or moving.
5. Poor management
6. Lack of advancement
7. Burnout
8. Compensation package differences Job and employee skill mismatch: the job may be less or more satisfying and challenging according to the employee.
9. Inferior facilities, tools, etc.
10. Less recognition
11. Less or no appreciation for work done
12. Less growth opportunities
13. Poor training
14. Poor supervision less work and life balance practices there are some other numbers of factors that contribute to employee turnover.

We explore some of these factors in more detail below:-

1. **The economy** - In exit interviews one of the most common reasons given for leaving is the availability of higher paying jobs. Some minimum wage workers report leaving one job for another that pays only 50 cents an hour more. Obviously, in a better economy the availability of alternative jobs plays a role in turnover, but this tends to be overstated in exit interviews.

2. **The performance of the organization** - an organization perceived to be in economic difficulty will also raise the spectre of impending layoffs. Workers believe that it is rational to seek other employment.

3. **The organizational culture** - much has been written about organizational culture. It is sufficient to note here that the reward system, the strength of leadership, the ability of the organizations to elicit a sense of commitment on the part of workers, and its development of a sense of shared goals, among other factors, will influence such indices of job satisfaction as turnover intentions and turnover rate.

4. **The characteristics of the job** - some jobs are intrinsically more attractive than others. A job's attractiveness will be affected by many characteristics, including its repetitiveness, challenge, danger, perceived importance, and capacity to elicit a sense of accomplishment. A job's status is also important, as are many other factors.

5. **Unrealistic expectations** - Another factor is the unrealistic expectations and general lack of knowledge that many job applicants has about the job at the time that they receive an offer. When these unrealistic expectations are not realized, the worker becomes disillusioned and decides to quit.

6. **Demographics** - empirical studies have demonstrated that turnover is associated in particular situations with demographic and biographical characteristics of workers. But to use lifestyle factors (e.g. smoking) or past employment history (e.g. many job changes) as an explicit basis for screening applicants, it is important for legality and fairness to job applicants to verify such bio data empirically.

V. MEASURE TO REDUCE THE RATE OF TURNOVER:

1. **Benefits:** Offering employees an affordable medical, dental, and vision package in this day and age is a great way to keep employees happy. Healthy employees are happy employees, and being able to provide affordable health care for their spouses and families as well is something every company should offer.

2. **Added Bonus:** An added bonus could be vacation time, sick leave. On-site child care would be extremely helpful for parents who have to work long or late hours—especially single parents.

3. **Higher pay:** Giving employees regular raises and paying well over minimum wage 20 would be an incentive for them to stay.

4. **A set schedule:** In food service and retail, and most service industries as well as health care, employees are forced to work six or more—even up to ten days in a row without a day off. Days off may even be split up, so the employees never really get a chance to rest. Giving them the opportunity to choose which days off they want, or at least giving them two in a row and not working them more than five, would be extremely beneficial in employee retention. This would also increase productivity and would be beneficial to the company.

5. **Job variation:** Employees get burned out on performing the same job every hour of every day, day in and day out for years, even months. Cross-training should be done, especially in food service and retail, in order to avoid burnout.

6. **A positive attitude from superiors:** Most employees don't like negativity from their superiors. Instead of always being told what they're doing wrong, they need positive reinforcement as well as constructive criticism. Managers and supervisors should always have a positive attitude toward their employees and never insult, criticize, or berate them.

Proper training for management: Managers should be trained thoroughly and consistently. The policies from location to location should be the same, and every manager and supervisor in the company should be trained the same way and be in agreement and consistent with company policies. Managers should be trained to treat their employees with respect, because without those employees, the business could not operate

VI. LITERATURE SURVEY:

“Factors that lead to Employee Attrition:

Uncover the factors that lead to employee attrition and explore important questions such as 'show me a breakdown of distance from home by job role and attrition' or 'compare average monthly income by education and attrition'. The various other factors are Salary hike, over-time, age, gender etc. This is a fictional data set created by IBM data scientists. “[1]

“Classification algorithms:

a) Logistic regression: It is a machine learning algorithm for classification. In this algorithm, the probabilities describing the possible outcomes of a single trial are modeled using a logistical function.

b) Naive Bayes: This algorithm is based on Bayes theorem with the assumption of independence between every pair of features. Naive Bayes classifiers work well in many real-world situations such as document classification and spam filtering.

c) Decision tree: It gives data of attributes together with its classes; a decision tree produces a sequence of rules that can be used to classify the data.

d) Random forest: Random forest classifier is a meta-estimator that fits a number of decision trees on various sub-samples of datasets and uses average to improve the predictive accuracy of the model and controls over-fitting. The sub-sample size is always the same as the original input sample size but the samples are drawn with replacement. “[2]

“Stacking Classifier:

Stacking is an ensemble learning technique to combine multiple classification models via a meta-classifier. The individual classification models are trained based on the complete training set; then, the meta-classifier is fitted based on the outputs-meta-features of the individual classification models in the ensemble. The meta-classifier can neither be trained on the predicted class labels or probabilities from the ensemble. Please note that this type of stacking is prone to over fitting due to information leakage. The related StackingCVClassifier.md does not derive the predictions for the 2nd-level classifier from the same dataset that was used for training the level-1 classifiers and is recommended instead.

a) Simple staked classification

b) Using probabilities as meta-features

c) Stacked classification and Grid search

d) Stacking of classifier that operates on different feature subsets.” [3]

“Ensemble methods:

