

# Performance Evaluation For Employee Selection Using Data Mining Techniques

Nikita Suhas Jangam  
C.K.Thakur College Mumbai,India  
nikitajangam23@gmail.com

Prarthana Pramod Gawand  
C.K.Thakur College Mumbai,India  
prarthana8797@gmail.com

Prof.Isha Sachin Thakre  
C.K.Thakur College Mumbai,India

**Abstract:** Worker determination is the layout by which a business assesses data about the pool of candidates originated during the enlistment stage. In the final of weigh the competitors, the organization chooses which candidate should be offered the position. Human capital is of a high worry for organizations; administration where their most intrigue is in contracting the profoundly qualified staff which are relied upon to perform exceptionally also. Several classification association rules were generated for selection process. This leads towards the congregation of best worker through productive representative determination. Concerning the increasing skill demands of the change in times, the candidates can be filtered by using various data mining techniques. In this research paper, data mining techniques were utilized to build a classification model to select the suitable employees.

*Key terms – Performance Evaluation, Data Mining, Classification, Association Rules.*

## I. INTRODUCTION

Nowadays, in a numerous organizations representatives are choose based on reference, on account of this the skilled candidates didn't get the opportunity and subsequently it causes the exhibition of the company. To maintain the strategic distance from this issue candidates can be selected using the various algorithms.

The aim of this paper is to collect the data candidates and investigate the relationship between the criteria for selecting the transcendent candidate using data mining techniques. Data mining is the way of unsheathing and investigate the significant data from the gigantic database, as to investigate patterns or relationships. It tends to be demonstrated as another amazing innovation with incredible potential to help the HR and company manager for selecting the best candidate. Data mining techniques are classified into predictive and descriptive models.

A predictive model makes a prediction about values of data using known realized outcomes found from various data. Predictive model is classified into classification, regression, time series analysis and prediction.

A descriptive model identifies patterns and relationships in data by examining the existing properties of data. Descriptive

model includes Clustering, Summarization, Association rules and Sequence Discovery[2].

Among all the above techniques of Data Mining, we have used the Classification and Association rules for analyzing the data. The goal of the classification technique is to predict the target class for each case in the data. Association rules are the if/then statements which help to show the relationship between the data items within large datasets in various types of databases. The association rule can be explain with an example as "If a customer visits to the mobile shop to buy the mobile then it is predicted that he will also buy tempered glass."

## II. DATA COLLECTION AND PRE-PROCESSING

Based on the collected data, some attributes are considered to evaluate the candidates performance for recruitment. The attributes used for classifying the candidates performance are mentioned in table 1.

TABLE I. CANDIDATE PERFORMANCE

Attributes	Description	Values
Age	The age of the employee	Between 22 years to 30 years
Location	The location of the employee	Mumbai, New Mumbai
Experience	Employee's experience of job calculated in years	1-5 years
Aptitude	Score of the aptitude test	minimum 20 marks
Technical	Score of the technical round	minimum 20 marks
HR	Result of the HR round	minimum 30 marks

TABLE II. SAMPLE OF CANDIDATES DATA COLLECTION

Name	Age	Location	Experience	Tech	HR
Ajay Kumar	24	Pune	2	23	35
Prapti Patil	32	Mumbai	1	24	20

Vinay Joshi	30	Vashi	6	28	36
Dipali Patil	28	Vashi	0	21	27
Akash Jadhav	26	Mumbai	3	25	36
Pankaj Thakur	25	Mumbai	2	20	38
Harshad Mokal	30	Pune	5	25	25
Satish More	31	Panvel	3	19	30
Girish A More	29	Mumbai	2	23	21
Aruna S Patil	23	Pune	0	21	30

III. RESEARCH METHODOLOGY

Association Rules: Association is defined as a data mining function that finds the probability of the co- occurrence of items in a collection. The relationships between co-occurring items are communicated as Association rules.

