

Employee Performance Prediction using Different Supervised Classifiers

Merry Grace T. Li, Macrina Lazo, Ariel Kelly Balan and Joel de Goma

School of Information Technology

Mapua University

Makati, Philippines

mgtli@mymail.mapua.edu.ph, mplazo@mymailmapuaedu.onmicrosoft.com,

akdbalan@mapua.edu.ph, jcdegoma@mapua.edu.ph

Abstract

Employee performance evaluation intends to measure the commitment of everyone in the company. Predicting employee performance is essential for company's success. This paper presented employee performance prediction in a company using machine learning. The researcher follows Cross-industry standard process for data mining (CRISP-DM). Logistic Regression, Decision Tree, and Naïve Bayes classification method are used to develop the prediction model. The result shows that Logistic Regression has higher accuracy with the other two classifier used.

Keywords

Logistic Regression, Employee Performance, and Data Mining

1. Introduction

Employee performance evaluation is designed to assess everyone's contribution to the organization. The performance of individuals against organizational goals determines whether the organization meets its goals. The basic objectives of performance evaluations are two-fold: firstly, to reward employees for meeting organizational objectives and secondly to identify which objectives are not met and to develop action plans to ensure they are achieved in the future (Islam and Rasad, 2006). It also helps remind workers what their managers expect in the workplace. They provide employers with information to use when making employment decisions, such as promotions, pay raises, and layoffs.

Moreover, the purpose of any performance review is to be sure employees know how they can develop their talents and energies and use them more effectively to contribute to the organization's success. Also, for the employees to grow and contribute effectively requires a good match between the employee's talents and energies and the demands of the position (Latham, 2016).

Company A is one of the world's oldest publishing house and the second-largest university press in the world. It is the publishing press for its university.

The operation in the Philippines started in 2003. The business is mainly technology hub (Software Development, QA, Infrastructure, and Service Management) among its 50+ offices worldwide. The company is not doing service for any local entity.

To sustain this company's ongoing business, they want to determine the characteristics of their performing employees and help the underperforming employees to be better.

To help the Company, the researchers wants to analyze the performance of their internal operations. It will give some insights how the employees being monitored and evaluated on their work.

The objective of the study is to determine the characteristics of the excellent and good performing employees. Also, to create a model that will predict employees who will perform well.

2. Related Studies

2.1 Employee Performance Evaluation

Employee Performance is the successful completion of tasks by a selected individual or individuals, asset and measured by a supervisor or organization, to pre-defined acceptable standards while efficiently and effectively utilizing available resources within a changing environment (Thao and Hwang, 2015). They are often directly connected with employee compensation and promotions. Every employee makes an individual contribution to the performance of the organizational unit and thus to the entire organization. If the performance of an organizational unit or individual employee falls behind for too long, this will lead to reorganization or redundancy (Noordzij, 2017). Moreover, these are crucial to the success of the company, as well as to the individual success of employees because consistent constructive feedback allows employees to see an unbiased opinion of their performance and then decide how to improve (Greene, 2018).

2.2 Factors of Employees Performance

Many variables affect the performance of employees in the workplace. These variables include leadership, motivation, and training (Thao and Hwang, 2015).

Leadership is a process whereby an individual influences a group of individuals to achieve common goals (Northouse, 2007). They are the key people for employee evaluation. They should be able to monitor and evaluate them properly.

Motivation is a key determinant of job performance and a poorly motivating force will be costly in terms of excessive staff turnover, higher expenses, negative morale, and increased use of managements' time (Jobber and Lee, 1994). Therefore, supervisors must know how to motivate their staff so that there will be no dissatisfaction among employees. According to Denton (Denton, 1991), a motivated workforce will lead to greater understanding, acceptance, commitment to implementation, understanding of objectives, and decision making between management and employees.

Training is a type of activity that is planned, systematic and it results in an enhanced level of skill, knowledge, and competency that is necessary to perform work effectively (Lappan, 1995). This has been proven to increase employee's performance.

2.3 High Performing Employees

High performing employees are the biggest asset to the company. High performers stand out from average performers in any organization. They consistently exceed expectations and are management's go-to people for difficult projects because they have a track record of getting the job done. They're great at their job and take pride in their accomplishments but may not have the potential (or the desire) to succeed in a higher-level role or to tackle more advanced work (Westfall, 2020).

