USE OF SOCIAL NETWORK FOR RECOMMENDING JOB BY APPLYING MACHINE LEARNING TECHNIQUES

Binny Parida[1], Prashanta Kumar Patra[2], Sthita Pragyan Mohanty[3]

[1]Student, Department of Computer Science and Engineering, College of Engineering and Technology, Bhubaneswar, Odisha, India
[2]Department of Computer Science and Engineering, College of Engineering and Technology, Bhubaneswar, Odisha, India
[3]Department of Information Technology, College of Engineering and Technology, Bhubaneswar, Odisha, India

---***---

Abstract: *With increase in the use of Internet, the rate of increase of social networks is becoming ubiquitous in recent years. As huge amount of user's data is available on the internet, it is necessary to use recommender engine to suggest relevant jobs using user's data. This paper focuses on the job portal website. The job recommender system takes the skills of the users from that website and makes recommendation to them with the job offers whose contents are quite relatable to the user's profile. At first, data is cleaned by removing the dirty data in form of extra space and duplicates. Then jobs are recommended to the applicants using their preferences. In this paper Machine Learning techniques are employed to improve the results. Based on these algorithms, comparisons are done for accurate and precise results. The outcome shows that Random Forest Classifier (RFC) gives the highest accuracy as compared to other techniques.*

Key words- Recommender System, Machine Learning, Random Forest Classifier (RFC), Job Recommender System

1.INTRODUCTION

The job recommender framework is used to search the job boards and suggest the most appropriate job offer that are relevant for the job seekers. It provides a tool to connect with the candidates and the job offers automatically by using relevant algorithm of machine learning. It saves a lot of time by matching the candidate with proper job position. Job Recommender System plays a vital role to help the job seekers for finding the appropriate job using candidate skills. It reduces time conflict as compared to Traditional based System. A Recommender System is an Information filtering technique that extracts the user preferences form large pool of data to recommend the top items. It acts as an information retrieval tool that helps to filter out and prioritize the data. Information retrieval is a term defined as the craft of looking through the data from gigantic measure of data. Based on the filtering approach of

information, recommender systems are basically divided into three major categories: Content based filtering method [12], Collaborative filtering method [13] and Hybrid filtering method.

Content based filtering method shown in Fig.1.use personal preferences to match and filter items and is able to recommend user unique taste. E.g. what sort of books do I like?
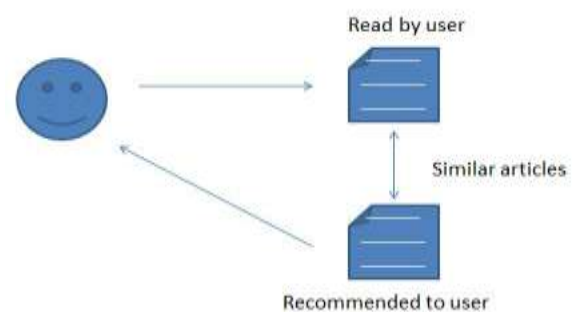

Fig.1. Content based filtering method

Collaborative filtering method in Fig.2.matches `like-minded' users with similar interests and common preferences E.g. if two people have similar 'taste' they can recommend items to each other.
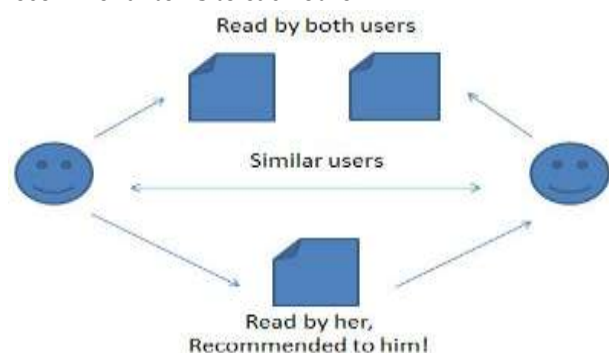

