

Predicting Employee Performance and discovering its correlation with various work environment factors

¹Gupta Rahil M, ²Shah Vishakha R, ³S. Senthil

^{1,2}MCA Student, School of CSA, REVA University, Bangalore

³Professor and Director, School of CSA, REVA University, Bangalore

Article Info

Volume 83

Page Number: 3626-3631

Publication Issue:

May-June 2020

Abstract

Major source of judgement for human resource managers is the performance of its employees in company. Many methods of evaluating and predicting the employee performance has evolved over recent period. In any organisation, employees play a major role in the success or failure of the organisation. Today, every organisation is taking innumerable measures to improve the performance of their employees. They try to use brute force approach and expect an increase in performance of employees. This also leads them to false predictions and estimations of their success goals when it is concerned to productivity of employees. Their primary goal is to achieve their goals on time, by improving employee productivity and efficiency. There are many research going on for applying machine learning to solve the concerned problem. Many classification algorithm are worked upon to predict the future performance rating of a particular employee. There is not much information available on what parameters are actually responsible for this performance rating of the employee. This paper is about the machine learning methods which helped in finding the factors that are responsible for the performance rating of an employee and also predict their future performance rating based on certain conditions. These conditions are dependent on the work environment of an employee and there concerned priorities about work. Upon discovering the factors responsible, the research gives a study about recommendations to be given to an employee or the organisation on how they can improve the metrics and increase employee productivity.

Article History

Article Received: 19 August 2019

Revised: 27 November 2019

Accepted: 29 January 2019

Publication: 12 May 2020

Keywords: Employee performance, Pearson Correlation, Rating, Random Forest

1. Introduction

Each organisation works with employees and employees' performance rating is directly related to success criteria of an organisation. If employees are given enough attention to help them improve their performance, it can lead them to increase in their productivity and grow in their career. Only giving focus to training employees on the technical side of their work is not enough. This can improve their skills but not improve their productivity. So, various factors are responsible for affecting the performance of an employee. Some of the factors may include their work

location distance, work load, office environment, job satisfaction, higher authorities support, working hours, number of leaves and so on. Not all factors may equally contribute to an employee's performance. It differs from person to person and also the priorities related to them. These factors have to be carefully evaluated for each case, to be able to find out the root cause of an employee performance rating. If all employees are treated with same brute force approach then, programme carried out for increasing the employee performance productivity will fail and there will not be any benefit to the employee and

eventually no benefit to the organisation also. Therefore, it is very important to use right techniques at right place to find out the solution to a give problem.

Employee Performance is the most important part of the organisation’s success criteria. Machine Learning is an evolving field in the area of research about performance of an employee. It is supervised learning where the target variable is known as performance rating of an employee. There are various classification algorithms which can be applied to predict employee performance. Support Vector Machines, Random Forest Classifier, K- Nearest Neighbour Classification, Gradient Boost Classifier etc. are widely used algorithms for supervised machine learning approach towards Employee Performance Rating.

2. Review of Literature

Logistic Regression

Logistic Regression, a supervised learning approach[1], an algorithm that works on the principle of probability of occurrence of the classes such as yes/no, pass/ fail, better/ worse etc. It can also be more than two classes also, such as, in this case, 2, 3, or 4 rating for an employee’s performance rating.

Support Vector Machine

Vladimir N. Vapnik and Alexey Ya.Chervonenkis [2] in 1995 proposed support vector machine algorithm through the theory of support vector networks. This machine learning algorithm is a supervised learning model. It works on the principle of finding a hyperplane in an N dimensional vector i.e. the number of attributes in the dataset applied.

Bernoulli Naive Bayes

This machine learning algorithm is based on the Bayes’ theorem[3] which is a supervised machine learning model. It works on the assumption that there is independence with conditions among every pair of attributes in the dataset when provided with a target variable to predict. So, Bernoulli NB[4] derives conclusion through multivariate Bernoulli distributions.

Artificial Neural Network

Also known as “Connectionist systems” [7] is derived from the biological neural networks of human brains. It learns on its own from the conclusions of existing examples provided to it. For this study, MLP [8] i.e. Multilayer Perceptron is used which is a class of feed forward artificial neural networks.

