

Impact Of Machine Learning In Human Resource Management: Towards The Modernization Of Leadership

Rodrigo Miranda Cabrera¹, Manuel Leon Ganchozo², Rodríguez Salazar Oswaldo Rene³, Rafael Wilfredo Rojas Bujaico⁴, Héctor Huamán Samaniego⁵, John Fredy Rojas Bujaico⁶

¹Universidad Tecnológica Latinoamericana en Línea-UTEL México, rodimc@gmail.com

²Facultad de Ciencias de la Industria y Producción Universidad Técnica Estatal de Quevedo
<https://orcid.org/0000-0003-3265-1257>, mleon@uteq.edu.ec

³Universidad Católica Santa María Arequipa Perú, <https://orcid.org/0000-0002-7275-3605>,
orodriguez@ucsm.edu.pe

⁴Universidad Nacional de Huancavelica, <https://orcid.org/0000-0002-8426-1333>, rafael.rojas@unh.edu.pe

⁵Universidad Nacional del Centro del Perú, <https://orcid.org/0000-0003-0761-5000>, hhuaman@uncp.edu.pe

⁶Universidad Nacional de Huancavelica, <https://orcid.org/0000-0001-6614-9615>, john.rojas@unh.edu.pe

Abstract

In the era of fifth industrial revolution, organizations can definitely rely on technology to play a major role in human resource management. Internet of Things (IOT), Big Data, Cloud computing, Robotics and Artificial Intelligence are some of the technologies which are making this revolution a reality. Only a satisfied employee can be a productive employee and the retention rate is comparatively high for employees belonging to that category. Identifying the factors leading to satisfaction of employee will definitely assist management in introducing feasible factors among the ones identified. This is where machine learning comes into action. Machine learning techniques like classification and clustering play a major role in analyzing data as well as making effective predictions. In this paper an attempt is made to discuss the impact of machine learning on human resource management to control turnover of employees.

Keywords: Employee attrition, Machine Learning, Clustering, Classification.

1. Introduction

Employee turnover can be defined as “The proportion of the employees who leave an organization over a set period” [1], which is definitely unwelcoming for organizations. Every effort is taken by organizations to increase employee engagement and thus retain them once they prove their worth to the organization. In the era of fifth industrial revolution, organizations can definitely rely on technology to play a major role in this aspect. Internet of Things (IOT), Big Data, Cloud computing, Robotics and Artificial Intelligence

are some of the technologies which are making this revolution a reality.

Only a satisfied employee can be a productive employee and the retention rate is comparatively high for employees belonging to that category. Identifying the factors leading to satisfaction of employee will definitely assist management in introducing feasible factors among the ones identified. This is where machine learning comes into action. Machine learning techniques like classification and clustering play a major role in analyzing data as well as making effective predictions.

In classification algorithms, a computer program studies the existing data in order to understand their behavior and make predictions about the outcome of new set of data. The set of input data provided is divided into different classes or categories. The algorithm predicts the class into which the new data belongs.[3,9,10] Classification is considered as a supervised algorithm because the categories are already known. On the other hand, clustering is categorized as an unsupervised algorithm. Here the groups are formed by the algorithm based on similarity between the data items. In this paper an attempt is made to discuss the impact of machine learning on human resource management to control turnover of employees.

2. Literature review

In the article, “Predicting employee attrition using machine learning”, the authors are explaining about the research conducted to predict employee attrition using machine learning models. The researchers had used a dataset created by IBM. The study compared the performance of the models developed using Support Vector Machine, Random Forest and K Nearest Neighbor in the prediction of human resource attrition.[14]

In the article, “Applications of machine learning algorithms in online recruitment systems”, the authors are explaining about leveraging machine learning algorithms for evaluating the candidates in an online recruitment system. An e-recruitment system is proposed which incorporates machine learning for the task. The article explains the use of regression model in predicting the extraversion characteristic of the candidate. Text analytics is used on data obtained from personal blogs and LinkedIn profiles. A regression tree is created. Candidate ranking is thus automated. [16]

In the article, “On the Application of Fuzzy Clustering for Crime Hot Spot Detection”, a partitioning method called as Fuzzy clustering is

used for detecting the crime hotspot. The performance is compared with two hard clustering methods, k means and medoid. In fuzzy clustering, the elements can belong to more than one cluster. The degree of belonging of a member in a cluster is denoted by membership coefficient whose values range from 0 to 1. The experiments in the study is proving that fuzzy clustering is more appropriate for handling spatial data that nonhierarchical methods. [17]