2.4 Promotion

Promotion is the vertical movement of an employee within the organization. In other words, promotion refers to the upward movement of an employee from one job to another higher one, with an increase in salary, status, and responsibilities. Promotion may be temporary or permanent, depending upon the needs of the organization. The promotion has an in-built motivational value as it elevates the authority, power, and status of an employee within an organization. It is considered good personnel policy to fill vacancies in a higher job through promotions from within because such promotions provide an inducement and motivation to the employees and remove feelings of stagnation-and frustration (Westfall, 2020). For pages other than the first page, start at the top of the page, and continue in double-column format. The two columns on the last page should be as close to equal length as possible.

2.5 Promotion

A paper (Jayadi, 2019) used Naïve Bayes Classification for employee performance prediction in a company. The research followed Cross-industry standard process for data mining (CRIPS-DM). The result of their study has an accuracy as high as 95.48 percent.

A study (Li et al, 2016) proposed to use an enhanced K-nearest neighbor (KNN) algorithm to cope with problems encountered in human performance prediction. The study focuses in three aspects, which are the neighboring distance calculation based on entropy, the classification determination strategy, and the quantitative description method of human performance. The improved KNN algorithm was proved better compared to other researched that uses KNN algorithm with different features.

A research (Sarker et al, 2018) used data clustering method for evaluating the employee's performance and decision making process. The proposed method will predict number of employee that will be promoted and discharged according to their performance.

A research paper (Anu et al, 2019) used Support Vector Machines, Random Forest, Naive Bayes, Neural Networks and Logistic Regression to create a model that provides insights about the performance and commitment of employees. The employees are classified into 3 classes indicating the level of their performance. Based on the result, Support Vector Machines has a better accuracy compared to the other classifiers.

Another research (Xue et al, 2019) aims to improve prediction performance. The researchers proposed a novel framework for personnel performance prediction. Their approach is combination of hybrid convolutional recurrent neural network (CRNN) model and KNN. The experimental results shows that the proposed approach has a better result compared to existing methods.

3. Methodology

The researcher used CRISP-DM (Cross-Industry Standard Process for Data Mining), a cross-industry process for data mining, provides a structured approach to planning a data mining project. It is a robust and well-proven methodology, to discuss the analysis (Jayadi, 2019).

3.1 Promotion

To understand the business objectives and requirements first, the researcher had a one-on-one interview with the HR Manager of the Company from the business perspective, and then converting this knowledge into a data mining problem definition and designed a preliminary plan to achieve it.

Second, the researcher went into the Company website to see what kind of business they are handling, what kind of services they give, and the culture that they want to impart to their clients. This has given the researcher an overview of the company.

Thirdly, the researcher studied and researched performance evaluation and promotion, how this is being measured and monitored, and its drivers.

3.2 Data Understanding

The data was extracted internally from the HRIS Team of the Company. The data given was in Excel format and is composed of structured data, and the definitions have been discussed and validated by the H.R. Manager. The date when the data extracts as of September 2019. The extracted data composed of 3 datasets. Table 1 shows sample features considered in developing the model

Table 1. Sample Features

	Feature	Count	Data Type
1	Age Group	982	String
2	Gender	982	Character
3	Tenure Group	982	String
4	Marital Status	982	String
5	Company	982	String
6	Department	982	String
7	Compensation Grade Profile	982	String
8	State	982	Character
9	Promoted in the last 3 years?	982	Boolean
10	Business Unit	982	Character
11	Training 2018	982	Boolean
12	Performance 2015	982	Character
13	Performance 2016	982	Character
14	Performance 2017	982	Character
15	Performance 2018	982	Character
17	Training Title	982	String
18	Service Provider	982	String
19	Type	982	String
20	Program Cost/Head	982	Integer
21	Employee ID	982	Integer

Employee Dataset

This dataset contains a full list of all employees, including active and resigned employees, along with their company details. It is composed of 18 fields with 858 observations. The fields are Employee ID, Age, Gender, Date of Birth, Hire Date, Marital Status, Business Title, Company, Department, Cost Center, Rating - Most Recent, Compensation Grade Profile, Active Flag, Termination Date - All, Termination Reason, Regrettable, Address City and Address State.