Given a set of items  $I=\{I_1, I_2, \dots, I_m\}$  and a database of transactions  $D=\{t_1, t_2, \dots, t_n\}$  where  $t_i=\{I_{i1}, I_{i2}, \dots, I_{ik}\}$  and  $I_{ij} \in I$ , an association rule is an implication of the form  $X \Rightarrow Y$  where  $X, Y \subset I$  are sets of items called Itemsets and  $X \cap Y = \emptyset$ [1]. The strength of the Association Rule can be measured in terms of its support and confidence.

The Support(s) for an association rule  $X \Rightarrow Y$  is the percentage of transaction in the database that contains  $X \cup Y$ . The Confidence ( $\alpha$ ) for an association rule  $X \Rightarrow Y$  is the ratio of the number of transactions that contain  $X \cup Y$  to the number of transactions that contain  $X$ [4].

The association between various data items can be discovered by mining Multilevel Association Rules, Multidimensional Association Rules and/or Quantitative Association Rules. Multilevel  $A_{quan1} \wedge A_{quan2} \rightarrow A_{cat}$

We have mined quantitative association rules from the pre-processed data as mentioned in the table 3.

TABLE III. ASSOCIATION RULES

Association Rule	Sup- port (%)	Confi- Dence (%)
age(C1,22..30)^location("pune")^ experience(c1,1..5)^aptitude(20..30)^ technical(20..30)^hr(30..40) ->Candidate_grade(C1,"Excellent")	30	100
age(C2,30..40)^location("mumbai")^ experience(c1,1..5)^aptitude(20..30)^ technical(20..30)^hr(20..30) ->Candidate_grade(C2,"poor")	10	25
age(C3,22..30)^location("airoli")^ experience(c1,>5)^aptitude(20..30)^ technical(20..30)^hr(30..40) ->Candidate_grade(C3,"Excellent")	10	100
age(C4,22..30)^location("vashi")^ experience(c1,1..5)^aptitude(20..30)^ technical(20..30)^hr(20..30) ->Candidate_grade(C4,"poor")	10	100

age(C5,22..30)^location("mumbai")^ experience(c1,1..5)^aptitude(10..20)^ technical(20..30)^hr(30..40) ->Candidate_grade(C5,"Excellent")	30	75
age(C6,22..30)^location("mumbai")^ experience(c1,1..5)^aptitude(20..30)^ technical(10..20)^hr(30..40) ->Candidate_grade(C6,"Excellent")	30	75
age(C7,22..30)^location("pune")^ experience(c1,1..5)^aptitude(20..30)^ technical(20..30)^hr(20..30) ->Candidate_grade(C7,"Average")	30	100
age(C8,22..30)^location("panvel")^ experience(c1,1..5)^aptitude(10..20)^ technical(10..20)^hr(30..40) ->Candidate_grade(C8,"poor")	10	100
age(C9,22..30)^location("mumbai")^ experience(c1,1..5)^aptitude(20..30)^ technical(20..30)^hr(30..40) ->Candidate_grade(C9,"Average")	30	75
age(C10,22..30)^location("pune")^ experience(c1,1..5)^aptitude(10..20)^ technical(20..30)^hr(30..40) ->Candidate_grade(C10,"poor")	30	100

IV. CLASSIFICATION

Classification is an analytical task where the classifier is used to assume the categorical labels i.e. Classes.

Given a database  $D=\{t_1, t_2, \dots, t_n\}$  of tuples (items, records) and a set of classes  $C=\{c_1, \dots, c_m\}$ . The

classification problem is to define a mapping  $f : D \rightarrow C$  where each  $t_i$  is assigned to one class. A class,  $C_j$ , contains accurately those tuples mapped to it; i.e.  $C_j = \{t_i \mid f(t_i) = C_j, 1 \leq i \leq n, \text{ and } t_i \in D\}$ [2].

Classification is a process that works in two step in which data are interpreted for descriptive goals. The first step consist of the learning step (or training phase) in which the classifier is built describing a predetermined set of data classes or concepts. A classification algorithm constructs the classifier by analyzing or "learning from" a training set made up of database tuples and their associated class labels. A tuples,  $X$ , is spoken to by a n-dimensional property vector  $X=(x_1, x_2, \dots, x_n)$ , depicting n estimations made on the tuple from n database properties, individually,  $A_1, A_2, \dots, A_n$ . each tuple,  $X$ , is expected to have a place with a predefined class.

