****IBM HR Employee Attrition Rate**

**ABSTRACT:** Employees are the most valuable assets to any organization. They are the fulcrum and laying foundation of an organisation. It is they who add value to the organization in terms of quantity as well as quality. To find, attract, develop and retain the right talent is a critical task for the management. Therefore, it is indispensable to maintain a permanent and promising workforce which over the years has become a tough task for employers and thereby resulting in increased attrition. This comprehensive analysis on the IBM dataset aims to make a detailed study on attrition, comprehend to various interesting features, to forecast attrition so employers are prepared well in advance and most importantly implement ways to retract rates of attrition in the future.

**Need Of The Study:**

The study was mainly undertaken to identify the level of employee’s attitude, the dissatisfaction factors they face in the organization and for what reasons they prefer to change their job. Once the levels of employee’s attitude are identified, it would be possible for the management to take necessary actions to reduce attrition level. Since they are considered as the backbone of any company, their progression and development will lead to exponential and consistent growth of companies in the longer run.

**Scope Of The Study:**

* To determine effect of attrition on the business.
* Determination of solutions to avoid or to control attrition.
* To understand the extent of job satisfaction among the employees.
* To suggest proper measures.
* This study helps the company to understand more on the attrition rate in the company,
* The study also helps to find the drawbacks of the current retention strategies.

**Objectives of The Study:**

* To foster a pattern during the stint of an employee as to when and what factors lead to attrition based on available facts and historical data.
* To analyse the parameters of a dissatisfied employee and adopt curated set of innovative ideas focusing on employee needs and desires.
* To consistently retain a low percentage of attrition to not only acquire top notch employees but discretely invite the interests of potential investors.
* To deploy an ascertained model which forecasts attrition beforehand and directly revamps a smooth decision-making process.

To constitute state of the art retention policies and framework that would mutually benefit both the organization and employees while fixating on the reasons of attrition mentioned above

**Dataset and Domain**

Data Rows: 1470 entries Data Columns :35 entries

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No** | **Feature Name** | **Data Type** | **Missing Value Count** |
| 1 | Age | Numerical | 0 |
| 2 | Attrition | Categorical | 0 |
| 3 | Business Travel | Categorical | 0 |
| 4 | Daily Rate | Numerical | 0 |
| 5 | Department | Categorical | 0 |
| 6 | Distance From Home | Numerical | 0 |
| 7 | Education | Numerical | 0 |
| 8 | Education Field | Categorical | 0 |
| 9 | Employee Count | Numerical | 0 |
| 10 | Employee Number | Numerical | 0 |
| 11 | Environment Satisfaction | Numerical | 0 |
| 12 | Gender | Categorical | 0 |
| 13 | Hourly Rate | Numerical | 0 |
| 14 | Job Involvement | Numerical | 0 |
| 15 | Job Level | Numerical | 0 |
| 16 | Job Role | Categorical | 0 |
| 17 | Job Satisfaction | Numerical | 0 |
| 18 | Marital Status | Categorical | 0 |
| 19 | Monthly Income | Numerical | 0 |
| 20 | Monthly Rate | Numerical | 0 |
| 21 | Num Companies Worked | Numerical | 0 |
| 22 | Over18 | Categorical | 0 |
| 23 | Overtime | Categorical | 0 |
| 24 | Percentage Salary Hike | Numerical | 0 |
| 25 | Performance Rating | Numerical | 0 |
| 26 | Relationship Satisfaction | Numerical | 0 |
| 27 | Standard Hours | Numerical | 0 |
| 28 | Stock Option Level | Numerical | 0 |
| 29 | Total Working Years | Numerical | 0 |
| 30 | Training Times Last Year | Numerical | 0 |
| 31 | Work Life Balance | Numerical | 0 |
| 32 | Years At Company | Numerical | 0 |
| 33 | Years In Current Role | Numerical | 0 |
| 34 | Years Since Last Promotion | Numerical | 0 |
| 35 | Years With Curr Manager | Numerical | 0 |

**The dataset contains :**

* 26 **Numerical Variables (Discrete:**26  **Continuous :**0 **)**
* 9 **Categorical Variables (Ordinal: Nominal:**9 **)**
* **The dataset has no missing values in any of the features.**