Historical Performance Review

This dataset contains the rating of employees from the years 2015 to 2019. It is composed of 8 fields with 524 observations. The fields are Employee ID, FY15 EOY Rating, FY16 EOY Rating, FY17 EOY Rating, FY18 EOY Rating, FY19 EOY Rating, and Promoted in the last three years and Business Unit.

Trainings 2018

This dataset contains the training detail that the employee has availed for 2018. It is composed of 28 fields, with 794 observations. The fields are W.D. Number, Business Title, Gender, Hay Band, Still Active?, date Hired, Tenure, Company, Division, Cost Center, Department, Training Code, Training Title, Service Provider, Type, Program

Cost/Head, Incidentals/Head, Total Cost/Head, Venue, Line Manager, Position, Sponsorship, Present, Training Date, End Date, Number of Days, Bond Period (Months) and Bond End.

The unit of analysis for the study is the employees of the Philippine business of the Company as of September 20, 2019. There is a total of 858 unique employees identified through the employee dataset.

The target variable for predicting the characteristics of excellent and good performing employees is an excellent or good rating in employee performance

3.3 Data Integration

Data Diagnosis was done by checking the incorrect values, inconsistent data format, outliers, missing values, high cardinality, duplicate records, and highly correlated variables. Data Cleansing and Transformation was made by having a primary key which is the Employee ID and created calculated fields such as tenure group from the hire date, created bins for age, and made a uniform tagging for the performance evaluation.

3.4 Modelling

The researcher used Supervised Classification and Predictive Modelling and White Box Modelling Technique for predicting the characteristics of Excellent and Good performing employees, the proposed method is Two-Class Logistic Regression, it is a well-known method in statistics that is used to predict the probability of an outcome and is especially popular for classification tasks. The algorithm predicts the probability of occurrence of an event by fitting data to a logistic function. In Logistic Regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.) (Jayadi, 2019).

3.5 Model Evaluation

The metrics used by the researcher are accuracy, precision, recall, F1 score, and AUC.

4. Results and Discussion

The data set used from the original dataset, which eliminates inactive workers from the dataset. The target variable used is Excellent and Good in the Performance Rating, and predictive variables are Gender, Tenure Group, Marital Status, Business Title, Department, Compensation Grade Profile, Address City, Program Cost/Head, Service Provider, Training Title and Type.

The researcher compared the results using Microsoft Azure Machine Learning to combine the results. As shown in Figure 1, the models of the Logistic Regression, Decision Tree, and Naïve Bayes gave somewhat similar results, but the Two-Class Logistic Regression model was the best when it came to precision.

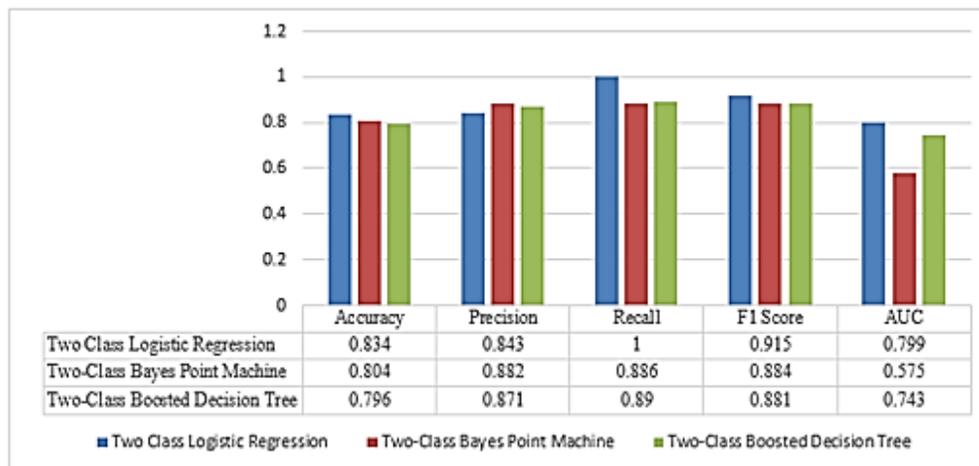


Figure 1. Validation Result

Permutation Feature Importance used to identify the right variables. These scores help to identify the best features used in a model. The trait values are shuffled randomly, one column at a time, and the model's performance is measured before and after. The scores returned by the module, after permutation, represent the change in a trained model's performance. Significant features are typically more prone to the shuffling process, resulting in higher significance ratings.

