International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

FORECASTING EMPLOYEE TURNOVER IN THE IT INDUSTRY USING MACHINE LEARNING

[1]Mrs.N.V.N. SOWJANYA,[2]T. SAHITHI, [3]GOVINDULA AJAY,

[4]P. SATYA KARTHIKEYA

[1](Assistant Professor) , CSE. Teegala Krishna Reddy Engineering College Hyderabad.
[2,3,4]B,tech , scholar , CSE. Teegala Krishna Reddy Engineering College Hyderabad.

ABSTRACT

Now a days, Employee turnover becoming a biggest reason for the IT companies loss . The companies are in the lowest position in the market because of shortened number of employees . The employee turnover is causing this problem of having deficient number of employees in the organization which leads to incomplete work . These all affects the performance of the company or organization in getting loss. With the consideration of the problems , the company staff members have to identify the problems facing by the employees in their company and also find the reasons of not letting employees to leave the company. This will useful for the company that they can understand the employee intension and the problems they are facing . By knowing the employee intension to leave the company , The company will help the employee problems they are facing . Then this will change the employee’s decision and motivating them to do work for their company . So, the turnover will be reduced.

This research paper uses machine learning algorithms, like Naïve bayes, Random forest and KNN to forecast an employee’s intention to leave the organization in the near future and identifies the significant features impacting the employee’s intention to leave the organization. The highest accuracy is gained by the Naïve Bayes classifier. The aim of this research paper is to develop a model to forecast employee attrition and provide the organizations opportunities to address any issues and improve retention

1. INTRODUCTION

Employee turnover is one of the prime concerns of businesses of all size. Employee turnover, or employee turnover rate, calculates the total number of employees who leave an organization during a given period. If a company is experiencing a high employee turnover, they may experience a range of negative impacts. It's crucial for companies to track their employee turnover, uncover the reasons why employees leave and enact employee retention strategies. Information Technology in India is a vast industry comprises technology services, consulting, and outsourcing. At present days, IT sector has a huge growth by covering all technologies and computing . As the IT sector is increasing, the competition between the software companies are also increasing . IT industry and IT sector has a highest employee turnover rate . It is very critical for the company management team for loosing the employees because they invested lot of money for them in training and on boarding resources by directly or indirectly . This Research paper gives the brief idea of why employee turnover is increasing and what are the reasons behind this . It focuses mainly on the reasons which affecting the employees from the company to leave . The information technology (IT) sector in India has grown to a great extent to cover several aspects of technology and computing.

This Research paper helps to forecast the intension of the employee who are leaving so that the company or organization can make some decisions which satisfy the employee to do work in their organization .

This will reduce the employee turnover in many companies . By this , so many IT companies gets benefited. The Supervised machine learning algorithms are used and forecasted the employee turnover in the major cities like Bengaluru, Chennai, Mumbai . The classification algorithms are used to find out the factors that affects the employees. The distinguishing features of this research are that data is collected from employees directly to forecast the turnover intention during this pandemic time rather than relying on available data sets and also explores various contributing factors that may impact the turnover intent. The IT industry has also played a prime role in the Indian economy by promoting exports, improving standards of living and generating revenues.

In recent years, there has been a massive increase in the competition among companies in sustaining in the business. Information Technology industry is long plagued by high turnover rate and hence it is very critical for the management to reduce the turnover intention of the employees considering the fact huge amount of direct and indirect expenses spent on training and onboarding resources. According to the 2020 Talent Technology Outlook study conducted by SCIKEY which was in turn published in economic times in December 2019[12], the attrition rate has increased beyond 22%. According to financial express report in February 2020[13], the IT and BPM industry has contributed to about 7.9% to Indian GDP in 2019. This Research paper focuses on identifying the factors within an organization which impacts the employee switchover intention to a significant extent and hence by identifying these factors the organizations can improve satisfaction levels of the employee by providing them with improved working conditions which would reduce their intention to go for a job hunt. So, by considering these factors the employee turnover intention in major cities like Chennai, Coimbatore and Bangalore is forecasted by using supervised machine learning algorithm and bringing insights and decisions about the reduction of turnover thoughts of employees. The classification algorithms have been used to find out the insights from the respondents by using various tools.