Fig.2. Collaborative filtering method

The Hybrid filtering method shown in Fig.3 is the combination of content based filtering method and Collaborative filtering method.
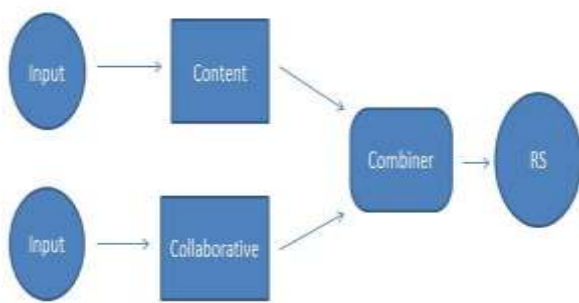
Fig.3. Hybrid filtering method

1.1 Impact of Recommender System in Machine Learning

To minimize the information overload, we use Machine Learning techniques. These techniques [9] help the system for better optimization results with accurate values. The proposed model is shown in Fig.4.
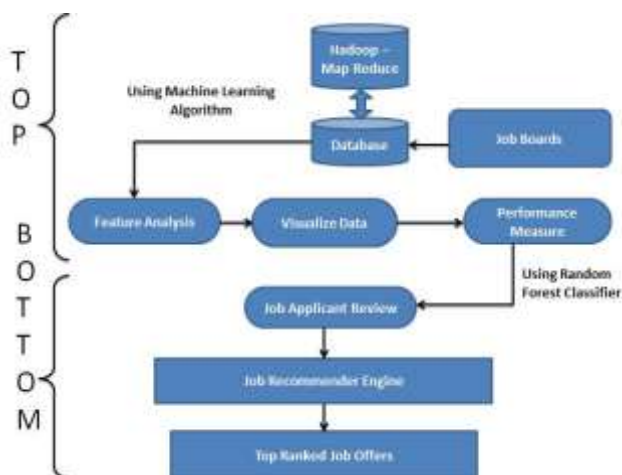

Fig.4. Proposed Model

Many research works has been done by researchers in this field. Some of them have made comparison with other methodologies whereas some proposed their own model for better performance.

Sidahmed Benabderrahmane et al. [1] proposed a decision making tool to give a proper channel to recruiters to find the efficient job seekers. Shuo Yang et al. [2] discussed the various aspects related to recommender system. It focuses on two major filtering techniques of recommendation system [14]. First one is content based filtering technique and second one is collaborative filtering technique. It combines these techniques to construct a hybrid filtering technique using extensive feature engineering. Nedra Mellouli et al. [3] show a job board recommendation system named as Smart4Job. This system deals with different job hunting websites for job aspiring candidates. This then, circulates new job offers with proper clarification to job seekers. The hybrid filtering technique is used in this case. Finally by using the click history, the best job board is envisaged for job seekers using the time series analysis module. Miao Jiang et al. [4] introduces a real world career recommendation system. Here the job applicants have to accept the email alerts to get their job offers. By examining the click behavior of individual users, it forms a set of specific attribute that helps the recruiters to go for the right job seeker. The probabilistic model is proposed in this system. Shibbir Ahmed et al. [5] introduce a job recommender system that is useful for the recruiters to get their suitable job seekers. Job recommender system uses an algorithm to estimate similarity for user and item. Collaborative filtering technique is used here. Hybrid filtering technique is also used to calculate user-item interaction for recommendation of job boards. Then it selects the top job [11] offers as a list from the user-item interaction. Comparison is done based on the three filtering techniques. Priscila Valdiviezo Diaz et al. [6] used a Bayesian model for the recommendation of job boards to the suitable job applicants. The system uses the similar users and items information for recommendation of job boards. This model is based on collaborative filtering technique. Four different datasets are used here to get the good performance. Normalized Discounted Cumulative Gain (nDCG) is used for better prediction accuracy in the datasets. Shiqiang Guo et al. [7] propose a resume matching system named as Resu Matcher. It extracts the qualitative information and experience of the job seekers without the involvement of active user's data. Basically it uses the machine learning techniques for extraction of data. It uses a novel statistical similarity index method for suggestion of jobs using applicants qualification, academics, experience, how much the job [10] seeker technically sound etc. without less active users input. Mamadou Diaby et al. [8] proposed a system that collects information about job seekers from face book and LinkedIn and provides the job based on their given data. Currently it is used by San Francisco-based Software Company that basically deals with Face book and LinkedIn users by offering them the jobs using their bio data. A taxonomy based vector model is used here. It also used OR fuzzy logic operation and AND fuzzy logic operation.

