IT-ROCKET SCIENCE

R PROJECT - PREDICTIVE MODELLING

DATA MINING II

Project description

Business needs

IT-Rocket Science is a multinational company in the field of information technologies specialized in developing and offer technologic and consulting services for several markets.

In 2017, the company noticed that several of their employees were leaving the company without a perceptible major reason.

Although the company recognizes that some staff turnover is inevitable, a high rate of churn is costly for them. The processes of recruitment, hire and train require always a financial outlay, and usually new hires are associated with low productivity and profit in a short term. An unusual high rate of churn could also be a sign of existing problems inside the organization.

The main purpose of this project is to try to develop efficient and accurate predictive models that will allow the company to understand what the main and hidden reasons behind the employee churn are.

For that purpose, employee historical data from 1450 employees was acquired and a survey regarding to employee satisfaction in the several departments were carried out through the organization. Besides that, Human Resources also create a small file with what could be pertinent aspects regarding to the employees. Using those three files and the file where is indicated if an employee churned or not the company, a team of data scientists should perform analysis of that data and build a predictive model in order to detect what are the main reasons behind the churn, the most accurate possible, and to apply the same model in future detection of possible churn situations.

Variable description

HistoricalData.xls

Variable	Description
EmployeeID	Employee unique identification
BirthDate	Employee date of birth
Gender	Employee gender
MaritalStatus	Employee's marital status
Dependents	Number of dependents
JobType	Job type: remote, non-remote and midterm
Department	Working department: IT, Human Resources, Marketing
JobLevel	Seniority of employee's job: 1 - Entry/Junior; 2 - Intermediate; 3 - First Level Management; 4 - Middle Level Management; 5 - Senior
JobRole	Employee's role on the company
TypeContract	What is the type of appointment associated to the employee?
Education	Level of education: College, Associate Degree, Bachelor Degree, Master Degree, PhD
EducationArea	Field of education
NumCompaniesWorked	Number of companies the employee worked during his lifetime
SalaryRise(%)	Percentage of salary rise since the first employee's contract
WeekHours	Number of standard working hours per week
TenureWorking	Number of years working
NumberProjectsLastYear	Number of projects involved last year
TenureCompany	Number of years working at the company
TenureRole	Number of years in the current role
LastPromotion	Number of years since the last promotion
TenureManager	Number of years with the current manager
MonthlyIncome	Employee's monthly income

SatisfactionSurvey.txt

Variable	Description
EmployeeID	Employee unique identification
FacilitiesSatisfaction	Employee's perception to facilities conditions: 1 - Bad; 2 - Satisfactory; 3 - Good; 4 - Very good
RoleSatisfaction	Employee's role satisfaction: 1 - Low; 2 - Medium; 3 - High; 4 - Very High
HierarchySatisfaction	Emplyee's satisfaction regarding to hierarchical structure in the organization: 1 - Low; 2 - Medium; 3 - High; 4 - Very High
BalanceWork&Life	Balance rating between work and life aspects: Bad, Medium, Good, Great

ChurnIndicator.csv

Variable	Description
EmployeeID	Employee unique identification
Churn	Dependent Variable

HumanResourcesEvaluation.csv

Variable	Description
EmployeeID	Employee unique identification
DistanceHomeOffice	Number of Kms from home to office
JobDedication	Job Dedication: 1 - Low; 2 - Medium; 3 - High; 4 - Very High
AfterHours	Does the employee work after hours?
JobPerformance	Employee's performance rating: 1 - Very Low; 2 - Low; 3 - Medium; 4 - High

Evaluation Criteria – R project

The present document summarizes the evaluation guidelines for the Data Mining project in R. Please read it carefully to understand how your group work will be evaluated.

Predictive Models

Build a predictive model for a dependent variable representing the employee's churn of a specific company.

Deliverables

The following material must be submitted through Moodle platform within deadlines till **31th May**:
a) A written report, in pdf, that summarizes the analytical process and business conclusions, with a maximum of 30 pages.
b) The code scripts that allowed the students to reach to their main conclusions.

Guidelines for Project Evaluation

Standard criteria (18 points)

This component aims to evaluate the quality of the group work around what has been presented and discussed in practical classes:
a) Performance of the model
b) Quality and synthesis of the written report.
c) Number of variables present in the model used to score the customers. The fewer the better.
d) Quality of proposed transformations
e) Pre-processing steps
f) Quality and adequacy of visualizations

Advanced criteria (remaining 2 points)

To stimulate curiosity and self-learning skills, this component awards a bonus to groups willing to investigate themes not covered/ presented in practical and theoretical classes. This translates to the use of predictive techniques not presented during classes.

If groups choose different techniques, an additional appendix must be provided in the pdf report. This appendix will contain a theoretical presentation of the algorithm(s) used and a justification of the configurations made. If a group chooses to go for the advanced criteria, this component will be a MANDATORY discussion topic in the oral defenses.

Discount criteria

a) Delay in the delivery of the project (1 point for each day of delay)