3. Machine learning in human resource management

3.1 Recruitment

Selecting the right person for the right job is a very crucial function in human resource management. Highly successful companies like Citigroup, Morgan Stanley and Goldman Sachs are discovering the application using machine learning to predict the performance of new hires. The assistance of a machine tool which can predict the best choice of hire will lead to an appropriate selection. Machine learning based tools are used for screening the Curriculum Vitae of employees.[7] The machine learning models like association rule mining can correlate the skills possessed to the skills required for the job which increases the quality of candidate hiring. In association rule mining we can predict the happening of some event or item depending on the happening of other events or item. In the case of Curriculum vitae screening, association rule mining can predict the performance of employees based on certain specific skills possessed by them.

“Koru” is software that uses predictive analytics to forecast whether a candidate is correctly matching the requirement of the organization. The team identified the performance traits and shortlisted seven most important among them. The candidates were made to attend a test based on the identified parameters. The drivers of performance for each job were found using

association rule mining and customized tests were framed for specific jobs. The matching of the results were done using machine learning models for predicting the performance of employees. Machine learning algorithms like classification and clustering proved that the competencies required for performance is different for different jobs and also is different organization wise also. Companies like airbnb, deutsche bank are utilizing such software for recruitment.[2]

Data can be collected from employees and the test results are stored. The level of performance of the employees will be collected from employers. Using classification algorithm data is classified as low performers and high performers. Decision tree can be formed which can give idea about the factors having maximum influence on performance using measures like information gain.

In certain recruitment systems, web analytics is employed using machine language to extract details from profiles like LinkedIn and blogs to predict the personality of prospective candidates. Linguistic analysis will help to extract meaning from texts which might be structured or unstructured. Text analytics can be applied on this to understand the sentiments in that using the techniques of sentiment analysis. Sentiment analysis helps in identifying the sentiments behind the text. This will help in understanding the character of a prospective candidate by analyzing data from social network profiles. Using Natural language processing companies can understand the pulse of employees from the comments they put in various social media platforms. Sentiment analysis can identify the feelings of individuals from the analysis of the unstructured data.

Candidate ranking can be done using supervised machine learning algorithms like classification where the existing employees data where the feature set can be a labeled data set denoting the features of employees as well as whether the

employee was a good performer or not. Once the classification is done, the performance chance of test data can be performed to predict whether the prospective candidate is fit for the job. [6]

Using association rule mining algorithm in machine learning, job-finding sites can recommend opportunities to job seekers. This algorithm works by finding the association of recommended jobs with the searches done by the job seeker and also by tracking the clicks and various posts put up by them Phenom people is an example of machine learning based tool that connects right talents and right jobs.

3.2 Employee segmentation

We have heard a lot on customer segmentation where they are divided into segments according to which they are catered with their customized requirements. Same way employees can also be segmented based on criteria like demographic features, performance or gender and the like. [4] This can be done with the help of machine learning algorithms like k-means algorithm. Clusters can be formed using k-means algorithm according to the characteristics of employees which are stored in the dataset. [5] The organization can make advantage of this clustering by understanding the requirements of various clusters of employees. This understanding will help the organization in looking into genuine requirements and hence ensuring retention of employees.

3.3 Attrition prediction

Employee Attrition is a major problem faced by an organization, which turns out to be costly for organizations. It will be highly beneficial for organizations if the probability of separation of employees can be predicted. It can support organizations in devising plans for retention of those employees. This prediction can be effectively carried out by using machine learning algorithms like classification and clustering. In classification, the data regarding employees can be divided into training data and

test data. Machine algorithms are applied on training data to understand to classify the employees to different categories or classes based on the target variable. The test data can be used to check whether these classifications are right. Thus, a model can be developed to predict the results of new data sets. Using clustering algorithm, we can group employee data sets into different clusters based on the characteristics of employees. When a new employee data arrives, the algorithm can predict to which cluster it actually belongs. The factors leading to satisfaction of each cluster can be identified. This will give an insight into the methods to be adopted to improve retention and thus reduce attrition rate in organization.

Factors like distance from home, monthly salary, performance appraisals and the like have an enormous impact on the decision of employee to stay back in the organization. The impact of each factor on employee will be different. Readymade tools are available in software to find such correlations and interdependencies. Python which is a general-purpose programming language has many algorithms used for implementation of machine learning. The use of Python can make this process very easy. Python has got in built libraries which can discover the correlations between the features of employees and attrition.