The goal of ensemble methods is to combine the prediction of several base estimators built with a given learning algorithm in order to improve generalizability/ robustness over a single estimator. Two families of ensemble methods re usually distinguished: one is the averaging method and the other is boosting method. In general, a tree of depth h can capture interactions of order h. There are two ways in which the size of the individual regression trees can be controlled. Regression: It supports a number of different loss functions for regression, the default loss function for regression is at least squares.” [4]

“Stacking with probability distributions:

Stacking with probability distributions and multi-response linear regression

Ting and Witten (1999) stack base-level classifiers whose predictions are probability distributions (PDs) over the set of class values, rather than single class values. The meta-level

attributes are thus the probabilities of each of the class values returned by each of the base level classifiers. The authors argue that this allows them to use not only the predictions, but also the confidence of the base-level classifiers.” [5]



Fig.2.1. Reasons for Employee Attrition

Future scope: Currently, the dataset size is 1470 rows. If this size is increased, the accuracy can be further improved by the stacked classifier. The logic here is that if there is more data, it will enable training more exhaustively and one could also implement deep learning algorithms. Furthermore, feature engineering will not be required as principal component analysis could work on the 35 features of the dataset and use the important ones which contribute to the attrition value.

Limitations: In the existing systems they used only few of data mining techniques for data prediction. Employee attrition effects in financial, time and effort loss for organizations. It is a big issue since a trained and experienced employee is difficult to substitute and it is cost effective.

Comparison: We compare our scheme with related work in terms of features, computation, and space requirement. We exclude the cost and space for those certificates verification in this comparison, as it may vary in different scenarios. In the earlier inventions of this project, the use of stacking classifier was not introduces whereas in this project the stacking classifier is employed here which performs the computation, by comparing all the algorithms and gives its result to meta classifier. The Meta Classifier gives the best algorithm depending on the high accuracy of such algorithm. Hence higher accuracy is achieved here. [6]

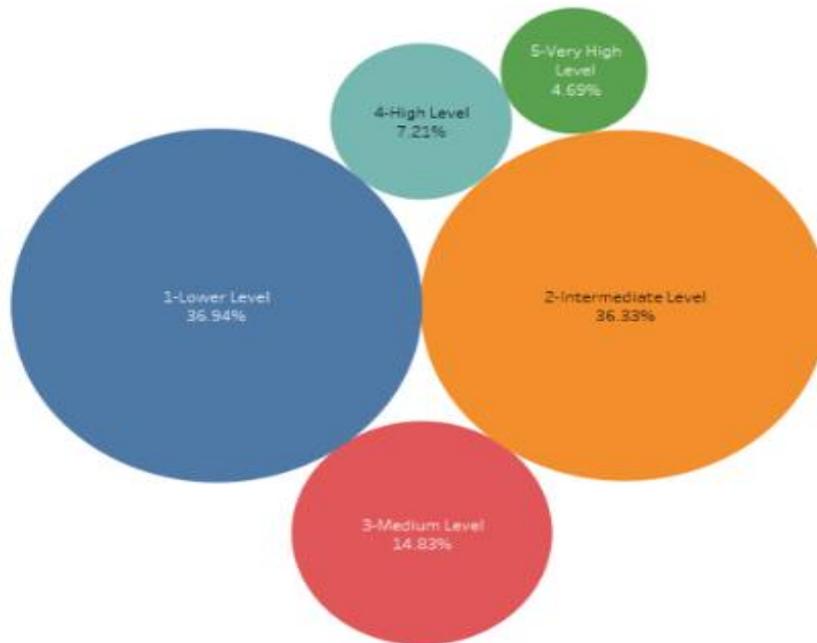


Fig.2.2. Attrition by job level

VII. SYSTEM ANALYSIS

7.1 Existing Problem:

In any organization, managing Human Resources is an important task. Loss of employees lowers the overall productivity of the team and is also financially costly. Attrition of employees leaves behind a void that is costly to fill Machine Learning can be utilized for predicting an employee’s attrition. Adding unstructured, textual data into a conventional attrition identification. The outcome is raise performance in attrition identification analysis. This study supportive for marketing decision makers to improved recognize customer those have probability to attrition.

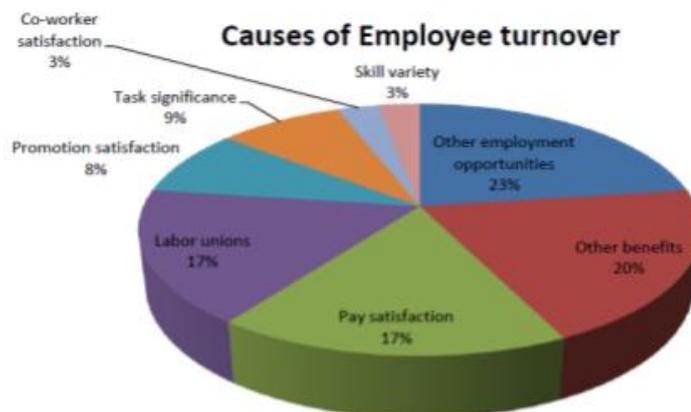


Fig.3.1. Employee turnovers and its causes

7.2 Proposed Work:

The proposed system has the following steps:

1. Data set:
2. Data pre-processing:

3. Test dataset and training dataset:

4. Data classification techniques:

1. Data set:

Data set is a collection of data. Most commonly a data set corresponds to the contents of a single database, where every column of the table represents a particular variable, and each row corresponds to a member of the dataset. For our project we take employee data from IBM which contains 1470 records and 35 fields including categorical and numeric features. Each record in the employee data set represents single employee information and each field in the record represents a feature of that particular employee.

