Computer Engineering and Applications Vol. 8, No. 3, October 2019

# Predicting the Occurrence and Causes of Employee Turnover with Machine Learning

Xiaojun Ma[1*], Shengjun Zhai[2], Yingxian Fu[3], Leonard Yoonjae Lee[4], and Jingxuan Shen[5]

[1]*Carnegie Mellon University, Pennsylvania, USA*
[2]*The University of Chicago, Illinois, USA*
[3]*Temple University, Pennsylvania, USA*
[4]*Seoul International School, Seongnam, Korea*
[5]*Dalian Royal School, Dalian, China*
[*]*villaroelandrei@gmail.com*

## ABSTRACT

This paper looks at the problem of employee turnover, which has considerable influence on organizational productivity and healthy working environments. Using a publicly available dataset, key factors capable of predicting employee churn are identified. Six machine learning algorithms including decision trees, random forests, naïve Bayes and multi-layer perceptron are used to predict employees who are prone to churn. A good level of predictive accuracy is observed, and a comparison is made with previous findings. It is found that while the simplest correlation and regression tree (CART) algorithm gives the best accuracy or F1-score, the alternating decision tree (ADT) gives the best area under the ROC curve. Rules extracted in the if-then form enable successful identification of the probable causes of churning.

**Keywords**: Employee turnover; machine learning; decision tree; multi-layer perceptron; rules.

## 1. INTRODUCTION

Employee turnover is an important issue in all organizations. Voluntary or forced employee turnover impedes the company's growth. A high turnover rate increases human resource costs [1], and employee loss can adversely affect organizational productivity by necessitating additional time and effort to replace skilled workers.

Several issues arise from the employee perspective. There may be diverse reasons for employees quitting their organizations. Better offers including higher, better career opportunities, and more attractive locations can be favorable reasons [2], and negative ones may include unhealthy personal relations with supervisors or peers and bad and unsafe workplace environments. Cotton and Tuttle [3] suggested age, tenure, pay, overall job satisfaction, and perceived fairness as major causes of employee turnover. Supervision, recognition, and growth potential are also considered key factors influencing turnover [4-7].

A good relationship between supervisors and employees can foster healthy workplace environments by reflecting trust, respect and loyalty, producing a stronger team feeling and higher job satisfaction and ultimately reducing turnover intentions. Job satisfaction is considered one of the most important factors in voluntary employee turnover [8]. Job satisfaction is a key predictor of employee turnover [9]. Employee turnover may be reduced through direct or indirect leader

support [10]. The Leader-Member Exchange prediction model [11] considers the supervisor-subordinate relationship quality.

Employee turnover is problem not only for individual organizations but also for the economy and society as a whole since it can adversely affect long-term growth. Churning employees can affect workplace morale and disrupt work [12]. The multidimensional nature of employee churn has thus gained considerable research attention. In contrast, little work has been done in the field of employee churn prediction [1]. This calls for the need to better predict employee churn to better prevent its occurrence.

Predicting human behavior accurately is hard. However, information technology and its rapid advances and accumulating data from the workplace, together with machine learning algorithms, enable the mapping of behavioral patterns of employees to certain categories and predict their turnover intentions. Collecting data from employees through a friendly interface and selecting an appropriate predictive algorithm is challenging. Here discrete examples include the "happiness app" [13] for collecting feedback from employees following each day's work, and a gradient boosting algorithm for generalizing and classifying feelings on noisy data from numerous employees from various organizations [1].

Human Resource Predictive Analytics (HRPA) is a multidimensional approach to assimilating information and tools for predictive modeling for forecasting employee turnover. HRPA creates a path from data to insights for employees' behavioral intentions [14]. Planned behavior theory [15] suggests that behavioral intentions predict actual behaviors. This suggests that predictive analytics from human resource management and machine learning can be used to predict turnover intentions. Ajit [12] used the extreme gradient boost technique to predict employee churn by collecting data from a global retailer and comparing algorithm performance with various classifiers.

The present paper uses six different techniques to predict employee churn and identifies key features and a necessary and sufficient subset of features for equivalent prediction for the whole dataset. Two rule sets obtained from two algorithms are considered to better understand the causes of churn among different employees.

