A Project Report

on

EMPLOYEE PERFORMANCE PREDICTION

Submitted in partial fulfillment of the
requirement for the award of the degree of

School of Computer Science and Engineering

Under The Supervision of
E. Goutham
Assistant Professor

Submitted By

SAKSHI GAUR
19SCSE1180106

CHIRAG KAUSHIK
19SCSE1010729

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING /
DEPARTMENT OF COMPUTERAPPLICATION
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA
December, 2021

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **"EMPLOYEE PERFORMANCE PREDICTION"** in partial fulfillment of the requirements for the award of the **BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING** submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of **JULY-2021 to DECEMBER-2021**, under the supervision of **Mr. E. GOUTHAM, Assistant Professor, Department of Computer Science and Engineering**, Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

19SCSE1180106 – SAKSHI GAUR

19SCSE1010729 – CHIRAG KAUSHIK

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Supervisor

(Mr. E. Goutham, Assistant Professor)

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of 19SCSE1180106 – SAKSHI GAUR, 19SCSE1010729 – CHIRAG KAUSHIK has been held on _____ and his/her work is recommended for the award of **BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING.**

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date: November, 2013

Place: Greater Noida

ABSTRACT

These days, human capital is a big concern for companies' management where their most interest is in hiring the highly qualified personnel which are expected to perform highly as well. Human Resources Management (HRM) has become one of the important interests of managers and decision makers in almost all types of businesses to adopt plans for correctly finding highly qualified individuals. Accordingly, managements have become more interested about the performance of employees working in their companies. Some of the researches show that professional skill development programs are required in order to prepare employees to perform their tasks more efficiently. The knowledge flow model of the Open-source tool is also used to for the same purpose. To get a highly accurate model, several experiments are executing these days, based on the previous techniques that are implemented tool used to help decision makers and human resources professionals to predict and enhance the performance of their employees.

Problem Formulation: There is no standard way to evaluate the human capital. You can take any organization no matter the level of that, one thing that you will find common is that many employees are wrongly credited for the project. Suppose a project has been given to a group of employees, tasks are assigned to peers but the back-end story is that one person has contributed more than others. But in the final presentation all are given equal credit. This adds up and leads to the problem that the organization will not be able to take fair decisions to promote/demote the employees. Therefore, it becomes crucial to keep the actual track of employees. This project is the solution to the problem.

Proposed Solution: Data mining is a field of information and knowledge discovery. With data mining techniques, first we will pre-process the dataset and will try to extract such knowledge from features.

We will use Decision Trees as it is one of the most used techniques. It creates the decision tree from the data given using simple equations depending mainly on calculation of the gain ratio, which gives automatically some sort of weights to attributes used to implicitly recognize the most effective attributes on the predicted target. A decision tree would be built with classification rules generated from it. Here Decision tree is used for predict Employee Performance.

For the prediction we will use Naïve Bayes algorithm, based on that all the independent attributes are selected from the value of the conditional independence. The naïve bayes algorithm computes its learning model from the set of conditional independences and its frequency from the dataset.

Tools and Technology Used:

- Data Mining: It is the process of discovering hidden patterns and extracting knowledge from the database.

It will also include the data pre-processing that is, manipulation of data before it is used in order to ensure enhanced performance.

- Feature Selection: It is one of the important steps involved in any model. There can be lot of features associated with any idea but selecting those which affects it the most are few. Example – Filter Method, Embedded Method, etc.

• Machine Learning: Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

Algorithms – Decision Trees, Naïve Bayes, Clustering, etc.

Result and Output: Human resource can play an essential role in company growth. A human resource department need an assessment whether the employee would comply company's wants. They can use of machine learning technology to predict employee's resignment before it happens and can decide in advance how to face it.

Naïve Bayes method along with confusion matrix will measure an effective performance score. This will give output in the percentage form. The Naïve Bayes model is also time efficient.

Future Scope: This will resolve the very problem of false crediting to employees. The effect of this model can be long-term. It clears the image of every working employee with minimum effort.

This model can be associated with the employee database and will serve as an extra feature accountable for the employee promotion.

