Application of Machine Learning in Employee Performance Prediction

Archana Boob[1*], Sandeep Sharma[2], Saurabh Singh[3], Rafsan Ali[4]

School of C&IT REVA University Rukmini Knowledge Park Yelahanka Bangarore. 560064

Corresponding Author: rafsanaliakaash@gmail.com,tel-98801505700

Abstract— In emerging developing countries such as India, companies heavily rely on their human workforce for services. That is why employee performance management at the individual level is must and the business case for implementing a system to measure and improve employee performance should be strong. The concept of the project is: Today majority of the giant retail companies are facing a lot of issues in their current assessment planning of their employees. This wrong assessment planning leads to employees not being used to the fullest potential which causes loss to businesses and major capital loss in man hours, also this assessment planning requires a lot of manual strategies which are very costly and hence these assessment strategies then turn out to be costly, time taking, biased and working on mostly non relevant data. We used the Machine learning classification technique for the extraction of knowledge significant for predicting employee performance using a .csv file sourced from (INX Future Inc.).

Keywords- Employee Performance Analysis,SVM, Machine Learning, Algorithm K-NN algorithm, random forest.

I. INTRODUCTION

The growth of unlimited data which is being used efficiently to predict or make certain decisions is because of machine learning.Their is a vast availability of data which is being used by the computers to make decisions and accurate ones based on previous data.More the data,better is the decision making.They are used for improving processes to a larger extent with the help of the data fed to the model [8].

There are many classification techniques in machine learning such as Decision Tree, Neural Network, Decision tree is among the popular classification techniques, which can produce the interpretable rules or logic statement[7]. Tools are easy to use and results can be found within a short span of time.If data is too big then the model will take time to give its prediction but nevertheless it will give accurate prediction based on the attributes given to it,the data needs to be cleaned so it can be understood by the machine making it easier for the computer to understand..

The REVA University is a private co-educational engineering college in Bangalore Karnataka (India). The mandate of the University is to provide learning, consultancy and research services designed to inform promote national development and standards of competence and integrity in a result-oriented service. In this knowledge economy, the University relies heavily on the human capital to build value. Consequently, performance management at the individual employee level is essential and the business case for implementing a system to measure and improve employee performance at the University is strong.Businss organisations sometimes cant figure out what is wrong with the company employees,so a particular model is built with the help of machine learning to check the performance of the individual and based on that the employee is judged if he/she is fit to be in the company.[4].

The University introduced performance monitoring and evaluation system to see the performance of its employees in a proper manner to see which individual was working in what format. Employee performance evaluation is systematically carried out, according to a definite plan, typically by a corresponding manager or supervisor. Employee appraisal results are useful in compensation decisions, promotions, training and development programs, feedback, and personal development. Employee performance evaluation at the School therefore forms a basis for many HR decisions. To maintain consistency and disdain from partiality, this research sought to propose a classifier model for employee performance prediction. With proper machine learning model,the HR can select the individual best on the attributes that make an individual the perfect employee based on previous employee results.As a result the efficiency of the employee increases followed by the increase in efficiency of the whole company.[11].

II. RELATED WORK

Machine learning has been used in various fields all over the globe such as in companies,artificial intelligence and medical fields to make prdictions or classify something based on the data presented to them.The data provided is in vast number which helps for the predictions to be even more accurate so that way it becomes easier to make the model and get a proper prediction.The classification part is mostly used in medical fields to detect a disease. patterns from large data sets in order to find observable patterns in HR[3].

Used rule-based classification Machine learning technique to extract knowledge significant for predicting training needs of newly-hired members in order to devise the necessary development programs. They used the Cross Industry Standard Process for Machine learning in discovering significant models needed for predictive analysis and demonstrated the required professional trainings to prepare faculty members to perform their tasks effectively.

Used k-nearest neighbors algorithm (k-NN) is a non-parametric method used in regression and classification. .In both cases, the input consists of the k closest neighbours. The output depends on whether k-NN is used for classification or regression[2].

Used Support Vector Machine" (SVM) which is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the different classes very well.

III. METHODOLOGY

3.1 Research Design
The development of this research followed the random forest model was best suited for this research because it provided a generic guide to develop Machine learning project lifecycle. The employee performance data was collected from INX future.IncA number of models were tried out to see which one gave better accuracy based on the data given by the company.The best attributes were recorded and the rest were droped by using feature selection. The classification process was carried out with three different Machine learning algorithms, random forest classifer,K-NN , support vector machine to identify the best and most suitable classification algorithm. The chosen algorithm was then improvised to obtain the best classification rate. The sequences of steps followed by this research are illustrated in the figure 1[6].

