1. INTRODUCTION

1.1 Overview

Today attrition is one of the major problems faced by industry across the world. It is the most burning issue for the industry, and high attrition rates lead to many issues in the boundary of the organization like losing the talents and knowledge, cost related to training and administration, and recruitment. It is observed that many attributes lead to the attrition of an employee. Which includes working environment, job satisfaction, employer's behavior, job timing, and most important is salary or incentives. Also, the prediction model plays an essential role in finding the behavior of employees. Timely delivery of any service or product is the primary goal of any organization in recent days due to high competition in industries. If a talented employee leaves unexpectedly, the company is not able to complete the task at defined times. It may become the reason for the loss of that company. Therefore, companies are interested in knowing the employee's attrition. They can make a proper substitute or arrangements earlier. There may be various reasons for employee attrition, which include less salary, job satisfaction, personal reasons, or environmental issues if the employer terminates an employee for any reason. It is known as involuntary attrition (Kaur & Vijay, 2016). On the other hand, voluntary attrition is known as the left of an employee by their side. This kind of attrition is a loss for the company if he or she is a talented employee. In the present scenario, everyone wants a higher salary and job security. Therefore, employees leave jobs immediately if they got a better chance in other places. In the recent era of computer science, machine learning approaches play an important role in employee attrition prediction. These approaches provide predictions based on historical information of the employee, such as age, experience, education, last promotion, and so on. Based on the prediction results HR

department have prior knowledge about employee attrition. The HR department also has preplanned recruiting employees as a substitute for the employee who is interested in leaving in the coming days. Various researches have also studied the performance of different machine learning approaches (Ajit, 2016; Sikaroudi et al., 2015). Kaur et al. (Kaur & Vijay, 2016) have discussed various reasons or factors that are involved in employee attrition. They have also investigated that talented employee replacement is a time-consuming and challenging task. It is also a significant factor in loss in business. Compensation is one solution to decreasing the attrition rate. Moncaarz et al. (Moncarz et al., 2009) have discussed how attrition can be decreased by providing better compensation. Punnoose and Ajit (Ajit, 2016) have provided a comparative analysis of various machine learning approaches for employee turnover. Tree-based approaches are also used to predict employee attrition (Alao & Adeyemo, 2013). Jantan et al. (Jantan et al., 2010) have compared tree-based methods with other traditional machine learning approaches. Radaideh and Nagi (Al-Radaideh & Al Nagi, 2012) uses the decision tree for employee attrition prediction. In their work, they have found that job title is an essential feature of attrition, whereas age is not a very important feature. Saradhi (Saradhi & Palshikar, 2011) uses various machine learning approaches for employee attrition prediction. They have taken a database of 1575 records with 25 features of employee and applied various classification approaches to predict attrition. They have shown that SVM has higher accuracy, which is 84.12%. Due to confidentiality and noisy HR data, sometimes prediction has higher accuracy. It is difficult to generalized predictions for different organizations and employee roles (Zhao et al., 2018). Previous studies presented accuracy as a primary evaluation standard for attrition prediction. Various machine learning approaches are used and evaluated in different datasets. It is challenging to conclude that which model is best for attrition prediction. The rate of employee attrition is always less than the employee who stays in the organization. Therefore, datasets are

always imbalanced. Accuracy measures are not reliable for imbalanced datasets (Sexton et al., 2005; Sikaroudi et al., 2015; Tzeng et al., 2004). So that it is desired to have an accurate model to enhance the prediction accuracy of the models. Which provides better results to employers. Based on the accurate prediction results employers and HR department know the behavior of their employee

1.2 Purpose

ML depends heavily on data, without data, it is impossible for an “AI” to learn. It is the most crucial aspect that makes algorithm training possible. In Machine Learning projects, we need a training data set. It is the actual data set used to train the model for performing various actions.