The distinguishing features of this research are that data is collected from employees directly to forecast the turnover intention during this pandemic time rather than relying on available data sets and also explores various contributing factors that may impact the turnover intent.

Employee turnover has drawn researchers' and human resource managers' attention because organizations lack niche skill sets and resources, which require time and planning to acquire at crucial times. The hiring lead time is often long, particularly when special skills are involved. In some organizations, like U.S. national laboratories, the process can take months because of security-clearance requirements. Therefore, a good employee-turnover forecastion at the firm and departmental levels is essential for effective human resource planning (HRP),budgeting, and recruiting.

Human resource planning is an ongoing systematic planning process optimize the human resource pool. For organizations to efficiently and effectively execute tasks, the rightpeople must be available at the right places at the right time (Khoong, 1996).

Over the years, organizations have scaled up their efforts in manufacturing, marketing and financing. However, organizations have always struggled to develop sustainable HRP models (Henemanet al., 1993), whose objective is to match employees and jobs to avoid manpower shortages or surpluses (Camb´al et al. ˇ , 2011). To achieve this balance, employee turnover is often central to organizational workforce planning and strategy.

As summarized in Table 2.2, researchers developed employee turnover models using various statistical methods. Previous studies have identified employee-turnover explanatory forecastors. For instance, Bluedorn (1982) related turnover to the individual's routine, age, service length, and perception of environmental opportunities. Balfour and Neff(1993) suggested that caseworkers with more education, less experience, and less stake in an organization are more likely to turnover. Wright and Cropanzano (1998) associated emotional exhaustion with job performance and subsequent turnover, but not with job satisfaction. forecasting employee turnover based on employee absenteeism and performance, Morrow et al. (1999)'s study showed a positively correlated absenteeism and voluntary turnover as well as negatively correlated performance ratings and voluntary turnover. Thaden et al. (2010) indicated that organizational culture might potentially be an importantfactor for retaining workers.

Other insights have been gained from more recent research. For instance, according to Tews et al. (2014), personal events, professional events, internal work events and constituent attachment are highly related to turnover. Collini et al. (2015) found that the interaction between interpersonal respect, mission fulfillment, and engagement are statistically significant forecastors for turnover in health care. However, these researchers found that diversity climate is not related to turnover.

Finally, only Sexton et al. (2005)considered outside economic variables, unemployment index, and consumer price index in the employee-turnover forecasting model. However, their final model did not include these variables. Ferrara and van Dijk (2014) in the International Journal of Forecasting revealed a new interest in forecasting business cycles with some complex

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

methodologies. However, forecasting business-cycle turning points is quite difficult, and Hamilton (2011) suggested that ”the best econometricians can do is probably to now cast recessions; that is, to recognize a turning point as soon as it occurs, or soon thereafter.”

An outside variable might facilitate this situation. Wright and Cropanzano associated emotional exhaustion with job performance and subsequent turnover, but not with job satisfaction. forecasting employee turnover based on employee absenteeism and performance, Morrow et al. (1999)’s study showed a positively correlated absenteeism and voluntary turnover as well as negatively correlated performance ratings and voluntary turnover. Thaden et al indicated that organizational culture might potentially be an important factor for retaining workers.

2.LITERATURE SURVEY

Rohit Punnoose and Pankaj Ajit have used several machine learning algorithms to forecast the employee’s attrition in the orgaminzation. They used data from the Human resource information system of a global retailer and used to compare XGBoost against other six algorithms like Logistic Regression, Naïve Bayes, Random forest, KNN, SVM, LDA. The highest accuracy was obtained by XGBoost. Grid search technique was used for calculating the performance and for selecting the best value for the hyperparameter.

S K Monisaa Tharani and S N Vivek Raj have collected data from 416 employee’s using convenience sampling and structure questionnaire. They used several algorithm’s like logistic regression, KNN, Naïve Bayes, XGBoost, among which the highest accuracy is attained by XGBoost. Text mining and sentimental analysis are other techniques used. Different features like marital status, age, job stress, alternative job opportunities, attitude towards covid, willing to relocate, gender, education were found as the reasons for employee’s disintergration.