To minimize the information overload, we use Machine Learning techniques. These techniques help the system for better optimization results with accurate values. A recommender system gives recommendations to the users through an ltering process that depends on user's inclinations and perusing history. The data about

the candidate is taken as an input. The data is taken from the information that is through perusing and browsing information. The proposal framework is an execution of the machine learning calculations [15]. It is a stage that gives its users different preferences dependent on their inclinations and likings. To limit the data over-burden, we use Machine Learning procedures. A recommender system thus, takes the data about the user as information. These strategies help the framework for better advancement results with exact qualities.

2. DATA PREPROCESSING

Data Preprocessing is the method of converting raw and unused data into relevant data. It includes data cleaning, extracting, selecting and reducing to get an appropriate data.

2. 1 FEATURE EXTRACTION

Machine Learning is a type of learning that automatically learns from experience. There is no need to specifically write each and every instruction for the machine to perform. It is an analytical method used in the analysis of data for building of models. Machine learning is a subset of Artificial Intelligence [16]. Feature extraction is the process of removing extra features from the raw dataset for processing. Following Algorithms and methods are used for Feature extraction i.e. Histogram and Heat map using seaborn.

2.1.1. Histogram

A frequency histogram is a type of bar graph that represents the number of times the particular feature occurred in the dataset. It helps to visualize the distribution of feature in an efficient manner. Estimation occurs by using various parameters. Fig.5. and Fig.6.represents the frequency of age estimate and company staff counts occurred in the particular feature.
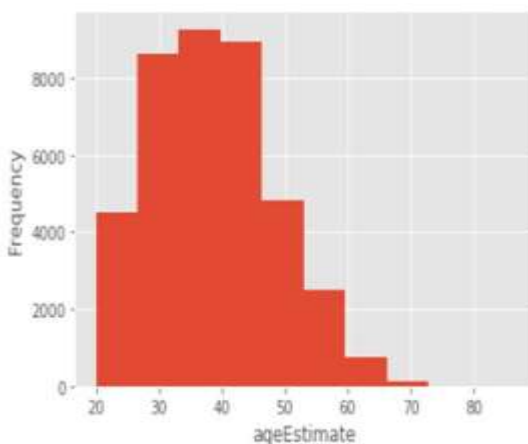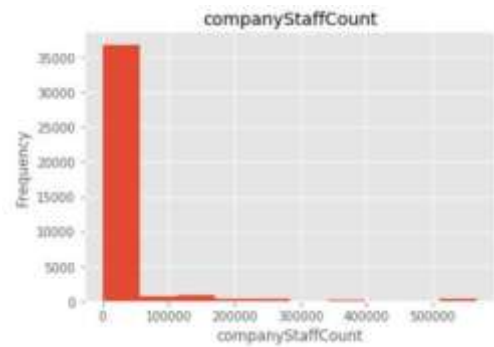

Fig.5. Frequency of Age Estimate

Fig.6. Frequency of Company Staff Count

Fig.7. and Fig.8.represents the frequency of connections count and Position Id
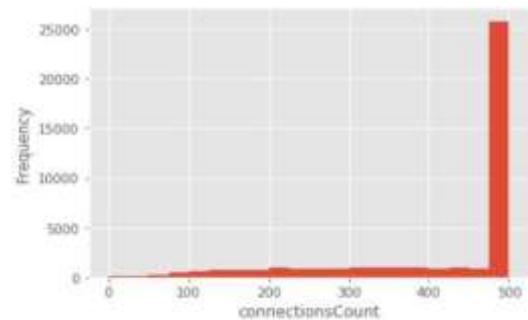