The learning step expands with use of model for classification which is the second step. First, the predictive accuracy of the classifier is calculated. A test set is use, made up of test tuples and their associated class labels. They are independent of the training tuples, meaning that they were not used to construct the classifier. The accuracy of a classifier on a given test set is the percentage of test set tuples that are effectively classified by the classifier. The associated class name of each test tuple is compared with the learned classifier's class prediction for that tuple. If the accuracy of the classifier is considered acceptable, the classifier can be used to classify future data tuples for which the class label is not known[2].

Here the classifier is used to assume the categorical labels among “Selected” and “Rejected”. The corresponding class label is provided to each training tuple. The classification rules obtained at end of the training phase are as mentioned in table 4.

The classification rules mentioned in table 4 predict employee selection for the company. The outcome of the classification process is the set of classification rules which predict the inclusion of any candidate in the team. The rejected candidate can approach for the next recruitment to enhance their performance.

TABLE IV. TABLE IV.CLASSIFICATION RULES

If Candidate_grade = poor AND location = mumbai AND experience >= 1 then Employee = rejected
If Candidate_grade = Excellent AND location = mumbai AND experience >= 1 then Employee = selected
If Candidate_grade = Average AND location = pune AND experience >= 1 then Employee = rejected
If Candidate_grade = poor AND location = mumbai AND experience = 0 then Employee = rejected
If Candidate_grade = Average AND location = mumbai AND experience = 0 then Employee = rejected
If Candidate_grade = Excellent AND location = pune AND experience >=1 then Employee = selected
If Candidate_grade = Excellent AND location = pune AND experience = 0 then Employee = rejected
If Candidate_grade = Average AND location = mumbai AND experience >5 then Employee = selected
If Candidate_grade = poor AND location = mumbai AND experience >5 then Employee = rejected
If Candidate_grade = poor AND location = pune AND experience >5 then Employee = rejected
If Candidate_grade = poor AND location = pune AND experience >=1 then Employee = rejected
If Candidate_grade = Average AND location = vashi AND experience >=1 then Employee = selected

If Candidate_grade = poor AND location = panvel AND experience >=1 then Employee = rejected
If Candidate_grade = Average AND location = vashi AND experience >=1 then Employee = selected
If Candidate_grade = poor AND location = vashi AND experience >=1 then Employee = rejected
If Candidate_grade = Excellent AND location = vashi AND experience =0 then Employee = rejected
If Candidate_grade = Average AND location = panvel AND experience >=1 then Employee = selected
If Candidate_grade = Average AND location = panvel AND experience = 0 then Employee = rejected

V. CONCLUSION

This paper prospects the imminent proficiency of data mining techniques for upgrading the candidate selection process. The association rules are utilized to discover the applicants execution and through the classification rules we have selected best candidates for the organization. By applying above mentioned techniques on the candidates data, the selected applicants for the company listed in the table 5.

TABLE V. FINAL SELECTED TEAM

Name	age	Location	Experi-ence	Apti-tude	Tech-nical	HR
Ajay Kumar	24	pune	2	22	23	35
Vinay Joshi	30	vashi	6	25	28	36
Akash jadhav	26	mumbai	3	19	25	36
Pankaj thakur	25	mumbai	2	21	20	38
Harshad mokal	30	pune	5	27	25	25
Girish more	29	mumbai	2	24	23	21

REFERENCES

- [1] Study on Recruitment Selection Process with Reference” Researchgate.net/publication/268210664 November 2014
- [2] Maet D”DMining: Introduadvanced Topics”,P
- [3] Han,ne,Mini Ce
- [4] Vipin Kumar, Pang-Ning Tan, Michael Steinbach,”Introduction to Data Mining”,Addison-Wesley,2005,ISBN: 0321321367
- [5] Rajan chattamvelli,”Data Mining Methods” Narosa 2009
- [6] ”The impact of recruitment and selection criteria on organizational performance” Global Journal of Human Resource Management Vol.3,No.2,pp.22- 33, March.2015