An analysis of employee attrition by age

Quy Vu

****

1. Motivation, data, and research questions

Employee attrition has always been one of the most costly and difficult challenges to employers. Not only does it negatively impact performance, (Tonz and Huckman, 2008) it also raises recruitment and training costs. In the UK, attrition costs businesses at least £4.13bn every year to train new employees to work at optimum productivity (Oxford Economics, 2017).

There has been numerous research on factors affecting attrition. However, the findings were not consistent due to the diversity of methods applied (Ongori, 2007). On the other hand, quantitative research mostly implemented linear regression as analytical method with less than 300 observations (Michaels and Spector, 1982; Cotton et al., 1986; Hom and Kinicki, 2001; Morrell et al., 2004; Ahituv and Lerman, 2005; Ekkehart, 2006; Cohen and Golan, 2007). In addition, recent analyses dedicated to attrition (which should be separated from turnover) were not made available to the wide public. This situation where limited research is conducted on this topic can be explained by the increasing confidentiality of human resources data (Armstrong et al., 2014).

In 2017, IBM has released a dataset on employee attrition with all personal identifiers removed for confidentiality purpose (Watson Analytics Forum, 2017). It contains information on the demographics, career progression, job specification of 1,470 employees from sales, research, and human resources departments working in the pharmaceutical industry. The population aged from 18 to 60 and took charge of all job levels, from executive to manager. As a result, this dataset may give interesting insight to employee attrition, especially in research-involved companies. This is also an opportunity to revisit empirical findings and to reconsider their applicability in the current context.

The aim of this analysis is to revisit previous research findings, identify then interpret the most significant factors that help explain employee attrition. With an interest in investigating how the explanation varies across age groups, this analysis seeks to answer to the following questions:

- Are previous research findings help explain attrition in this case?
- What other factors best explain employee attrition?
- Are the explanations difference across age groups?

2. Tasks and approach

2.1. Characterise attrition by age

To gain an overview of the distribution of employee attrition by age, a kernel density estimation plot was utilised (Figure 1 - left) to represent those who left and those who stayed. To take into account of the population size of each age group, the attrition rate was visualised with scatter plot in which the size of each dot corresponds to the number of employee of that age (Figure 1 - right). Subsequently, age groups with similar trend in attrition rate were consolidated to larger groups to facilitate subsequent analyses. Stacked bar plot and multidimensional scaling (where more similar

observations were closely represented) was utilised to assess the grouping (Figure 2b). The aim of this was also to provide input for the modelling stage where a separate model is fitted to each group.

2.2. *Characterise age groups with selected features*

Available features that had been agreed as correlated to employee attrition by numerous studies were selected to describe the 3 age groups, which were: Performance (-), job and workplace satisfaction (-), age (-), job level (-), marital status (Married negative), education (+), income (-), promotion (-) (Ongori, 2007). Categorical features were then binary encoded, then the average value of each feature was visualised with a radar plot.

This method highlighted the most significant differences between the groups, facilitating high level explanation of how attrition varies by age. This averaging method, in general, is not robust against outliers, but provides a representative employee profile. In addition, this method works with binary variables while median as its popular robust alternative would output either 0 or 1.

2.3. *Identify the variation of relationship with attrition by age*

For each age group, radar plot was again selected to characterise leaving and staying employees. This technique allows the investigation of how each employee characteristic related to attrition in general, and in each age group. In addition, analysing by age groups instead of the whole population ensures patterns in the smaller groups are not overwhelmed by larger ones.

2.4. *Develop explanatory models*

To assess how each feature contributed to attrition in each age group, logistic regression was selected. This method is particularly suitable for binary prediction tasks, facilitates interpretability but still produces decent prediction result comparing to other machine learning algorithms (Pailet et al., 2009). In addition, comparing to linear regression, it does not rely on assumptions such as normally distributed data, homogeneity of variance, but would function best with in cases where data are linearly separable (Kotsiantis et al., 2007).

Logistic regression's multicollinearity assumption was addressed with feature selection steps. Feature correlations were visualised with heatmap (Figure 5 – Appendix 1), in which highly correlated features were identified, and pair plot was used to spot any non-linear pattern. To minimise manual feature selection, only features containing information that can be derived from other features and are highly correlated with other features in the dataset; or containing no meaning were removed. Automatic feature selection was then performed with LASSO, allowing less important features to be excluded from the model.

Categorical features were binary encoded. The dataset was then normalised to zero mean and unit variance to facilitates convergence of gradient descent - which was used as the optimisation method (Ioffe and Szegedy, 2015). To address class imbalance (237 left/1470 records, SMOTE oversampling technique was used to generate “synthetic examples along the line segments joining any/all of the k minority class nearest neighbours”. This method helps improves classifier performance, particularly in ROC space (Chawla et al., 2002).