In future this can be of great use to all the organizations as with the rise in number of employees the chance of miscredit will also increase and it will become difficult to keep honest track of all employees' performance.

Table of Contents

List of Figures

Acronyms

KPI	Key Performance Indicator
NB	Naive Bayes

CHAPTER-1

Introduction

For all organizations in the world, the turnover rate of employees is gradually increasing and becoming one of the significant issues for the organization, companies etc. The turnover problem consumes a lot of things like cost for the company including the cost of hiring new employee, cost severance pays and training them to make them prepare for the best. Companies, organizations and institutions are more concerned about how to reduce these turnover as it cost a lot of resource for them. The first step to reduce this problem is by identifying which employee will resign. The main cause of employee's reassignment is when the company can't provide enough satisfaction to the employee. Each employee's performance boosts their company competitive advantage among all other organizations. Human resource in each organization is tasked to assure their employee's satisfaction so that it can provide good performance causing a more profit to the organization. The successful companies, organizations and institutions are those who can identify their employee performance. Another way to get an idea about employee resignation is by monitoring their performance by using KPI. KPI is an important tool to determine and measure the success of the company. Using KPIs, Company performance toward its visions and how well it executed missions and strategies can be monitored easily. These KPIs gives an idea about the success ratio of a company is going to achieve according to the performance of their employees. An employee must be positioned according to his/her performance in the organization. Employee performance boosts company success to achieve its goals. In future, predicting employee performance will be a necessity for companies to success. The employee performance can be predicted based on these key performance indicator and satisfaction. So, the employee turnover can also be predicted by using Machine Learning. Machine learning had their algorithms and techniques. Classification is used in this kind of prediction under machine learning. Human Resources etc. are using machine learning these days. Naive bayes is one of the best and purest forms of Bayesian type. The algorithm is based on that all the independent attributes are appointed from the value of the conditional independence which makes it perfect for this task. This algorithm computes its learning model from the set of conditional independences and its frequency from the dataset. A Bayes proportion is used in Naive Bayes Algorithm. This algorithm is well-known in the area of text mining as a good algorithm to solve classification problems from the beginning. Here, a classification methodology is proposed to predict the performance of employees using machine learning. In this project, we proposed a Chi-squared test for feature selection and naïve bayes based prediction methodology classification which is proposed to tell us the performance rating of employees using machine learning. Prediction helps the company to decide which employees are deserved to behold or not.

CHAPTER-2 Literature Survey/Project Design

Research Papers: Various research papers have revealed the fact that machine learning can do miracles for the project. Different types of machine learning techniques like k-means clustering, decision trees, random forest, naïve bayes etc., are there for the classification problems. The need is to select the right technique with the right features.

1. Employee's Performance Analysis and Prediction using K-Means Clustering & Decision Tree Algorithm - by Ananya Sarker, S.M. Shamim, Dr. Md. Shahiduz Zama & Md. Mustafizur Rahman.
2. Employee Performance Prediction System using Data Mining - Tejas Raut, Priya Kale, Rashmi Sonkusare, A. K. Gaikwad.

Articles: There can be a lot of features associated with an employee. Here comes the need of data mining. Articles on data mining clears the way for the model to use the database effectively (to reduce the missing rows, filling null values and discovering patterns etc).

1. https://www.ibm.com/cloud/learn/data-mining
2. https://www.britannica.com/technology/data-mining

Books: Machine learning book will be of great use as it gives the idea about the various models and techniques along with their advantages and disadvantages.