The predictive attributes were extracted from the employee records form and were used as performance attributes. They include:

1. EmpNumber
2. Age
3. Gender
4. MaritalStatus
5. EducationBackground
6. EmpDepartment
7. EmpJobRole
8. BusinessTravelFrequency
9. DistanceFromHome
10. EmpEducationLevel
11. EmpEnvironmentSatisfaction
12. EmpHourlyRate
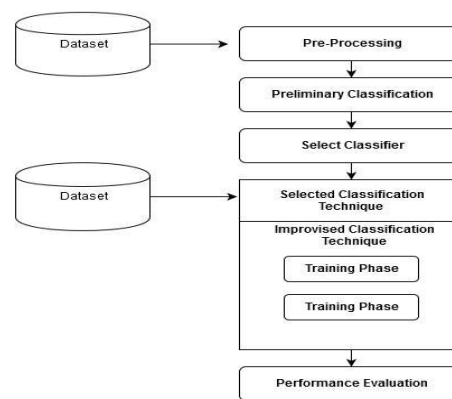13. EmpJobInvolvement
14. EmpJobLevel

Figure 3.1: Steps of building the classifier

There are many attributes but the prediction is to be made with respect to the target value.

3.2 Data Understanding
The data for training the model was collected from the companyvINX Future.Inc in a new .csv file. The data collected included 1200 employee's performance appraisal records, described by 28 parameters. the parameters show the different attributes of an employee based on which the prediction is to be made. This data was divided into two datasets. we split the dataset into test set and training data set to get a proper prediction[5].

3.3 Data Cleaning
The raw data we got was completely clean. The categorical values had to be dealt with which was done while cleaning the data. The types of data were then reviewed and modified. Data cleaning was done, we checked for missing values but data contained no missing values. To get proper prediction we did label encoding for categorical values and then visualized the data to check the main attributes which would help in proper prediction of the model.

3.4 Modeling and Experiment
After cleaning the data we then checked for the attributes which would effect the model more by using visualization tools. We also did feature selection. We split the dataset into test and train data set in a ratio of 70:30.Then we fit the model and used random forest classifier, K-NN, svm to check the accuracy. We also saw the confusion matrix to check if the values were classified properly.
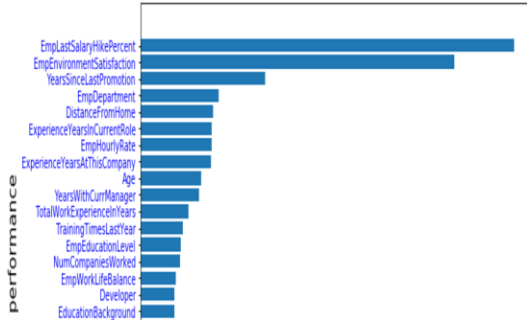

Figure 3.4.1: factors effecting employee performance

For this type of classification the random forrest was most suitable as it uses decision tree concept which was very good in finding insights into the data as to which attribute had more effect on the employee performance.The models were compared based on their accuracy as shown below:

Table 1 : Classification Algorithms Efficiency Rate

NO	MODEL	EFFICIENCY
1	random_forest_model 1	0.920833
2	random_forest model 2	0.933333
3	knn_model3	0.812500
4	SVM	0.86

As seen from the table 1, random_forest_model 2 algorithm had the highest accuracy of 93.33% and was therefore best suited for the training and development of the classification model.

3.5 Classification Model
Used random forest classifier and used k-nearest neighbors algorithm (k-NN) which is a non-parametric method used in classification and regression .In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression. To establish an efficient trained model to predict employee performance, as first step, building a Random forest model without feature selection which showed an acceptable efficiency of 91% however, to push the possibilities, K-Nearest Neighbor model was built which had 81% efficiency, which was lower than our previous model. The final step of using Random Forest along with feature selection which provided 94% efficiency, which was the best and hence was considered for selection.We also checked with svm but that gave 86% accuracy.

IV. RESULTS AND DISCUSSION

4.1 Classification Model Results
The research found out that several factors had a great effect on employee performance . One of the most effective factors was the e salary hike percentage which had the maximum effect. Other attributes that participated in the decision tree were EmpNumber, Gender, EducationBackground, MaritalStatus, EmpDepartment, EmpJobRole, BusinessTravelFrequency, OverTime, Attrition, EmpDepartment.