2.5. *Evaluate models, assess findings and unconsidered features*

The main metric to evaluate the model’s performance was the average under the precision-recall curve score, ranging from 0.5 to 1 with 1 representing perfect performance and vice versa. This

metric is more informative and considered more helpful in evaluating classification tasks in the presence of class imbalance where we are more interested in investigating the rare class i.e leaving employees in this case (Davis and Goadrich, 2006; Saito and Rehmsmeier, 2015). The output coefficients were represented in heatmap to facilitate feature-wise and group-wise comparison. Interpretation of the coefficient was then compared with the findings in previous section and that of empirical studies.

3. Analytical steps

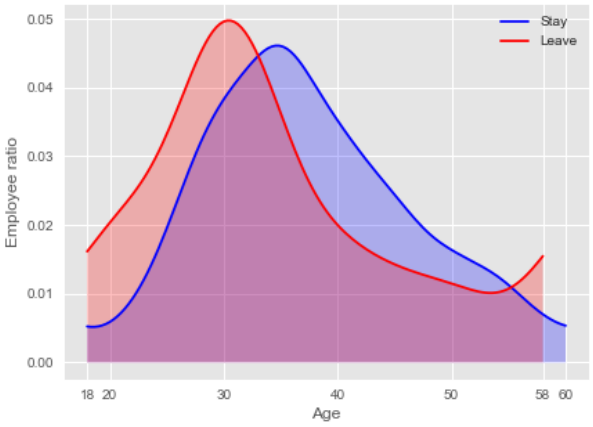
3.1. Characterise attrition by age

Figure 1-left shows the distribution of leaving employee across ages in red and its counterpart in blue. The majority of employees in this dataset aged from 25 to 45, meaning that patterns noticed when analysing the population as a whole is likely to be dominated by this age group.

Figure 1-right scatters the attrition rate by age, with the size of each dot corresponds to the size of that age group. 3 clear trends can be noticed: Attrition decreased quickly until 24, then showed a steadily decreasing trend before rising from 55 onwards. This suggested that employees aged between 18-24, 25 – 55, and 56 – 60 should be investigated separately.

Figure 2 displays the attrition rate of the 3 large age groups, suggesting that the above trend was adequately maintained. Figure 2b shows the output of applying multidimensional scaling to the data, suggesting the similarity of employees in the second age group.

Figure 1 - Employee attrition by age



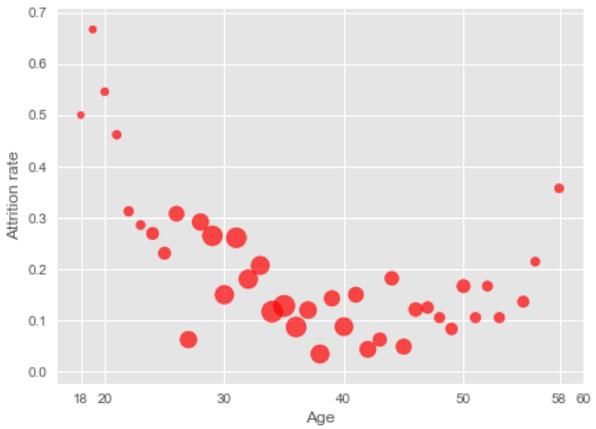


Figure 2 - Employee attrition by age group

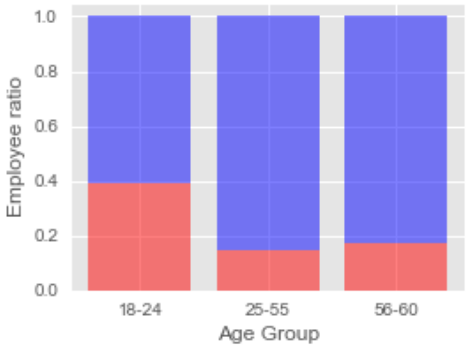
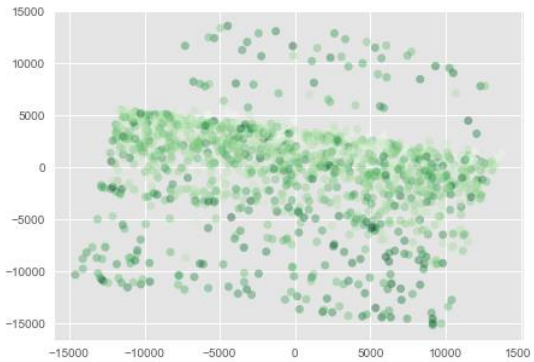