## 2. MATERIAL AND METHODS

### 2.1. OVERVIEW OF THE DATASET

The data used was obtained from feedback of 3887 employees from 34 different organizations in Barcelona, Spain. The feedback was collected from 10$^{th}$ July 2014 to 8$^{th}$ March 2017. These companies were multinational as well as Spanish and belonged to various sectors including consulting, information technology, electronic payment, manufacturing, retail, tourism, and education. The companies were participating in quality improvement programs or Kaizen, from where job satisfaction data were collected. The dataset was anonymous to protect private information. This is a publicly available dataset [16] consisting of 4 tables: churn, posts, comments and votes. There is churn when an employee leaves (or is terminated from) the organization. Each employee was asked to open a mobile app [13] and answer the question "How happy are you at work today?" The answer could be chosen from one of four icons (Figure 1).


ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

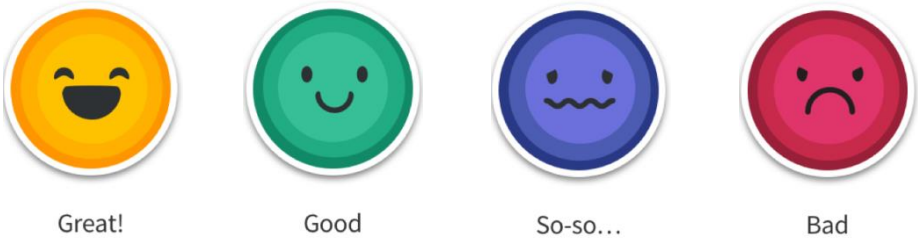


FIGURE 1. Four icons to choose from when voting on current state of happiness

After voting, the employee was shown another screen to record comments or suggestions. Finally, the third screen allowed the employee to see coworkers' comments and suggestions and indicated "like" or "dislike."

## 2.2. FEATURES IN THE DATASET

There are 35 input features and 1 target attribute in the dataset (Table 1). All input features are numeric and fall into one of three categories: (i) employee, (ii) company, and (iii) employee-company. Here 13 employee features are unique to each employee, whereas 18 company features are typical of a company, retaining the same values for all employees of the company. Features 1–13 denote the sum, mean, standard deviation (sd) and counts of happiness (hap), the comment length (len), number of comments (cmt) or characters in a comment (char), and likes, total likes and dislikes (totld) and ratio of likes (*i.e.* likes/ totld). The corresponding features for the company are represented in features 14–26 of Table 1, along with five other company-level features including no. of employees in that company, days since last interaction, average daily app use by employees, average number of daily posts per employee and total hap votes received per employee. The third category of features is employee values normalized by corresponding company features. For example, "e/c hap mean" denotes employee mean happiness divided by company mean happiness. There are four such features 32–35 in Table 1, denoting the mean and standard deviations of happiness, average length of a comment and average comment frequency for every employee divided by the corresponding values for the employed company.

Feature 36 is the class variable requiring to be learned and/or predicted for a value of either "yes" or "no." The positive class is "yes" since it indicates employees who churn and represents about 6% of total employees. The dataset with these 36 features plus employee ID and company ID as two features [16] is an open domain dataset preprocessed using the original four data files [17].

TABLE 1.
Features used for prediction

| No. | Var. short name | Feature type | No. | Var. short name | Feature type |
|---|---|---|---|---|---|
| 1. | emp hap mean | Employee | 19. | com len sd | Company |
| 2. | emp hap sd | Employee | 20. | com daily cmt | Company |
| 3. | emp hap count | Employee | 21. | com daily char | Company |
| 4. | emp len sum | Employee | 22. | com like sum | Company |
| 5. | emp len mean | Employee | 23. | com like mean | Company |
| 6. | emp len sd | Employee | 24. | com like sd | Company |
| 7. | emp daily cmt | Employee | 25. | com totld | Company |
| 8. | emp daily char | Employee | 26. | com lik/totld | Company |
| 9. | emp like sum | Employee | 27. | no. of emp | Company |
| 10. | emp like mean | Employee | 28. | days last int | Company |
| 11. | emp like sd | Employee | 29. | daily app use | Company |
| 12. | emp totld | Employee | 30. | daily posts | Company |
| 13. | emp lik/totld | Employee | 31. | total hap votes | Company |
| 14. | com hap mean | Company | 32. | e/c hap mean | Employee-Company |
| 15. | com hap sd | Company | 33. | e/c hap sd | Employee-Company |
| 16. | com hap count | Company | 34. | e/c cmt mean | Employee-Company |
| 17. | com len sum | Company | 35. | e/c cmt freq | Employee-Company |
| 18. | com len mean | Company | 36. | Churn | Output (class) |

## 2.3. FEATURE SELECTION

Assessing the importance of variables in a dataset most closely related to the target or class variable is crucial for data introspection or exploratory data analysis. It is also considered a required preprocessing step in many machine learning algorithms since a selected subset of features can significantly improve prediction accuracy. In addition to selecting key features, an appropriate subset of features effectively addresses redundancy in variables by ignoring a dependent variable, rendering equivalence relations unique, not dubious.