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	Gender	HourlyRate	JobInvolvement
41	Yes	Travel_Rarely	1102	Sales		1	2 Life Sciences	1	1		2 Female	94	
49	No	Travel_Frequently	279	Research & Development		8	1 Life Sciences	1	2		3 Male	61	
37	Yes	Travel_Rarely	1373	Research & Development		2	2 Other	1	4		4 Male	92	
33	No	Travel_Frequently	1392	Research & Development		3	4 Life Sciences	1	5		4 Female	56	
27	No	Travel_Rarely	591	Research & Development		2	1 Medical	1	7		1 Male	40	
32	No	Travel_Frequently	1005	Research & Development		2	2 Life Sciences	1	8		4 Male	79	
59	No	Travel_Rarely	1324	Research & Development		3	3 Medical	1	10		3 Female	81	
30	No	Travel_Rarely	1358	Research & Development		24	1 Life Sciences	1	11		4 Male	67	
38	No	Travel_Frequently	216	Research & Development		23	3 Life Sciences	1	12		4 Male	44	
36	No	Travel_Rarely	1299	Research & Development		27	3 Medical	1	13		3 Male	94	
35	No	Travel_Rarely	809	Research & Development		16	3 Medical	1	14		1 Male	84	
29	No	Travel_Rarely	153	Research & Development		15	2 Life Sciences	1	15		4 Female	49	
31	No	Travel_Rarely	670	Research & Development		26	1 Life Sciences	1	16		1 Male	31	
34	No	Travel_Rarely	1346	Research & Development		19	2 Medical	1	18		2 Male	93	
28	Yes	Travel_Rarely	103	Research & Development		24	3 Life Sciences	1	19		3 Male	50	
29	No	Travel_Rarely	1389	Research & Development		21	4 Life Sciences	1	20		2 Female	51	
32	No	Travel_Rarely	334	Research & Development		5	2 Life Sciences	1	21		1 Male	80	
22	No	Non-Travel	1123	Research & Development		16	2 Medical	1	22		4 Male	96	
53	No	Travel_Rarely	1219	Sales		2	4 Life Sciences	1	23		1 Female	78	
38	No	Travel_Rarely	371	Research & Development		2	3 Life Sciences	1	24		4 Male	45	
24	No	Non-Travel	673	Research & Development		11	2 Other	1	26		1 Female	96	
36	Yes	Travel_Rarely	1218	Sales		9	4 Life Sciences	1	27		3 Male	82	
34	No	Travel_Rarely	419	Research & Development		7	4 Life Sciences	1	28		1 Female	53	
21	No	Travel_Rarely	391	Research & Development		15	2 Life Sciences	1	30		3 Male	96	
34	Yes	Travel_Rarely	699	Research & Development		6	1 Medical	1	31		2 Male	83	

....Includes more 1,436 rows

Table 1 Data Set used for predictions

2. Data pre-processing:

From the IBM employee dataset we implement a feature selection method to select the most important features of the dataset and divide total dataset into two sub datasets. One is test dataset another one is training dataset. That is if suppose any feature value in the record contain any null value or undefined or irrelevant value then separate that entire record from the original dataset and place that record into training dataset, else if the record contain perfect data with all features then place that into test dataset. Test dataset contain all important features to predict employee attrition or employee attrition and training dataset contain irrelevant data.

