International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Team Formation and Exploration of Employee Absenteeism in Teams Through Machine Learning and Multilayer Perceptrons (MLP)

Amrita Dutt[1], Vijaishri Tewari[2], Shashikant Rai[3]

[1]Research Scholar, Indian Institute of Information Technology
[2]Professor, Indian Institute of Information Technology
[3]Assistant Professor, Indian Institute of Information Technology

Abstract:
Organizations are based on employee productivity and team production. The challenge in team production is the formation of teams and absenteeism of employees in teams results in lower satisfaction of employees. Researchers today have found the importance of machine learning in our daily lives to improve quality. The Present study is the prediction of employees that come under the same manager and analysis of employee absenteeism through multi-layer perceptron (MLP) neural networks from the HRM dataset of 311 unique instances for the analysis of employee performance. After training the data with the aforementioned parameters, the results of linear regression obtained by testing data is a Mean absolute error: 1.44, the Mean absolute error: 1.68 of Bayesian Regression.

Keywords: Team formation; Employee absenteeism; Machine learning

Introduction:

Formation of the team is a problem for many years in various organizations, the challenge in the formation of teams is to create small teams that have the required set of skills for the success of the team (Keane et al., 2020) team formation from any database is a fundamental problem such as under same managers, employees are considered as a team or not. To measure the rate of organization success and customer satisfaction, there is one important factor that comes with great effect is the performance of employees. employee engagement is an important factor for the success of an organization (Arai et al., 2019) An employee is an important asset for any organization (Mohbey, 2020) as labor is an important factor in the economy, put in human efforts to ensure the success of the organization. An employee with ineffective performance will disrupt the production which in turn affects the profit that may harm the success of the organization. However, every organization expects maximum profit and reduced cost management which create a problem for employee effectiveness which is an important factor for the growth of an organization (Lather et al., 2019).one of the important factors that affects the effectiveness of employee is absenteeism (Qaisar, 2019). understanding the cause and effect of absenteeism may help to enhance the success of the organization to maximize profit and customer satisfaction.

This paper aims to help the human resource department of any organization in finding factors responsible for absenteeism using Multilayer perceptron (MLP) neural network, through the formation of teams. The

research was conducted on the dataset of HRM provided on Kaggle and explored the same factors that have a good correlation with absenteeism that helped in gathering results.

Employee Absenteeism in Teams:

Related work:

Machine learning has been used in many fields, in organizations nowadays such as predicting and analyzing employee performance (Jayadi, 2019), also a lot of research efforts have recently been devoted to the problem of formation of the teams (Lappas et al., 2009) for Human resource department in any organization. the utility of Human Resource management in operating, planning, developing, and identity connection between the employee performance and teams has been demonstrated by a lot of researchers (Radha K, 2021). Moreover, many of these studies concentrate on analyzing and evaluating employee behavior and attitudes for the performance, do not discuss the organization's main system, as portrayed by employees. As every organization, are built up around the teamwork system, many studies are on the formation of teams, Recommendations for teams (Petkovic et al., 2012), the existing study found the employee's trends and factors responsible for absenteeism while working in the team, additionally, author concentrated solely on forming Teams from Human Resource data set of the organization for predicting and analyzing employee absenteeism.

Data dictionary:

HR datasets are hard to find and it can be quite challenging to gather HR data from a company. However, Dr Carla Patalano and Dr Rich have created a publicly available HR dataset which is available at Kaggle. The dataset contains multiple attributes such as names, DOBs, age, gender, marital status, date of hire, reasons for termination, department, whether they are active or terminated, position title, pay rate, manager name, performance score, absences, Most Recent Performance Review Date, and Employee Engagement Score. This dataset contains 311 unique instances.

To form the teams out of the dataset, we have assumed that people working under the same manager form a team. We have used manager IDs to form the teams and then aggregate the attributes to get the team attributes. Before forming the teams, we have preprocessed data to remove those instances which have null values.