3.3.1 Attrition analysis on a fictitious data set

The above-mentioned application of machine learning in attrition analysis can be demonstrated little more in detail with a fictitious HR dataset prepared by IBM data scientists which is available in Kaggle. The dataset has features like age, business travel, department, distance from home, monthly salary, performance rating and the like of which majority are numeric variables and some of them are categorical variables. The variable attrition represents whether the employee has left the organization or not. The variable which has values 'Yes' and 'No' are replaced by 1 and

0. Other categorical variables showing department, gender, job role etc. are changed to dummy indicator variables using the function `get_dummies()` from pandas library. For example, consider the variable gender which assumes values male or female which are categorical variables. While converting that to numerical variable, two fields as `gender_male` and `gender_female` are created. If the employee is female, `gender_male` will have value 0 and `gender_female` will have value 1. Same way for department also, if there are departments as hr, sales etc., there will be fields created as `department_hr`, `department_sales` etc. which assumes value 1 if the employee belongs to that particular department.

Some attributes in the data set may be having large range of values and some attributes might be having small range of values. This incongruity may lead to inappropriate predictions of attrition. For example, salary may be having a high range and `gender_male` may be having binary values 0 and 1. Normalisation process will scale the data so that they will fall in between a range like 0 to 1. Min max normalization is one of the methods for performing normalization. In Min-max normalization new value is found out by using the formula [11]

$$\text{Val1} = \frac{\text{val} - \text{minval}}{\text{maxval} - \text{minval}} (\text{new}_{\text{maxval}} - \text{new}_{\text{minval}}) + \text{new}_{\text{minval}}$$

Where `val` is the original value, `maxval` and `minval` are the maximum and minimum value of the attribute, `newmaxval` and `newminval` are the maximum and minimum in the new range specified.

For example, if salary is between 10,000 and 50,000, and the new range is [0,1] then the scaling for the value 12,000 happens as

$$\text{Val1} = \frac{12000 - 10000}{50000 - 10000} (1 - 0) + 0$$

The answer will be 0.05

There are various factors that have an impact on the turnover of employees. Demographic features like age, gender, income and the like affect the decision of attrition to a great extent. There exist positive correlations for certain variable and negative correlation of yet another set of variables with the target field, i.e. attrition. There are inbuilt functions available in python libraries to calculate numerically the correlation values from the available dataset. Function to find correlation will work only for numeric variables in python. This problem is taken care

of by using get dummies which does the conversion to numerical variable as mentioned earlier. for example, the field Marital Status which indicates the marital status of employee assumes three values as married, single and divorced. This is replaced by three fields Marital Status single, MaritalStatus_Married and MaritalStatus_Divorced. If the employee is not married, the field MaritalStatus_single will have value 1 and other two fields will be zero. The same is followed for other MaritalStatus values also. This method increases the number of fields, but all numerical operations can be performed by the conversion by adding dummies.

Training received over the last 12 months	Learning organisation	Lean organisation	Taylorist organisation	Simple organisation	All organisations
None	52.6	62.5	83.5	83.4	65.5
1–9 days	29.8	23.1	11.2	12.7	22.1
At least 10 days	17.6	14.4	5.3	3.9	12.4

Table.I Method of Increasing Field

The variables Environment Satisfaction, Job Involvement, Job level, Job Satisfaction, Monthly Income, Percent Salary Hike, Total Working Years, Years In Current Role are having a negative correlation. When there is an increase in these factors the chance of attrition decreases.

The correlation also shows that the chance of attrition is high for employees who travel frequently, employees from sales department, male employees, laboratory technicians, sales representatives, employees with marital status as single and for employees who do more overtime.

The information thus obtained through analysis is very vital for organizations for planning employee retention programs to reduce the turnover of employees. Machine language is paving way of getting right information at the right time for a right decision.

While doing an analysis, visual representation gives a clearer picture for further action. The bar graph below represents the relationship between marital status and attrition. The plot shows that attrition is high for employees who are not married. The plotting is done with the help of **seaborn** which is a statistical data visualization library available in python.

