

Employee Performance Appraisal For Salary Hike using Decision Tree

¹V. Venkata Sai Nitesh, ²Roseline Villuri, ³Surya Karteek Reddy, ⁴R. John Wesley, ⁵G.V.Gayathri
^{1,2,3,4}Student, ⁵Assistant Professor, ANITS, Visakhapatnam, India.

¹niteshvirothi19@gmail.com, ²roselinerosie@gmail.com, ³suryakarteek1998@gmail.com,
⁴johnwesleyrachaprolu45@gmail.com, ⁵gayathri.ganivada@gmail.com

Abstract - The progress and productivity of an organisation depends on the capability of the employees. In a company the critical decision for managers is to assess performance of an employee. Employees are the key of an organisation so, finding the better one is always a hectic task. Capability of the employee depends on potential of the employee, the performance of the workers rely on personality traits. This system will evaluate each employees performance depending on attributes mentioned by company and then after evaluating, these results are used in calculation of salary hike of each employee. So here in order to evaluate the performance and calculate hike we use data mining concept clustering.

Keywords — Employee Performance, Hike Calculation, Decision tree, Data Mining, Organization, Salary Hike.

I. INTRODUCTION

An organization growth depends on the capabilities of their employees. Performance appraisal must be as an intrinsic part of manager's responsibilities and should not be a time consuming and in addition to that this process has to improve performance and effectiveness. In order to do this we need huge amount of data and processing. So, what is performance appraisal, it means evaluating the behaviour of the employees, including quantitative and qualitative aspects of job performance. As already mentioned here clustering concept is used to evaluate the performance. So how do we assess their abilities we need some characteristics or attributes to decide. So here are few attributes that can used to evaluate the performance.

- a. Quality: How well the employee is fulfilling the task correctly and thoroughly assigned to him.
- b. Productivity: How well employee produce a significant volume of work in a specified period of time.
- c. Job Knowledge: To what extent employee displays understanding of the work, equipment, process required to perform a task and what practical and technical knowledge he has for the job.
- d. Attendance: It is an important attribute as it results in punctuality their willingness to work extra time when required.
- e. Cooperation: How well is he communicating with his subordinates, co-workers, supervisors in positive manner. Being responsible for his job assignments and

performance.

- f. Creativity: How imaginative and innovative he is in his work. Suggesting new ideas, proposing new work methods to eradicate waste and find better ways in doing things.
- g. Lead: The judgement and proper decision making skills while directing others. Directing work in an effective way to meet required goals.
- h. Communication: How well is he interacting with staff in an intelligent way.
- i. Character: Accepting criticism politely and being good with co-workers.
- j. Teamwork: Not only completing his own task working in a team will always be very important for organisation growth.

These are some of the important attributes that are vital during evaluating an employee performance. These attributes are given some values around 0 to 100 to calculate hike.

II. LITERATURE SURVEY

There has been many research carried in the fields of this data mining for employee performance evaluation, researchers used many kinds of algorithms for calculating the performance of an employee. [2] In this paper they used ID3 algorithm for evaluating the performance of an employee where initially they take the data of an organization with employees and then perform data pre-processing to that data then followed by algorithm where

primarily they calculate total entropy for the given set followed by calculating gain i.e. Information gain for each data attribute. The attribute with highest gain will be considered as root node then again gain and entropy are calculated for all attributes. In this way a tree is formed, each level is given a priority or value by which performance of an employee can be calculated. [5] In this system performance is measured with respect to a factor. They use cognitive maps and analytical hierarchy to indicate factors. The criteria taken here are employees attitude, clinical skills and communication skills where the criteria used is excellent as 4, good as 3, fair as 2 and poor as 1. [7] In this system they considered previous year database where they used decision tree to predict the performance of an employee. This system as they used previous year database they construct the tree for the database present they use ID3 algorithm where gain and entropy are calculated for each attribute now after obtaining the required tree an analysis will be made where root node is given as higher priority followed by other nodes now based on this performance is calculated which is used for predicting performance of an employee based on their previous year database. [4] Here the process is quite different compared to previous systems where they used time factor for evaluating it is distinguished according to time giving importance on past targeted at past occurrence where focusing on present that calculate present situation focusing on future. They use method where they focus on past event with the work that is already done. The drawback of this system is result modification. [9] Here this system uses decision tree for classifying data into variable types of classes that is used for prediction purpose here the task is performed by the admin where he adds data set for evaluating entropy and information gain just like other systems but it is used in predicting performance of employees accurately. [8] This system is also similar to the above mentioned systems where they used ID3 algorithm with gain ratio here it is used to predict student's performance in an higher institute in order to improve quality of teaching in those institution. By this process we will come to know the requirements of students and so we can fulfil them for best results here they also used WEKA tool for J48 and Naïve Bayes algorithm for analysis purpose finally these results are compared and represented

III. PROPOSED ALGORITHM

There are four phases

- 3.1 Data Evaluating and Pre-processing.
- 3.2 Algorithm Selection.
- 3.3 Entropy and Information Gain calculation.
- 3.4 Hike calculation.