This paper uses a supervised subset evaluation technique, namely CFS subset evaluation [18], to select a set of variables that can predict the class (*i.e.* churn) with nearly the same accuracy as the whole dataset.

## 2.4. CLASSIFICATION

The main objective of this paper is to predict which employees are likely to churn. For this, 6 different algorithms are used as follows:

#### Classification and Regression Trees (CART)

CART is a simple but effective classification algorithm [19] that constructs tree-like checks to determine whether the value of a particular variable is less than a given value. If yes, then it branches off to one side, and if not, the branching goes to the other side. This is recursive until a decision is finally reached. The most suitable variable and its corresponding value for branching are selected so as to maximize the Gini purity [20] at its branch nodes:


ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

$$\text{Gini purity Index} = \sum_{i=1}^{C} \wp_i^2 \quad (1)$$

where $\wp_i$ is the probability of finding an observation of class $i$ in a node and $C$ is the number of classes in the data. The algorithm is nonparametric and hence insensitive to scale or shift. CART is also least affected by the presence of a few outliers, making it a good classification algorithm.

**Random Forest (RF)**

A single tree may lead to general rules and overlook the presence of certain pockets in the dataset revealing significant patterns or relations in input-output variables. To address this limitation, many trees can be built such that each tree is based on a sample of the whole dataset instead of all observations. This makes the RF algorithm [20] filter out possible noise in the dataset, revealing important patterns in knowledge discovery. Another advantage is that the time needed for constructing a tree is proportional to the cube of the number of instances [21]. That is, a 25% sample allows the construction of 64 different trees in the same amount of time as one tree from the whole data, where each observation participates 16 times, thereby providing more dynamic rules and stable predictions than a single decision tree.

**Alternating Decision Tree (ADT)**

Boosting in decision trees is a special method relying on voting from a number of parallelly grown trees, but it can make classification complicated, lengthy and cumbersome, causing resulting trees to be difficult to visualize and rules to be difficult to interpret. By contrast, the alternating decision tree algorithm [22] generates smaller and easier rule sets, and it is a generalization of simple decision trees, voted decision trees and stumps in which stumps are a simplistic version of trees with just one branch and two leaves that ends at the first level [23].

**Pruned Decision Tree (PDT)**

Pruning is another healthy practice in building or post-processing decision trees in which branches that do not contribute to classification accuracy are removed, making rules simpler and more interpretable but reducing rule set cardinality. As additional benefit, pruning can effectively combat overfitting.

The pruning algorithm in this paper is J48, which is composed of two types of pruning, namely subtree replacement and raising [24]. In the first method, intermediate nodes are replaced by leaves, reducing the required number of tests for a decision. The process starts from leaves of a complete tree upwards. In the second type, a node (along with sub-branches) is moved upwards towards the tree root, cutting off parts of upper branches. The first method is quite simple and effective, and the second method has a lesser known effect on decision tree efficacy. Out-of-bag error rates are typically used to determine the branch or subtree to be pruned and

thus is known as reduced-error pruning. This helps to generalize the rules and render them more flexible for data outliers or noise, reducing the overfitting problem.

**Naïve Bayes Classifier (NB)**

Naïve Bayes is a simple and popular classifier [25] based on the Bayes theorem of conditional probability. The algorithm assumes all features to be statistically independent of one another. To estimate the class of an observation (*i.e.* whether an employee will churn), the class is predicted to be the one with maximum posterior probability:

$$\hat{y} = \underset{k \in \{1,2,\ldots,K\}}{\arg\max}\, \wp(C_k) \prod_{i=1}^{n} \wp(x_i \mid C_k) \tag{2}$$

where $x_i$ is the value of the $i^{th}$ input variable, and $C_k$ is an observation of class $k$ from one of the $K$ classes.

**Multi-Layer Perceptron (MLP)**

The MLP is an artificial neural network (ANN) composed of an input node layer (generally equal to the number of input variables), an output node and several intermediate nodes across multiple layers. Each layer is connected to nodes in the subsequent layer through weights and activation functions, and the output is calculated as 0 or 1 in a binary classification problem. The MLP employs backpropagation as a supervised learning technique. With multi-layers of artificial neurons, the MLP can be seen as a deep learning technique.

