iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** I **Month of publication:** January 2022

DOI: https://doi.org/10.22214/ijraset.2022.39886

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

Recommendation System for Human Resource Department

Ulka Khobragade
Electronics and Telecommunication, Yeshwantrao Chavan College of Engineering,Wanadongri, Nagpur, India

Abstract: ***The objective is to find suitable skilled employees for the job among different departments within the organization. For finding the quality of an applicant or even the already employed employee, the HRs of companies goes through a lot of hectic schedule, time consuming processes, decision making, etc. In this case, Recommendation System, which is a part of Machine Learning, proves to be effective in making decisions on behalf of the HRs if an employee or an applicant is suitable enough for the job. The aim of the project is to predict whether the already employed employees, who belong to different department within the organization can perform well or not if assigned to a different department.***
Keywords: Recommendation system, Collaborative Learning, K-NN, Similarity, Similarity Correlation, Cosine etc.

I. INTRODUCTION

The aim of the project is to predict whether the already employed employees, who belong to different department within the organization can perform well or not if assigned to a different department. For this we have used a dataset with an approximate shape of 200x25. Here the rows consist of *employee id number* and columns or features corresponds to competencies, i.e., the skills. The values in the dataset are the ratings given by the manager under whom the employees work for respective competency. The rating scale is based on the standards of the organization and based on the type of dataset and the objective, Singular Value Decomposition (SVD), K-nearest neighbors (KNN), Matrix-Factorization, Stochastic Gradient Descent (SGD), Collaborative filtering similarities are some of the approaches used and analyzed. Depending upon the performance score, we can select one of the shortlisted algorithms and approaches.

With the growing number of industries, there is a hike in employment or more precisely, with change in technology and demand in the industry there is requirement for manpower, not just in number but in quality. For finding the quality of an applicant or even the already employed employee, the HRs of companies goes through a lot of hectic schedule, time consuming processes, decision making, etc. As it is already known to us all, machine learning is what makes our lives easier, also helping us predict the future outcomes. In this case, recommendation system, which is a part of machine learning, proves to be effective in making decisions on behalf of the HRs if an employee or an applicant is suitable enough for the job.

The aim of the project is to predict whether the already recruited employees, who belong to a different department within the organization can perform well or not if assigned to another department. I, in this project, have also tried to predict who among them, irrespective of the department, would be suitable for a job if there is a vacancy in any department due to any reasons, say a resignation from an employee. In this proposal, we have come up with several algorithms and approaches which can be proven effective to achieve the objective of the project.

II. PROBLEM STATEMENT

The objective is to find suitable skilled employees for the job among different departments within the organization. For finding the quality of an applicant or even the employee, the HRs of companies goes through a lot of hectic schedule, time consuming processes, decision making, etc.

This system will save time of the HR representatives. Not only will this help in having track on performance of the employees, but also bifurcating them based of their skillset in suitable departments, recommending best suitable employee available for a particular job/department, etc. This in turn can provide high performance rate of a department as a whole and carrying out the objectives in department and ultimately in the organization, efficiently and early. Being able to develop such a system can really be beneficial to the organization. If fortunate enough, might be able to predict the performance of employees not only in department but also within the organization in future.

©IJRASET: All Rights are Reserved

III. METHODOLOGY

The overview of the proposed methodology is shown in the fig. (1). As the content, form and quantity of the original data are different. The recommendation algorithm is generally based on the specific business scenario. For the human resources case, some common recommendation algorithm recommended effect is not obvious [1].

As described in the abstracts, we have used recommendation system for internal job posting. Let us understand the typical meaning of a recommendation engine as shown in Fig. 1. A recommendation engine filters the data using different algorithms and recommends the most relevant items to users. It first captures the past behavior of a customer and based on that, recommends products which the users are likely to buy. Before applying any kind of algorithm following steps are performed:

A. Data Collection

This is the first and most crucial step for building a recommendation engine. The data can be collected by two means: explicitly and implicitly. Explicit data is information that is provided intentionally, i.e., input from the users such as movie ratings. Implicit data is information that is not provided intentionally but gathered from available data streams like search history, clicks, order history, etc.