R Dishala Sandamini and R Gayashini Shyanka have created a system for forecasting employee turnover intention with reference to the apparel industry. Several data mining techniques like clustering, classification analysis and regression analysis were used.

S. N. Khera and Divya have studied the reason for employee turnover intention. The highest accuracy gained by support vector machine (SVM) algorithm. They concentrated on who will leave the organization rather than who will not. Characteristic features that were used are age, gender, marital status, job level, job profile, job satisfaction.

R. Saranya and S. Muthumani Impact of Perceived Organisation Support and Organisation Commitment on Turnover Intention of Women Employees in IT Industry. Using preserved organization support, the intention for attrition is determined.

S. Yadav, A. Jain, and D. Singh have used logistic regression, support vector machine, random forest, decision tree and adaboost to forecast the attrition rate of employee and provided the framework for forecasting the

IJARST
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

attrition rate of employees by analyzing the attitude and behaviour of an employee.

P. Ramachandran, M. Baranidharan by using ASP.NET software has found out that the effect on job satisfaction was insignificant where a decrease in job satisfaction and organizational commitment leads the employee to quit the company. Campbell, Jesse & Im, Tobin & Jeong, Jisu studied the Internal coherence and switchover intention and found that public service motivation is related to job performance and other behaviours. A decrease in organizational efficiency and the least motivation results in turnover intention.

NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely.NumPy stands for Numerical Python. In Python we have lists that serve the purpose of arrays, but they are slow to process. NumPy aims to provide an array object that is up to 50x faster than traditional Python lists. The array object in NumPy is called ndarray, it provides a lot of supporting functions that make working with ndarray very easy. Arrays are very frequently used in data science, where speed and resources are very important.

Pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with “relational” or “labeled” data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python. Additionally, it has the broader goal of becoming the most powerful and flexible open source data analysis/manipulation tool available in any language. It is already well on its way toward this goal. Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas data structures. Sea born helps you explore and understand your data. Its plotting functions operate on data frames and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots. Its dataset- 13 oriented, declarative API lets you focus on what the different elements of your plots mean, rather than on the details of how to draw them. Seaborn helps you explore and understand your data. Its plotting functions operate on dataframes and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots. Its dataset-oriented, declarative API lets you focus on what the different elements of your plots mean, rather than on the details of how to draw them. Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy and Matplotlib.

IJARST
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

3.SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

The system architecture in the image you provided is a process for utilizing machine learning to forecast and implement strategic policies to reduce employee turnover. The process begins with an employee dataset that undergoes data pre-processing. The cleaned data is then split into testing and training datasets. Machine learning algorithms are applied to these datasets for training and testing purposes, leading to forecasting results that inform the creation of strategic policies to reduce turnover.

The architecture is illustrating a machine learning process for reducing the employee turnover. “Employee data set”is represented by an icon of buildings and documents, indicating the collection of employee data. An arrow leads from the dataset to “Data preprocessing, ”represented by gear icons, indicating the cleaning and preparation of data. The processed data goes through “Dataset splitting,” dividing it into testing and training sets, each set is represented with distinct icons. “Training” involves stacks of coins icon, symbolizing the investment in processing power or resources for machine learning algorithms. “Testing” is connected back to training, forming a feedback loop for continuous improvement. The outcome leads to “Forecasting” depicted by people icons and graphs, indicating forecastions related to employees. Finally, forecasts inform “Strategic policies to reduce turnover” illustrated by people discussing around a table, showing collaborative decision-making.

