Employee Attrition Model

Authors: Levin Reichmuth, Jorit Studer and Taejun Moon

Module: Machine Learning I

Submitted on June 10th, 2022

Supervisor: Dr. Matteo Tanadini, Daniel Meister and Dr. Alberto Paganini

Lucerne University of Applied Sciences and Arts

Table of Contents

1 Introduction

Employees, according to Swaminathan & Hagarty (2020), are the foundation of any business. Its success is largely determined by the quality of its employees and their ability to stay with the company. Organizations confront a number of issues as a result of staff attrition:

1. Training new personnel is costly in terms of both money and time.
2. Potential to lose experienced employees
3. Productivity impact
4. Profitability impact

Therefore, IBM data scientists created a fictitious data set as a challenge for data scientists. Among the data types are metrics such as education level, job satisfaction, and commute distance. The dataset can be found on the company's GitHub account (IBM, 2019).

2 Methodology

The following topics are layed out through out this paper:

1. Linear Models
2. Extending the Linear Model: Non-linearity
3. Extending the Linear Model: Generalised Linear Models
4. Support Vector Machines
5. Neural Networks
6. Optimisation

3 Data preparation

3.1 Data Transformation and Sanity Check

The following variables are available:

Table 1: Summary Numeric Variables

	N	Mean	SD	Min	Q1	Median	Q3	Max
Age	1470	36.92	9.14	18	30	36.0	43	60
DailyRate	1470	802.49	403.51	102	465	802.0	1157	1499
DistanceFromHome	1470	9.19	8.11	1	2	7.0	14	29
Education	1470	2.91	1.02	1	2	3.0	4	5
EmployeeCount	1470	1.00	0.00	1	1	1.0	1	1
EmployeeNumber	1470	1024.87	602.02	1	491	1020.5	1556	2068
HourlyRate	1470	65.89	20.33	30	48	66.0	84	100
MonthlyIncome	1470	6502.93	4707.96	1009	2911	4919.0	8380	19999
MonthlyRate	1470	14313.10	7117.79	2094	8045	14235.5	20462	26999
PercentSalaryHike	1470	15.21	3.66	11	12	14.0	18	25
StandardHours	1470	80.00	0.00	80	80	80.0	80	80
TotalWorkingYears	1470	11.28	7.78	0	6	10.0	15	40
TrainingTimesLastYear	1470	2.80	1.29	0	2	3.0	3	6
YearsAtCompany	1470	7.01	6.13	0	3	5.0	9	40
YearsInCurrentRole	1470	4.23	3.62	0	2	3.0	7	18
YearsSinceLastPromotion	1470	2.19	3.22	0	0	1.0	3	15
YearsWithCurrManager	1470	4.12	3.57	0	2	3.0	7	17

Table 2: Summary Factor Variables

	Level	N	%
Attrition	No	1233	83.9
	Yes	237	16.1
BusinessTravel	Rarely	150	10.2
	Frequently	277	18.8
	None	1043	71.0
Department	Sales	63	4.3
	R&D	961	65.4
	HR	446	30.3
EducationField	Human Resources	27	1.8
	Life Sciences	606	41.2
	Marketing	159	10.8
	Medical	464	31.6
	Other	82	5.6
	Technical Degree	132	9.0
EnvironmentSatisfaction	1	284	19.3
	2	287	19.5
	3	453	30.8
	4	446	30.3
Gender	Female	588	40.0
	Male	882	60.0
JobInvolvement	1	83	5.6
	2	375	25.5
	3	868	59.0
	4	144	9.8
JobLevel	1	543	36.9
	2	534	36.3
	3	218	14.8
	4	106	7.2
	5	69	4.7
JobRole	Healthcare Representative	131	8.9
	Human Resources	52	3.5
	Laboratory Technician	259	17.6
	Manager	102	6.9
	Manufacturing Director	145	9.9
	Research Director	80	5.4
	Research Scientist	292	19.9
	Sales Executive	326	22.2
	Sales Representative	83	5.6
JobSatisfaction	1	289	19.7
	2	280	19.0
	3	442	30.1
	4	459	31.2
MaritalStatus	Divorced	327	22.2
	Married	673	45.8
	Single	470	32.0
NumCompaniesWorked	0	197	13.4

Table 2: Summary Factor Variables *(continued)*

	Level	N	%
	1	521	35.4
	2	146	9.9
	3	159	10.8
	4	139	9.5
	5	63	4.3
	6	70	4.8
	7	74	5.0
	8	49	3.3
	9	52	3.5
OverTime	No	1054	71.7
	Yes	416	28.3
PerformanceRating	3	1244	84.6
	4	226	15.4
RelationshipSatisfaction	1	276	18.8
	2	303	20.6
	3	459	31.2
	4	432	29.4
StockOptionLevel	0	631	42.9
	1	596	40.5
	2	158	10.7
	3	85	5.8
WorkLifeBalance	1	80	5.4
	2	344	23.4
	3	893	60.7
	4	153	10.4

3.2 Missing Value Check

```
# Do we have any missing values?
sapply(emp_attrition, function(x) all(is.na(x) | x == '' ))
```

There are no missing values in this dataset.

4 Conclusion

References

IBM. (2019). *IBM HR analytics employee attrition & performance*. https://github.com/IBM/employee-attrition-aif360/blob/master/data/emp_attrition.csv

Swaminathan, S., & Hagarty, R. (2020). *IBM HR analytics employee attrition & performance* (2nd ed.). IBM. https://developer.ibm.com/patterns/data-science-life-cycle-in-action-to-solve-employee-attrition-problem/