1. Introduction to Machine Learning with Python by Andreas C. Müller and Sarah Guido

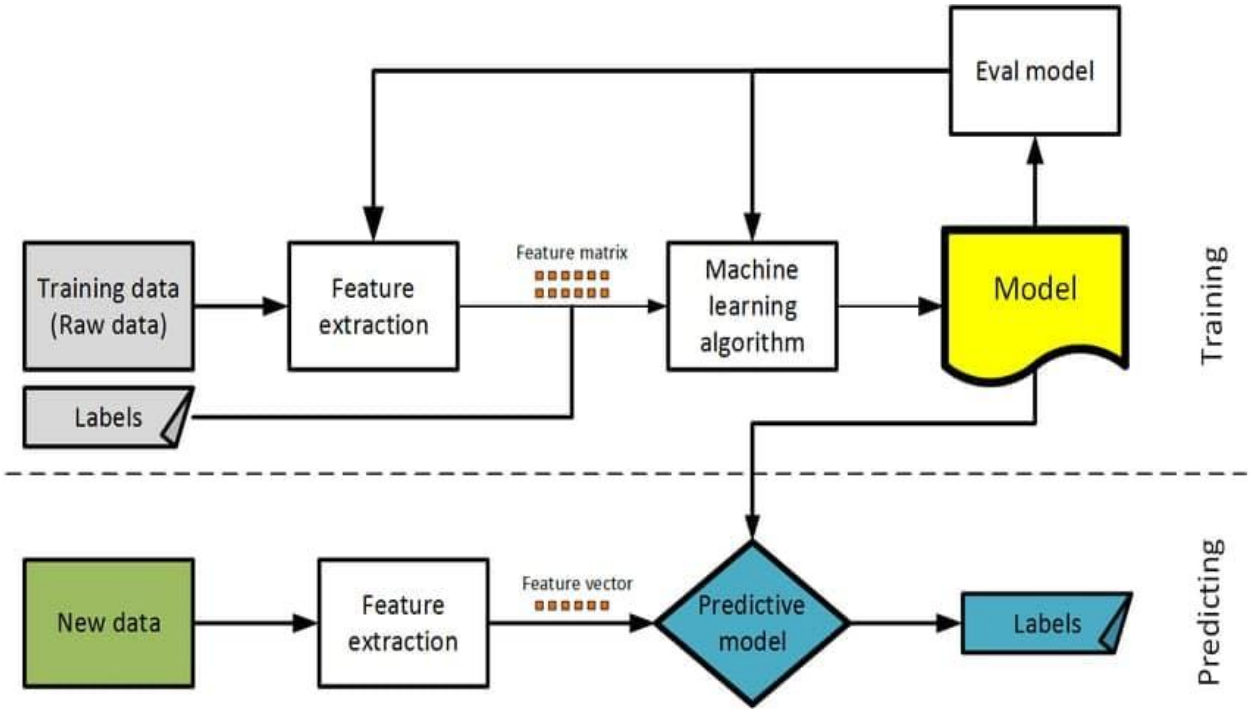

Fig. 2.1. Design of the model

CHAPTER-3 Working of the Model

Working Steps:

- The user needs to introduce the model with the employee with his information of the ten features which are: Job Involvement, Monthly Income, Monthly Rate, Years in current role, Year since last promotion, Years with current Manager, Relationship Satisfaction, Percent salary hike, Number of companies worked, Distance from home.
- The model does the processing.
- It will produce the result in the form of performance rating.

Processing Methods: Data mining is a field of information and knowledge discovery. With data mining techniques, first we will pre-process the dataset and will try to extract such knowledge from features.

We will use Decision Trees as it is one of the most used techniques. It creates the decision tree from the data given using simple equations depending mainly on calculation of the gain ratio, which gives automatically some sort of weights to attributes used to implicitly recognize the most effective attributes on the predicted target. A decision tree would be built with classification rules generated from it. Here Decision tree is used for predict Employee Performance.

For the prediction we will use Naïve Bayes algorithm, based on that all the independent attributes are selected from the value of the conditional independence. The naïve bayes algorithm computes its learning model from the set of conditional independences and its frequency from the dataset.

- Data Mining: It is the process of discovering hidden patterns and extracting knowledge from the database.

  It will also include the data pre-processing that is,
  manipulation of data before it is used in order to ensure
  enhanced performance.
- Feature Selection: It is one of the important steps involved in any model. There can be lot of features associated with any idea but selecting those which affects it the most are few. Example – Filter Method, Embedded Method, chi-square test etc.

  A chi-square test is used in statistics to test the independence of two events. Given the data of two variables, we can get observed count O and expected count E. Chi-Square measures how expected count E and observed count O deviates each other.

