**ECE 143 Final Project – Group 14:**

**Employee Attrition. Why do workers quit?**

**Problem:**

**Attrition is basically the employee turnover rate in an organization.**

**This can happen for many reasons, namely:**

* **Employees looking for better opportunities**
* **A negative working environment**
* **Bad Management**
* **Sickness/Death**
* **Excessive working hours**

**By analyzing the data we shall build a predictive model to determine how likely the**

**employee is going to quit.**

**We shall also suggest certain steps that can be taken to reduce attrition rate.**

**Dataset:**

* <https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>

Contains employee attrition data. Consists of 1470 observations with 35 features for each observation including but not limited to Education Field, marital status, age, gender etc.

The final label is whether the employee in question left the company or not which is a binary target.

**Significance of the issue and popular analysis methods:**

* Employees are the key elements of a company, each of them makes contribution by her/his own effort. Therefore, a sudden resignation, which is unfortunately pretty common, of an important stuff usually leads to a huge crisis. However, by analysing several attributes of people who quit or not quit form their own company, we could acquire several statistical laws, thus predict whether a subject will resign in the future given her/his attributes. From the perspective of the company, they could utilize the prediction model to make precautions to the potential loss of human resource.
* Commonly, for constructing such prediction models, linear or logistic regressions are popular methods to determine the positive correlations between different parameters and the final consequence. Once the importance of attributes are known, a decision tree could be built by probability theory. Or in a simpler way, a Gaussian mixture model could be made, if several parameters with greatest importance fitted well by Gaussian distribution.

**Proposed Solution and Real-World Application:**

* Firstly we need to parse the data, find the missing values and drop useless features. Then we’ll be analyzing the data thoroughly and visualize how each feature affects attrition and also emphasize why a certain feature is important to determine attrition. To facilitate this, we shall be analyzing the data based on those significantly impacted attributes like Gender, EducationLevel, WorkingEnvironment, and Income. Also, we will try to create feature distribution histograms between attrition and other parameters such as DailyRate, BusinessTravel, DistanceFromHome, HourlyRate, and so on, to figure out whether those are also correlated.
* By analyzing this data we shall build a predictive model to determine if an employee will quit a certain organization or not. For classification purposes we will be trying out logistic regression and gaussian mixture models.
* Split and train and validate dataset, apply some machine learning methods like XGBoost to optimize yes/no attrition hyperparameters, report accuracy, precision, and recall parameters of builted XGBoost model.
* We believe that with this model, companies should be able to analyze their employee data and effectively figure out what aspects need to be improved to prevent employees from quitting. Also, thay could get started to recruit new staff in case of a foreseen post vacancy.

|  |  |  |
| --- | --- | --- |
| Step | Estimated completion time | Person(s) in charge (among the group of 14) |
| 1. Extracting and cleaning up data. | One week |  |
| 2. Data Analysis and visualization | Two weeks |  |
| 3. Building the predictive model. | One week |  |