B. Data Storage

The amount of data dictates how good the recommendations of the model can get. For example, in a movie recommendation system, the more ratings users give to movies, the better the recommendations get for other users. The type of data plays an important role in deciding the type of storage that has to be used. This type of storage could include a standard SQL database, a NoSQL database or some kind of object storage.

C. Data Filtering

After collecting and storing the data, we have to filter it so as to extract the relevant information required to make the final recommendations. There are various algorithms that help us make the filtering process easier.

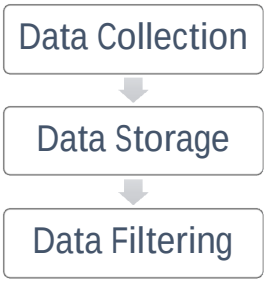


Fig. 1 Flowchart for data preprocessing

The methodology used in this research follows the steps in Fig. 2.

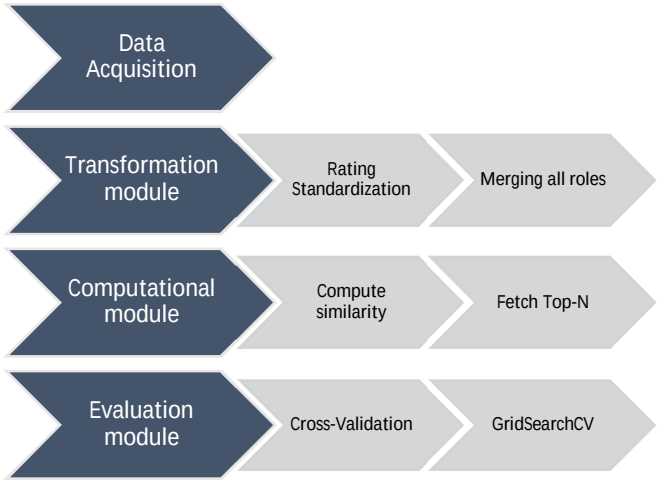


Fig. 2 Flow of model building and evaluation

A. Data Acquisition

As described earlier, information of the employees is collected. This data is collected from the managers handling different units of employees and provides rating to their employees. The data is acquired from different branches of the same company. The data has employee name, the competences (functional and behavioral) and its values are the rating of the employee for respective competencies.

B. Transformation Module

In this module we transform our data in the desired format as compatible with the algorithm we will be using.

1) *Rating Standardization:* The rating provided by the Manager to his/her employees of different branches might vary in rating scale. To bring all samples on the same rating scale, the company uses some confidential standards. The output being a rating scale of 1 to 5, one being the lowest and 5 being the highest.
2) *Merging all Roles:* There are employees perusing different roles within the same branch, so the data acquired will be on separate files. We have to merge all these files into one dataset. Also perform preprocessing on the same.

C. Computational Module

In this paper we will apply K-NN algorithm and Similarity Correlation. Let us first understand Collaborative filtering. The term "collaborative filtering" (CF) was coined in 1992 by Goldberg et al., who proposed that "information filtering can be more effective when humans are involved in the filtering process" [2]. The concept of collaborative filtering as it is understood today was introduced two years later by Resnick et al. [3]. Their theory was that users like what like-minded users like, where two users were considered like-minded when they rated items alike. When like-minded users were identified, items that one user rated positively were recommended to the other user, and vice versa. Compared to CBF, CF offers three advantages [4]. Collaborative filtering, also referred to as social filtering, filters information by using the recommendations of other people. It is based on the idea that people who agreed in their evaluation of certain items in the past are likely to agree again in the future. A person who wants to see a movie, for example, might ask for recommendations from friends. The recommendations of some friends who have similar interests are trusted more than recommendations from others. This information is used on deciding which movie to see.