**Feature Attributes:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No** | **Features** | **Unique elements in each feature** | **Data Type** |
| 1 | Attrition | * Yes * No | Categorical |
| 2 | Business Travel | * Non-Travel * Travel-Frequently * Travel-Rarely | Categorical |
| 3 | Department | * Sales * Research & Development * HR | Categorical |
| 4 | Education | * 1(Middle School) * 2(High School) * 3(UG) * 4(PG) * 5(Ph.D.) | Numerical |
| 5 | Education Field | * 1(Human Resources) * 2(Life Sciences) * 3(Medical) * 4(Marketing) * 5(Technical Degree) * 6(Other) | Categorical |
| 6 | Environment Satisfaction | * 1(Low) * 2(Medium) * 3(High) * 4(Very High) | Numerical |
| 7 | Gender | * Male * Female | Categorical |
| 8 | Job Involvement | * 1(Low) * 2(Medium) * 3(High) * 4(Very High) | Numerical |
| 9 | Job Level | * 1(Junior) * 2(Intermediate) * 3(Senior) * 4(Vice-President) * 5(President) | Numerical |
| 11 | Job Satisfaction | * 1(Low) * 2(Medium) * 3(High) * 4(Very High) | Numerical |
| 12 | Marital Status | * Single * Married * Divorced | Categorical |
| 13 | Over18 | * Yes | Categorical |
| 14 | Overtime | * Yes * No | Categorical |
| 15 | Performance Rating | * 3(Excellent) * 4(Outstanding) | Numerical |
| 16 | Relationship Satisfaction | * 1(Low) * 2(Medium) * 3(High) * 4(Very High) | Numerical |
| 17 | Stock Option Level | * 0 * 1 * 2 * 3 | Numerical |
| 18 | Work Life Balance | * 1(Bad) * 2(Good) * 3(Better) * 4(Best) | Numerical |
| 10 | Job Role | * 1(Sales Executive) * 2( Research Scientist) * 3( Laboratory Technician) * 4( Manufacturing Director) * 5( Healthcare Representative) * 6( Manager) * 7( Sales Representative) * 8(Research Director) * 9(Human Resources) | Categorical |

**Irrelevant Columns:**

There are few features which are redundant and do not contribute in the prediction of the target variable. Hence, we drop them.

|  |  |
| --- | --- |
| **Features** | **Justification** |
| Standard Hours | 80 Standard Hours is same for all the records |
| Over18 | ‘YES’ is redundant to all records |
| Employee Count | Pointless feature with ‘1’ for all records |
| Employee Number | Using this feature as the index colum |

**Justification**:

1. This is a fictional dataset created by IBM data scientists and contains employee details which focuses primarily on HR analytics.
2. This is a classic classification problem and the dependant variable is **Attrition.**

We used all classification algorithms like Logistic Regression, Decision Tree, and Random Forest boosting techniques to increase the accuracy and performance of the model

**Project Complexity**

1. Complexity is substantial as the dataset is well crafted with no missing values and enough features to comprehensively analyse the pain points for attrition.
2. Feature extraction and feature engineering was minimal as most features were already pre-defined and on point.
3. Most numerical features were normally distributed therefore minimal data transformation techniques were required.

**Project Outcome**

1. With a good prediction score companies will have a better understanding of their respective employees on a commercial level.
2. Companies can now formulate smarter retention systems and programs to maintain employee satisfaction.
3. Academically speaking, the dataset is fantastic to refresh graphical and visual representation of important features. Features like employee, environment and manger satisfaction levels are considered. Work life balance and job role are important factors to all employees and using these variables the company can weigh the probabilities of retention precisely.
4. Towards the end of the analyses, we found that factors like overtime, marital status and distance from home are critical pain points and highly influence attrition rates.

**Model Building**

Here, we split the entire data into 85-15% ratio, apply a Decision Tree Classifier model , and then train the model with some parameters.  
  
We then, measure the importance of features included in the model and plot them in a graph.

Next, a Random Forest model is applied here, with some specific parameters. K-fold cross-validation has been applied to achieve an unbiased result without overfitting.

**Model Evaluation:**

1**.**Random Forest Classifier

Training accuracy:0.79

2.Decission Tree classifier

Training accuaracy:0.728

**Limitations**

Since our data set had very few records, we were confined to create a model within that. In future, the dataset can include more detailed data so that we can dive deeper and based on the attrition rate, we can give benefits to the employees so that the employee attrition rate gets decreased.

**Conclusion**:

The study has set out to be a real time application in the organizations where the management can predict the future actions of the employees based on

their records and observations. The main focus was to build a model which can

efficiently predict the employees that might leave the company in future, and

considering the real scenario, the higher management will be more interested to know the potential employees who might actually leave so that they can set

their attention on them to stop them to do so.

Hence, with the help of our model, companies can now have a smooth flow of

human talent without any major hiccups. All factors are immensely scrutinized and not only will companies benefit but employees too. Companies have the

ability to forecast and take measures well in advance while smarter retention

policies , ESOPs and incentives can be formulated for their respective

employees. **It’s a win-win situation all round.**

