**Predicting Employee Attrition**

In recent years, attention has increasingly been paid to human resources (HR), since worker quality and skills represent a growth factor and a real competitive advantage for companies. After proving its mettle in sales and marketing, artificial intelligence is also becoming central to employee-related decisions within HR management. Organizational growth largely depends on staff retention. Losing employees frequently impacts the morale of the organization and hiring new employees is more expensive than retaining existing ones.

You are working as a data scientist with HR Department of a large insurance company focused on sales team attrition. Insurance sales teams help insurance companies generate new business by contacting potential customers and selling one or more types of insurance. The department generally sees high attrition and thus staffing becomes a crucial aspect.

To aid staffing, you are provided with the monthly information for a segment of employees for 2016 and 2017 and tasked to predict whether a current employee will be leaving the organization in the upcoming two quarters (01 Jan 2018 - 01 July 2018) or not, given:

1. Demographics of the employee (city, age, gender etc.)  
2. Tenure information (joining date, Last Date)  
3. Historical data regarding the performance of the employee (Quarterly rating, Monthly business acquired, designation, salary)

**Data Dictionary**

**Train Data**

|  |  |
| --- | --- |
| Variable | Definition |
| MMMM-YY | Reporting Date (Monthly) |
| Emp\_ID | Unique id for employees |
| Age | Age of the employee |
| Gender | Gender of the employee |
| City | City Code of the employee |
| Education\_Level | Education level : Bachelor, Master or College |
| Salary | Salary of the employee |
| Dateofjoining | Joining date for the employee |
| LastWorkingDate | Last date of working for the employee |
| Joining Designation | Designation of the employee at the time of joining |
| Designation | Designation of the employee at the time of reporting |
| Total\_Business\_Value | The total business value acquired by the employee in a month (negative business indicates cancellation/refund of sold insurance policies) |
| Quarterly Rating | Quarterly rating of the employee: 1,2,3,4 (higher is better) |

**Test Data**

|  |  |
| --- | --- |
| Variable | Definition |
| Emp\_ID | Unique Id for the employees |

**Sample Submission**

This file contains the exact submission format for the predictions. Please submit the CSV file only.

|  |  |
| --- | --- |
| Variable | Definition |
| Emp\_ID | Employee ID |
| Target | 0: if the employee does not leave the organization, 1: if the employee leaves the organization |

**Evaluation Metric**

The evaluation metric for this competition is macro [f1\_score](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html)

**Guidelines for Final Submission**

Please ensure that your final submission includes the following:

1. Solution file containing the predictions for the Emp\_ID in the test set. (format is given in sample submission CSV)
2. A Zipped file containing code & approach (Note that both code and approach document are mandatory for shortlisting)

* Code: Clean code with comments on each part
* Approach: Please share your approach to solve the problem (doc/ppt/pdf format). It should cover the following topics:
  1. A brief on the approach, which you have used to solve the problem.
  2. What data-preprocessing / feature engineering ideas really worked? How did you discover them?
  3. What does your final model look like? How did you reach it?