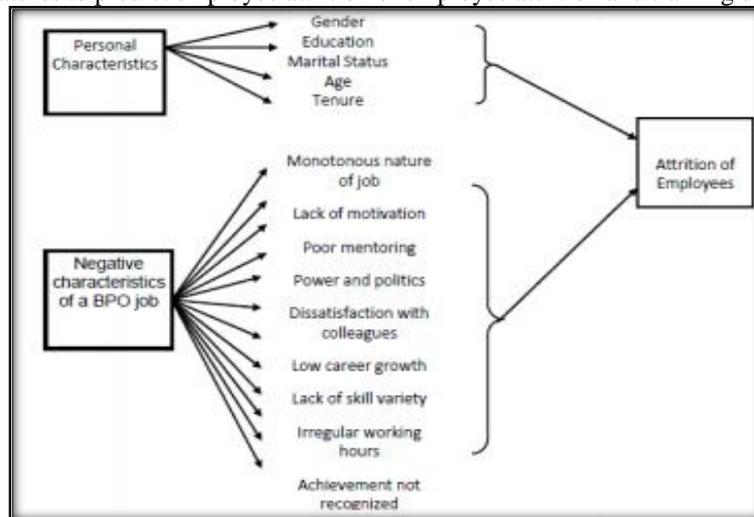


Fig.3.1.1. Possible characteristics for Employee Attrition

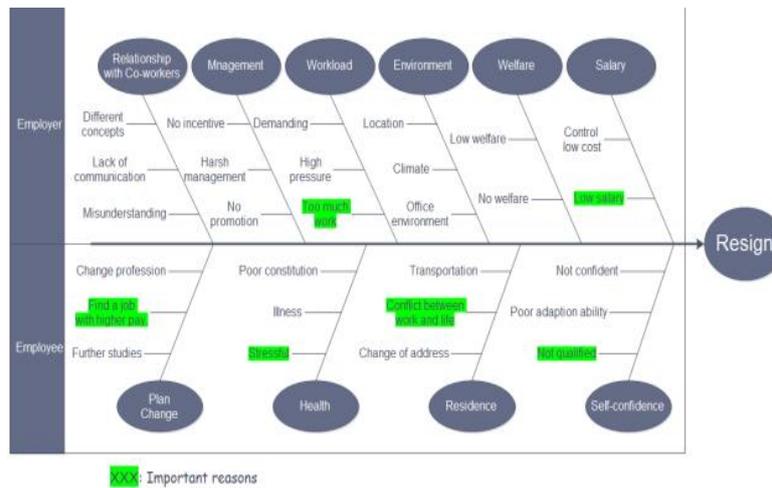


Fig.3.1.2. Reasons of resign using Fishbone Diagram

Flowchart for System Architecture

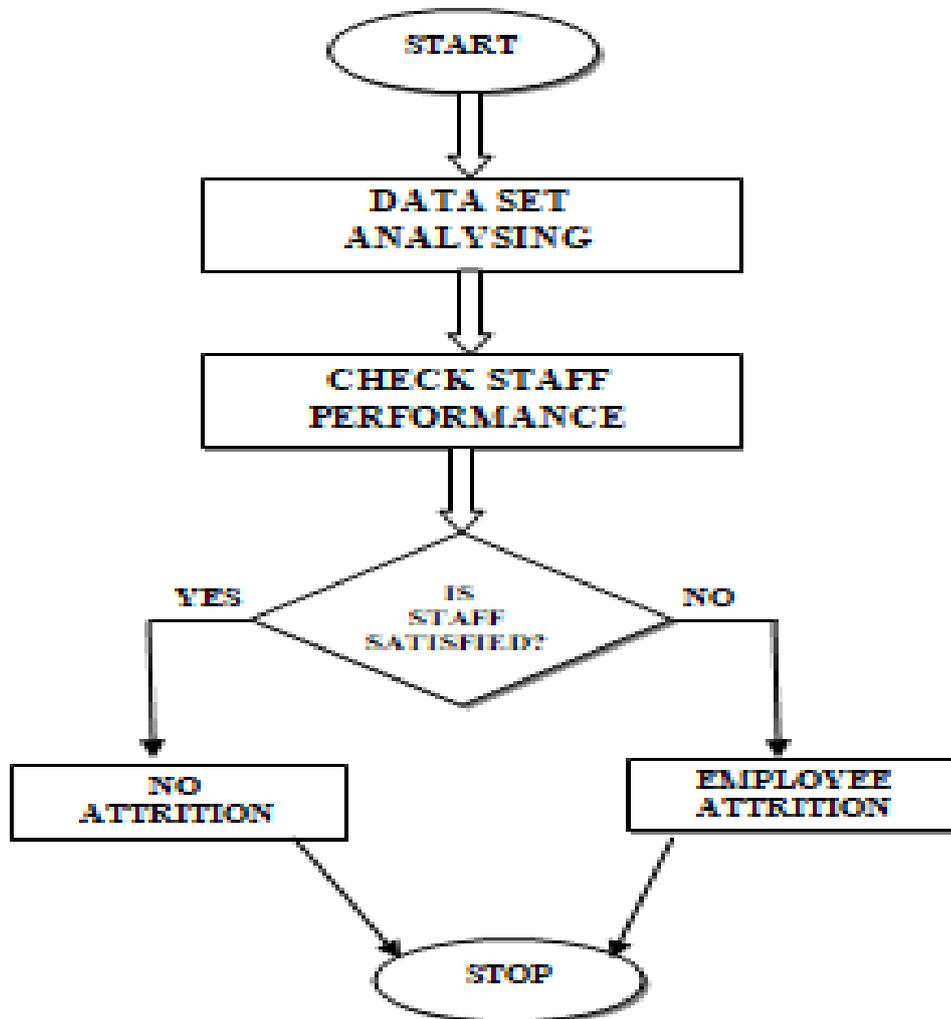


Fig.3.2 Employee Attrition System

3. Test dataset and training dataset:

Separating data into test datasets and training datasets is an important part of evaluating data mining models. By this separation of total data set into two data sets we can minimize the effects of data inconsistency and better understand the characteristics of the model. The test data set contains all the required data for data prediction and training data set contains all irrelevant data. Here we have 788 records in test dataset and 682 records in training dataset. We apply data classification and data prediction on the test dataset of 788 records.

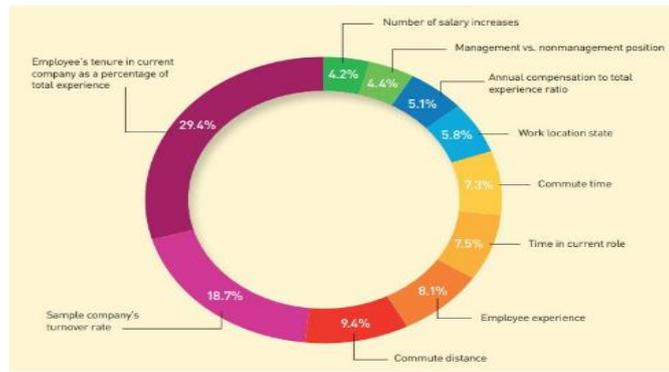


Fig.3.3. Revelations from Work-force turnover study

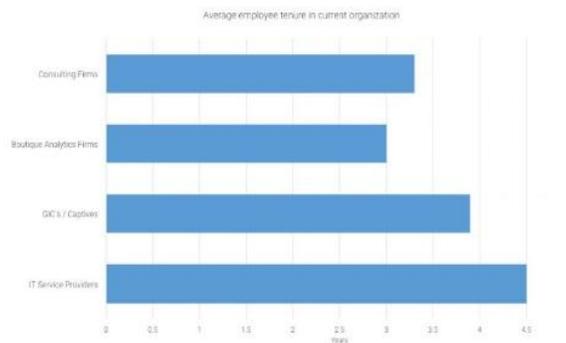


Fig.3.4. Attrition performance evolution

4. Data classification techniques:

Data classification is the process of organizing data into categories for its most effective and efficient use. Data classification techniques are Decision tree

a) **Decision Tree:** It is tree structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branched notes the outcome of a test, and each leaf node holds a class label.