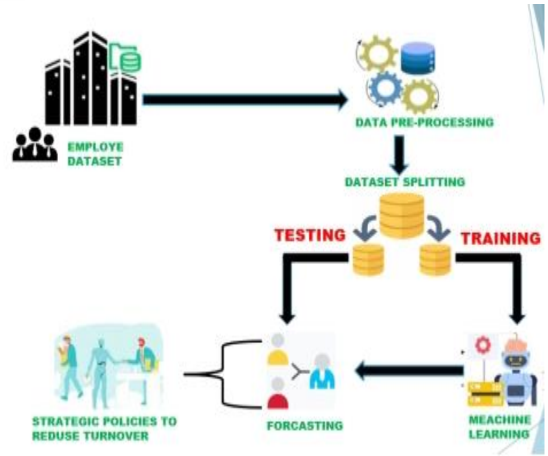


fig 1 system architecture

ACTIVITY DIAGRAM

Activity Diagrams in UML serve to visually represent dynamic workflows, showcasing the sequence and conditions of activities within a system or business process. The key components include nodes, representing actions or decisions, and transitions, illustrating the flow between these nodes. Initial and final nodes mark the activity's start and end. Control flows connect actions, specifying the order of execution, while decision nodes enable branching based on conditions. Forks and joins manage parallel flows, and swim lanes partition activities among different entities for clarity.

- **Nodes**: Represent actions or decisions.
- **Transitions:** Illustrate flow between nodes.
- **Initial and Final Nodes:** Indicate activity start and end.
- **Control Flows:** Connect actions, defining execution order.
- **Decision Nodes:** Facilitate

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

branching based on conditions.

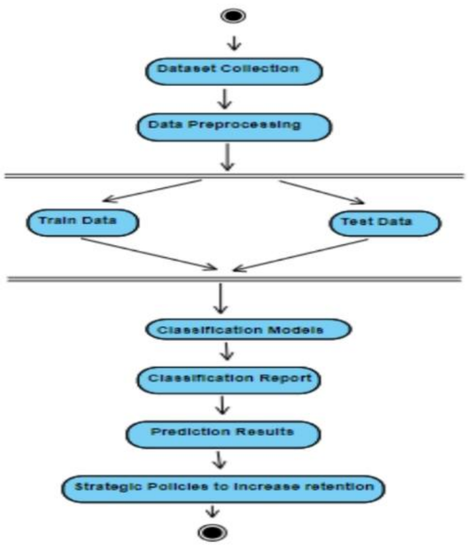
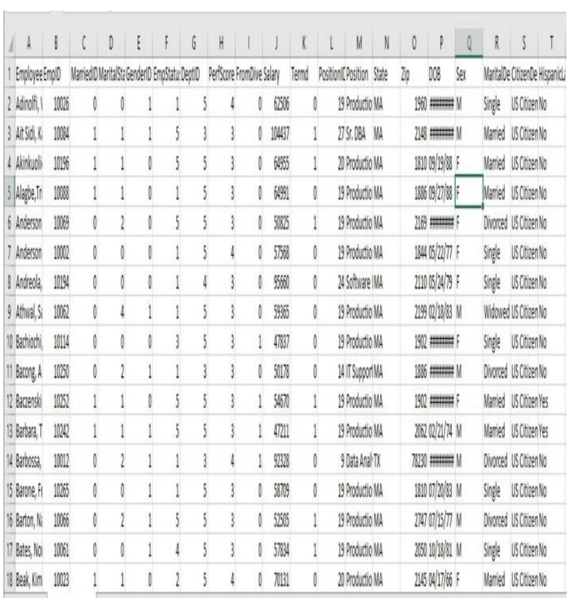
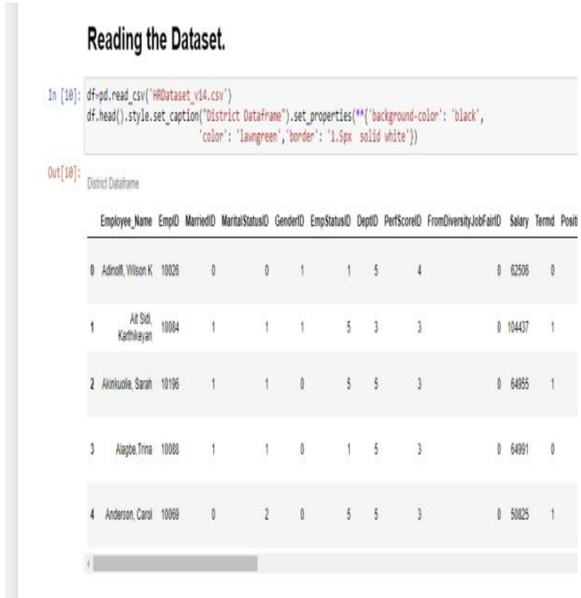


Fig 2 Represents Activity Diagram

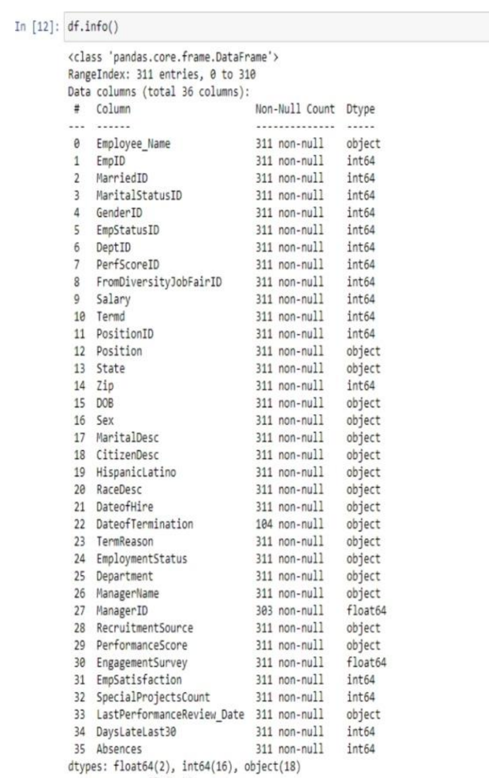
4.OUTPUT SCREENS



Ss 4.1. Dataset

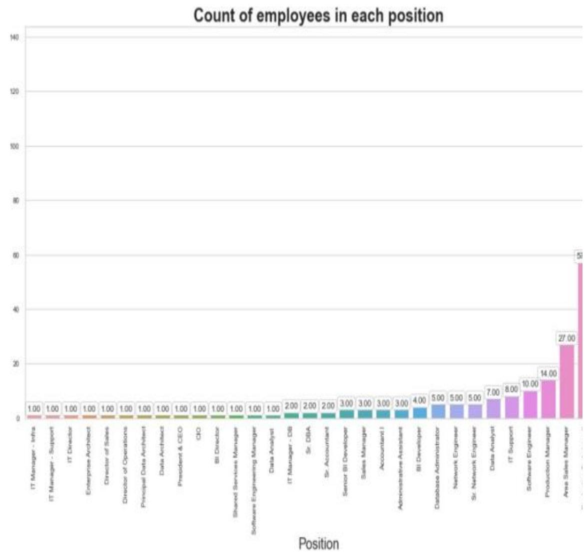


Ss 4.2. Reading the Dataset

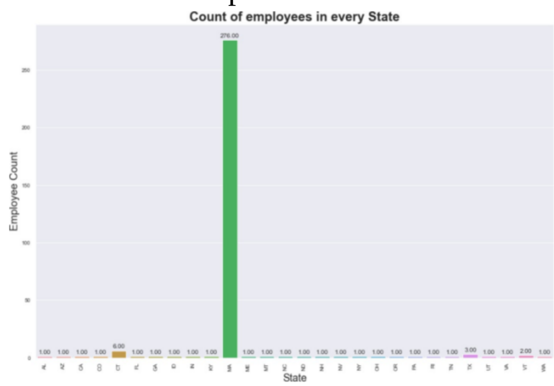


Ss 4.3. Dataset Information

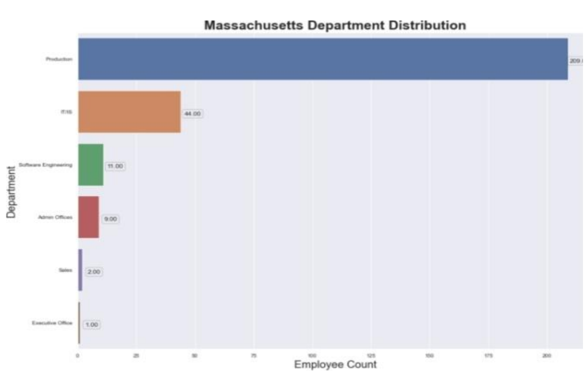
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362



Ss 4.4 Number of employees in each position

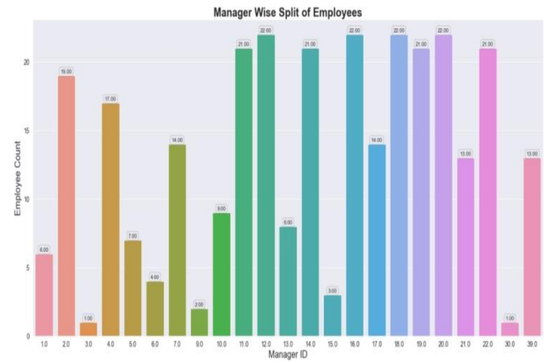


Ss 4.5 Number of employees in each state

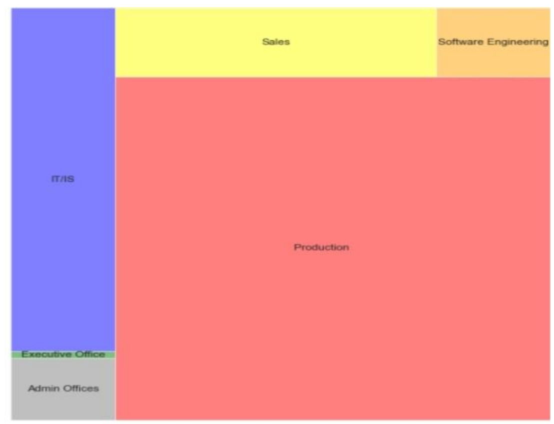


Ss 4.6 Department

Distribution

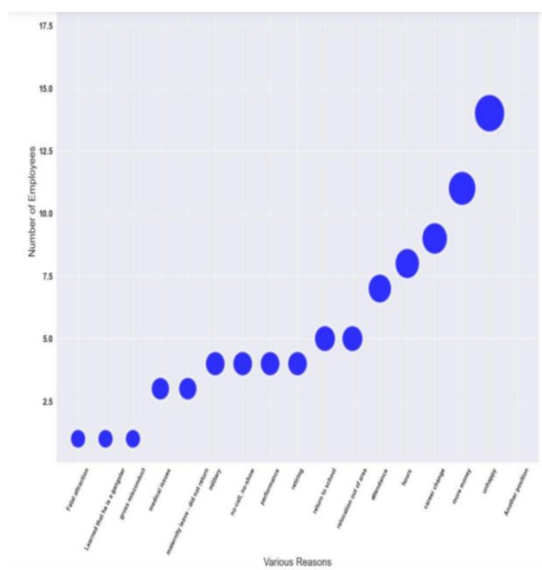


Ss 4.7 Manager – Employee count



Majority of Employees belong to the Production Wing of the company.

Ss 4.8 Department wise count



Ss 4.9 Reasons for turnover

.

```
              precision    recall  f1-score   support

           0       0.67      0.67      0.67         3
           1       0.67      0.67      0.67         3

    accuracy                           0.67         6
   macro avg       0.67      0.67      0.67         6
weighted avg       0.67      0.67      0.67         6

9]: print(confusion_matrix(y_test, y_predict))

[[ 1  5  0  0  0  0  0  0  0  0  0  0  0]
 [ 2 40  0  0  0  0  0  0  1  0  0  0  0]
 [ 0  1  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  2  1  0  0  0  0  0  0  0  0  0  0]
 [ 0  1  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  1  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  1  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  2  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  1  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  1  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  1  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  2  0  0  0  0  0  0  0  0  0  0  0]]

5]: print(classification_report(y_test, y_predict))

                                  precision    recall  f1-score   support

                Another position       0.00      0.00      0.00         7
    Learned that he is a gangster       0.00      0.00      0.00         1
               N/A-StillEmployed       0.63      0.93      0.75        40
                   career change       0.00      0.00      0.00         3
maternity leave - did not return       0.00      0.00      0.00         1
                  medical issues       0.00      0.00      0.00         1
                        military       0.00      0.00      0.00         1
                      more money       0.00      0.00      0.00         2
                no-call, no-show       0.00      0.00      0.00         1
                     performance       0.00      0.00      0.00         1
          relocation out of area       0.00      0.00      0.00         2
                return to school       0.00      0.00      0.00         1
                         unhappy       0.00      0.00      0.00         2

                        accuracy                           0.59        63
                       macro avg       0.05      0.07      0.06        63
                    weighted avg       0.40      0.59      0.47        63
```

Ss 4.10. Confusion matrix

5.CONCLUSION

Employee attrition is a major problem that the Indian IT industry and other industries have been experiencing in the last decade. This research was aimed to develop a forecast model using machine learning, to tackle the problem of employee turnover in the Indian information technology industry and to help find a solution for the turnover and take significant measures to reduce the employee turnover in the industry. In this paper we used different algorithms like naïve bayes, KNN, and random forest among these classifiers the Naïve bayes is giving the highest accuracy.

6.FUTURE ENHANCEMENTS

Further research can be conducted by examining the employees who have switched over from an organization rather than forecasting the turnover intent across the entire industry. This may give provide new insights into the forecaster variables. The study can also be conducted in other sectors to forecast the switchover intent of the employees.

The employee-turnover model is based on employee records exported from HR's management system. However, some key information, like salary, is not accessible. Nevertheless, many researchers have proved that employee salary is a key factor in employee-turnover. The forecasting model will continue being developed and improved if additional information, like salary is available.Regardless of the statistical methods used in building a model, interview and survey methods can be more precisely designed for certain factors among different targeted groups. These factors are managers' greatest concerns.For example, when the turnover rate suddenly increases, interview and survey data can quickly identify the important changing factor, like leadership, organizational structure, or job satisfaction.Finally, the optimal employee hiring strategy can be explored in industrial and academic areas. Instead of simply applying an EOQ model to determine an optimal hiring number,a more complex optimization model can be developed by considering how to redistribute and retrain current employees to the opening position. This may provide additional value to both academic and industry practitioners.

7.REFERENCES

1. R. Punnoose and P. Ajit, "forecastion of Employee Turnover in Organizations using MachineLearningAlgorithms,"ijarai

,vol.5,no.9,2016,doi: 10.14569/IJARAI.2016.050904.

2. S K Monisaa Tharani and S N Vivek Raj, "forecasting employee turnover intention in IT&ITeS industry using machine learning algorithms", 2020, doi: 10.1109/ISMAC49090. 2020 .9243552.
3. Dishala Sandamini R and Gayashini Shyanka R, "Employee Turnover forecastion System: With Special Reference to Apparel Industry in Sri Lanka", 2021, doi: 10.1109/I2CT51068.2021.9418108.
4. S. N. Khera and Divya, "forecastive Modelling of Employee Turnover in Indian IT Industry Using Machine Learning Techniques," Vision, vol. 23, no. 1, pp. 12–21, Mar. 2019, doi: 10.1177/0972262918821221.
5. R. Saranya and S. Muthumani, "Impact of Perceived Organisation Support and Organisation Commitment on Turnover Intention of Women Employees in IT Industry," Journal of Theoretical and Applied Information Technology, vol. 25, no. 2, pp. 256-261, may 2015.
6. S. Yadav, A. Jain, and D. Singh, "Early forecastion of Employee Attrition using Data Mining Techniques," in 2018 IEEE 8th International Advance Computing Conference (IACC), Greater Noida, India, Dec. 2018, pp. 349–354, doi: 10.1109/IADCC.2018.8692137.
7. P. Ramachandran, M. Baranidharan, V. Ramesh, and S. Prabhakaran, "An Interactive Mining Approach to find the job Satisfaction and Staff Turnover Intentions," International Journal of Computer Science Engineering & Technology, vol. 1, no. 2, pp. 67-69, Mar. 2011
8. Abelson, M. A. and Baysinger, B. D. (1984). Optimal and dysfunctional turnover: Towards an organizational level model. Academy of management Review, 9(2):331–341.
9. Agresti, A. (1996). Categorical data analysis, volume 996. New York: John Wiley & Sons.
10. Alao, D. and Adeyemo, A. (2013). Analyzing employee attrition using decision tree algorithms.