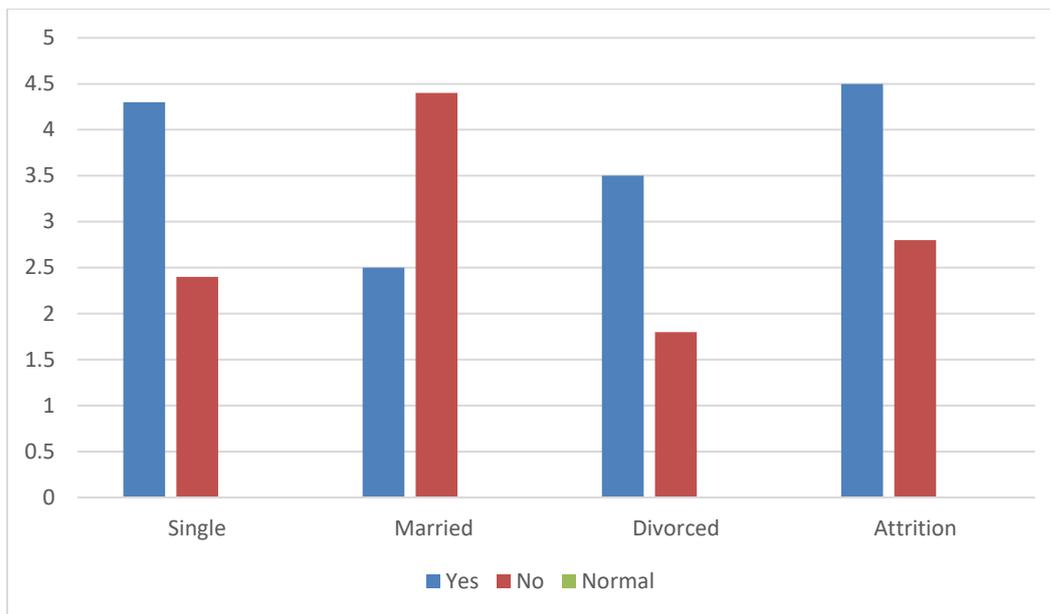


Figure 2: Plot showing the relationship between employees leaving the company with respect to marital status.

Business travel is another factor having an impact on attrition of employees. The figure below shows visual representation of number of employees where attrition = 1(yes) and number

of employees where attrition=0(No) for each category of Business travel as Travel_Rarely, Travel_Frequently and Non_travel. The figure depicts that the attrition percentage is more for employees who travel frequently. The plot can again be drawn with seaborn.

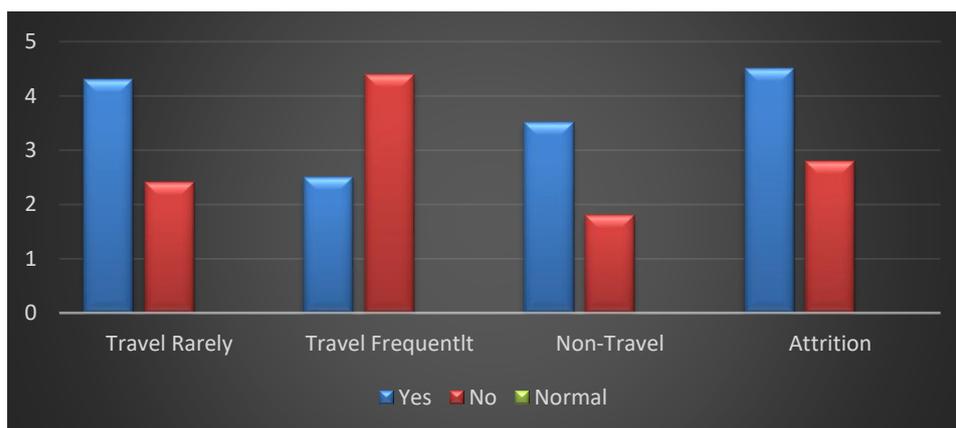


Figure 3: Plot showing the relationship between employees leaving the company with respect to business travel.

Job role also has impact on attrition of employees. The figure below demonstrates

visual representation of number of employees where attrition = 1(yes) and number of employees where attrition=0(No) for each category of JobRole. The figure depicts that the attrition percentage is more for employees who work as Sales Representatives.