- Candidates between 31years to 40 years of age tend to
- perform better at Ranking .
- Candidates with +15 years total experience tend to be
- average performers.
- Candidates with 12 to 15 years of experience show higher probability of better performance .
- Candidates with 1 year of experience have very high
- probability of high-performance. However, these candidates also have at least 50% probability of attrition.
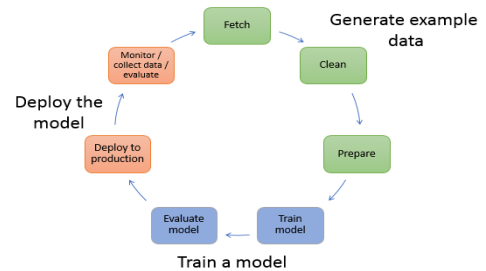

Figure 4.1 : Training model

As we can see from the above diagram first we need to take the dataset from the company, then we clean the data, then we find the target value and other attributes. Then we make a model based on which the prediction is to be made. We split the data set into training and testing dataset. We fit the model with the algorithm which we want to use and let the model predict based on the target value.
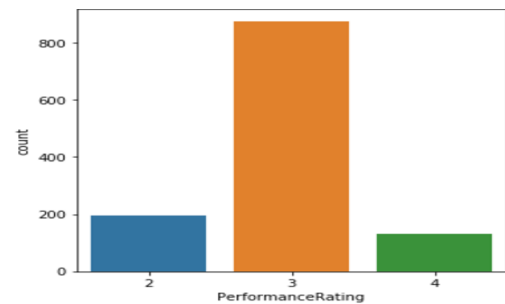

Figure 4.2: count to performance rating

Here we see the count of the no of employees with different ratings with respect to performance.
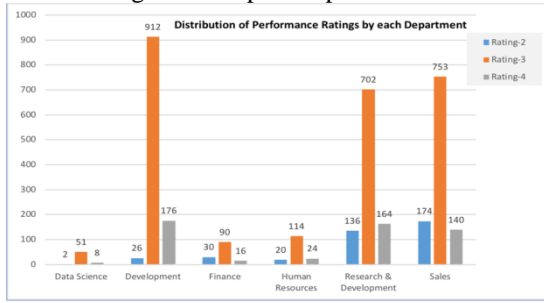

Figure 4.3: Distribution of performance rating by each department.

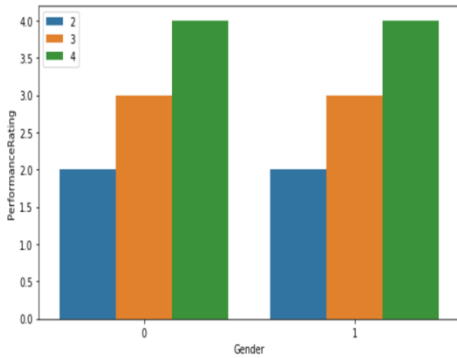Here we see department wise performance by each department.

Figure 4.4: This graph shows gender with respect to performance rating. There is no change in particular for men and women.
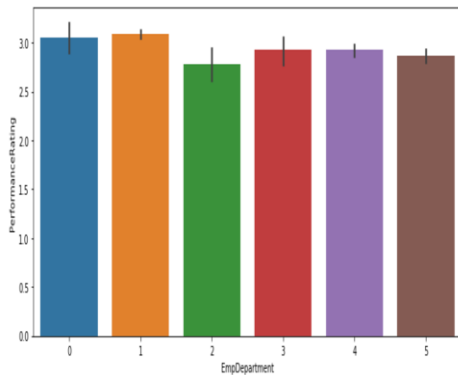

Figure 4.5: This shows each different department with respect to performance rating
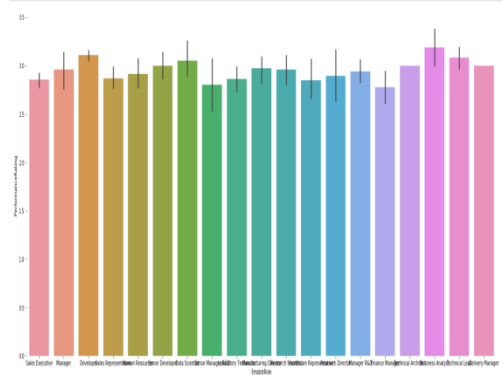