*Extracting Rules from Decision Trees*

In addition to predicting employee churn, it is also important to understand their reasons for churning. These reasons may differ across employees and vary over time.

If-then rules in the conjunctive normal form (CNF) in which descriptors are separated by "and" are considered one of the simplest and most suitable form of knowledge extracted from classifiers such as decision trees.

Each leaf traverses from the root represents a rule. A typical rule takes the form:

$$\text{If } x_1 = v_1 \text{ and } x_2 = v_2 \text{ and} \ldots \; x_m = v_m \text{ then } C_k \tag{3}$$

where the $m$ variables $x_1, x_2, \ldots, x_m$ taking up values $v_1, v_2, \ldots, v_m$ would indicate that the class would be $C_k$. Each $x_i = v_i$ is called a descriptor of the rule, and $m$ is the length of the rule.

## 3. RESULTS

First, a subset of features is selected using the CFS subset evaluator [26]. The subset contains the following 8 features:


ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

- com len mean
- com len sd
- com totld
- com daily char
- total hap votes
- e/c hap mean
- e/c cmt freq
- emp lik/totld

Six different classification algorithms are considered to identify employees who churn. Each algorithm is run using the whole dataset by taking all 35 descriptor variables and using 10-fold cross validation to test classifier performance. The average performance from each of the 10 folds are reported in Table 2 below. Three performance measures are used to assess classification accuracy of algorithms in terms of their ability to predict employee churn. The first two metrics may be expressed as:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$
$$\text{F1 Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

The third is calculated from the graph of Precision vs. Recall (called the Receiving Operator Characteristic or ROC curve) by taking the area under the curve (AUC) as the measure of classifier performance. AUC-ROC in a binary class problem can denote the probability of accurately ranking two observations from each class. The Precision and Recall are calculated as:

$$\text{Precision} = \frac{TP}{TP + FP}$$
$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

with *TP*, *TN*, *FP* and *FN* denoting the number of true positive, true negative, false positive and false negative observations classified during the prediction, where Churn = Yes is taken as the positive class. All feature selection, classification and rule induction are performed in Weka, an open source platform for machine learning [26].

The results are presented in Table 2. The 2nd and 3rd columns provide the true negative rate (TNR) and true positive rate (TPR), while the 4th and 5th columns give the false negative rate (FNR) and false positive rate (FPR). While F1 score is rather conservative, Accuracy favours majority classes. However, AUC-ROC is considered more acceptable for imbalanced classes.

TABLE 2.
Performance of different classifiers

| Algorithm | TNR | TPR | FNR | FPR | F1 score | Accuracy | AUC-ROC |
|---|---|---|---|---|---|---|---|
| CART | 0.983 | 0.634 | 0.366 | 0.017 | 0.668 | 0.961 | 0.849 |
| ADT | 0.987 | 0.504 | 0.496 | 0.013 | 0.591 | 0.957 | 0.905 |
| RF | 0.982 | 0.639 | 0.361 | 0.018 | 0.667 | 0.961 | 0.895 |
| PDT | 0.984 | 0.601 | 0.399 | 0.016 | 0.653 | 0.961 | 0.852 |
| NB | 0.582 | 0.874 | 0.126 | 0.418 | 0.211 | 0.600 | 0.814 |
| MLP | 0.982 | 0.563 | 0.437 | 0.018 | 0.610 | 0.956 | 0.877 |

Apart from the classification of employees as Churn and Not-churn (Yes and No), CART and PDT algorithms can produce rule sets for insights into clear symptoms of churn. Knowing such rules can provide managers to better manage employee turnover by calculating some simple features. Two sets of rules (one from each of these two algorithms) are presented in Tables 3 and 4 below. Only rules with support of 10 or more observations are presented for generality. The last two columns of each table provide the number of observations in support of ('Sup') and contradicting ('Con') the rule.