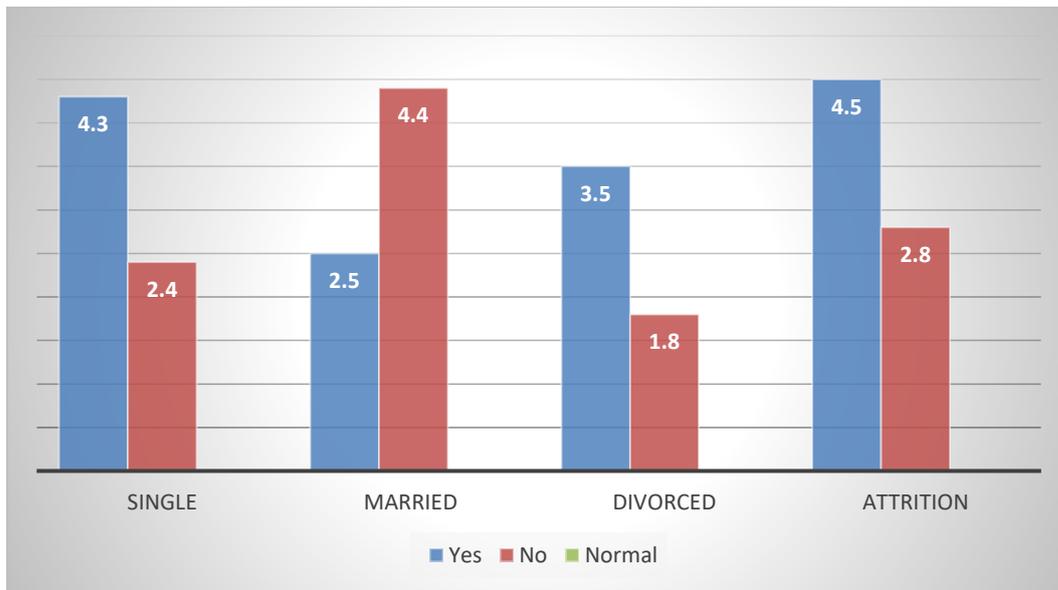


Figure 3: Plot showing the relationship between employees leaving the company with respect to job role.

3.3.1.1 Finding and visualizing important features using random forest

Random forest is a supervised machine learning algorithm. Random forest proposes certain methods for extracting the most important

features which are having more impact on the target variable. In our example we can extract important features having an impact on the attrition of an employee. The understanding of features of priority will help the employer to concentrate on improvement of these features so that the employee will stay back in the organization.



Figure 5: Plot showing most important 20 attributes affecting Attrition as given by random forest algorithm

It is evident from the following figure that Monthly Income is having a huge effect on the level of attrition. Other most important nineteen more features are also depicted in the figure. This will provide insight to the strategic planning members of organization for attaching required priorities to the benefits provided.

3.3.1.2 Building an attrition prediction model

Apart from understanding the features regarding employees and its impact on attrition, machine learning algorithms can also build attrition prediction models which can predict whether an employee would prefer to continue in the organization or there exists a probability for the employees to leave. These models can be developed using programming languages like R or Python. Data mining techniques like classification or clustering is performed in the models.

While implementing classification, the whole dataset can be divided into training data and test data. Training data identifies a pattern of behavior to classify the dataset based on the attrition value as 1 or 0 denoting whether employee has resigned or is still continuing in organization. The test data can be tested to see the accuracy of the proposed classification. This classification will help in the prediction of attrition in a new dataset.[9]

Naïve Bayes, Support Vector Machine, K Nearest Neighbour, Logistic Regression, MLP Classifier, XG Boost are some of the classification algorithms used for building prediction models. The performance of the algorithms shows variation in different data sets. XGBoost has exhibited good performance with attrition data sets. Performance of algorithms can be compared by measuring accuracy, model

runtime and memory utilization.[12][13][14][15]

Clustering techniques creates clusters of datasets showing certain patterns of behavior. It is an unsupervised algorithm. [10] The attrition prediction of new data can be predicted by identifying the cluster to which it belongs. K means algorithm is a very popular clustering technique where the entire data set is divided into k clusters. A data record will be belonging to one of the k clusters. K means is a non-hierarchical clustering technique. Another popular clustering is Fuzzy Hierarchical clustering where a data record can belong to more than one cluster. A membership value is associated with each data record to identify the level of proximity to clusters. [17] Fuzzy algorithms can work on categorical and numerical data whereas in k means only numerical data is accepted. But because of the calculations involved it might consume more time than K Means.[18]

4. Conclusion

Machine learning can change the way the human resource management domain functions in an organization. It is making changes in all aspects of human resource management starting from human resource planning. Enormous data is available in human resource information systems (HRIS) available in organizations. Personal data of employees are also available in their social networking sites like LinkedIn, Facebook and also in blogs. It is up to the organization to unleash the potential of such data and convert that into valuable insights and predictions for a better functioning of organization.