The researcher used Classification - Average Log Loss as the metric to measure the performance of the feature, and the top three resulting variables is shown in Table 2.

Table 2. Top three features

Feature	Score
Compensation Grade Profile	0.051082
Tenure Group	0.043811
Department	0.016671

The objective of the project is to create a platform that will aid credit card sales agents to accomplish their business objectives using the count of customers as their primary metric. The researcher determined first the current workflow of the financial institution. The current workflow will guide the researcher on how to create the data visualization.

5. Conclusion

The researcher concluded that the profiles of Excellent and Good performing employees are based from Compensation Grade Profile, Tenure Group and Service Department. Among the classifier, the two Class Logistic Regression has a better result. To further improve the model, additional features should be included.

References

- Islam, R. and Rasad, S.M., Employee performance evaluation by the AHP: A case study. *Asia Pacific Management Review*, pp 163-176, 2006.
- Latham, A., Why Your Employee Performance Ratings Are Hurting Your Organization, Available: <https://www.forbes.com/sites/annlatham/2016/02/07/why-your-employee-performance-ratings-are-hurting-your-organization/>, 2016
- Thao, L.T., and Hwang, C.J., Factors Affecting Employee Performance– Evidence from Petrovietnam Engineering Consultancy J.S.C, 2015.
- Noordzij, L., What is employee performance? Retrieved, Available: <https://www.effactory.com/knowledge/themes/what-is-employee-performance>, July 4, 2017.
- Greene, L., Seven tips for effective employee performance evaluation, Available: <https://www.trainingjournal.com/articles/features/seven-tips-effective-employee-performance-evaluation>, 2018.
- Northouse, G., *Leadership Theory and Practice*, 3rd Edition, Thousand Oaks, CA: Sage Publications, 2007.
- Jobber, D. and Lee, R., A comparison of the perceptions of sales management and salespeople towards sales force motivation and demotivation, *Journal of Marketing Management*, pp. 325-332, DOI: 10.1080/0267257X.1994.9964278, 1994.
- Denton, D. K., *Horizontal management: Beyond total customer satisfaction*. Oxford: Maxwell Macmillan, 1991.
- Lappan, S., Book Reviews : Gordon McBeath (1992) *The Handbook of Human Resource Planning* Oxford: Blackwell Business Publishers, 1995.
- Westfall, B., High Potentials Vs. High Performers: A Manager'S Guide, Available: <https://www.softwareadvice.com/resources/high-potentials-vs-high-performers-a-managers-guide>, 2020.
- Jayadi, R., Employee Performance Prediction using Naïve Bayes. *International Journal of Advanced Trends in Computer Science and Engineering*, pp 3031-3035. 10.30534/ijatcse/2019/59862019, 2019.
- Li, N., Kong, H., Ma, Y., Gong, G., and Huai, W., Human performance modeling for manufacturing based on an improved KNN algorithm, *The International Journal of Advanced Manufacturing Technology*, DOI:10.1007/s00170-016-8418-6, 2016.

- Sarker, A., Shamim, S. M., Zaman, S., and Rahman, M., Employee's Performance Analysis and Prediction Using K-means Clustering & Decision Tree Algorithm, *Global Journal of Computer Science and Technology*, 2018.
- Lather, A. S., Malhotra, R., Saloni, P., Singh, P., and Mittal, S., Prediction of employee performance using machine learning techniques, Proceedings of the International Conference on Advanced Information Science and System, DOI=<https://doi.org/10.1145/3373477.3373696>, 2019.
- Xue, X., Feng, J., Gao, Y., Liu, M., Zhang, W., Sun, X., Zhao, A., and Guo, S., Convolutional Recurrent Neural Networks with a Self-Attention Mechanism for Personnel Performance Prediction, *Entropy*, 2019.

Biography

Merry Grace Li is a Master's student at Mapua University

Macrina Lazo is a part-time instructor at School of Information, Mapua University

Ariel Kelly Balan is the Dean of School of Information Technology, Mapua University

Joel de Goma is a full-time faculty member at School of Information, Mapua University