Fig. 7. Frequency of Connections Count
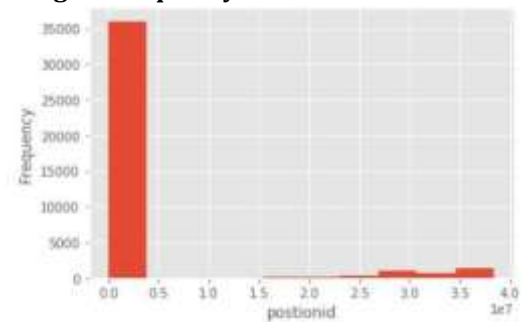

Fig.8. Frequency of Position Id

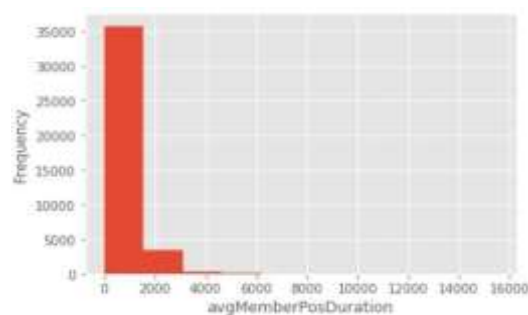Fig.9. and Fig.10.Shows the frequency of the average member position duration and average company position duration

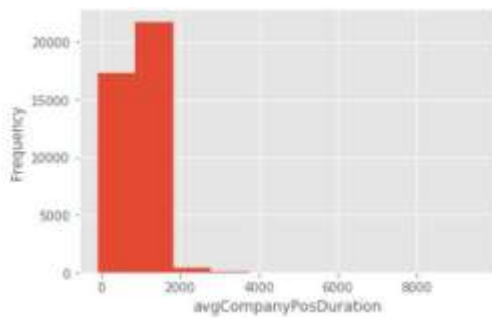**Fig 9.Frequency of Average Member Position Duration**

Fig.10. Frequency of Average Company Position Duration

2.1.2. Heat map using Seaborn:

A Heat map is a two-dimensional graphical representation of features which is represented in a matrix format. This a color coding format to represent user's history related to features [18]. Dark color shows the most visited feature whereas light color shows the less importance given to that feature. Fig.11. represents Heat map for Job features using the given dataset.
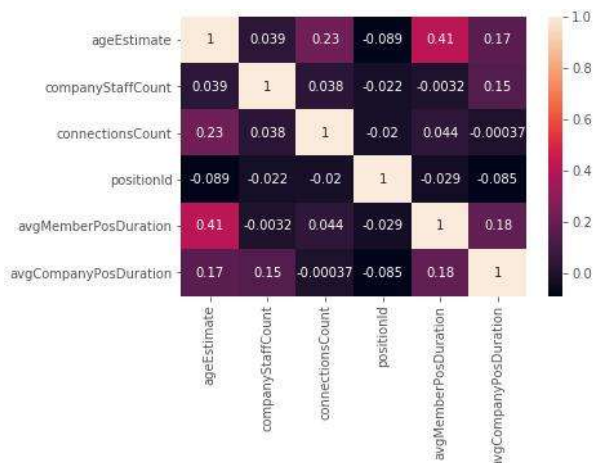

Fig.11. Heat map for Job features

2.2. FEATURE SELECTION

At first, data is cleaned by removing the noisy or dirty data. These data are present in the form of extra spaces and duplicates. Thus, it corrects the inaccurate or irrelevant data which helps the system to increase the performance. Job portal website dataset is used in this paper. The recommender system takes the user's information from the given features to make recommendation to them accordingly. Among various features available on this dataset, relevant features are selected for the recommendation process. These are age estimate, company staff count, connections count, position id, average member position duration and average company position duration. The mean values of each feature are calculated below in table 1.

Table 1: Mean Values

Features	Mean values
Age Estimate	38.43425
Company Staff Count	20451.81618
Connections Count	424.57316
Position Id	2987946.54374
Average Member Position Duration	875.18698
Average Company Position Duration	888.25753

3. METHODS APPLIED

This dataset consists of 39537 candidates and 8 job features after the feature selection method. Python is used as implementation language for recommendation process. Following are the machine learning [17] classification algorithms that are used to analyze the prediction accuracy.

I) Logistic Regression (LR):

Logistic Regression is supervised machine learning algorithm that learns from training dataset. In case of classification, it consists of target variable (or output), Y and set of features (or input), X. The resultant output is only two possible outcomes. The goal of logistic regression is to get the best fit model that describe the bond between X and Y. The accuracy for the given dataset using logistic regression is 87.14%.