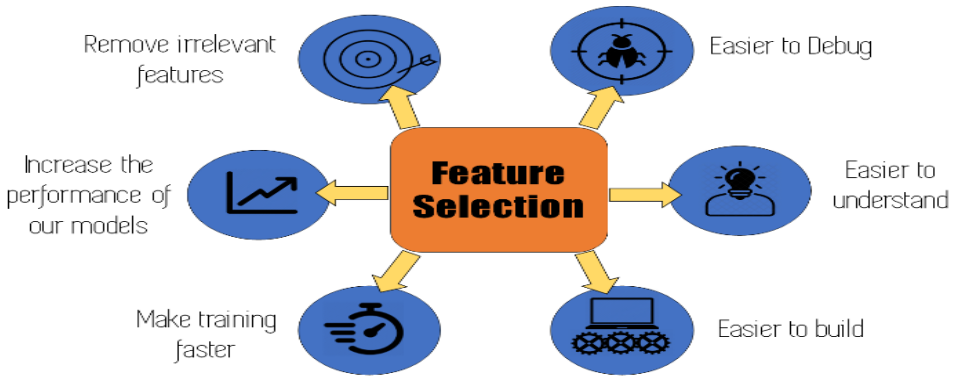

Fig. 3.1. Need of Feature Selection

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.

Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which help in building the fast machine learning models that can make quick predictions.

It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

Naïve: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features.

Hence each feature individually contributes to identify that it is an apple without depending on each other.

Bayes: It is called Bayes because it depends on the principle of Bayes' Theorem.

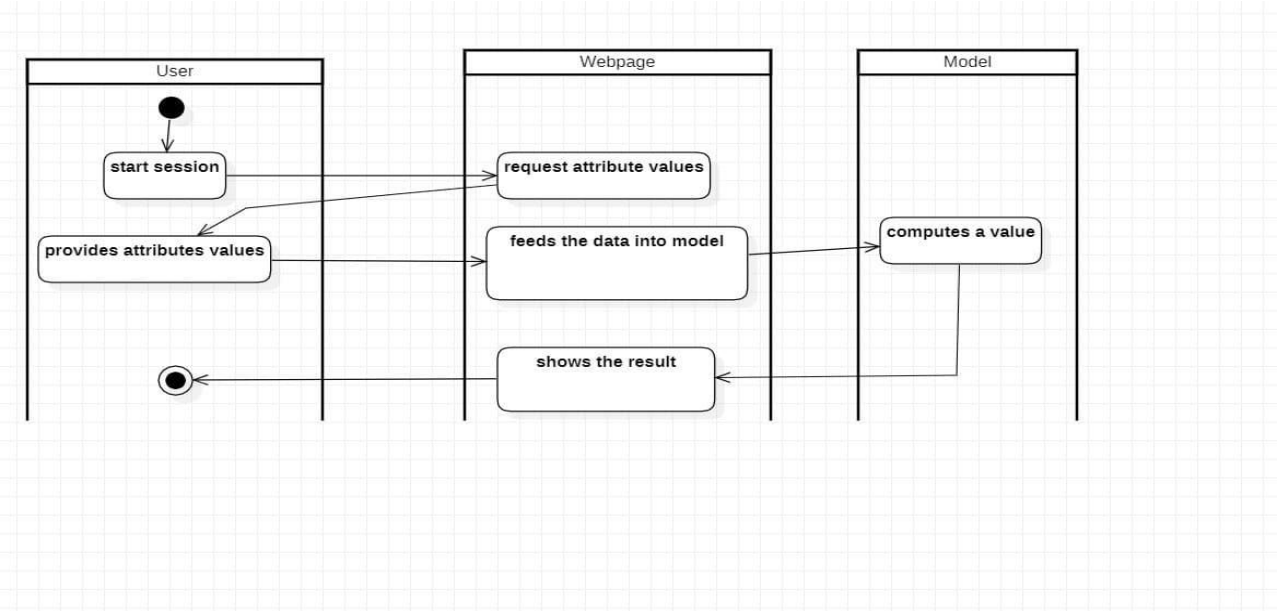

Fig. 3.2. Sequence Diagram

CHAPTER-4 Result and Discussion

With the association of Chi-squared test and NB method, it successfully measures an updated performance rating as the objective variable, with 99.72% accuracy. This Naïve Bayes model takes very less time to calculate with minimum hardware specifications. Figure 4.1. displays the screenshot of model accuracy.