Random Forest Classifier

A supervised machine learning model was first proposed byHo, Tin Kam (1995) [9]. This is again a

supervised machine learning model which is a CART (Classification and Regression Trees) based algorithm. Random Forest concept is derived from the Decision trees Algorithm which splits every possibility as a branch of decision statements. [10]Random forest is a group of randomly selected decision trees with different possibilities of categorisation to derive the target variable. It gives the combination with best possible accuracy to predict the target variable.

Bootstrap Aggregation (Bagging) in decision trees results are very prone to deviations in accuracy due to the data they are trained on. Random Forest algorithm uses this as its strength to randomly sample the data with replacements and form different trees to derive various results and accuracy to predict target variable.

3. Step By Step Machine Learning Approach For Employee Performance Analysis

Out of many popular machine learning algorithms, we are applying following algorithms to predict employee performance rating and also know various factors related to their performance.

Data Exploratory Analysis

Dataset used in the research is collected through daily surveys and form filling in different start-ups in Bangalore and one to one interaction with their employees. The dataset consists of 27 attributes and 1200 records. Basic information about the attributes of dataset are as follows:

Attribute	Data Type
EmpNumber	1200 non-null object
Age	1200 non-null int64
Gender	1200 non-null object
EducationBackground	1200 non-null object
MaritalStatus	1200 non-null object
EmpDepartment	1200 non-null object
EmpJobRole	1200 non-null object
BusinessTravelFrequency	1200 non-null object

DistanceFromHome	1200 int64	non-null
EmpEducationLevel	1200 int64	non-null
EmpEnvironmentSatisfaction	1200 int64	non-null
EmpHourlyRate	1200 int64	non-null
EmpJobInvolvement	1200 int64	non-null
EmpJobLevel	1200 int64	non-null
EmpJobSatisfaction	1200 int64	non-null
NumCompaniesWorked	1200 int64	non-null
OverTime	1200 object	non-null
EmpLastSalaryHikePercent	1200 int64	non-null
EmpRelationshipSatisfaction	1200 int64	non-null
TotalWorkExperienceInYears	1200 int64	non-null
TrainingTimesLastYear	1200 int64	non-null
EmpWorkLifeBalance	1200 int64	non-null
ExperienceYearsAtThisCompany	1200 int64	non-null
ExperienceYearsInCurrentRole	1200 int64	non-null
YearsSinceLastPromotion	1200 int64	non-null
YearsWithCurrManager	1200 int64	non-null
Attrition	1200 object	non-null
PerformanceRating	1200 int64	non-null

Department wise Performance Analysis

It is observed that an Employee's performance varies from department to department and it is essential to analyse the environmental conditions subject to each department.

So, department wise data exploratory analysis for employee performance is as follows:

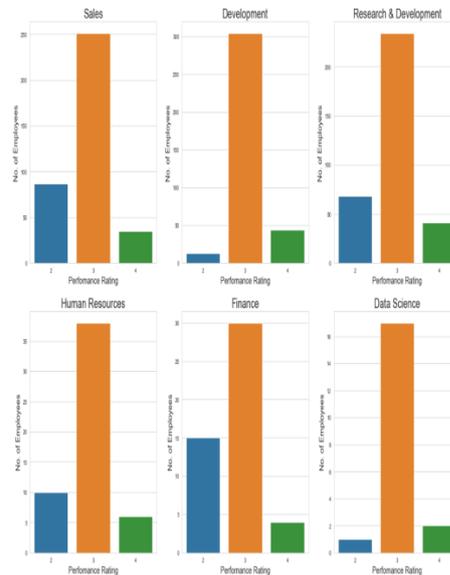


Figure 3.2.1: Department wise average performance rating

It is clear from the above graph that the average employee performance rating of "Development" department proves to be better than any other departments. This is because, it has higher number of employees with 3 and 4 rating as compared to other department employees.

4. Feature Selection

Next step is select the features which have largely contributed to the performance of an employee. For this there are two major methods being widely used, "Pearson Correlation Method" and "Feature Importance". In this study, Pearson Correlation method has given useful results for further analysis.

Pearson Correlation Heat Map

Pearson Correlation is basically a number in the ranger of 1 to -1 which shows the relation between two attributes. Product moment Correlation Coefficient is an alternative name to Pearson Correlation factor. Here, 0 represents no relation among two attributes. There exists a lower level of association among attributes if the coefficient is in range of .1 to .3 both positive and negative, medium association if it is in range of .3 to .5 and high if in range of .5 to 1.0 .

