أكاديمية سدايا
SDAIA Academy

PREDICTING EMPLOYEES ATTRITION

Classification project

A PROPOSAL BY
MUNEERA ALSHUNAIFI

DATE
20 OCTOBER 2021

OVERVIEW

Employee attrition occurs when an employee departs a company for a variety of reasons, however, Several companies lost their employees without an obvious reason, in which it affect negatively to the company's productivity and performance. In this project we will try to predict the appropriate reasons that led to the attrition by using machine learning classification algorithms: Logistic regression, Decision trees, SVM, Adaboost classifier. Also, we will focus on some performance measure metrics to evaluate the algorithms such as: Precision, Recall, F1-score and we will see the accuracy confusion matrix.

DATA DESCRIPTION

Fictional data set created by IBM data scientists. It consist of 1470 rows and 35 columns. The columns are:

Numeric columns

- Related to personal information: age, distance_from_home, employee_number (id variable)
- Related to income: hourly_rate, daily_rate, monthly_rate, monthly_income, percent_salary_hike
- Related to time in company: years_at_company, years_in_current_role, years_since_last_promotion, years_with_curr_manager, total_working_years
- other: num_companies_worked, standard_hours(to delete), training_times_last_year, employee_count (to delete)

Categorical columns

- Binary : Attrition(target variable), gender, over18 (to delete), over_time
- Nominal : department, education_field, job_role, marital_status
- Ordinal :
  - Ordinal regarding satisfaction and performance : environment_satisfaction, job_satisfaction, relationship_satisfaction, work_life_balance,job_involvement,performance_rating
  - Other ordinal: business_travel, education, job_level, stock_option_level

TOOLS

To explore and analyze the data and do the prediction models in python, we will use Jupyter notebook and Python packages, such as: Pandas and NumPy Matplotlib, seaborn and SKLearn for modeling. Also we may use extra tools such as PowerBi for visualization and Flask framework for deployment.

CONCLUSION

We aim to know the reasons behind the attrition of the employees within a company based on external or internal impacts by using machine learning classification algorithms. In this document we went through the problem we are aiming to solve, the description of the data that we will work on and the tools that we will use.