II) K nearest Neighbors (KNN):

K Nearest Neighbors is a supervised classification algorithm. It deals with labeled points to train label other points. It labels the new point by selecting the points that are close to that point. K is the number of nearest neighbors (points) used for accuracy prediction. The accuracy for the given dataset using K Nearest Neighbors is 87.75%.

III) Stochastic Gradient Decent Classifier (SGD):

Stochastic Gradient is an efficient approach to discriminate learning of classifiers. The implementation process is easy in this type of approach. The goal of this algorithm to find minima or maxima by iteration. The accuracy for the given dataset using Stochastic Gradient Decent Classifier is 86.56%.

IV) Random Forest Classifier (RFC):

Random Forest Classifier is an ensemble method used in machine learning. It control over-fitting using Bootstrap Aggregator or bagging. The goal of Random Forest Classifier is to combine multiple decision trees to get the final output. The accuracy for the given dataset using Random Forest Classifier is 94.54%.

4. RESULT ANALYSIS

After using four different types of machine learning algorithms Random Forest Classifier results in highest prediction accuracy with highest F1 scores as compared to other algorithms. Table 2 represents a comparison table among various machine learning classification algorithms.

Table 2: Machine Learning Classification Algorithms

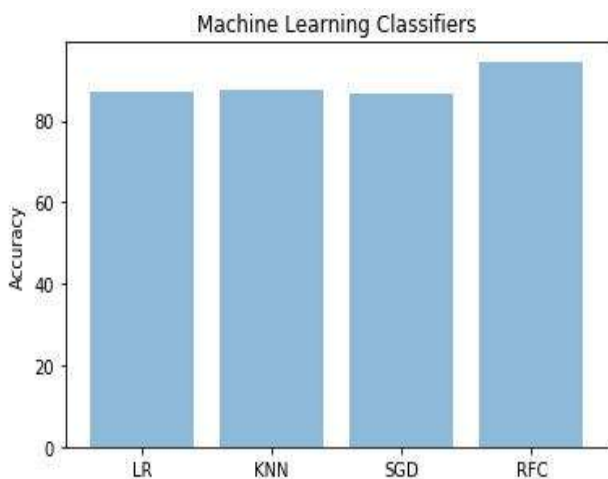| Sl. No.	Machine Learning Classifiers	Accuracy	Precision	Recall	F1 store
1.	Logistic Regression	87.14%	1.0	0.871	0.931
2.	K Nearest Neighbors	87.75%	0.945	0.916	0.930
3.	Stochastic Gradient Descent	86.56%	0.991	0.871	0.927
4.	Random Forest	94.54%	0.995	0.944	0.969

Fig.12. Accuracy prediction using different machine learning classifier algorithms.

Fig.12.represents the graph which compare the prediction accuracies among different machine learning classifier algorithms. It clearly shows that the Random Forest Classifier (RFC) have the highest accuracy value and can give accurate result as compared to other classification algorithms.

Fig. 13 represents the flow chart diagram of Random Forest Classifier. It describes the process of forming the model by gaining the highest accuracy from Random Forest Classifier algorithm.
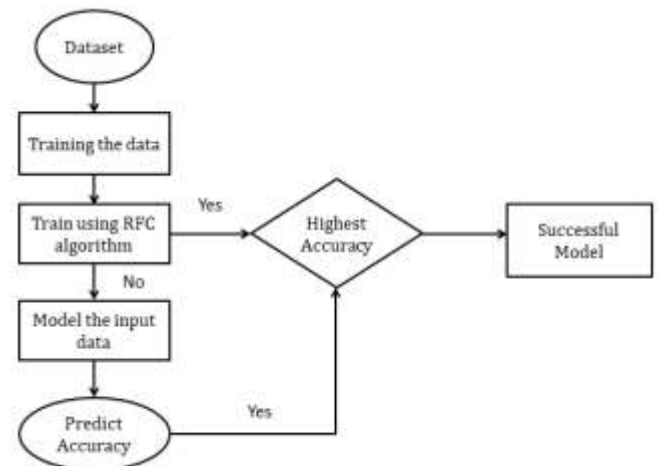

Fig. 13. Flow chart diagram of Random Forest Classifier

Thus, the recruiting companies all over the world can use Random forest classification algorithm in their job recommender system to recommend suitable jobs to the job seekers.