A Decision Tree

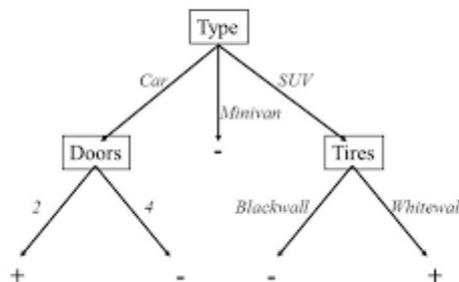


Fig 3.5. Working of Decision Tree

b) **Support Vector Machine:** In machine learning, support-vector machines (SVMs also support-vector networks) are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier

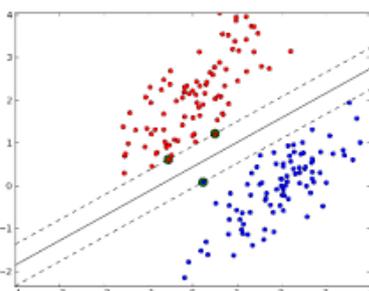


Fig 3.6. Working of Support Vector Machine

c) Adaptive Boosting:

The Adaptive Boosting (Ada Boost) algorithm works on the core principle of fitting a sequence of weak learners [8]. This algorithm uses a particular boost classifier as shown in Eq. 1. $f_t(x) = \sum_{T=1}^t (x)$ Eq. (1) f_t represents a weak learner which takes an object x input and accordingly returns a value which indicates the class of that object. $E_t = \sum [F_{t-1}(x_i) + \alpha h(x_i)]$ Eq. (2)

The sum training error E_t is given as Eq. 2 and it is minimized as each weak learner produces $h(x_i)$ the output hypothesis. For each iteration t , the α coefficient is assigned. F_{t-1} is the boosted classifier. Eq. 1 and 2 are used for training. The weights w_1, w_2, \dots, w_N are applied to each of the training samples, this is known as boosting iteration. Initially all the weights are assigned by Eq. 3. $w_i = 1 / N$ Eq. (3)

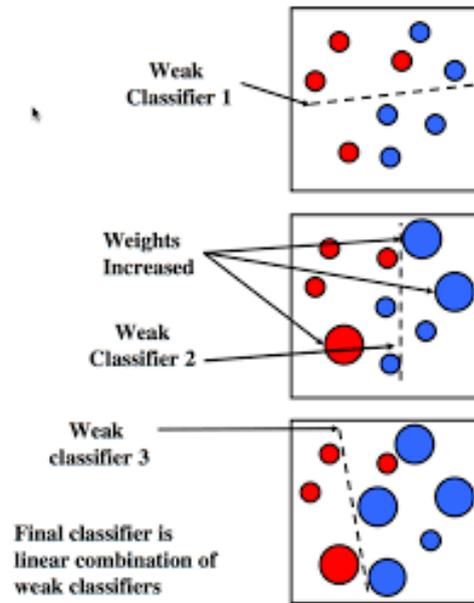


Fig 3.7. Working of Adaptive Boosting

d) Random Forest: Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where you join different types of algorithms or same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithm of the same type i.e. multiple decision trees, resulting in a forest of trees, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks.

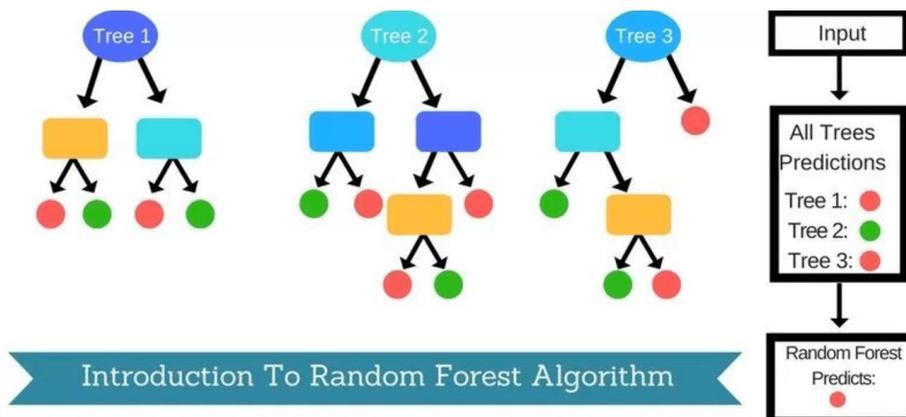


Fig 3.8. Working of Random Forest

VIII. CONCLUSION and RESULT:

After performing various empirical tests on the IBM HR Analytics Employee Attrition & Performance dataset from Kaggle with 35 features for employee attrition prediction in an organization, the results from the table in the above section depict that the Stacked Classifier provides the most optimal solution with a score accuracy of >90%. Hence, it can be stated that multiple trained models learned on different classification algorithms can be cross validated, averaged and passed into a Meta Classifier to improve the accuracies of the individual classification models.