Best way to represent a Correlation Coefficient is visualization through a heatmap. The Correlation

Coefficient Heatmap for the following dataset is shown in below figure:

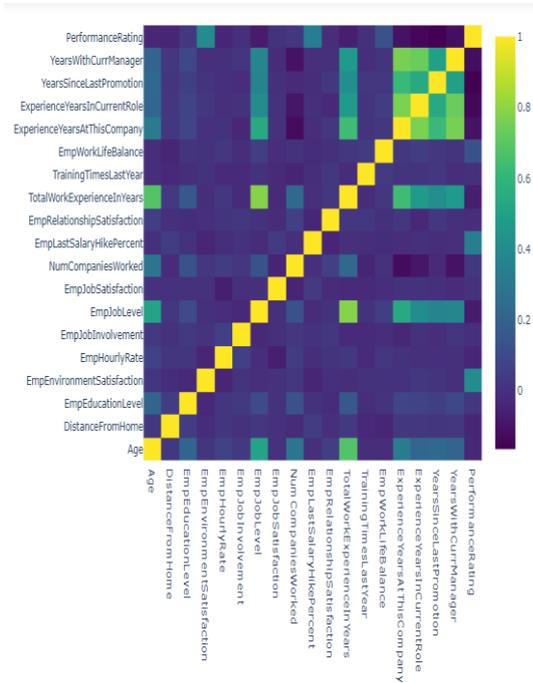


Figure 4.1.1: Correlation Coefficient Heatmap on Employee Dataset

Top 5 factors contributing to Employee Performance Rating:

The feature showing highest correlation with the target variable i.e. Employee performance rating

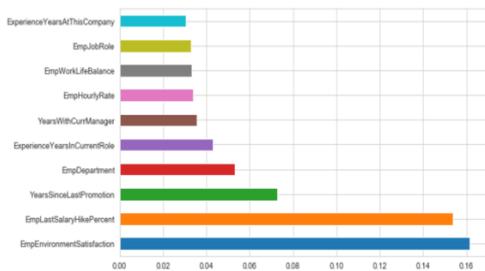


Figure 5.1: Top 5 factors contributing to employee performance rating

It is derived from the above studies that the top 5 factors which are affecting Employee Performance Rating are “Employee Environment Satisfaction”, “Years since Last Promotion”, “Employee Last Salary Hike Percent”, “Employee Department” and “Experience years in Current Role”. These factors majorly help an employee to perform better or vice versa in their respective roles.

Every organization should pay attention to improving these factors for improving their employees’ productivity rather than just focussing on

intensive training sessions for strengthening its employees’ skills.

5. Model Building

Next step is to evaluate an employee performance rating and predict an employee’s performance based on the important factors observed from the above study. The major machine learning algorithms used for this study are: Logistic Regression, Support Vector Machine, Random Forest Classifier, Naive Bayes Bernoulli and Artificial Neural Network.

Algorithms applied for predicting Employee Performance Rating:

Results for employee performance rating using Logistic Regression:

	precision	recall	f1-score
2	0.54	0.37	0.43
3	0.77	0.92	0.84
4	0.60	0.21	0.31

Results for employee performance rating using Support Vector Machine:

	precision	recall	f1-score
2	0.73	0.54	0.62
3	0.84	0.96	0.90
4	0.88	0.52	0.65

Results for employee performance rating using Bernoulli Naive Bayes:

	precision	recall	f1-score
2	0.00	0.00	0.00
3	0.71	1.00	0.83
4	0.00	0.00	0.00

Results for employee performance rating using Artificial Neural Network:

	precision	recall	f1-score
2	0.71	0.73	0.72
3	0.88	0.91	0.90
4	0.73	0.55	0.63

Results for employee performance rating using Artificial Neural Network:

	precision	recall	f1-score
2	0.90	0.88	0.89
3	0.95	0.97	0.96
4	0.96	0.90	0.93

6. Prediction of Employee Performance - A Comparative Study

In the above study, Random Forest Classifier gives the highest accuracy in the prediction of performance rating of an employee.