There are many features which are responsible for Employee’s Attrition, e.g. poor working conditions, retirement or resignation, unfair pay, lack of professional growth, or other personal reasons.etc. For better prediction of the Employee’s Attrition, we should consider as many relevant features as possible.

2. LITERATURE SURVEY

2.1 Existing problem

Among all employee related problems, employee attrition is one of the key problem in the today's scenario despite the changes in the external environment. Attrition is said to be gradual reduction in number of employees through resignation, death and retirement. The other name given for Attrition is attrition. When a well-trained and well-adapted employee leaves the organization for any of the reason, it creates an

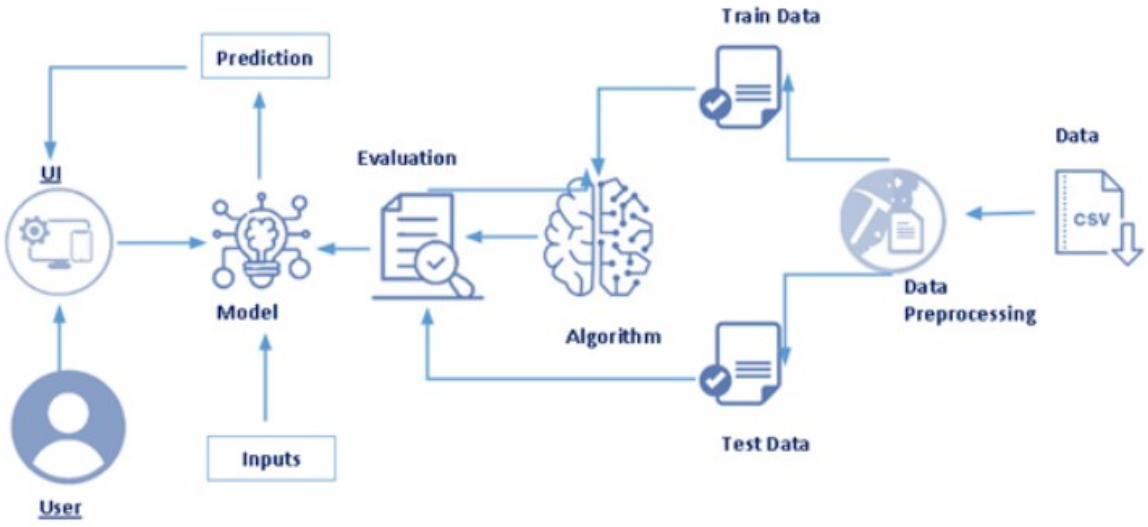
empty space in an organization (i.e) there occurs a vacuum in the organization. It creates a great difficulty for a Human resource personnel to fill the gap that has occurred. Modern Human resource managers is taking various steps to reduce the employee attrition rate and it has been a pivotal challenge for today's Managers. Many of the employees may also tend to leave the job for various undisclosed factors such as lack of job security, lack of career advancement, desire for change in new opportunities, anticipating higher pay, problems with supervisors and few other personal reasons. This study helps in knowing why attrition occurs, reasons for employee attrition, challenges faced by managers in retaining employees and also suggest some measures in retaining employees.

2.2 Proposed solution

The intension is to build a model that predicts the Attrition of the Employees based on the given factors of an employee using Machine Learning.

3. THEORITICALANALYSIS

3.1 Block diagram



.

3.2 Hardware / Software designing

a. Install Anaconda Software

To install Anaconda on your local system, go through the below links according to your system requirements. After Anaconda is installed, run the .exe folder.

b. Run Jupyter

Search Anaconda Navigator and open a Jupyter notebook.

c. Numpy

Using Anaconda Navigator: conda install numpy iv)

Pandas

Using Anaconda Navigator: conda install pandas v)

Matplotlib

Using Anaconda Navigator: conda install matplotlib

OR

Using command prompt: pip install matplotlib vi)

Scikit-Learn

Using Anaconda Navigator: conda install -c conda-forge scikit-learn

OR

Using command prompt: pip install -U scikit-learn

4. EXPERIMENTAL INVESTIGATIONS

i) Dataset Collection

Collect the dataset or create the dataset.