5. CONCLUSION

Recommender system helps us to get what we need from the massive amount of information. It recommends the exact item rather than an overload of information. Recommender system gained popularity in recent years due to its success over a huge amount of data. As one of the domain, online recruiting system utilizes the recommender systems by providing suitable jobs to users using the social network user's data. This paper provides a brief detail of various categories of recommender system. Implementation work is done using python language by comparing various machine learning classification algorithms. Results show that Random forest classifier gains highest prediction accuracy with high F1 score as compared to other algorithms.

REFERENCES

[1] S. Benabderrahmane, N. Mellouli and M. Lamolle, "On the predictive analysis of behavioral massive job data using embedded clustering and deep recurrent neural networks", Knowledge-Based System, vol. 151, 2018, pp. 95-113.

[2] Y. Shuo, Md Korayem, K. AlJadda, T. Grainger and S. "Natarajan, Combining content-based and collaborative filtering for job recommendation system: A cost-sensitive Statistical Relational Learning approach", Knowledge-Based Systems, vol. 136, 2017, pp.37-45.

[3] N. Mellouli, S. Benabderrahmane, M. Lamolle and P. Paroubek, "Smart4job: A big data framework for

intelligent job offers broadcasting using time series forecasting and semantic classification", International Journal of Engineering & Science Research, vol. 5, 2017.

[4] J. Miao, Y. Fang, H. Xie, J. Chong and M. Meng, "User click prediction for personalized job recommendation", World Wide Web, vol. 22 (1), 2019, pp. 325-345.

[5] S. Ahmed, M. Hasan, Md N. Hoq and Md A. Adnan, "User interaction analysis to recommend suitable jobs in career-oriented social networking sites, International Conference on Data and Software Engineering (ICoDSE)", IEEE, 2016, pp. 1-6.

[6] V. D. Priscila, F. Ortega, E. Cobos and R. L. Cabrera, "A Collaborative Filtering Approach Based on Naïve Bayes Classifier", IEEE, vol. 7, 2019.

[7] S. Guo, F. Alamudun and T. Hammond, "RésuMatcher: A personalized résumé-job matching system", Expert Systems with Applications", vol. 60, 2016, 169-182.

[8] M. Diaby and E. Viennet, "Taxonomy-based job recommender systems on Facebook and LinkedIn profiles, Eighth International Conference on Research Challenges in Information Science" (RCIS), IEEE, 2014, pp. 1-6.

[9] A. Gupta and D. Garg, "Applying data mining techniques in job recommender system for considering candidate job preferences", International Conference on Advances in Computing, Communications and Informatics, IEEE, 2014, pp. 1458-1465.

[10] V.L. Singh and M. Singh, "A burnout model of job crafting: Multiple mediator effects on job performance", IIMB management review, vol. 30, 2018, pp. 305-315.

[11] S. K. Singh, "Estimation of Calorie Content in a Diet Using Big Data" International Research Journal of Engineering and Technology (IRJET) https://irjet.net/archives/V7/i4/IRJET-V7I4733.pdf

[12] L. N. Tondji, "Web recommender system for job seeking and recruiting", 2018.

[13] A. Anandhan, L. Shuib, M. A. Ismail and G. Muijtaba, "Social media recommender systems: review and open research issues", IEEE, vol. 6, 2018, pp. 15608-15628.

[14] M. Narayanan and A. K. Cherukuri, "A study and analysis of recommendation systems for location-based social network with big data", IIMB Management review, vol. 28, 2016, 25-30.

[15] P. Kaur, "E-recruitment: A conceptual study, International journal of applied research" 2015, pp. 78-82.

[16] S. K. Singh, "Rain Fall Prediction Using Bigdata Analytics," International Journal of Innovations in Engineering and Technology (IJIET) http://dx.doi.org/10.21172/ijiet.151.07

[17] D. Wang, Y. Liang, D. Xu, X. Feng and R. Guan, "A content-based recommender system for computer science publications", Knowledge-based systems, vol. 157, 2018, pp. 1-9.

[18] K. Haruna , M. A. Ismail, D. Damiasih, J. Sutupo and T. Herawan, "A collaborative approach for research paper recommender system", vol. 12, 2017.