TABLE 3.
Rules from the *CART* algorithm

| # | com daily cmt | emp lik/totld | no.of emp | com hap mean | emp like mean | e/c hap mean | e/c cmt mean | Churn | Sup | Con |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | $< 9.94\times10^{-5}$ | < 0.077 | ≥ 93 | ≥ 2.83 | < 0.67 | | | **Yes** | 72 | 5 |
| 2. | $< 9.94\times10^{-5}$ | < 0.077 | ≥ 93 | < 2.83 | | < 0.0089 | < 13.8 | **Yes** | 86 | 52 |
| 3. | $< 9.94\times10^{-5}$ | < 0.077 | < 93 | | | | | **No** | 92 | 0 |
| 4. | $< 9.94\times10^{-5}$ | < 0.077 | ≥ 93 | < 2.83 | | ≥ 0.0089 | | **No** | 45 | 11 |
| 5. | $< 9.94\times10^{-5}$ | > 0.078 | | | | | | **No** | 575 | 23 |
| 6. | $\geq 9.94\times10^{-5}$ | | | | | | | **No** | 2872 | 46 |

TABLE 4.
Rules from the *PDT* algorithm

| # | com daily cmt | emp lik/totld | no.of emp | e/c cmt mean | emp hap mean | Churn | Sup | Con |
|---|---|---|---|---|---|---|---|---|
| 1. | ∈ (0.000086, 0.000099] | ≤ 0.39 | | ≤ 1.85 | | **Yes** | 76 | 5 |
| 2. | ≤ 0.000086 | ≤ 0.39 | > 340 | ≤ 1.85 | ≤ 1.62 | **Yes** | 122 | 50 |
| 3. | ≤ 0.000086 | ≤ 0.39 | ≤ 123 | | | **No** | 92 | 0 |
| 4. | ≤ 0.000086 | ≤ 0.39 | > 340 | ≤ 1.85 | > 1.62 | **No** | 64 | 25 |
| 5. | ≤ 0.000086 | ≤ 0.39 | > 340 | > 1.85 | | **No** | 19 | 0 |
| 6. | ∈ (0.000086, 0.000099] | > 0.39 | | | | **No** | 44 | 3 |
| 7. | ≤ 0.000086 | > 0.39 | | | | **No** | 546 | 20 |
| 8. | > 0.000099 | | | | | **No** | 2918 | 46 |


ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

## 4. DISCUSSION

Out of 8 features, 5 are company-level variables and indicate how well respective management was able to involve its employees in the app, mainly to write comments or like or dislike others' comments. The next 2 variables are company-normalized employee metrics indicating employee involvement relative to peers. Finally, 1 variable is an individual measure of how well one's comments are liked or disliked by peers. This key observation is also noted in Berengueres et al. (2017), who indicate that company management is primarily responsible for creating an environment conducive for happy employees.

The overall prediction accuracy in Table 2 indicates that most algorithms (except Naïve Bayes) generally performed equally well for the employee churn dataset. While overall accuracy reached 96%, the True Positive rates (TPR) are low (50 – 64%), with Random Forest as the best performer. However, Naive Bayes showed a high TPR by correctly classifying 87% of churned employees (while misclassifying 40% of non-churned employees). If the area under the ROC curve can be seen as the single reliable measure of classifier performance, then the Alternating Decision Tree (ADT) is the best classifier (90.5% AUC). This is well in agreement to Berengueres et al. (2017), who obtained 80% accuracy for non-churned employees and 96% accuracy for churned ones.

For the two rule sets, the pruned decision tree (PDT) gave not only *more* but also *shorter* rules. A closer look and a comparison of the first two rules in each set (the rules for Churn = Yes and indicating roughly the same set of observations) show that CART rule 1 (with 5 descriptors) is longer than PDT rule 1 (with just 3 descriptors), although the number of support and that of conflict are nearly the same for each. For rule 2, the pruned set rule had one descriptor less and roughly 50% more support than the CART rule, with about the same number of contradicting observations. In sum, the PDT algorithm gave much better rules than the CART algorithm, both in terms of brevity and generality.

## 5. CONCLUSION

The occurrence of employee churn is predicted, and its causes are identified using six machine learning techniques together with feature selection and rule extraction to reveal important patterns in the dataset. Using a publicly available dataset, features most closely related to employee churn are selected. From the set of selected features, company features are inferred to be most responsible for employee churn, followed by employee behavior relative to peers. Accuracy of 96% is observed in most of the prediction algorithms considered, but 60% of churn is correctly identified, giving an area under the ROC curve as about 90%. These results are consistent with previous findings.

Rules extracted from two different decision trees offer important insights for management to better manage and prevent employee churn. A pruned decision tree provides simpler and more accurate rules than a simple CART tree. Accuracy measures support findings of previous studies employing the same dataset to predict employee churn.

## ACKNOWLEDGEMENTS

The authors acknowledge Zhiyue Liu of St. Johnsbury Academy for his participation in this research.