An employee's attrition value prediction could benefit the organization by knowing where they are going wrong and with which employees. The organization could accordingly start looking for people to fill the positions which are predicted to be empty by using this stacked classifier model.

Test case id	Test cases	Preconditions	Input test data	Steps to be executed	Expected result	Actual result	Pass/Fail
1	Using Employee id	Performs Predictive Analysis	Reading data from the input .csv file	Choosing any id from the given list	If employee id exists shows attrition rate	Checks whether attrition occurs or not	Pass

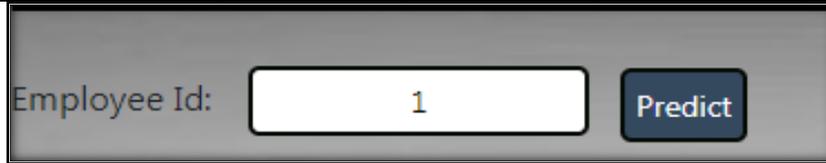


Fig.4.1. Using Employee ID

2	Confidence Level	Performs Predictive Analysis denotes confidence of the algorithm	Based on all the features shows how accurate the prediction is	Choosing any id from the given list and examining the features	To have maximum confidence level for each employee record	Estimating with maximum confidence	Pass
---	------------------	--	--	--	---	------------------------------------	------

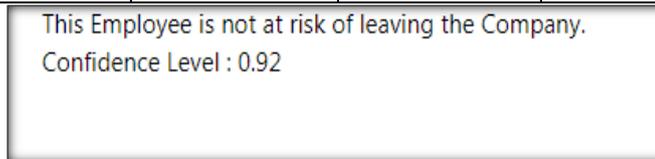


Fig.4.2. Determining the Confidence Level

3	Descriptive Analysis	Determines the visual representation of the dataset for all features	The dataset provided containing features of the employees	Examini-ng the changes in the features	Graphic represent-ation of the data	Graphic represent-ation of the data involving different features	Pass
---	----------------------	--	---	--	-------------------------------------	--	------