Since not all attributes are useful, we have filtered out extra attributes and have kept only the following attributes in the final data.

These are columns in the dataset that is used for analysis:

- PerfScoreID
- Salary
- EngagementSurvey
- EmpSatisfaction
- MarriedID
- MaritalStatusID
- GenderID
- EmpStatusID
- PositionID

We have aggregated different attributes differently to obtain the results for teams. We have taken the mean of Salary and Absences while for all other attributes we have taken the mode of the data.

Analysis

To carry out the analysis we performed descriptive analysis and generated a correlation heatmap for different attributes.

The correlation attribute is given as follows.

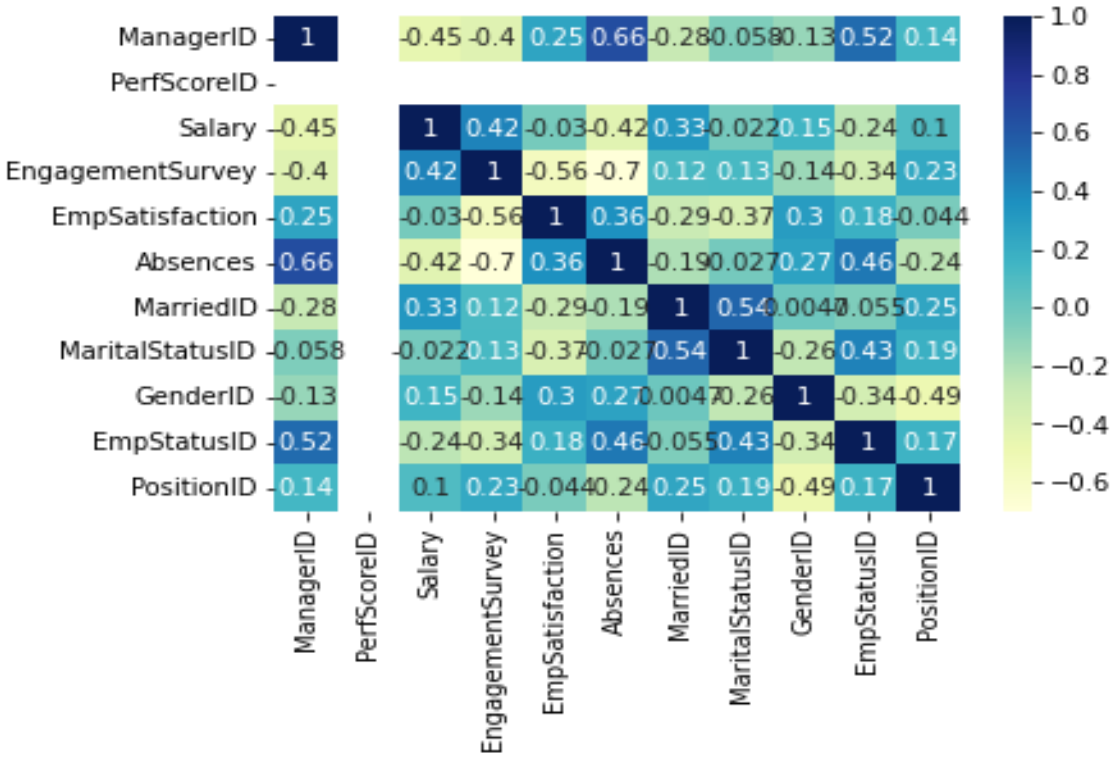
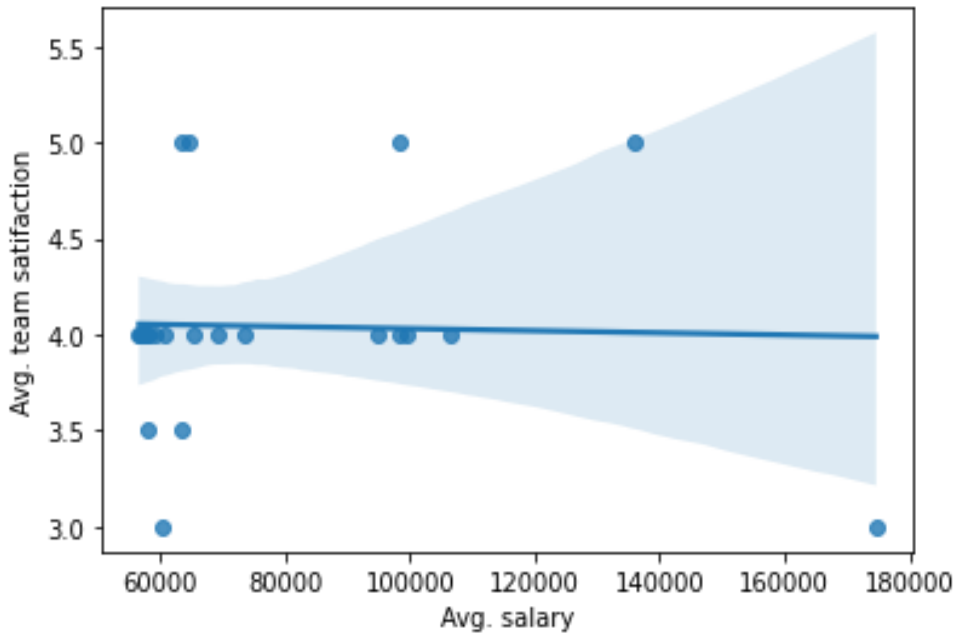


