International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

# EMPLOYEE TRACKING

**Joseph Zen**

*Student, Dept. of Dual Degree Computer Applications, Sree Narayana guru institute of science and technology, Kerala, India*

---

**Abstract -** Now a day's Employee Attrition prediction become a serious problem within the organizations. Employee Attrition may be a big issue for the organizations specially when trained, technical and key employees leave for a far better opportunity from the organization. This leads to financial loss to exchange a trained employee. Therefore, we use the present and past employee data to research the common reasons for employee attrition or attrition. For the prevention of employee attrition, we applied a documented classification methods, that is, Decision tree, Logistic Regression, SVM, KNN, Random Forest, Naive bayes methods on the human resource data. For this we implement feature selection method on the info and analysis the results to stop employee attrition. This is helpful to companies to predict employee attrition, and also helpful to their economic process by reducing their human resource cost.

**Keywords— Attrition, Decision Tree, Logistic Regression, SVM, KNN, Random Forest, Naïve bayes**

## 1. INTRODUCTION

An employee would prefer to join or depart a corporation counting on many causes i.e. work environment, work place, gender equity, pay equity and lots of other. the remainder of the workers might imagine about personal reasons as an example relocation thanks to family, maternity, health, issues with the managers or co-workers during a team. Employee attrition may be a major problem for the organizations particularly when trained, technical and key employees leave for best opportunities from the organizations. This finally results into monetary loss to substitute a trained employee. Consequently, we utilize this and past employee data to assess the familiar issues for employee attrition. the worker attrition identification helps in predicting and resolving the problems of attrition. we will use this data to prevent the rate of attrition of the workers . For this working we use some methodologies of knowledge classification. Those methodologies are Decision Tree (it is tree structure that comprises a branches, root node and leaf nodes. every internal node indicates a test on an attribute, every branch indicates the results of a test, and each leaf node holds a category label), Naive Bayes (it may be a classification methodology counting on Bayes Theorem. A Navie Bayes classifier presumes that the existence of a selected during a class is unrelated to the existence of the other feature)

## 2. METHODOLOGY

### Data sets

Data set might also be a series of understanding . most primarily a understanding set corresponds to the contents of one database, the place each and every column of the desk represents a precise variable, and each and every row corresponds to a member of the dataset. For our task we take worker information from IBM which includes 1470 documents and 35 fields along with specific and numeric features. Each file inside the worker facts set represents one worker facts and each and every area inside the report represents a characteristic of that particular employee.

### Data Pre-Processing

From the IBM worker dataset we put in force a function choice approach to pick out the predominant vital aspects of the dataset and divide whole dataset into two sub datasets. One is check dataset any other one is education dataset. That is if think any function cost inside the report include any null cost or undefined or inappropriate cost then separate that whole file from the first dataset and vicinity that file into coaching dataset, else if the report incorporate best statistics with all facets then region that into check dataset. Test

dataset include all essential points to predict worker attrition or worker attrition and coaching dataset include beside the point data.

## Test dataset and training dataset:

Separating statistics into take a look at datasets and coaching datasets is a quintessential a section of evaluating statistics processing models. By this separation of whole information set into two facts units we will limit the penalties of expertise inconsistency and higher recognize the traits of the model. The take a look at records set consists of all the distinctive facts for records prediction and education facts set carries all beside the point data. Here we've got 788 archives in take a look at dataset and 682 documents in education dataset. We observe facts classification and information prediction on the take a look at dataset of 788 records

## Data classification techniques

Data classification is that the process of organizing data into categories for its best and efficient use. Data classification techniques are Decision tree, K nearest neighbour (KNN), Support vector machine (SVM), Logistic regression, Naive bayes.

## Decision Tree

It is tree shape that consists of a root node, branches, and leaf nodes. Each interior node denotes a take a look at on an attribute, every department denotes the result of a test, and every leaf node holds a category label.

## Naive Baye

It is a classification method primarily based on Bayes Theorem. A Navie Bayes classifier assumes that the presence of a unique in a category is unrelated to the presence of any different feature. For example, a fruit may additionally be viewed to be an apple if it is red, round, and about three inches in diameter. Even if these elements rely on every different or upon the existence of the different features, all these houses independently make a contribution to the chance that this fruit is an apple

## Logistic Regression:

It is a statistical technique for inspecting a dataset in which there are one or greater impartial variables that decide an outcome. For instance hours of reading will increase then the chance of passing checks will increase

## Data classification techniques

Data classification is that the process of organizing data into categories for its best and efficient use. Data classification techniques are Decision tree, K nearest neighbour (KNN), Support vector machine (SVM), Logistic regression, Naive bayes.

## Decision Tree

It is tree shape that consists of a root node, branches, and leaf nodes. Each interior node denotes a take a look at on an attribute, every department denotes the result of a test, and every leaf node holds a category label.