## REFERENCES

[1] J. Berengueres, G. Duran, and D. Castro, "Happiness, an inside job?: Turnover prediction using employee likeability, engagement and relative happiness," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. ACM, 2017, pp. 509–516.

[2] V. V. Saradhi and G. K. Palshikar, "Employee churn prediction," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1999–2006, 2011.

[3] J. L. Cotton and J. M. Tuttle, "Employee turnover: A meta-analysis and review with implications for research," *Academy of management Review*, vol. 11, no. 1, pp. 55–70, 1986.

[4] D. G. Allen and R. W. Griffeth, "Test of a mediated performance–turnover relationship highlighting the moderating roles of visibility and reward contingency." *Journal of Applied Psychology*, vol. 86, no. 5, p. 1014, 2001.

[5] D. Liu, T. R. Mitchell, T. W. Lee, B. C. Holtom, and T. R. Hinkin, "When employees are out of step with coworkers: How job satisfaction trajectory and dispersion influence individual-and unit-level voluntary turnover," *Academy of management journal*, vol. 55, no. 6, pp. 1360–1380, 2012.

[6] B. W. Swider and R. D. Zimmerman, "Born to burnout: A meta-analytic path model of personality, job burnout, and work outcomes," *Journal of Vocational Behavior*, vol. 76, no. 3, pp. 487–506, 2010.

[7] T. M. Heckert and A. M. Farabee, "Turnover intentions of the faculty at a teaching-focused university," *Psychological reports*, vol. 99, no. 1, pp. 39–45, 2006.

[8] C. O. Trevor, "Interactions among actual ease-of-movement determinants and job satisfaction in the prediction of voluntary turnover," *Academy of management journal*, vol. 44, no. 4, pp. 621–638, 2001.

[9] D. N. Dickter, M. Roznowski, and D. A. Harrison, "Temporal tempering: An event history analysis of the process of voluntary turnover," *Journal of Applied Psychology*, vol. 81, no. 6, p. 705, 1996.

[10] R. W. Griffeth and P. W. Hom, *Retaining valued employees*. Sage Publications, 2001.

[11] S. Wang and X. Yi, "It's happiness that counts: Full mediating effect of job satisfaction on the linkage from LMX to turnover intention in Chinese companies," *International Journal of Leadership Studies*, vol. 6, no. 3, pp. 337–356, 2011.

[12] P. Ajit, "Prediction of employee turnover in organizations using machine learning algorithms," *Algorithms*, vol. 4, no. 5, p. C5, 2016.

[13] "Happyforce." [Online]. Available: https://myhappyforce.com/en/

[14] S. N. Mishra, D. R. Lama, and Y. Pal, "Human resource predictive analytics (hrpa) for hr management in organizations," *International Journal of Scientific & Technology Research*, vol. 5, no. 5, pp. 33–35, 2016.


ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

[15] I. Ajzen, "The theory of planned behavior," *Organizational behavior and human decision processes,* vol. 50, no. 2, pp. 179–211, 1991.
[16] J. Berengueres, "Daily Happiness & Employee Turnover," 2017. [Online]. Available: https://kaggle.com/harriken/employeeturnover
[17] J. Berengueres, "How many unlikes it takes to get fired?" 2017. [Online]. Available: https://www.kaggle.com/harriken/how-many-unlikes-it-takes-to-get-fired
[18] M. A. Hall, "Correlation-based feature subset selection for machine learning," PhD Thesis, University of Waikato, 1998.
[19] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees.* Monterey, CA: Wadsworth and Brooks, 1984.
[20] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5–32, 2001.
[21] J. K. Martin and D. Hirschberg, "On the complexity of learning decision trees," in *International Symposium on Artificial Intelligence and Mathematics.* Citeseer, 1996, pp. 112–115.
[22] Y. Freund and L. Mason, "The alternating decision tree learning algorithm," in *ICML*, vol. 99, 1999, pp. 124–133.
[23] W.Iba and P.Langley, "Induction of one-level decision trees," in *Machine Learning Proceedings 1992*. Elsevier, 1992, pp. 233–240.
[24] T. R. Patil, S. Sherekar et al., "Performance analysis of Naïve Bayes and J48 classification algorithm for data classification," *International Journal of Computer Science and Applications,* vol. 6, no. 2, pp. 256–261, 2013.
[25] T. Mitchell, *Machine learning,* 2nd ed., McGraw Hill, USA, 1997.
[26] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter,* vol. 11, no. 1, pp. 10–18, 2009.