Fig 1. Correlation Heatmap

Some other important observations were drawn by the following graphs.

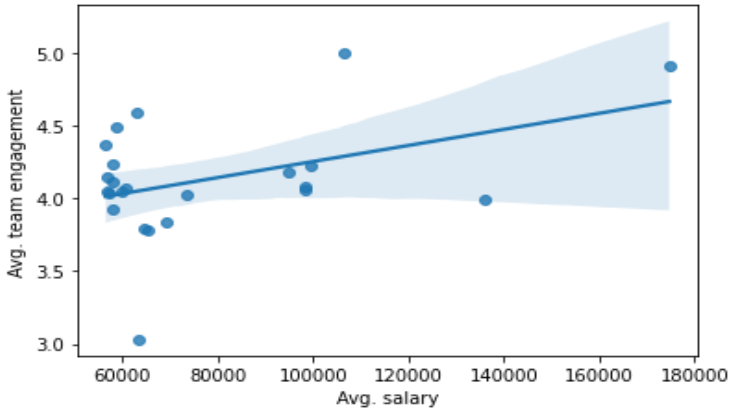
Effect of salary on Satisfaction:

The Average team satisfaction doesn't seem to vary significantly with the Average team salary.



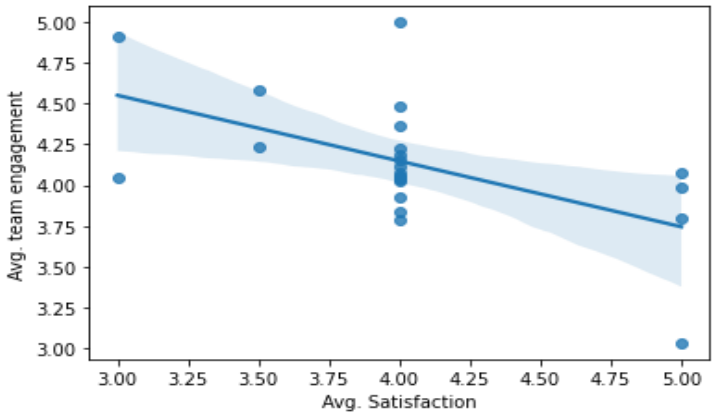
Effect of salary On Engagement:

Average team engagement increases with an average salary.