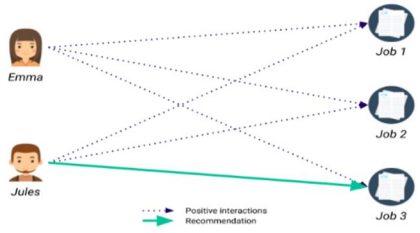


Fig. 3 Example of collaborative filtering

Let us know about the two approaches used in this research:

1) *Similarity Correlation:* The most commonly used measures of similarity, the cosine similarity. Consider the vectors U = *(3,4)* and V = *(1,1)*, you can easily tell how far apart these two vectors are (by using Euclidean distance or some other metric).

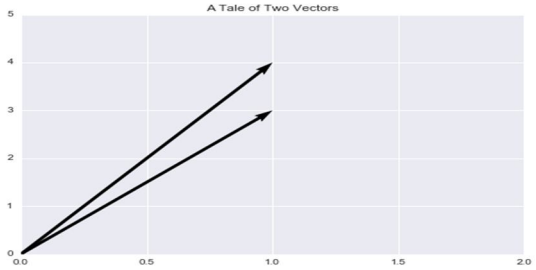


Fig. 4 Vector U and vector V 1Vector

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

With the cosine similarity, we are going to evaluate the similarity between two vectors based on the angle between them. The smaller the angle, the more similar the two vectors are. With the cosine similarity, we are going to evaluate the similarity between two vectors using the angle between them. The smaller the angle, the more similar the two vectors are [5]. Following is the angle meaning:

- Cosine (0°) = 1 (Maximum similarity)
- Cosine (90°) = 0 (No similarity)
- Cosine (180°) = -1

If we restrict our vectors to non-negative values (as in the case of movie ratings, usually going from a 1-5 scale), then the angle of separation between the two vectors is bound to be between 0° and 90°, corresponding to cosine similarities between 1 and 0, respectively. Therefore, for positive-valued vectors, the cosine similarity returns a value between 0 and 1, one of the 'ideal' criterions for a similarity metric. One important thing to note is the cosine similarity is a measure of orientation, not magnitude [5]. There are many similarity metrics we can use, namely Euclidean, Pearson's, Cosine, etc. The formula for getting the similarity score of cosine metrics(cos) and Pearson's (r) is as shown:

$$\text{Similarity } \cos(\theta) = \frac{A.B}{||A||\,||B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

In sklearn, NearestNeighbors method can be used to search for k nearest neighbors based on various similarity metrics.

2) *K-NN:* K-NN stands for K-Nearest Neighbors. In neighborhood-based techniques, a subset of users is chosen based on their similarity to the active user, and a weighted combination of their ratings is used to produce predictions for this user [6]. Most of the approach can be summarized as below:

- Assign a weight to all users with respect to the target employees
- Select *k* employees that have the highest similarity score with the target user
- Compute a prediction from a weighted combination of the selected neighbors' ratings

The most commonly used measure of similarity is the Pearson correlation coefficient between the ratings of the two users [7]. A significant number of recommender systems utilize the *k*-nearest neighbor (*k*NN) algorithm as the collaborative filtering core. This algorithm is simple; it utilizes updated data and facilitates the explanations of recommendations. Its greatest inconveniences are the amount of execution time that is required and the non-scalable nature of the algorithm. The algorithm is based on the repetitive execution of the selected similarity metric [8]. In the case of kNN, we measure the similarity or correlation between the target employee 's' and other employee vector 't' (where t ∈ T). The top k most similar employees to s are considered to be the neighbors for the employee s, which we denote by NB(s) (taking the size k of the neighbors to be implicit):

NB(s) = {t s1, t s2, ⋯, t sk}.