ML depends heavily on data, without data, it is impossible for an “AI” to learn. It is the most crucial aspect that makes algorithm training possible. In Machine Learning projects, we need a training data set. It is the actual data set used to train the model for performing various actions.

There are many features which are responsible for Employee’s Attrition, e.g. poor working conditions, retirement or resignation, unfair pay, lack of professional growth, or other personal reasons.etc. For better prediction of the Employee’s Attrition, we should consider as many relevant features as possible.

We can collect dataset from different open sources like kaggle.com, data.gov, UCI machine learning repository etc.

The kaggle repository link is :

https://www.kaggle.com/ashukr/exploring-

co2emission?select=Indicators.csv

ii) Data Preprocessing :

Here, we are reading the dataset(.csv) from the system using pandas and storing it in a variable ‘df’. It's time to begin building your text classifier! The data has been loaded into a DataFrame called df. The .head() method is particularly informative.

We might have your data in .csv files, .excel files or .tsv files or something else. But the goal is the same in all cases. If you want to analyse that data using pandas, the first step will be to read it into a data structure that’s compatible with pandas.

load a .csv data file into pandas. There is a function for it, called read_csv().We will need to locate the directory of the CSV file at first (it’s more efficient to keep the dataset in the same directory as your program).

iii) Finding The Missing Values

Sometimes you may find some data are missing in the dataset. We need to be equipped to handle the problem when we come across them.

One of the most common ideas to handle the problem is to take a mean of all the values of the same column and have it to replace the missing data.

We will be using isnull().any() method to see which column has missing values

iv) Handling Categorical Variables

The department column of the dataset has many categories and we need to

reduce the categories for better modeling.

a. Data Exploration

Let us find out the number of employees who left the company and those who didn't!

b. Data Visualization :

Data visualization is where a given data set is presented in a graphical format. It helps the detection of patterns, trends and correlations that might go undetected in text-based data.Understanding your data and the relationship present within it is just as important as any algorithm used to train your machine learning model. In fact, even the most sophisticated machine learning models will perform poorly on data that wasn't visualized and understood properly.

To visualize the dataset we need libraries called Matplotlib and Seaborn.The Matplotlib library is a Python 2D plotting library which allows you to generate plots, scatter plots, histograms, bar charts etc.

Let's visualize our data using Matplotlib and the Seaborn library. Let us visualize our data to get a much clearer picture of the data and the significant features

5. FLOWCHART

Train And Test The Model Using Random Forest Regressor

There are several Machine learning algorithms to be used depending on the data you are going to process such as images, sound, text, and numerical values. The algorithms that you can choose according to the objective that you might have may be Classification algorithms are Regression algorithms.

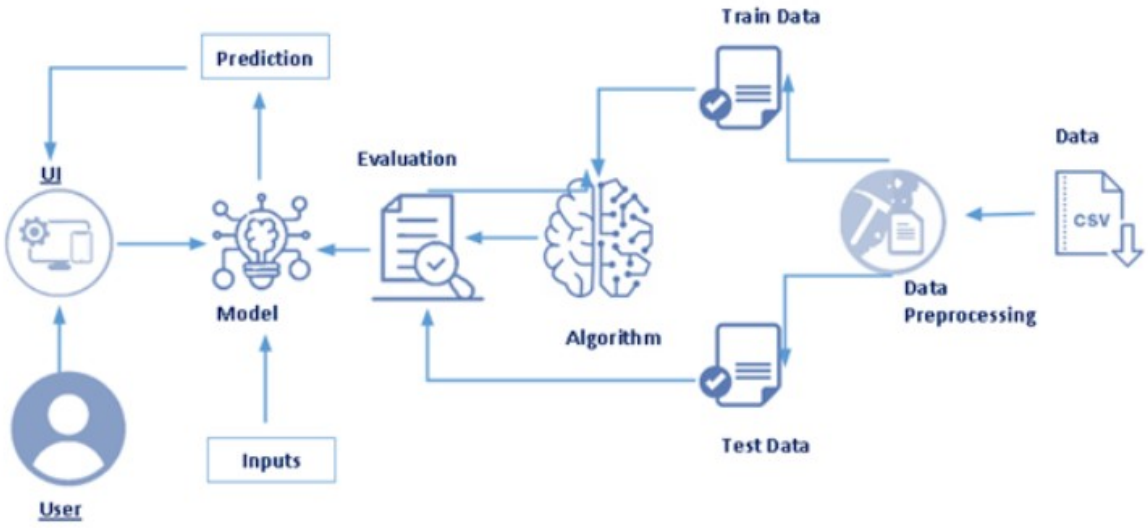
Example: 1. Linear Regression.