## Naive Baye

It is a classification method primarily based on Bayes Theorem. A Navie Bayes classifier assumes that the presence of a unique in a category is unrelated to the presence of any different feature. For example, a fruit may additionally be viewed to be an apple if it is red, round, and about three inches in diameter. Even if these elements rely on every different or upon the existence of the different features, all these houses independently make a contribution to the chance that this fruit is an apple

## Logistic Regression

It is a statistical approach for examining a dataset in which there are one or greater impartial variables that decide an outcome. For instance hours of analyzing will increase then the chance of passing assessments increases

## Support Vector Machine (SVM)

it performs classification by finding the hyperplane that completely separates the vector into two non overlapping classes. The vectors that outline the hyperplane are the support vectors

## K-Nearest Neighbour (KNN)

examine every price with the neighbour values or Nearest values. It is a non parametric approach used for classification. Here we practice classification strategies on the take a look at information set and categorize the information into exceptional departments. That is income department, human useful resource department, lookup and improvement department. After that we follow overall performance analysis, training analysis, and profits evaluation on all categorised departments statistics to predict the worker attrition with the aid of discovering the excellent personnel

## Performance analyzer

It is used to analyze the common overall performance of personnel in every department. Here we have common overall performance in income branch is 3.13677, common overall performance in lookup and improvement is 3.16233, and common overall performance in human aid branch is 3.14285. From the common performances we locate out the personnel who have the absolute best overall performance than common overall performance and predict these personnel from worker attrition

## Salary analyzer

It is used to analyze the personnel who have excessive profits and who have low salary. If any of the personnel getting low income even even though their overall performance is excessive then we become aware of these personnel and stop them from worker attrition by means of incrementing their salaries. The personnel who have their profits decrease than 6000 they belongs to low income class and who have greater than 6000 profits these personnel belongs to excessive revenue categories. Here we have 914 personnel getting low revenue and 556 personnel getting excessive salary. In these 556 personnel these getting excessive revenue 231 are girls and 325 are males.

## Education analyzer

It is used to analyze the personnel who have higher qualification and who have decrease qualification. Here we divide the whole personnel in 5 classes in accordance to their instructional qualification. Those 5 classes are personnel with single degree, personnel with double degree, triple degree, personnel with 4 degrees, and in the end personnel with 5 diploma qualification. Finally we locate out the common performance, common job delight and common month-to-month pay to the personnel of every category.

## Predicted data

By this complete evaluation we discover out the excellent personnel and we forestall these personnel from worker attrition by means of offering the all necessities

### 3.EXPERMENTAL ANALYSIS

In the present structures they used solely few of facts mining methods for facts prediction. In the proposed structures we use 5 algorithms that is KNN, SVM, logistic regression, choice tree, and navie bayes. Here we additionally used function choice approach on worker statistics set. Generally the worker records set carries worker records like skills, nature of work, salary, overall performance ratting etc. By the usage of characteristic determination we can choose some required aspects from worker records set for our analysis. In the beneath desk 5.1 we suggests the elements of datasets and their type.

| Features | Data type |
|---|---|
| Age | Number[10] |
| BusinessTravel | Varchar[20] |
| DailyRate | Number[10] |
| Education | Number[10] |
| DistanceFromHome | Number[10] |
| Department | Varchar[20] |
| EducationField | Varchar[20] |
| EmployeeNumber | Number[10] |
| Gender | Varchar[20] |
| EnvironmentSatisfaction | Number[10] |
| Hourly rate | Number[10] |
| Job level | Number[10] |
| Jobinvolvement | Number[10] |
| JobRole | Varchar[20] |
| Mirtalstatus | Varchar[20] |
| MonthlyRate | Number[10] |
| Monthly income | Number[10] |
| Jobsatisfaction | Number[10] |
| OverTime | Varchar[20] |
| NumCompaniesworked | Number[10] |
| PercentSalaryhike | Number[10] |
| PerfomanceRating | Number[10] |
| RelationshipSatisfaction | Number[10] |
| Stock option level | Number[10] |
| Totalworkingyears | Number[10] |
| TraningTimeslastYear | Number[10] |
| WorkLifeBalance | Number[10] |
| YearsAtCompany | Number[10] |
| Years In current Role | Number[10] |
| YearsSinceLastPromotion | Number[10] |