Fig. 4.1. Implementation Screenshot

CHAPTER-5 Conclusion and Future Scope

Conclusion:

- This study shows that human resource can plays an essential role in company growth.
- A human resource department need an assessment whether the employee would comply company's wants.
- They can use of machine learning technology to predict employee's resignment before it happens and can decide in advance how to face it. From the evaluation, correctly classified instance is 95.48% using the proposed model of Naïve Bayes.
- This is shows that the naïve bayes technique is very good at predicting.
- Alongside, based on the confusion matrix, it found a slight amount of false-positive result that means the cost of using the naïve bayes technique is small.
- Findings based on the insights and feedbacks have been done on the existing employee prediction system.
- The datasets were usually used for specific case studies and all the attributes were not used to their full extent.
- The data in the dataset can be used to a better extent by categorizing specific attributes and removing redundant data from the dataset.
- Further analysis can be done by including more attributes to which the employee attrition is related.
- Data visualization and numerical analysis can be done to gain better insight regarding the trends and patterns in predicting the attrition rate and the major attributes involved in prediction.

Future Scope:

This will resolve the very problem of false crediting to employees. The effect of this model can be long-term. It clears the image of every working employee with minimum effort.

This model can be associated with the employee database and will serve as an extra feature accountable for the employee promotion.

In future this can be of great use to all the organizations as with the rise in number of employees the chance of miscredit will also increase and it will become difficult to keep honest track of all employees' performance.

REFERENCES

1. Y. F. Safri, R. Arifudin, and M. A. Muslim, KNearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor, Sci. J. Informatics, vol. 5, no. 1, p. 18, 2018. https://doi.org/10.15294/sji.v5i1.12057
2. O. Olubiyi, G. Smiley, H. Luckel, and R. Melaragno, A qualitative case study of employee turnover in retail business, Heliyon, vol. 5, no. 6, p. e01796, 2019. https://doi.org/10.1016/j.heliyon.2019.e01796
3. P. W. Hom, T. W. Lee, J. D. Shaw, and J. P. Hausknecht, One hundred years of employee turnover theory and research, J. Appl. Psychol., vol. 102, no. 3, pp. 530–545, 2017. https://doi.org/10.1037/apl0000103
4. R. F. Cox, R. R. A. Issa, and D. Ahrens, Management's perception of key performance indicators for construction, J. Constr. Eng. Manag., vol. 129, no. 2, pp. 142–151, 2003.
5. J. M. Cramer and B. Roes, Total employee involvement: measures for success, Environ. Qual. Manag., vol. 3, no. 1, pp. 39–52, 1993. https://doi.org/10.1002/tqem.3310030105
6. R. Punnoose and P. Ajit, Prediction of Employee Turnover in Organizations using Machine Learning Algorithms, Int. J. Adv. Res. Artif. Intell., vol. 5, no. 9, pp. 22–26, 2016.
7. A. Idowu, Effectiveness of Performance Appraisal System and its Effect on Employee Motivation, Nile J. Bus. Econ., vol. 3, no. 5, pp. 15–39, 2017.
8. G. Cohen, R. S. Blake, and D. Goodman, Does Turnover Intention Matter? Evaluating the Usefulness of Turnover Intention Rate as a Predictor of Actual Turnover Rate, Rev. Public Pers. Adm., vol. 36, no. 3, pp. 240–263, 2016. https://doi.org/10.1177/0734371X15581850
9. Z. Ahmed, S. Sabir, M. Khosa, I. Ahmad, and M. A. Bilal, Impact of Employee Turnover on Organisational Effectiveness in Tele Communication Sector of Pakistan, IOSR J. Bus. Manag. Ver. IV, vol. 18, no. 11, pp. 2319– 7668, 2016.
10. Y. Zhao, M. K. Hryniewicki, F. Cheng, B. Fu, and X. Zhu, Employee turnover prediction with machine learning: A reliable approach, in Proceedings of SAI intelligent systems conference, 2018, pp. 737–758. https://doi.org/10.1007/978-3-030-01057-7_56