1. Logistic Regression.
2. Random Forest Regression / Classification.
3. Decision Tree Regression / Classification.

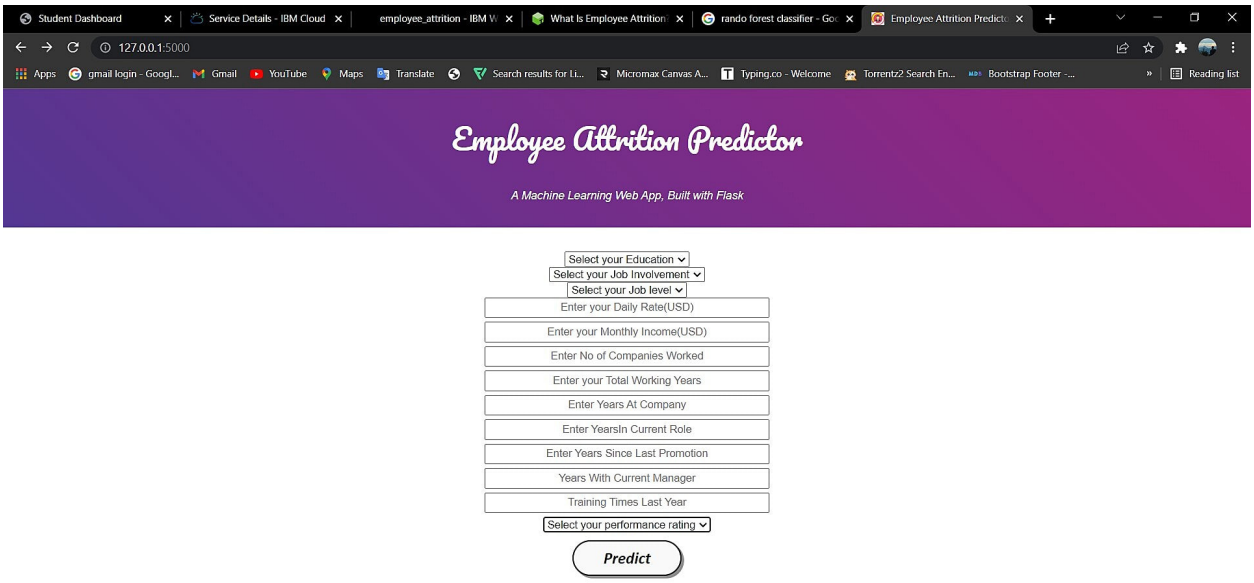
You will need to train the datasets to run smoothly and see an incremental improvement in the prediction rate.

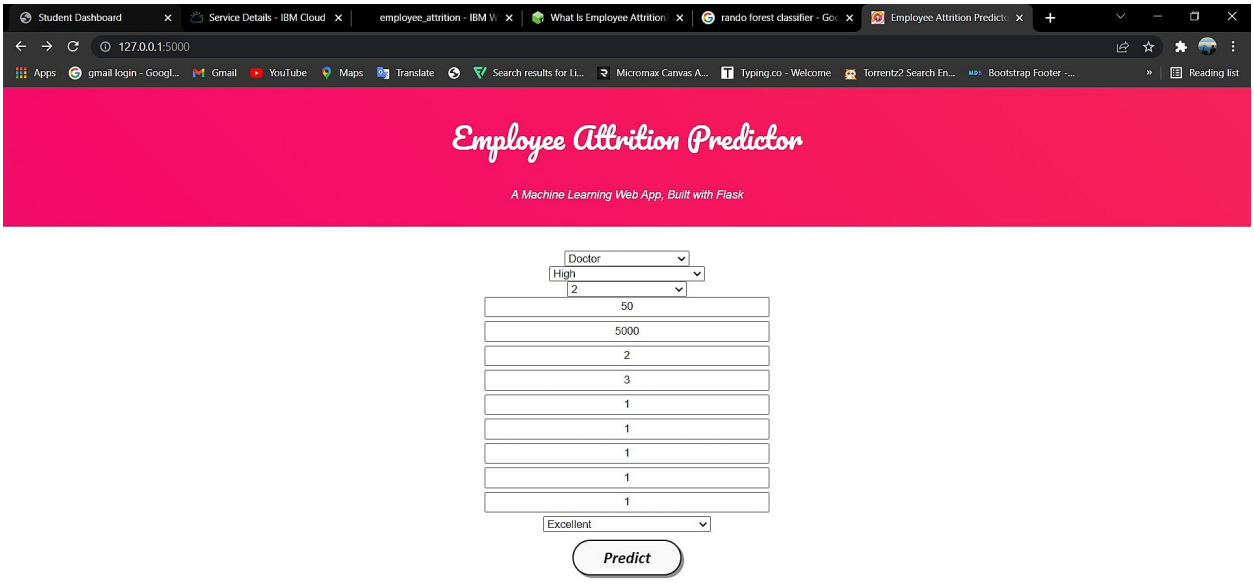
Now we apply the Random forest regressor algorithm on our dataset.

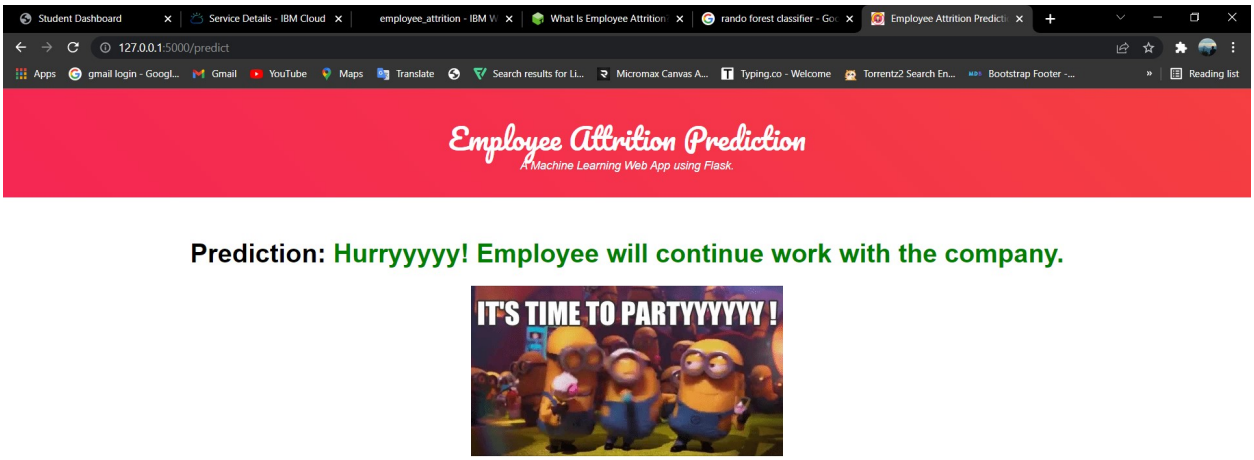
A Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as bagging. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.



6. RESULT


Student Dashboard
Service Details - IBM Cloud
employee_attrition - IBM W
What Is Employee Attrition
rando forest classifier - Goo
Employee Attrition Predicto
127.0.0.1:5000
Apps
gmail login - Googl...
Gmail
YouTube
Maps
Translate
Search results for Li...
Micromax Canvas A...
Typing.co - Welcome
Torrentz2 Search En...
Bootstrap Footer -...
Reading list
Employee Attrition Predictor
A Machine Learning Web App, Built with Flask
Select your Education
Select your Job Involvement
Select your Job level
Enter your Daily Rate(USD)
Enter your Monthly Income(USD)
Enter No of Companies Worked
Enter your Total Working Years
Enter Years At Company
Enter YearsIn Current Role
Enter Years Since Last Promotion
Years With Current Manager
Training Times Last Year
Select your performance rating
Predict


Student Dashboard
Service Details - IBM Cloud
employee_attrition - IBM W
What Is Employee Attrition
rando forest classifier - Goo
Employee Attrition Predicto
127.0.0.1:5000
Apps
gmail login - Googl...
Gmail
YouTube
Maps
Translate
Search results for Li...
Micromax Canvas A...
Typing.co - Welcome
Torrentz2 Search En...
Bootstrap Footer -...
Reading list
Employee Attrition Predictor
A Machine Learning Web App, Built with Flask
Doctor
High
2
50
5000
2
3
1
1
1
1
1
Excellent
Predict



---

7. ADVANTAGES & DISADVANTAGES

Minimize cost of new talent acquisition based on the employee profiling and company requirements. Analysis and assessment of the loss in expertise and skillsets. Measurement of financial and productivity loss due to attrition. Able to plan and minimize the loss.

Some attrition is predictable even without analyzing survey data. If 20% of the managers in an organization will reach retirement age in five years, the organization can start identifying employees who are good candidates for management and get them into the training pipeline; incentives to keep older employees on board after they reach retirement age might also be considered.

However, the analysis of engagement and exit survey data together reveals the less obvious red flags for attrition. Analysis of demographic data alone can highlight attrition hotspots linked to specific job types, work locations, and tenure levels; analysis of the survey data will reveal the reasons why turnover is higher in those hotspots. We can look at the areas of the experience that were failing employees who left according to engagement survey responses, and the reasons employees gave for leaving on exit surveys, and load all that information into the system to inform the model

8. APPLICATIONS

1. Various machine learning models have been used in employee attrition such as decision trees, random forests, naïve Bayes, logistic regression, and SVM. In this work, a deep learning prediction model is used to classify employee attrition.
2. banking and Financial services industry uses Machine Learning to detect and reduce fraud, measure market riska dn identify oppurtunities.
3. Machine Learning play a key part in security as they typically use predictive analysis to improve services and performance, but also to detect anamolies, fraud, understand consumer behavior and enhance data security.
4. ML Algorithms are also known for it's recommendation algorithms like in Facebook or YouTube where similar content will be suggested to engage the user

9. CONCLUSION

Predictive Attrition Model helps in not only taking preventive measures but also into making better hiring decisions. Deriving trends in the candidate’s performance out of past data is important in order to predict the future trends, as well as to board new employees. Moreover, HR can use the employee data to predict attrition, the possible reasons behind it and can take appropriate measures to prevent it.

We live in a data-driven world – from something as trivial as weather updates to complexity behind GPS navigations, data is constantly being generated every second in every field, which leaves us to decide how to turn it around for our advantage in our respective domains.

10. FUTURE SCOPE

In future it is possible to improve the analysis by considering new employees' opportunities as well as adverse working conditions (e.g., harm and hazard) and poor promotion prospects, discrimination and low social support, that are positively related to employees turnover intention

11. BIBILOGRAPHY

1. **https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.RandomFore stRegressor.html** 2. **https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.h tml**