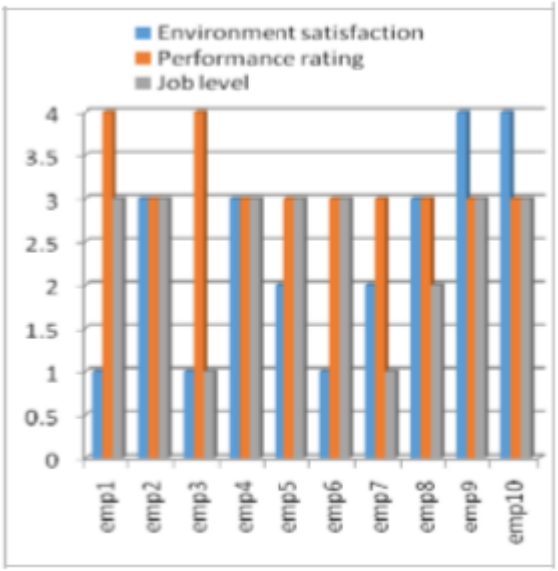
**5.1 Table Name: Employee Dataset**

We take solely 32 facets that is required for our analysis. This is very useful to make bigger accuracy of the system. In the under graphs we suggests the distinction between the accuracy of current and proposed system. The worker records is accrued from specific departments of an enterprise is saved in a database. Here we have viewed a pattern facts of ten personnel on the groundwork of Environment Satisfaction, Performance ranking and their Job levels. These document values in chosen information sets.

| emp Id | Environment satisfaction | Performance rating | Job level |
|---|---|---|---|
| emp1 | 1 | 4 | 3 |
| emp2 | 3 | 3 | 3 |
| emp3 | 1 | 4 | 1 |
| emp4 | 3 | 3 | 3 |
| emp5 | 2 | 3 | 3 |
| emp6 | 1 | 3 | 3 |
| emp7 | 2 | 3 | 1 |
| emp8 | 3 | 3 | 2 |
| emp9 | 4 | 3 | 3 |
| emp10 | 4 | 3 | 3 |

**5.2 A sample dataset of ten employees on the basis of Environment Satisfaction, Performance rating and their Job levels**

The below graph 5.3 will be displayed by using the sample data set of 10 employees in an organization. From the table we chosen the performance rating based on that we have generated the graph drawn below




**5.3 Graph for sample data set for 10 employees**

## 4. CONCLUSION

Employee attrition outcomes in financial, time and effort loss for organizations. It is a large problem considering the fact that a educated and skilled worker is challenging to replacement and it is value effective. We strive to discover to analyze the previous and current worker records to estimate the future attritioners and find out about the motives of worker

turnover. The effects of this getting to know describe that information extraction algorithms can be utilized to assemble dependable and correct predictive strategies for worker attrition. The problem of attrition identification is now not simply to depict attritioners from no attritioners. By the usage of tentative statistics find out about and facts extraction methids, we can depict the attrition likelihood for every one worker and furnish them rating to construct the retention methods

**ACKNOWLEDGEMENT**

In the name of almighty, I would like to extend my heartfelt thanks to our HOD Mrs. Kavitha C.R, Department of a Dual Degree Master of Computer Applications for the helps extended to me throughout my course of my study. I am deeply grateful to my guide Mr. Rajesh Basker Assistant Professor, Department of a Dual Degree Master of Computer Applications for the valuable guidance

**REFERENCES**

[1]. W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Vaesens,"New insights into a churn prediction in the telecommunication sector. An profit driven datamining approach," European journal of operational research, vol. 218, no. 1, pp. 211-229, 2012.

[2]. K.Coussement and D. VandenPoel, "Integrating the voice of customers through call center emails into a decision support system for attrition prediction," Information & Management, vol. 45, no. 3, pp. 164–174, 2008.

[3]. C.-P. Wei and I.-T. Chiu, "Turning telecommunications call details to attrition prediction: a data mining approach," Expert systems with applications, vol. 23, no. 2, pp. 103–112, 2002.

[4]. K. Coussement and D. Van den Poel, "Attrition prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," Expert systems with applications, vol. 34, no. 1, pp. 313–327, 2008.

[5]. J. Burez and D. Van den Poel, "Handling class imbalance in customer attrition prediction," Expert Systems with Applications, vol. 36,no. 3, pp. 4626–4636, 2009.

[6]. C.-F. Tsai and M. Y. Chen, "Variable selection by association rules for customer attrition prediction of multimedia on demand," Expert Systems with Applications, vol. 37, no. 3, pp. 2006–2015, 2010.

[7]. K. Coussement, D. F. Benoit, and D. Van den Poel, "Improved marketing decision making in a customer attrition prediction context using generalized additive models," Expert Systems with Applications, vol. 37, no. 3, pp. 2132–2143, 2010

[8]. B. Huang, M. T. Kechadi, and B. Buckley, "Customer attrition prediction in telecommunications," Expert Systems with Applications, vol. 39, no. 1, pp. 1414–1425, 2012

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

[10]. R. Khare, D. Kaloya, C. K. Choudhary, and G. Gupta, "Employee attrition risk assessment using logistic regression analysis,"

[11]. M. L. Kane-Sellers, Predictive models of employee voluntary turnover in a North American professional sales force using data-mining analysis.

[12]. X. Lin, F. Yang, L. Zhou, P. Yin, H. Kong, W. Xing, X. Lu, L. Jia, Q. Wang, and G. Xu, "A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables