COMM1190 A2 Team Report

Data, Insights and Decisions (University of New South Wales)

GLOBEX PHARMA REPORT

COMM1190 – Assessment 2

Numeer Imtiaz, Bhavik Aggarwal, Ricky Wang and Hailin Qin.

Date 23/07/2021 | Word count: 1989

1. Executive summary

Due to the growing concern of the staff turnover at Globex Pharma, consulting specialists have utilised the staff survey to develop predictive analysis models to identify factors that drive employee attrition and complacency in the workplace. These effectively determine the conditions of employees who are the most inclined to leave the firm. By accurately anticipating employee attrition, senior management of Globex can capitalise on the actionable insights to identify and take remedial action according to the suggestions made to target the driving forces as well as gearing their decision making to boost employee satisfaction and reduce turnover.

2. Predictive Data Analysis Models

Two data sets are used to randomly classify employee data: the training data set and the test data set (Containing the 30 entries). Two different models were used for prediction in each data set. By comparing the results of the two different models, the most accurate prediction can be obtained.

While we are confident about the results of our data and the advice we provide, we have to admit that our conclusions are limited. The first limitation is that our conclusions are based on the analysis of the data samples provided by Globex Pharma. Recommendations are also based on data analysis and extrapolation. Therefore, these conclusions and recommendations only apply to the sample data range, and should be applied in consideration of the organisation's size, existing cultures and procedures. As for the reasons for leaving employees who are not within the scope of the sample data, it is only for reference and not absolutely valid. The second limitation is that we used two models to compare and get more accurate results. However, the data models that can be used for prediction are not limited to these two. Different prediction models lead to different conclusions. Just looking at the two that we used, our results are correct and reliable. Finally, it is essential for Globex to recognise that our prospective models cannot be considered in isolation and other extrinsic factors such as job opportunities from other

companies must be accounted for when predicting attrition.

Reasoning for Models

Consultants have carried out an in-depth exploratory data analysis from the recent staff survey at Globex Pharma to understand the nature and scope of the turnover crisis at the company.

Through conducting exploratory data analysis, our consulting specialists have designed two comprehensive models which may allow Globex in anticipating employee attrition. ***Model 1*** utilises a logistic regression containing 15 predictors to determine the probability of staff turnover as a dependent variable. The model utilises a multi-linear model to predict the log-odds of attrition based on the data collected for the chosen variables, for each individual employee. Figure 1.1 demonstrates the use of the log-odds in predicting the probability of attrition. When the model predicts that the employee whose data is used did not leave the firm. However when , the model predicts that the employee did leave the company.

In choosing *Model 1*, the first step was to conduct a logistic regression using all 25 independent variables in proving the log-odds for Attrition. Then, we rationalised this model to 13 variables (*BusinessTravel* and *MaritalStatus* being categorical variables leading to 15 predictors) by using the Bayesian Information Criterion (BIC) method. This allowed us to extract only the significant variables from our initial model which disprove the null hypothesis:

$$H_0: \beta_p = 0$$

Model 2 utilises a classification tree in predicting whether employees stayed or left Globex within the last year. The model uses an intuitive set of rules which can be used by the company in forecasting staff turnover based on the information collected.

The parameters for *Model 2* were selected using the same 13 variables chosen in *Model 1* as determined through the BIC method. The same variables were chosen in order to maintain consistency within our findings and recommendations.

Factors that drive employee attrition

Both predictive analysis models effectively communicate the most prominent driving forces leading to employee attrition. *Model 1* (Refer to Table 4) suggests that the key parameters which have a positive effect of reducing the log-odds and therefore reducing P(Y=1) include 'MonthlyIncome' (-0.019), 'JobInvolvement' (-1.80), 'EnvironmentSatisfaction' (-1.306), 'JobSatisfaction' (-1.304), 'NumberCurrRole' (-0.637), 'EmployeeAge' (-0.591) and 'RelationshipSatisfaction' (-0.903). 'MonthlyIncome' and 'EmployeeAge' are indicated to have the greatest numerical impact on our model however their impact is unclear due to discrepancies in their 95% confidence intervals and as a result, cannot be addressed through recommendations. Conversely, the variables which are suggested to increase log-odds include 'YearsSinceLastPromotion' (0.595), 'OvertimeYes' (1.634), 'DistanceFromWork' (0.329), and 'NumCompaniesWorked' (0.491). The variables 'BusinessTravel' and 'MaritalStatus' are categorical variables with 3 or more options and thus have been separated into multiple predictors. 'BusinessTravel' suggests a negative correlation between the frequency of travel and the probability of employees staying.

Model 2 utilises the same variables employed in *Model 1* excepting for 'SalaryIncrease'. The classification tree identifies the most important factors affecting attrition to be 'Overtime' and then the next primary consideration is based on 'MonthlyIncome'. *Model 2* predicts that 7% of the given employees left the company. This resonates with our assessment of the test data which predicted that 2 out of 30 employees would leave.

Assessment of model's accuracy

To outline the reliability of the predictive data analysis framework for management at Globex Pharma, consultants have substantiated the accuracy of the models substantiated through observing the confusion matrices which relay high significance despite the accuracy for predicting positive attrition being approximately 35%. Nevertheless, the accuracy and value of the models is not negated as there is a prominent discrepancy in the data as less than 20% of the given data of attrition for all employees is yes. The available data is heavily skewed to the 'No' side, with only 204 'Yes' against 1108 'No', stemming as the underlying cause of discrepancy in the model.

Model 1 has returned considerably accurate readings of the data with overall accuracy of

87.58%. The model's accuracy is best showcased through the accuracy reading for the attrition of 'No' at 97.65%. The accuracy of 'Yes' being 32.84% can be partially accounted for due to the skewed data.

Model 2 has an overall accuracy reading of 87.88%, indicative of a reliable predictive framework to help gain insight into which employees are likely to experience attrition. Similar to Model 1, the classification tree has an accuracy of '97.47% for 'No' and only 35.87% for 'Yes' .

3. Recommendations

According to attained predictors and previous analysis, several recommendations can be concluded to improve employee retention. The first strategy is to introduce a training system, including induction, upskilling and job rotation. Bagga (2013) states that induction can help improve employee loyalty, allowing them to be better integrated into the organisation. This strategy may prove vital in improving job involvement, a key parameter improving the probability of job retention. Moreover, Km (2020) also points out that changing circumstances of business need constantly improving skills. Therefore, the implementation and incentivisation of training programs which allow employees to upskill in both their technical and interpersonal abilities can be beneficial in reducing anxiety and concerns of inadequate capability. This can further improve the flexibility of Globex's workforce and have a flow-on effect to key variables including job satisfaction as well as encouraging workers to pursue promotional opportunities within the organisation.

In addition, job rotation training can expand employees' synthetical ability effectively, stimulate innovation and problem-solving capacity (Alias et al. 2018). The company should implement a culture of acceptance in allowing employees to transition between roles smoothly, fostering an improvement in the key variable 'EnvironmentSatisfaction'. This policy can be implemented in parallel to the induction and upskilling programs and is targeted towards decreasing the impact of the variable 'YearsSinceLastPromotion' on attrition. Moreover, increased job rotation opportunities may assist in reducing employee attrition as implied in *Model 2*, which suggest that employees who have spent 3 or more years in their current role are more likely to leave the organisation.

The effect of distance from work travelled by employees and its subsequent impact on rates of attrition resulting in increased employee turnover is another key insight to be considered. *Models 1* and *2* outline that employees who travel longer distances are more inclined to experience attrition and leave Globex Pharma. Senior management can capitalise on this driving force behind employee dissatisfaction by taking remedial action

by implementing systems which minimise complacency associated with travel through providing employees with public transport concessions. Moreover, Globex can encourage remote work and 'work-from-home' opportunities where possible for employees who live considerably far away from Globex operating quarters to reduce overall travel times, thus reducing attrition stemming from excessive travel.

Additionally, 'Business Travel' has been found to be a predictor which influences the attrition of employees, allowing management of the firm to take advantage of the opportunity to easily reduce employee turnover through reducing travel during work hours. By communicating to management of the firm that employees who travel more are highly more likely to experience attrition, Globex can accordingly implement measures to minimise employees travelling during work through interacting with stakeholders such as executives meeting investors, maintaining supplier relations as well as through an online digital medium in place of travelling to meet essential business partners. Moreover, management can also take remedial action in the form of targeting younger employees who are more likely to leave the company than those who are older by infusing a change in the work culture by developing long-term career progression plans enticing younger workers to prolong their employment position at the firm, assisting management achieve the goal of reducing staff turnover.

Finally, 'Overtime' has been identified as a significant driver of employee attrition. In *Model 1*, when Overtime = Yes it increases the log-odds by 1.634. Similarly, in *Model 2*, a predicted 69% are retained when Overtime = No. In the given dataset, approximately 28% of Globex's workforce works overtime which may indicate a structural problem in workload assignment. We recommend Globex to conduct an assessment of working hours and workload which may provide key insights regarding which sectors are experiencing staff shortages and productivity losses. Globex should accordingly look at employing more staff and implement specialist training programs specifically targeted towards improving productivity in different sectors of the business. Our consultants recommend that reducing the overtime rate through spreading out business workload across a greater number of workers can lead to upskilling employees through exposure to diverse job roles.

Moreover, a reassessment of scheduling practices and the implementation of health protection programmes for people working in jobs involving overtime and extended hours can produce a work environment that attracts, motivates and retains hard-working individuals (Krishnamoorthy and Aravindan, 2020). These recommendations are key in addressing the variables *EnvironmentSatisfaction* and *JobSatisfaction* which are demonstrated to have a positive correlation with employee retention.

4. Employees who are likely and unlikely to leave the organisation

Our consultants have generated prognoses using our models, which predict the attrition for 30 employees who Globex is unsure will leave the organisation in upcoming months or not. The findings from both models as seen in Table 4 and Table 8 suggest that out of 30 employees, 2 are likely to leave. It is important to note that according to both models, the employee with the EmployeeNumber '2080' is likely to leave the organisation in upcoming months. This is confirmed by the probability generated via *Model 1* which suggests that this employee has a 68.2% chance of leaving the organisation.

Using *Model 1*, it is predicted that the employee with EmployeeNumber '2073' has a 64.7% probability of leaving the firm. This finding differs from *Model 2* which indicates that the other employee who is likely to leave Globex has the EmployeeNumber '2040'. Through assessment, it is predicted that, *ceteris paribus*, all other employees listed in the test data are less than 50% likely to leave Globex.

Appendix

Model 1: Logistic Regression to predict Attrition

Where:

Probability of Attrition being ‘Yes’

Intercept

Coefficient for the monthly incomes of each employee

Coefficient for the years since last promotion for each employee

Coefficient for whether employees work overtime or not (Where ‘Yes’=1, ‘No’ = 0)

Coefficient for satisfaction with work environment

Coefficient for job satisfaction

Coefficient for employees’ who are frequently involved in business travel

Coefficient for employees’ who are rarely involved in business travel

Coefficient for the number of years employees’ have been in their current role

Coefficient for employee age

Coefficient for relationship satisfaction with work colleagues

Coefficient for distance from work

Coefficient for job involvement

Coefficient for the number of other companies employees have worked for

Coefficient for employees with a marital status of single

Coefficient for employees with a marital status of married

Figure 1.1 Calculating Predictive Probability using Model 1

Table 1: Coefficients for Logistic Regression Model

Variables	*Coefficients*	*Std. Error*	*Lower 95%*	*Upper 95%*
	1.249 *	2.654	-0.172	2.66
	-0.019 ***	0.02	0	0
	0.595 ***	0.497	0.089	0.316
	1.634 ***	0.68	1.275	2.00
	-1.306 ***	1.096	-0.515	-0.725
	-1.304 ***	1.076	-0.512	-0.735
	1.764 ***	1.55	0.978	2.642
	1.042 **	1.447	0.316	1.870
	-0.637 ***	0.489	-0.246	-0.382
	-0.591 **	0.16	-0.067	-0.665
	-0.903 ***	1.102	-0.406	-1.166
	0.329 ***	0.321	0.012	0.500
	-1.80 ***	0.456	-0.729	-0.934
	0.491 ***	0.492	0.062	0.753
	1.409 ***	0.973	0.051	1.086
	2.04 *	0.968	0.905	1.94

This table contains the coefficients for each of the variables within Model 1. P-Values were generated using the standard errors and the significance codes

used are:

Table 2: Confusion Matrix for for Logistic Regression Model for Training Data (1312 Entries)

		ACTUAL ATTRITION	
	Variables	**No**	**Yes**
PREDICTED ATTRITION	**No**	1082	137
	Yes	26	67

Table 3: Accuracy of Logistic Regression Predictions for Training Data

Overall Accuracy	87.58%
Accuracy of Attrition = 'No'	97.65%
Accuracy of Attrition = 'Yes'	32.84%

Table 4: Prediction for Attrition in Test Data using Model 1

Entry No.	Employee No.	P(Y=1)	Predicted Attrition
1313	2037	0.014	No
1314	2038	0.103	No
1315	2040	0.379	No
1316	2041	0.02	No
1317	2044	0.026	No
1318	2045	0.089	No

1319	2046	0.105	No
1320	2048	0.108	No
1321	2052	0.042	No
1322	2053	0.49	No
1323	2054	0.337	No
1324	2055	0.262	No
1325	2056	0.068	No
1326	2057	0.043	No
1327	2060	0.087	No
1328	2061	0.086	No
1329	2062	0.04	No
1330	2064	0.218	No
1331	2065	0.019	No
1332	2068	0.054	No
1333	2071	0.173	No
1334	2072	0.173	No
1335	2073	0.647	Yes
1336	2074	0.058	No
1337	2075	0.145	No
1338	2076	0.256	No
1339	2077	0.076	No
1340	2078	0.188	No
1341	2079	0.009	No
1342	2080	0.682	Yes

Table 5: Confusion Matrix for for Logistic Regression Model for Test Data (30 Entries)

		ACTUAL ATTRITION	
	Variables	No	Yes
PREDICTED ATTRITION	No	25	3
	Yes	0	2

Model 2: Classification Tree to predict Attrition

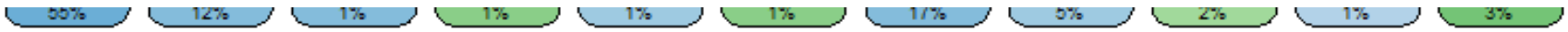

Table 6: Confusion Matrix for Classification Tree Model for Training Data (1312 Entries)

		ACTUAL ATTRITION	
	Variables	No	Yes
PREDICTED ATTRITION	No	1080	131
	Yes	28	73

Table 7: Accuracy of Classification Tree Predictions for Training Data

Overall Accuracy	87.88%
Accuracy of Attrition = 'No'	97.47%
Accuracy of Attrition = 'Yes'	35.78%

Table 8: Prediction for Attrition in Test Data using Model 2

Entry No.	Employee No.	Predicted Attrition
1313	2037	No
1314	2038	No
1315	2040	Yes
1316	2041	No
1317	2044	No
1318	2045	No
1319	2046	No
1320	2048	No
1321	2052	No

1322	2053	No
1323	2054	No
1324	2055	No
1325	2056	No
1326	2057	No
1327	2060	No
1328	2061	No
1329	2062	No
1330	2064	No
1331	2065	No
1332	2068	No
1333	2071	No
1334	2072	No
1335	2073	No
1336	2074	No
1337	2075	No
1338	2076	No
1339	2077	No
1340	2078	No
1341	2079	No
1342	2080	Yes

Table 9: Confusion Matrix for Classification Model for Test Data (30 Entries)

	ACTUAL ATTRITION

	Variables	No	Yes
PREDICTED ATTRITION	No	25	3
	Yes	0	2

Code

```
library(moments)
library(rpart)
library(rpart.plot)
library(leaps)
#STEP 1: Logistic Regression for all 26 Variables
A2_Dataset_Training$IfAttrition <- ifelse(A2_Dataset_Training$Attrition == "Yes", 1, 0)
logisticAttrition <- glm(IfAttrition ~ Age + BusinessTravel + Department + DistanceFromHome + Education +
                           EnvironmentSatisfaction + Gender + JobInvolvement + JobRole + JobSatisfaction +
                           MaritalStatus + MonthlyIncome + NumCompaniesWorked + OverTime + SalaryIncrease +
                           RelationshipSatisfaction + StockOptionLevel + TotalWorkingYears + TrainingTimesLastYear +
                           WorkLifeBalance + YearsAtCompany + YearsInCurrentRole + YearsSinceLastPromotion +
                           YearsWithCurrManager + HighPerformance, family = binomial(),
                         data = A2_Dataset_Training)
summary(logisticAttrition)

#STEP 2: Rationalise predictors using BIC - ends up with only the significant variables and a similar R^2
logisticAttritionBIC <- MASS::stepAIC(logisticAttrition, trace = 0, k = log(nrow(A2_Dataset_Training)))
summary(logisticAttritionBIC)
confint(logisticAttritionBIC)
```

***This code for using BIC with logistic regressions is extracted from**

```

#STEP 3: Confusion Matrix for Logistic Regression
probsAttritionLR <- predict(logisticAttritionBIC, newdata = A2_Dataset_Training, type = "response")
AttritionPredLogistic <- rep("No", 1342)
AttritionPredLogistic[probsAttrition > 0.5] <- "Yes"
table(AttritionPredLogistic, A2_Dataset_Training$Attrition)

#Number of 'Yes'/'No'in Attrition Training Data
totalYes <- sum(A2_Dataset_Training$Attrition =="Yes")
totalNo <- sum(A2_Dataset_Training$Attrition =="No")
totalNo; totalYes

OverallAccLR <- paste("Overall Accuracy of Training Data: ", ((1082+67)/1312)*100, "%")
AttritionNoAccLR <- paste("Accuracy of Attrition=No: ", (1082/(1082+26)*100), "%")
AttritionYesAccLR <- paste("Accuracy of Attrition=Yes: ", (67/(67+137)*100), "%")
OverallAccLR; AttritionNoAccLR; AttritionYesAccLR
```

https://bookdown.org/egarpor/PM-UC3M/glm-modsel.html

```r
#STEP 4: Classification Tree using only the variables identified in Step 2
treeAttrition <- rpart(Attrition ~ Age + BusinessTravel + DistanceFromHome +
                         EnvironmentSatisfaction + JobInvolvement + JobSatisfaction + MaritalStatus +
                         MonthlyIncome + NumCompaniesWorked + OverTime + SalaryIncrease + RelationshipSatisfaction +
                         StockOptionLevel + YearsInCurrentRole + YearsSinceLastPromotion,
                       data = A2_Dataset_Training)
rpart.plot(treeAttrition, Margin = -0.057)

#STEP 5: Confusion Matrix for Classification Tree
probsAttritionCT <- predict(treeAttrition, newdata = A2_Dataset_Training, type = "response")
table(probsAttritionCT, A2_Dataset_Training$Attrition)

OverallAccCT <- paste("Overall Accuracy of Training Data: ", ((1080+73)/1312)*100, "%")
AttritionNoAccCT <- paste("Accuracy of Attrition=No: ", (1080/(1080+28)*100), "%")
AttritionYesAccCT <- paste("Accuracy of Attrition=Yes: ", (73/(73+131)*100), "%")
OverallAccCT; AttritionNoAccCT; AttritionYesAccCT

#STEP 6: Predictions for Last 30 Entries Using Both Models (Q4)
print("Predicted Probability of Attrition for last 30 entries using Logistic Regression model:")
print(probsAttritionLR[1313:1342])

print("Predicted Probability of Attrition for last 30 entries using Classification Tree model:")
print(probsAttritionCT[1313:1342])
```

Reference List

Alias, N., Othaman, R., Hamid, L.A., Salwey, N.S., Romaiha, N.R. 2018, 'Managing Job Design: The Roles of Job Rotation, Job Enlargement and Job Enrichment on Job Satisfaction', *Journal of Economic & Management Perspectives*, vol. 12, no. 1, pp. 397–401. Available at: https://www-proquest-com.wwwproxy1.library.unsw.edu.au/docview/2266297444?pq-origsite=primo.

Bagga, G. 2013, 'How to keep the talent you have got', *Human Resource Management International Digest*, vol. 21, no.1, pp. 3–4. Available at: https://www-proquest-com.wwwproxy1.library.unsw.edu.au/docview/1282262731?pq-origsite=primo.

Km, R. K. 2020 'Employee Retention – Challenges and Realities Faced by Corporates for New Recruits as well as Existing Employees', *Ushus Journal of Business Management*, vol. 19, no.4, pp. 75–93. Available at: https://www-proquest-com.wwwproxy1.library.unsw.edu.au/publiccontent/docview/2499899311?pq-origsite=primo.

Krishnamoorthy, D. & Aravindan, S. 2020, 'A case study on overtime and its impacts on employees job satisfaction', *Journal of Contemporary Issues in Business and Government*, vol. 26, no. 2, pp. 911-919. Available at: https://cibg.org.au/article_7815_e3b88060de84619fa2f7a1a1b3094aff.pdf

Portugués, E. G. 2021, 'Predictive Modeling - 5.6 Model Selection', *Bookdown*, viewed 23 July 2021, <https://bookdown.org/egarpor/PM-UC3M/glm-modsel.html>.