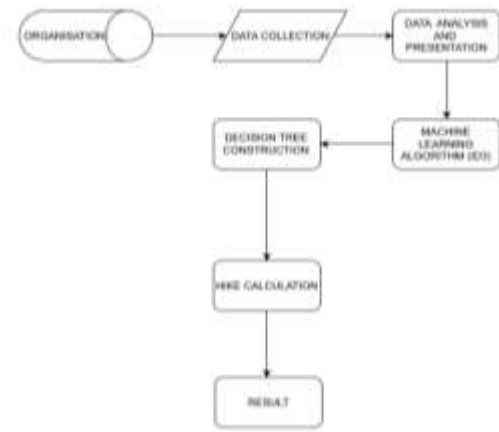


Fig. 1 Architecture

The architecture flow is represented for the given input datafile, for each data of an employee data pre-processing is performed then data evaluation then presentation followed by machine learning algorithm. The architecture flow is represented in the above figure. For the given input datafile, for each dataset data pre-processing is performed then data evaluation followed by presentation finally apply machine learning algorithm i.e., ID3 to calculate entropy and information gain for all the attributes where a decision tree is constructed then a root node will be assigned, now the attributes are given some specified values through which the hike of each employee is going to be calculated.

3.1 Data Evaluating and Pre-Processing –

Finding the right and proper data is always important, taking some random data is not advisable. Once the required data is obtained, processing that data is really important. It involves going through series of steps

- Data Cleaning: Data has to be cleansed like filling with missing values, smoothening the noisy data.
- Data Integration: Data with various representations are placed together and any issues within tare resolved.
- Data Transformation: Data here is generalized, normalized and aggregated.
- Data Reduction: It represents a reduced visualization of input data in data warehouse.
- Data Discretization: Reducing values of continuous attribute.

Now after going through these steps the data is evaluated and pre-processed.

3.2 Algorithm selection –

In order to evaluate this process, there are many algorithms and concepts that can be used such as decision tree and clustering. So, before selecting or choosing a specific algorithm, consider several characteristics such as why only this algorithm is being used, what are its advantages

over other algorithms, its efficiency and many more. The main objective is to find the right algorithm to the project. There are numerous algorithms that are available; choosing the apt one is a hectic task. The algorithm that is used is the decision tree ID3 algorithm, where calculation of the entropy and information gain to all the attributes that are present in the given dataset are performed.

3.3 Entropy and Information Gain calculation –

To decide or evaluate the performance of an employee, there are many attributes to be considered where calculation takes place from the resultant values i.e., 1 to 10 for each and every attribute. Then calculation of overall score is obtained.

To find this evaluation process, there are many algorithms like decision tree ID3 (Iterative Dichotomiser) algorithm, fuzzy algorithm.

Decision tree algorithm is used for accurate results, where this algorithm iteratively divides attributes into two groups which are the foremost governing attribute and remaining to construct a tree. Where we transform the data that is raw, to rule-based decision trees. ID3 is the most commonest decision tree algorithm. So, this begins with the first set S as root node and calculate entropy for whole dataset. The dataset is then split on the various attributes. The entropy for each and every branch is then calculated.

Then it's added which results the total entropy. For containing that attribute, make a decision tree node. Recurse on subsets using remaining attributes. The resulting entropy is subtracted from the entropy before the split.

The resultant is information gain or decrease in the entropy. The attribute which yields the larger information gain is resulted as the decision node.

$$\text{Entropy (Decision)} = -P(X) \cdot \log_2 P(X) - P(X1) \cdot \log_2 P(X1)$$

Then calculate the information gain for every attribute,

$$\text{Gain (Decision)} = \text{Entropy(Decision)} - \sum [P(\text{decision}/Y) \cdot \text{Entropy}(\text{Decision}/Y)]$$

Where calculate 'Y' attribute, now apply the same process for remaining other attributes.

Now the attribute that has highest information gain is chosen as root node. Then, entropy and gain scores are calculated again for the remaining attributes. Therefore, the next most important attribute is found. Lastly, this process continues till reaching a decision for that particular branch node. That is why, it's known as Iterative Dichotomiser.