Reference

1. For a useful survey of the ‘high performance work practice’ literature, see Wood (1999). Key contributions include Becker and Huselid (1998), Guest (1997),

- Huselid (1995), Ichiniowski et al. (1997) and Osterman (2000).
2. The initial findings of the survey are presented in a European Foundation report by Merllie and Paoli (2001).
 3. The factor analysis method used here is multiple correspondence analysis, which is designed for the analysis of categorical variables. See Benzecri (1973) and Greenacre (1993: 24–31).
 4. The positive coefficient for the use of the ‘learning’ forms in Sweden is no longer significant at the 0.05 level and the positive coefficient for the use of Taylorist forms in Spain is no longer significant at the 0.05 level. For a further discussion of these results, see Lorenz and Valeyre (2004).
 5. Denmark is clearly somewhat of an outlier in terms of the relation we are proposing between employment protection and the relative importance of the lean model of work organisation. A distinctive feature of the Danish institutional set-up is that while employment protection is relatively low, unemployment protection is among the highest in Europe. See Lundvall (2002) and Hall and Soskice (2001: 167–9).
 6. CEC ‘The 1998 Employment Guidelines—Council Resolution of 15 December 1997’ (1998).
 7. CEC ‘Partnership for a New organisation of Work—Green Paper’ (Bulletin of the European Union, Supplement 1/97, 1997).
 8. Often cited national programs include the New Work Organisation in Ireland Programme (Savage 1999) and the Finnish Workplace Development Programme (DGESA 2000).
 9. Muthukumar, V. and Ezhilmaran, D., 2020. A Cloud-Assisted Proxy Re-Encryption Scheme for Efficient Data Sharing Across IoT Systems. *International Journal of Information Technology and Web Engineering (IJITWE)*, 15(4), pp.18–36.
 10. MUTHUKUMARN, V., ADHIYAMAN, M. and EZHILMARAN, D., A SECURE AND ENHANCED PUBLIC KEY CRYPTOSYSTEM USING DOUBLE CONJUGACY SEARCH PROBLEM NEAR-RING, *Advances in Mathematics: Scientific Journal*, 9(3), pp. 1389–1395, 2020.
 11. Basheer, Shakila; Bivi, S Mariyam Aysha; Jayakumar, S; Rathore, Arpit; Jeyakumar, Balajee. Machine Learning Based Classification of Cervical Cancer Using K-Nearest Neighbour, Random Forest and Multilayer Perceptron Algorithms, *Journal of Computational and Theoretical Nanoscience*, Volume 16, Numbers 5-6, May 2019, pp. 2523-2527(5).
 12. Rohit Punnoose, PankajAjit, “Prediction of Employee Turnover in Organizations using Machine Learning Algorithms”, *(IJARAI) International Journal of Advanced Research in Artificial Intelligence*, Vol. 5, No. 9, 2016
 13. Ibrahim Onuralp Yiğit , Hamed Shourabizadeh , “An Approach for Predicting Employee Churn by Using Data Mining”, *IEEE, 2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*
 14. Sarah S. Alduayj, Kashif Rajpoot, “Predicting Employee Attrition using Machine Learning”, *IEEE, 2018 International Conference on Innovations in Information Technology (IIT)*
 15. Rachna Jai and Anand Nayyar ,”Predicting Employee Attrition using XGBoost Machine Learning Approach”, *IEEE, Proceedings of the SMART–2018, IEEE Conference ID: 44078, 2018 International Conference on System Modeling & Advancement in Research Trends, 23rd–24th November, 2018*
 16. Evanthia Faliagka, Kostas Ramantas, Athanasios Tsakalidis , Giannis Tzimas, “Applications of machine learning algorithms in online recruitment systems”, *ICIW 2012 : The Seventh International Conference on Internet and Web Applications and Services*
 17. Sultana, H Parveen; Shrivastava, Nirvishi; Dominic, Dhanapal Durai; Nalini, N; Balajee, J.M. Comparison of Machine

- Learning Algorithms to Build Optimized Network Intrusion Detection System, Journal of Computational and Theoretical Nanoscience, Volume 16, Numbers 5-6, May 2019, pp. 2541-2549(9).
18. ShuchitaMishra,TanviPradhan,PriyankaPar
mar,SuvarnaMaji,Ekta Sarda, “Crime
Prediction Using Fuzzy c-means
Algorithm”, International Research Journal
of Engineering and Technology (IRJET),
Volume: 05 Issue: 02 | Feb-2018