Given the active employee s and every other employee vector t, the similarity between them is obtained by:

$$sim(\vec{t}, \vec{s}) = \frac{\vec{t}\ .\ \vec{s}}{|\vec{t}| \times |\vec{s}|}$$

Following Formula is used in backend to predict ratings of the employees in different department of the company. $P_{a,i}$ is used to fill the missing ratings. Later KNN and cosine similarity metrices for getting the best fit employee for desired department by choosing the target employee from the wanted department.

$$p_{a,i} = \overline{r_a} + \frac{\sum_{u=1}^{k}(r_{u,i} - \overline{r}_u) \times sim(a,u)}{\sum_{u=1}^{k} sim(a,u)}$$

where;

$p_{a,i}$ = prediction of rating for target a of competency i

a = target employee

$\overline{r_a}$ = mean rating for an employee

$u_{1\ldots}u_k$ = k nearest-neighbor to employee a

$r_{u,i}$ = rating given to employee on i^{th} competency

sim (a, u) = Pearson correlation between a and u

©IJRASET: All Rights are Reserved

IV. RESULTS

Since our focus is on collaborative filtering with small trails on the Employee dataset we have kept it as simple as possible. However, both seemed to work fine with the available dataset. The given below are the final results for both the approaches.



Fig 5 Result of Similarity metric method

In the figure above, the integers are the employee_id and the corresponding is the similarity score, arranged in descending order.

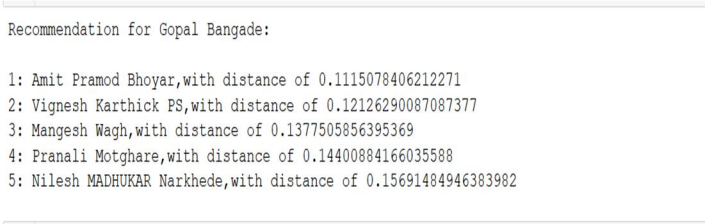
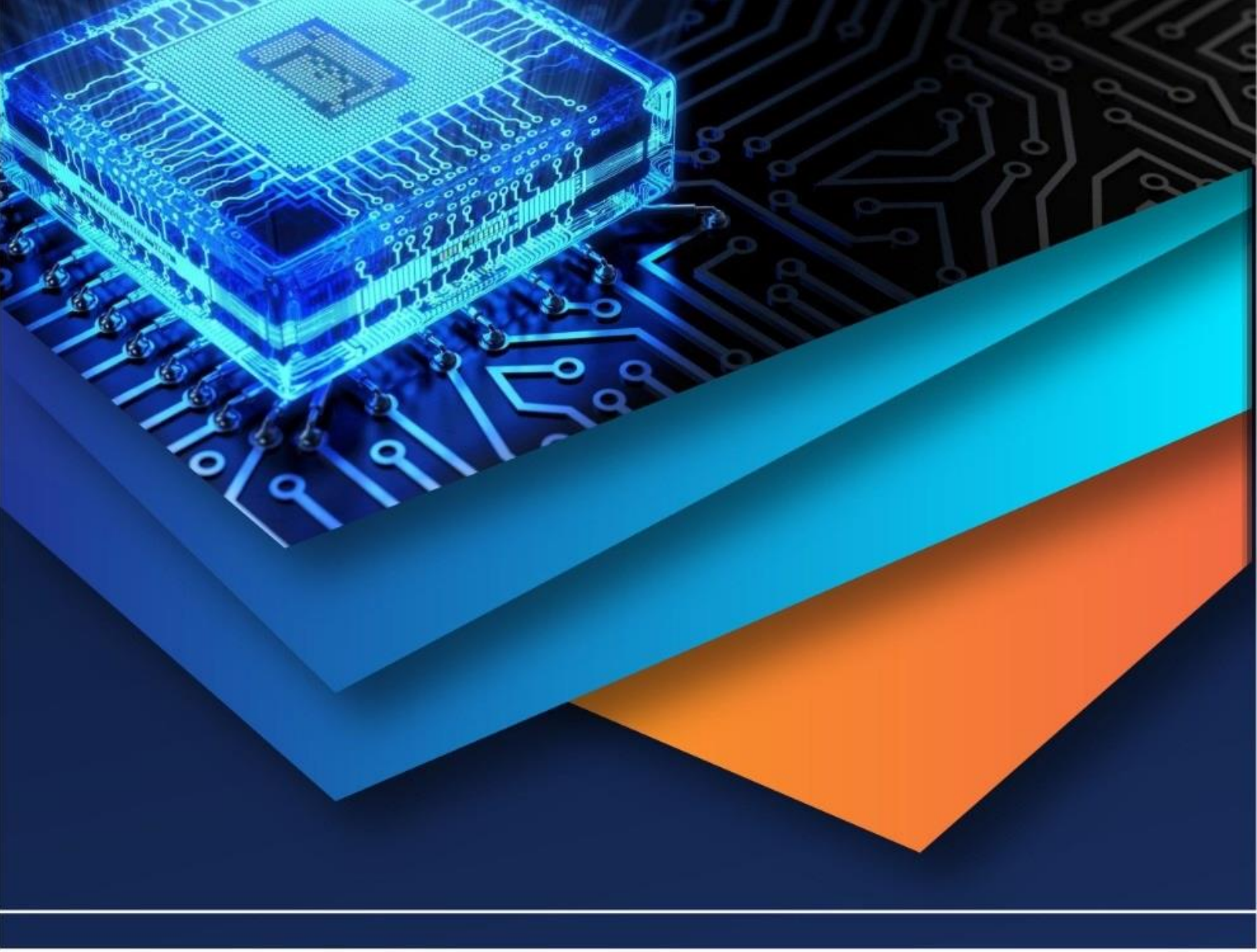