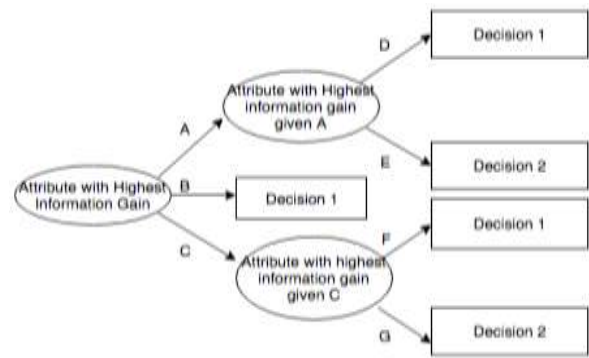


Fig 3.3 – Decision tree

3.4 Hike Calculation –

Hiking the salary of the employees, not only increases their salary but also results in increasing organisations income, so usually hiking the salary for the employees differs from organisation to organisation, where one might roll in the pay every 2 to 5 years, some might do it based on employee performance regularly. These labels depend on the organisation; some may consider 5 labels like poor, low, average, good, excellent. So, this is how the labels are assigned to each employee within the company that helps in differentiating the better employees within the company. A percentage hike is assigned. Now the present salaries of the employees are extracted from the dataset we've and a percentage salary hike that's applied to each employee and a new dataset is made with the updated salaries, this will be the final result.

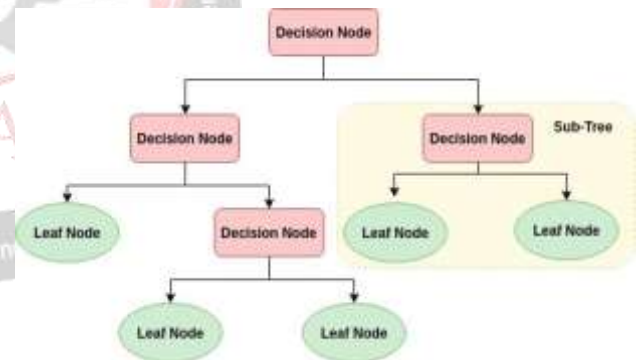


Fig 3.4- Final decision tree

IV. EXPERIMENT AND RESULT

To perform the experiment a dataset is required, the below is that dataset which can be used to perform the above-mentioned phases

MarketID	MarketStatus	GenderID	EmpStatusID	DeptID	FrontOffice	PayRate	Tenure	PositionID	ManagerID	EmpSalaries	SpecialProj	PerfScoreID
1	1	0	1	1	1	285	0	1	1	2	0	3
0	2	1	1	1	0	23	0	1	1	4	4	3
0	0	1	1	1	0	29	0	1	1	5	5	3
1	1	0	1	1	0	215	1	2	1	3	4	3
0	0	0	1	1	0	1850	0	3	1	3	5	3
1	1	0	5	1	1	205	1	2	1	4	4	3
1	1	0	5	0	0	95	1	3	17	5	0	3
0	0	0	1	6	0	55	0	3	17	5	0	3
0	0	0	1	0	0	55	0	3	17	1	0	1
1	1	1	1	0	0	50	0	3	17	5	0	3
0	0	1	1	0	1	955	0	3	17	5	0	3
0	4	0	4	0	0	55	1	3	17	4	0	3
1	1	1	1	0	0	955	0	3	17	3	0	3
0	0	1	5	0	0	55	1	3	17	5	0	3
0	3	1	1	0	0	55	0	3	17	3	0	3
0	0	1	1	0	0	55	0	3	17	3	0	1

Fig: 4

For the above dataset, the algorithm ID3 is performed, where every employee attributes are evaluated and the best among them is resulted below are some of the categories, for which all employees are categorized.



Fig: 4.1

4.1 Best Employee - As mentioned there are four categories of employees where 4 stands for best employee. Here is the representation of best employee result.



Fig: 4.2

4.2 Better Employee - The number 3 means employee who meets all the criteria and is regarded as an asset to the company.



Fig: 4.3

4.3 Moderate Employee - The number 2 means the employee performance is not up to the mark and need to improve his overall performance rate in order to avoid attrition which means they are likely to leave the company or get fired with this kind of application.