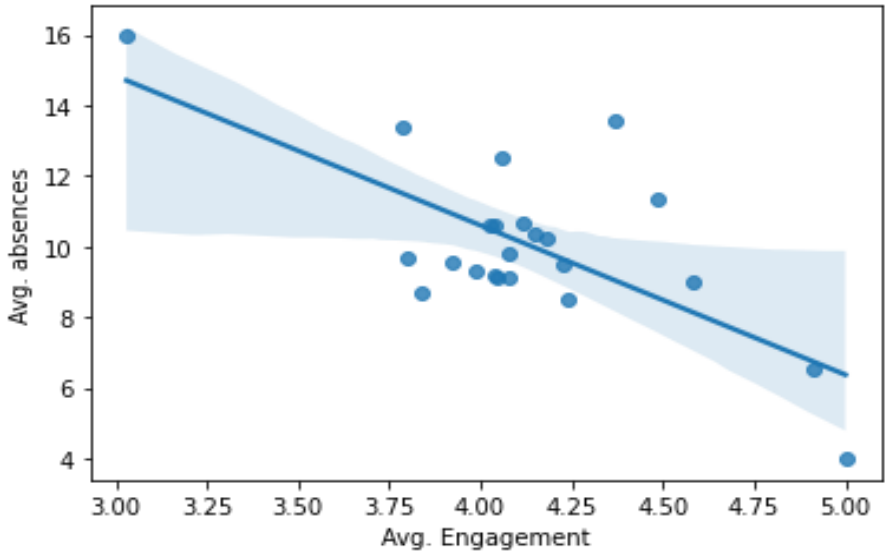
Effect of engagement on Satisfaction:

Average satisfaction increases when average team engagement decreases.



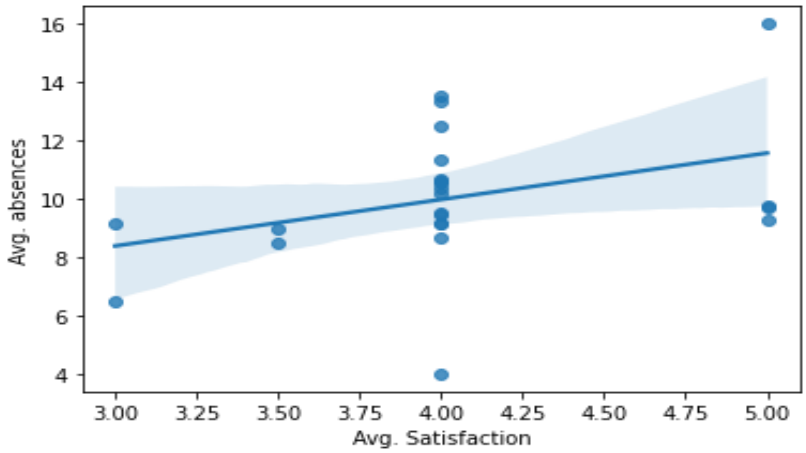
Effect of Engagement on absenteeism:

More Average Engagement leads to lesser absenteeism.



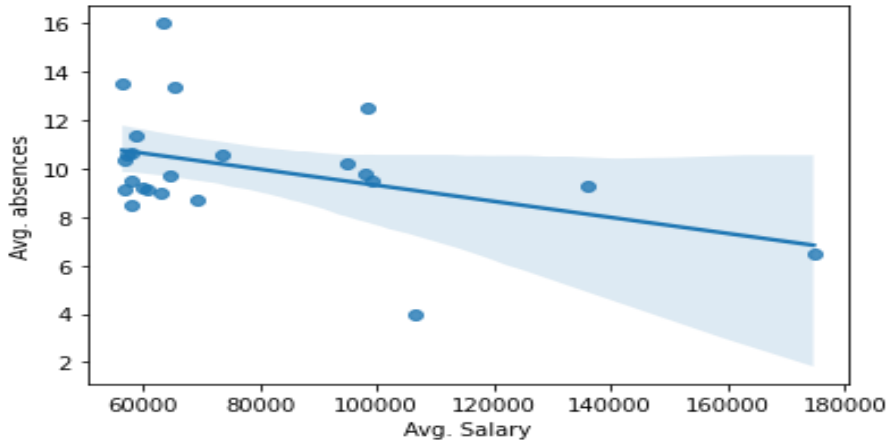
Relationship of absenteeism with satisfaction:

Surprisingly, average satisfaction increases with more absenteeism.



Effect of Salary on absenteeism:

Absenteeism decreases with an increasing average salary.



Effect of Gender on absenteeism:

Teams with more males have more absenteeism.

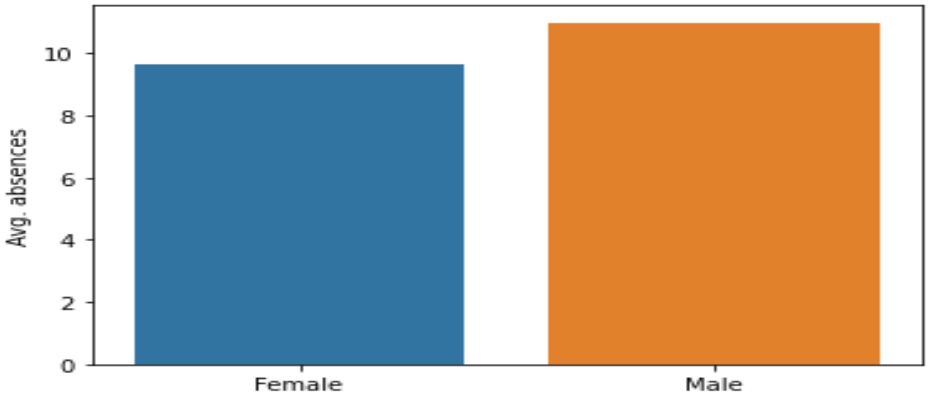


Fig: Seaborn Barplots between Gender and Absenteeism

Insight: From the graph, we could see that gender have an impact on absenteeism

Proposal: The employee should be motivated for more engagement at the workplace.

Effect of Marital status on absenteeism:

Teams with more married people have more absenteeism.

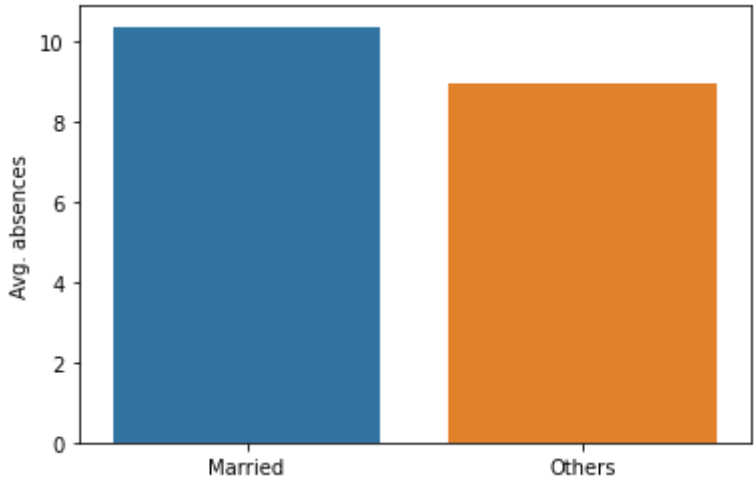


Fig : Seaborn Barplot between Marital status and absenteeism.

Insight: From the graph, it is clear that those employees who are married, record more absences at the workplace.