Algorithm	Accuracy (in %)
Logistic regression	73.75%
SVM	83.33%
Naïve Bayes	70.83%
ANN	83.75%
Random Forest	94.58%

7. Summary

Employee Performance is a crucial factor responsible for an organization's growth. Organizations should focus on improving employees' productivity by increasing their job satisfaction level, providing them timely promotions or peer recognition in their respective job roles, etc. Just training their employees to improve the skills is not enough to show change in employee productivity.

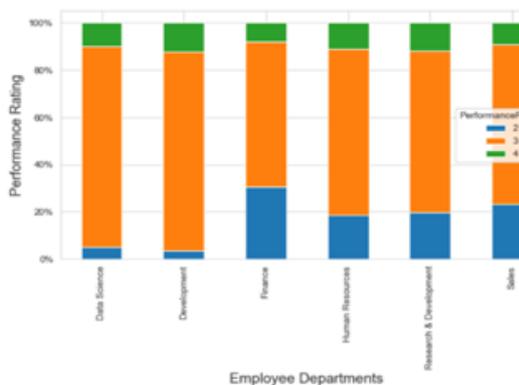


Figure 7.1: Department wise performance rating

In this study, the major five factors which were responsible for affecting employee performance are found. They are as follows:
Employee Environment Satisfaction
Years since Last Promotion

Employee Last Salary Hike Percent
Employee Department
Experience years in Current Role

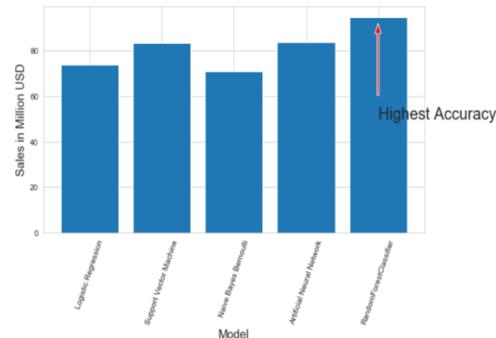


Figure 7.1.1: Performance of various algorithms in predicting employee performance rating

Prediction of Employee Performance Rating is another important aspect for human resource management department of an organization. This paper has a comparative study and analysis of various classification algorithms for prediction of employee performance rating.

Random Forest Classifier proves to give the best accuracy among all other classifiers in predicting the employee performance rating.

8. References

- [1] Tolles, Juliana; Meurer, William J (2016). "Logistic Regression Relating Patient Characteristics to Outcomes". *JAMA*. 316 (5): 533–4
- [2] Corinna Cortes, Vladimir Vapnik. "Support Vector Networks". In *Machine Learning*, 20, 273-297 (1995) © 1995 Kluwer Academic Publishers, Boston. Manufactured in The Netherlands.
- [3] McCallum, Andrew. "Graphical Models, Lecture2: Bayesian Network Representation" (PDF). Retrieved 22 October 2019.
- [4] Maron, M. E. (1961). "Automatic Indexing: An Experimental Inquiry" (PDF). *Journal of the ACM*. 8 (3): 404–417.
- [5] Chen, Yung-Yao; Lin, Yu-Hsiu; Kung, Chia-Ching; Chung, Ming-Han; Yen, I.-Hsuan (2019). "Design and Implementation of Cloud Analytics- assisted Smart Power Meters Considering Advanced Artificial Intelligence as Edge Analytics in Demand-Side Management for Smart Homes". *Sensors*. 19 (9): 2047.
- [6] Bottaci, Leonardo (1997). "Artificial Neural Networks Applied to Outcome Prediction for Colorectal Cancer Patients in Separate Institutions". *Lancet (London, England)*. *The Lancet*. 350(9076): 469–72.

- [7] Hastie, Trevor. Tibshirani, Robert. Friedman, Jerome. "The Elements of Statistical Learning: Data Mining, Inference, and Prediction". Springer, New York, NY, 2009.
- [8] Rosenblatt, Frank. x. "Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms". Spartan Books, Washington DC, 1961.
- [9] Ho, Tin Kam (1995). Random Decision Forests (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995
- [10] Ho TK (1998). "The Random Subspace Method for Constructing Decision Forests" (PDF). IEEE Transactions on Pattern Analysis and Machine Intelligence. 20 (8): 832–8