Figure 4.6: This graph shows every employee job role with respect to performance rating.
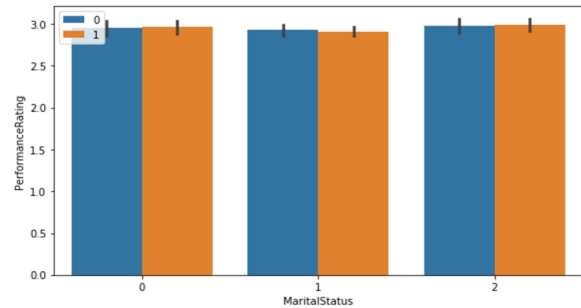

Figure 4.7: This shows performance rating with respect to marital status of employee.
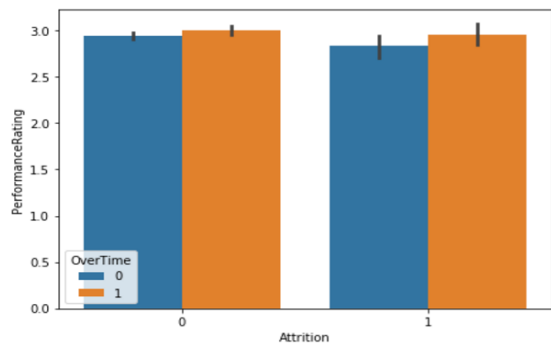

Figure 4.8: This graph shows performance rating with respect to attrition.

V. CONCLUSION AND FUTURE SCOPE

This paper is based on the probability of building a classification model for finding out employee performance. Many performance attributes were then tested using performance score for the data from INX future.inc. Some of the attributes were found very much affecting on the performance prediction.The emloyee job role and employee salary hike were the attributs that affected the model the most.The attributes that did not matter much was marital status attrition.

The salary hike attribute has shown alot of effect while the Marital Status and Gender have showed some small effect in predicting the performance. Some Educational factors like experience, Qualification and Professional Training have little affected the performance but not with that much clearity. Finally, the effect of salary hike on employee performance was clear where employees had better performance.

For management of the organizations and human resource Department, this model, can be used in predicting the employee performance. Several actions can be taken in this case to avoid any formulated risk related to hiring less performed employee.

In future work, it is advised to extend the prediction of employee as a continuous value instead of predicting performance category of the employee.

A comparative analysis of the category prediction model and a value prediction would help to choose a more robust model. When the appropriate model is generated, software could be developed to be used by the HR including the rules generated for predicting performance of employees.

REFERENCES

[1]. Al-Radideh, Q.P., Al-Naagi, E., (2011). Machine learning used for Classification Model for Employees Performance prediction, International Journal of Advanced sciences of computer and Applications, 3(2), pp 144 – 151

[2]. Anchetra, R.A, Cabautan, R.J.M., Lorenaa, B.T.T., Rabagoo, W., (2010). Predicting faculty trainings and performance development using rule-based algorithm classification, Asean Journal of Science of Computer and Information Technology 2: 7, pp 204 – 208.

[3]. heine, C.F., Chen, L.F., (2008). Learning machine to improve employee selection and increase human input: A case study in technology industry, Expert   applications on system, 34(1), pp 283–293

[4]. Delaavari, N., Phon-Amnnuaisuk S., (2009). Machine learning Application in Higher Learning Institutions, education on informatics, 7(1), pp. 32–55

[5]. Hamidah J., Abdul R.H., Zulaiha A.O., (2010). Discover Knowledge Techniques for Talent Forecasting in employee Resource Applications, World Academy of Engineering, Science and Technology.

[6]. Jantan, H., Hamdan, A. R., Othman, Z. A. (2011). applying Machine learning techniques for talent management. International Conference on Engineering of Computer and Applications IPCSIT Vol. 3.

[7]. Jantan, H., Hamdan, A. R., & Othman, Z. A. (2010). resource talent prediction in HRM using C4. 5 classification algorithm, International Journal on Science Computer and Engineering, 2(08-2011), pp 2525-2535.

[8]. Jantan, H., Hamdana, A. R., Othman, Z. A. (2008). Knowledge discovery techniques for talent forecasting in human resource application. World Academy of Science, Engineering and Technology, Penang, Malaysia, pp 803.

[9]. Jayanthi R., D.P. Goyal, S.I Ahson, (2008). Data techniques of Mining for better decisions in human resource systems, International Journal of Business Information Systems, 3(6), pp 468 – 484

[10]. Kotsiantiss, S.B., (2008). Supervised machine learning: a view of classification techniques, Informatics, 31, pp 247-268.

[11]. Kurgana, L.A., Musileke, P. (2008). A survey of knowledge discovery and Data Mining Models, The Knowledge Engineering Review, 21(2), pp 1 - 25