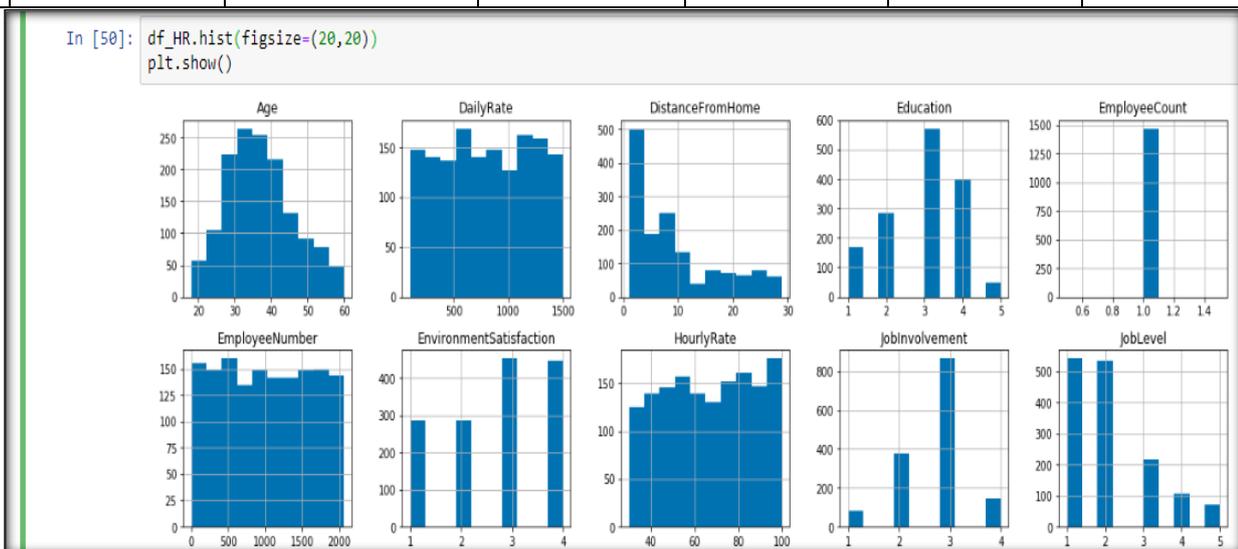


Fig.4.3. Graphical Representation of the data

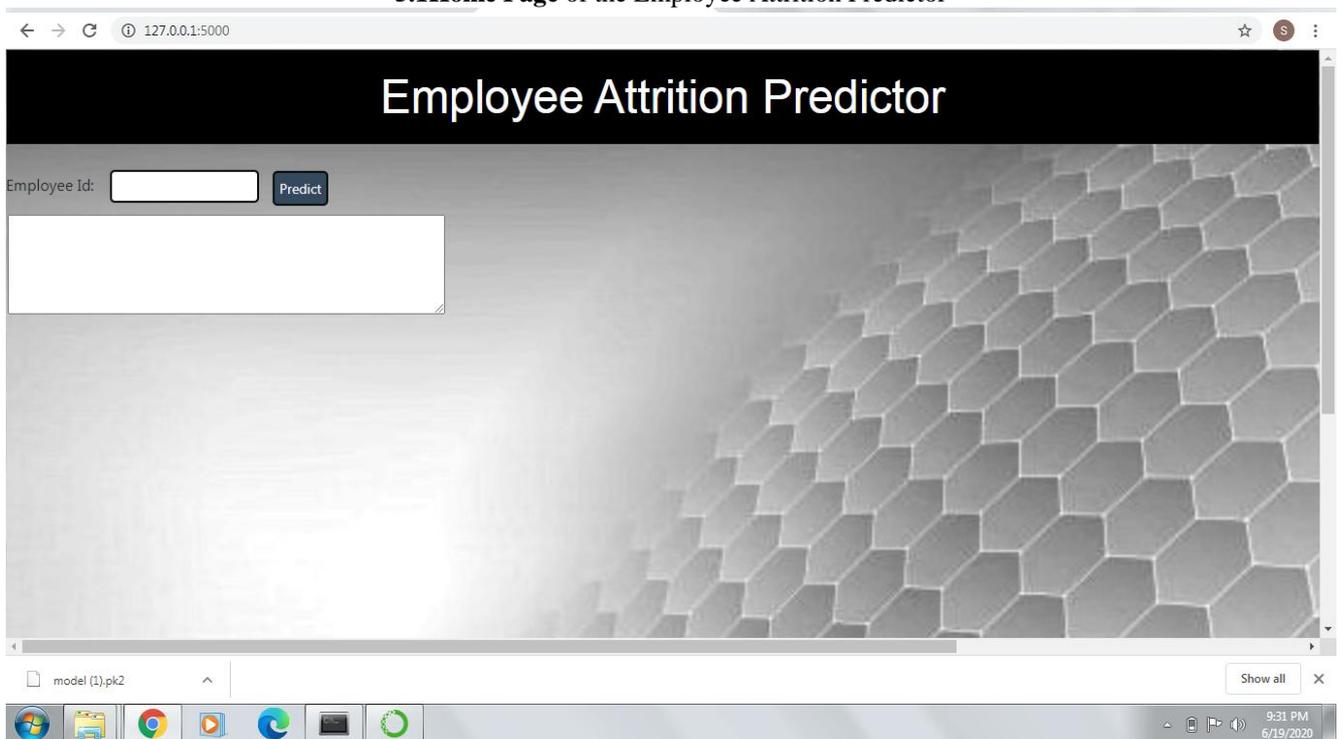
4	Predictive Analysis	Predicts if the employee leaves the organisation or not using various algorithms	Uses Features of the employees from the dataset	Applying algorithms on the employee features to predict turnover	Accurate-ly predict the employee attrition rate	Prediction of the employee attrition	Pass
---	---------------------	--	---	--	---	--------------------------------------	------



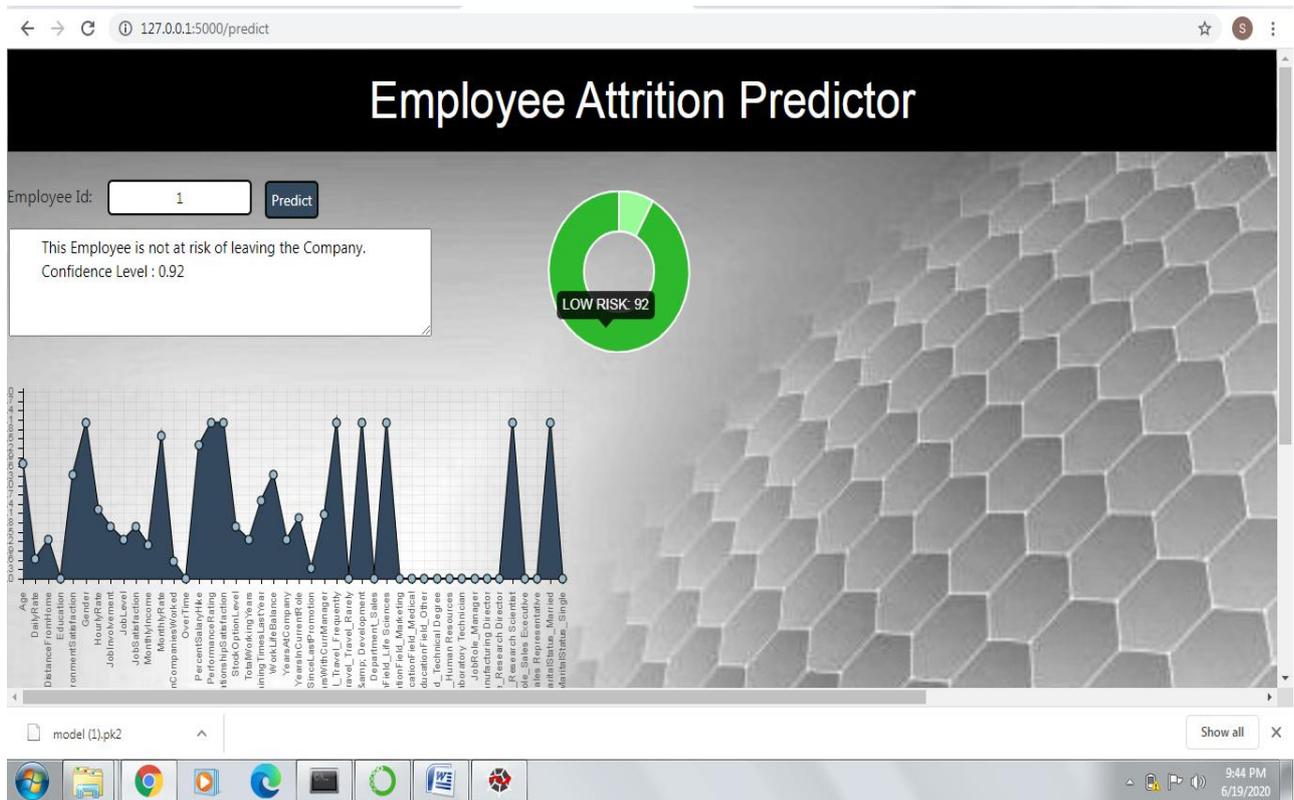
Fig.4.4. Prediction of Attrition

SCREENSHOTS:

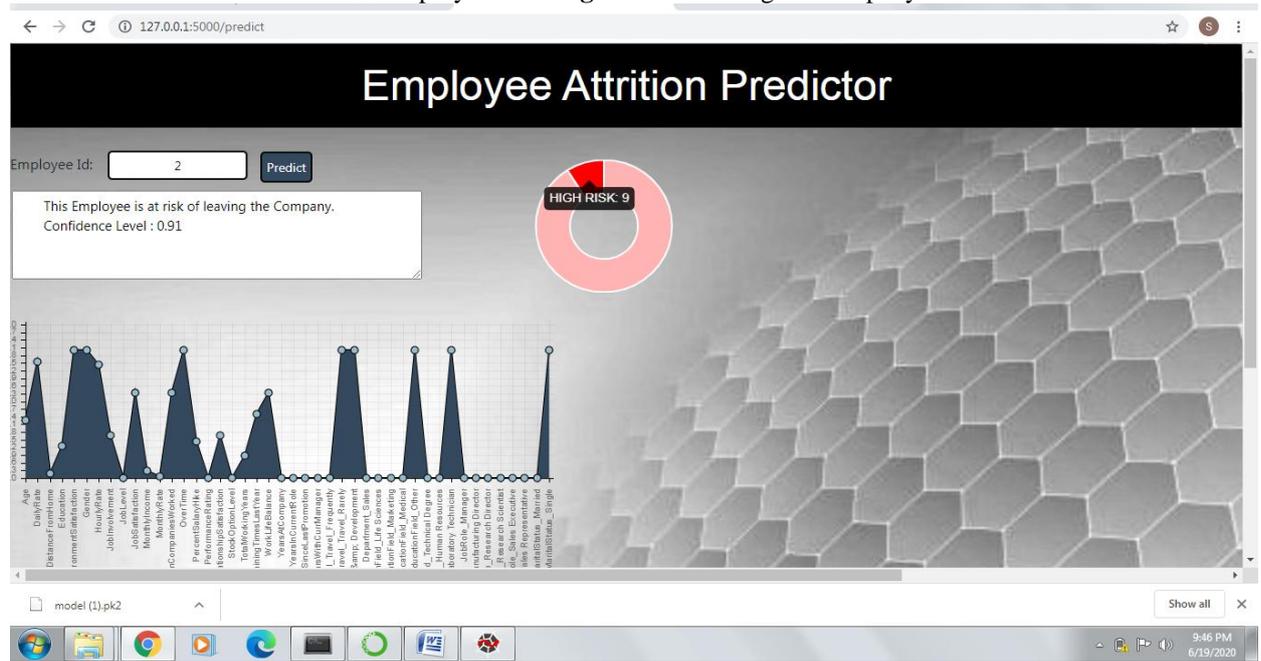
5.1 Home Page of the Employee Attrition Predictor



5.2 Employee with **Low risk** of leaving the company



5.3 Employee with High risk of leaving the company



REFERENCES

1. IBM HR Analytics Employee Attrition & Performance (2018, 28, August) Kaggle Inc. [Online] Available: <https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset/>
2. 7 Types of Classification Algorithms (2018, 6, September) Analytics India Magazine [Online] Available: analyticsindiamag.com/7-types-classification-algorithms/
3. Stacking Classifier (2018, 10, September) Github [Online] Available: https://rasbt.github.io/mlxtend/user_guide/classifier/StackingClassifier/
4. ENSEMBLE METHODS - ADAPTIVE BOOSTING (2018, 14, SEPTEMBER) SCIKIT LEARN [ONLINE] AVAILABLE: [HTTP://SCIKIT-LEARN.ORG/STABLE/MODULES/ENSEMBLE.HTML#ADABOOST](http://scikit-learn.org/stable/modules/ensemble.html#adaboost)
5. DECISION TREES (2018, 15, SEPTEMBER) SCIKIT LEARN [ONLINE] AVAILABLE: [HTTP://SCIKIT-LEARN.ORG/STABLE/MODULES/TREE.HTML#TREE](http://scikit-learn.org/stable/modules/tree.html#tree) [6] R. ANBU RANJITH KUMAR AND DR. V. ANTONY JOE RAJA
6. UNDERSTANDING SUPPORT VECTOR MACHINES (2018, 16, SEPTEMBER) ANALYTICS VIDHYA [ONLINE] AVAILABLE: [HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/09/UNDERSTANDING-SUPPORT-VECTOR-MACHINE-EXAMPLE-CODE/](https://www.analyticsvidhya.com/blog/2017/09/understanding-support-vector-machine-example-code/)
7. http://iaeme.com/MasterAdmin/UploadFolder/IJMHRM_07_01_001/IJMHRM__07_011.pdf
8. <https://www.ijedr.org/papers/IJEDR1804054.pdf>
9. <https://www.educba.com/data-science-and-its-growing-importance>
10. <https://towardsdatascience.com/people-analytics-with-attrition-predictions-12adcce9573f>