Proposal: Work-life balance policy should be opted by the organization ((Radha K, 2021)and also flexibility in working hours should be provided to the married employees to motivate them

Methodology

Based on the above observations, a pattern can be observed between different variables and it seems feasible to predict the absenteeism on the basis of other variables. We have employed several different models to predict the average absenteeism in teams. The following models are used.

- **Linear Regression:** Linear Regression is a supervised statistical model used to predict the unknown variable based on a number of variables. The model is trained on the collected data and can then be used to predict the unknown variable. We have used linear regression with L2 normalization.
- **Bayesian Regression:** Bayesian regression is a probabilistic model which is based on Bayes probability theory. It uses the assumption of class independence.
- **Multilayer perceptron:** MLP or Multi Layer Perceptron is a neural network with multiple layers of nodes. The model is trained on the known data in a supervised manner. During training, weights are adjusted which are used in the prediction of the unknown variable.

Usually with the regression models, R2 score is used as a metric of performance of the model but since the dataset is small, it was not feasible to use R2 score. We have therefore used 'average absolute error' to test the accuracy of the model. The formula for the same is given below.

$$e = \frac{1}{n}\sum_{i=1}^{n} |c - v|$$

where,

$$e : mean\ absolute\ error$$
$$c: actual\ data\ value$$
$$v: predicted\ data\ value$$

Before training the model, the data was normalized using *min-max normalizer*. The data was then divided into training and test sets.

Results

1. Linear Regression:

We have used the *sklearn* library to implement the linear regression. We have used the following parameters in the linear regression model.

- alpha: 100
- maximum iteration: 500
- After training the data with the aforementioned parameters, we obtained the following results on the test data.
- Mean absolute error: 1.44

2. Bayesian Regression:

We used the sklearn library's inbuilt bayesian ridge regression to implement the bayesian regression. The following parameters were set to the below mentioned values.

- alpha_1=1e-06
- alpha_2=1e-06
- fit_intercept=True
- lambda_1=1e-06
- lambda_2=1e-06
- n_iter=300

After training the data on the train set, we obtained the following results on the test set.

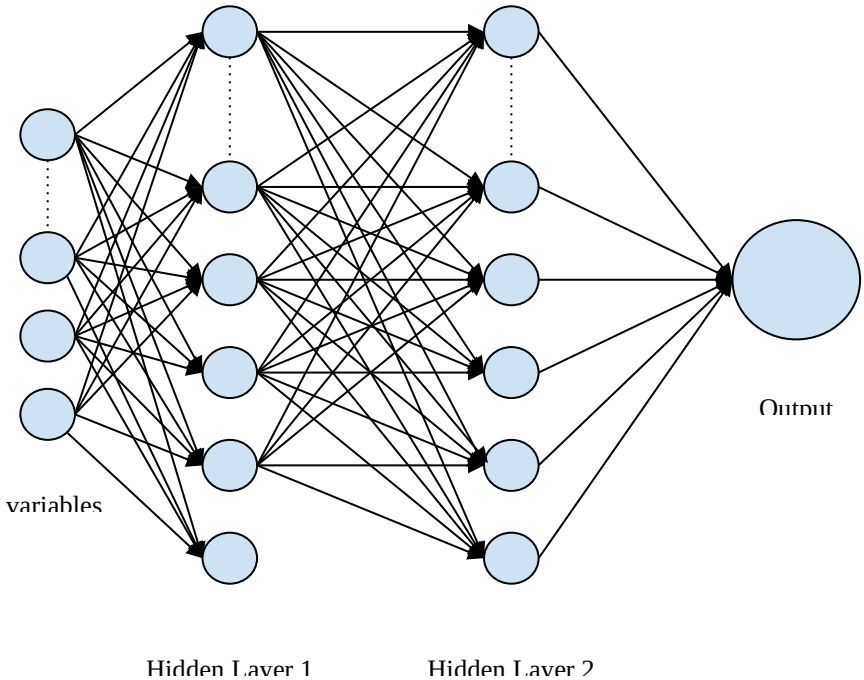
- Mean absolute error: 1.68

3. Multi-Layer Perceptron (MLP):

We used the sklearn library to implement the multilayer perceptron. The MLP is trained with the following parameters.