Fig 6 Result of KNN algorithm

In the picture above, those are the results of KNN, where the names are fo the employees and the float values are the distances of the target employee to every other employee in the dataset. The minimum distance being the closest.

REFERENCES

[1] Human Resource Recommendation algorithm based on ensemble learning and spark. By: Zihan Cong, Xingming Zhang, Haoxiang Wang and Hongjie Xu
[2] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, “Using collaborative filtering to weave an information Tapestry,” Communications of the ACM, vol. 35, no. 12, pp. 61–70, 1992.
[3] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, “GroupLens: an open architecture for collaborative filtering of netnews,” in Proceedings of the 1994 ACM conference on Computer supported cooperative work, 1994, pp. 175–186.
[4] Research-Paper Recommender Systems: A Literature Survey Joeran Beel1, Bela Gipp2, Stefan Langer1, and Corinna Breitinger1, 3 1Docear, Magdeburg, Germany {beel | langer | breitinger}@docear.org 2University of Konstanz, Department of Computer and Information Science, Konstanz, Germany bela.gipp@uni-konstanz.de 3Linnaeus University, School of Computer Science, Physics and Mathematics, Kalmar, Sweden
[5] Recomender System through Collaborative filtering, by Manojit Nandi on July 14, 2007
[6] Recommender Systems Prem Melville and Vikas Sindhwani IBM T.J. Watson Research Center, Yorktown Heights, NY 10598 {pmelvil,vsindhw}@us.ibm.com
[7] Paul Resnick, Neophytos Iacovou, Mitesh Sushak, Peter Bergstrom, and John Reidl. GroupLens: An open architecture for collaborative filtering of netnews. In Proceedings of the 1994 Computer Supported Cooperative Work Conference, New York, 1994. ACM.
[8] Improving the Effectiveness of Collaborative Filtering on Anonymous Web Usage Data Bamshad Mobasher,Honghua Dai,Tao Luo,Miki Nakagawa {mobasher,hdai,tluo,miki}@cs.depaul.edu School of Computer Science, Telecommunication, and Information Systems DePaul University, Chicago, Illinois, USA



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)