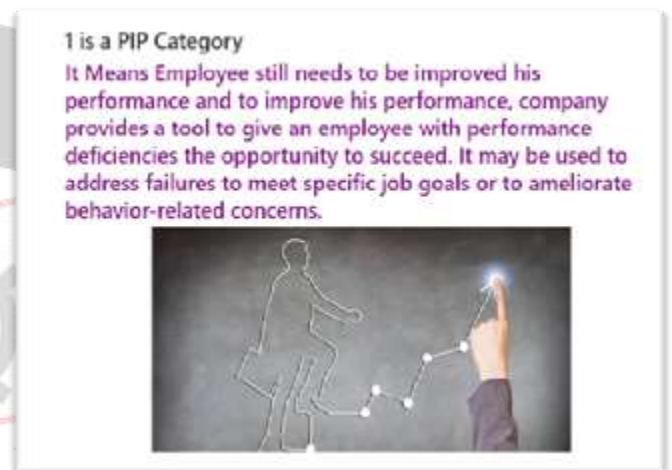


Fig: 4.4

4.4 Worst Employee - This is the least or worst performance of an employee where this kind employees are burden to the organization and they need to improve their performance and also need to work on their goals in order to stay in the company.

4.5 Comparative analysis

After experimenting with various algorithms like support vector machine, logistic regression, random forest tree and decision tree. After being using all the above mentioned algorithms it was found that decision tree accuracy is comparatively higher than other algorithms. Before starting the experiment each algorithm is used in order to find the best one starting with logistic regression where the accuracy is pretty good with 82% for given input and training sets. Then the other algorithm used that is support vector machine which is again performed really well with 80% of accuracy rate then random forest tree is implemented which performed quite accurately with 80%, finally decision tree is implemented which gave nearly an

accuracy of 94% which is highest compared to other all algorithms used.

SNO.	ALGORITHM	ACCURACY
1	LOGISTIC REGRESSION	82%
2	SUPPORT VECTOR MACHINE	80%
3	RANDOM FOREST TREE	80%
4	DECISION TREE	94%

V. CONCLUSION AND FUTURE SCOPE

This employee performance system helps in better understanding of employee responsibilities and their roles within the organization and also a review system for an employee job performance and their total contribution to a corporation. This has to be a positive experience and contributes to entire welfare of the company. This performance evaluation tool results improving productivity and performance for development of employees in the company, it also helps in doing best, motivating and uplift in confidence of their employees. It evaluates and also assess effectiveness and productivity of an employee to serve administrative and developmental use. Generally, this provides best productive feedback to identify the areas of advancement and to create a plan with higher officials like managers to enhance the skills, this motivates the workers if they are provided with quality-based compensation. This technique acts as a motivation for employees to extend and enhance their productivity further helps in advancement of their career, this also results in achieving their goals. Considering many aspects in evaluating the performance like giving adequate feedback to each and every employee based on their performance then by change in their behavior and their attitude towards their habits then by giving data to managers to make chances wherever required. All the above changes and results help a company increase in their revenue, Therefore, better the employees in a company better the organization, having best employees is always a greater sign. This system also helps in recruiting like their best employee qualities can be cross checked with newer employees which is going to have a greater impact in the coming years.

REFERENCES

- [1] Rohit Kotalwar, Raj Chavan, Sunny Gandhi, Vinayak Parmar, Data Mining: Evaluating Performance of Employee's using Classification Algorithm Based on Decision Tree
- [2] Nishant Mathur, Sumit Kumar, Santosh Kumar and Rajni Jindal "Employee Performance Evaluation System Using ID3 Algorithm", 2012.
- [3] Mathis, Jackson, "Classification of employee performance appraisal methods according to foreign authors", 2011.
- [4] Chen Xiaofan, Dvořáková, "Classification of employee performance appraisal methods based on the time factor." [J]. Nanchang Hangkong University, 2012.
- [5] Dr Lakshmi Rajamani, Mohd Mahmood Ali[J], "Quantitative Models for Performance Measurement Systems". University College of Engineering, Osmania University, Hyderabad.
- [6] Kalpesh Adhatrao, Aditya Gaykar, Amiraj Dhawan, Rohit Jha and Vipul Honrao "Performance evaluation using ID3 and C4.5 classification algorithm" 2013.
- [7] Vinayak Parmar, Rohit Kotalwar, Raj Chavan, Sunny Gandhi, "Performance Evaluation on previous year database". Department of CTPG, Kongu Engineering College, Perudurai, Tamilnadu, India.
- [8] Ramanathan, Saksham Dhanda, Suresh Kumar(2016), "ID3 algorithm with gain ratio", 2016.
- [9] Shubham tupe, ChetanMahajan, Dnyaneshwar Uplenchwar, PrathikDeo "Decision tree for classifying data into variable types of classes, 2017.
- [10] Mashaal A. Al-Barrak, Muna Al-Razgan "Predicting Students Final GPA Using Decision Trees: A Case Study", International Journal of Information and Education Technology, Vol. 6, No. 7, July 2016.
- [11] Dipak V. Patil, R. S. Bichkar, "Issues in Optimization of Decision Tree Learning", International Journal of Applied Information Systems 3(5):13-29, July 2012.