- Number of hidden layers: 2
- Number of nodes in first hidden layer: 100
- Number of nodes in second hidden layer: 100

The MLP is depicted in the figure given below.



The following results were obtained after training the MLP on training data and testing it on test data.

- Mean absolute error: 0.1829

Conclusion:

Performance analysis is an important factor to measure the growth of organizations well as individuals. The major factors related to enhancing the performance are job satisfaction and job engagement of employees. Since manual analysis through different tools is dull and tiresome work for the HR department. So, attempt towards machine learning to analyze the data related to HR can save time and give more accurate results to develop the relationship between important factors of employee performance. Important measures should be taken for factors like job satisfaction and employee engagement while working in a team to avoid absenteeism at the workplace. The method used for analysis is new and different for the formation of teams and analysis of absenteeism of employees, which gives a different perspective to the researchers in the future. In terms of job satisfaction of employees, the proper measures should be taken towards absenteeism to enhance the performance.

Reference:

1. Keane, P., Ghaffar, F., & Malone, D. (2020). Using machine learning to predict links and improve Steiner tree solutions to team formation problems - a cross company study. *Applied Network Science*, *5*(1), 57. https://doi.org/10.1007/s41109-020-00306-x
2. Al-Zibdeh, A.A., Hassanein, R.A.A., Al-Qassas, S.A. and Tir, F.N.A., 2021. Workplace Absenteeism Prediction using ANN.
3. Liu, H., Qiao, M., Greenia, D., Akkiraju, R., Dill, S., Nakamura, T., Song, Y. and Nezhad, H.M., 2014, December. A machine learning approach to combining individual strength and team features for team recommendation. In *2014 13th International Conference on Machine Learning and Applications* (pp. 213-218). IEEE.

4. Arai, K., Kapoor, S., & Bhatia, R. (Eds.). (2019). *Intelligent Systems and Applications* (Vol. 869). Springer International Publishing. https://doi.org/10.1007/978-3-030-01057-7
5. Mohbey, K. K. (2020). *Employee's Attrition Prediction Using Machine Learning Approaches* (pp. 121–128). https://doi.org/10.4018/978-1-7998-3095-5.ch005
6. Lather, A. S., Malhotra, R., Saloni, P., Singh, P., & Mittal, S. (2019). Prediction of employee performance using machine learning techniques. *Proceedings of the International Conference on Advanced Information Science and System*, 1–6. https://doi.org/10.1145/3373477.3373696
7. Qaisar, A. S. (2019). *Predicting Absenteeism at Work Using Machine Learning Algorithms*. https://doi.org/10.18081/2226-3284/16-12/1-12
8. Dataset : https://www.kaggle.com/rhuebner/human-resources-data-set
9. Radha K, R. m. (2021). An Experimental Analysis of Work-Life Balance Among The Employees using Machine Learning Classifiers. International Journal of Computer Trends and Technology. https://doi.org/https://www.ijcttjournal.org/archives/ijctt-v69i4p108
10. Jayadi, R. (2019). Employee Performance Prediction using Naïve Bayes. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(6), 3031–3035. https://doi.org/10.30534/ijatcse/2019/59862019
11. Lappas, T., Liu, K., & Terzi, E. (2009). Finding a team of experts in social networks. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '09*, 467. https://doi.org/10.1145/1557019.1557074
12. Petkovic, D., Okada, K., Sosnick, M., Iyer, A., Zhu, S., Todtenhoefer, R., & Huang, S. (2012). Work in progress: A machine learning approach for assessment and prediction of teamwork effectiveness in software engineering education. *2012 Frontiers in Education Conference Proceedings*, 1–3. https://doi.org/10.1109/FIE.2012.6462205