Figure 2b - MDS on numerical features, coloured by age

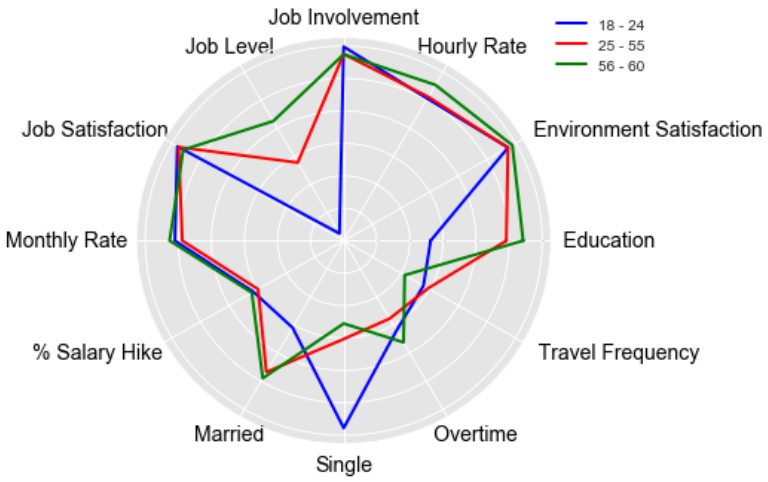


3.2. *Characterise age groups with selected features*

The findings of empirical studies were drawn upon to select a set of features highlighting the differences between age groups. The averages of each selected feature are visualised in Figure 4, which suggested that job level, education level, and marital status can be used to explain the difference in attrition across age groups. Since the majority of employees aged from 18-24 are yet to complete college and employers mostly prefer to recruit graduate students (Bureau of Labor and Statistics, 2017), they are limited to temporary or very junior position before resuming their studies (Bureau of Labor and Statistics, 2017), which can be related to high attrition rate.

The plot does not suggest any other significant difference between the two remaining groups.

Figure 3 - Comparison of age groups by selected features



3.3. *Identify the variation of relationship with attrition by age*

Figure 4 compares the average profile of leaving and staying employees by age groups. The first plot representing all age groups suggested that attrition is related to low job level, low satisfaction, being single, high travel frequency, having to work overtime, and low job involvement. This is consistent with numerous empirical studies, as suggested by the meta-analysis review conducted by John & Jeffrey in 1986. Table 1 provides a summary of explanation of this finding.

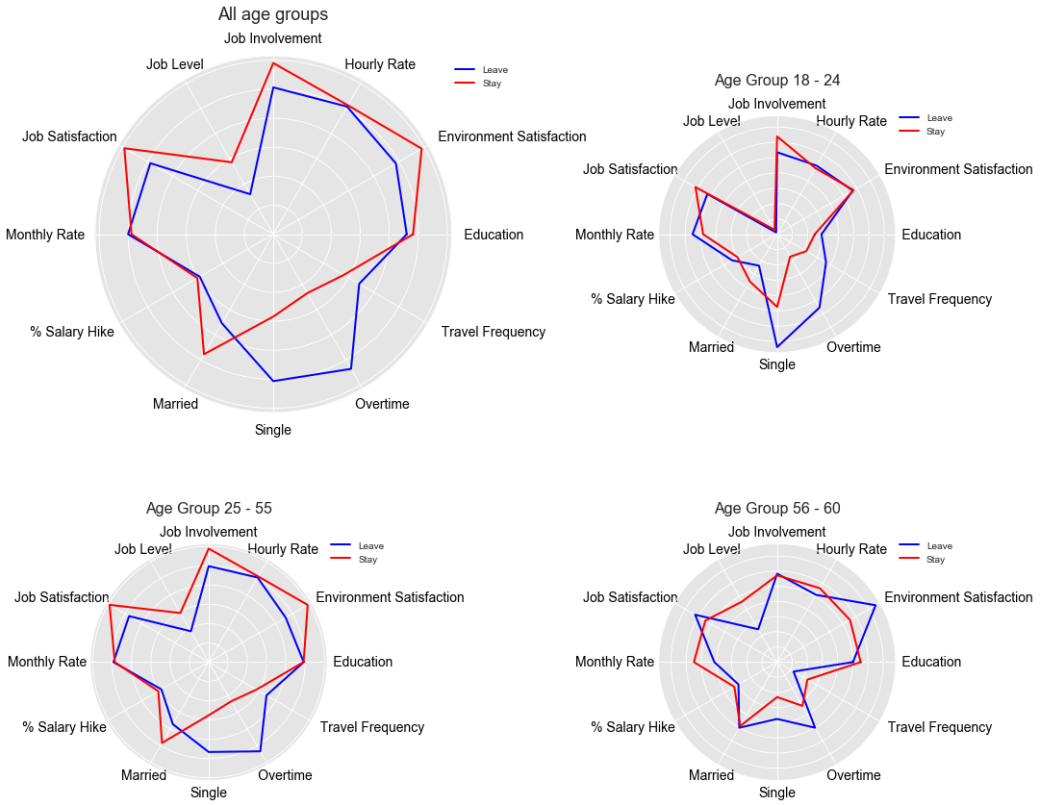
Figure 5 separates the average profile of leavers and the others. The similarity of the two figures on the left confirms that the overall trend is dominated by the 25-55 age group. Being single and having to worked overtime appear to be strong predictors of attrition, notably in the junior and the intermediate groups. However, while lower job satisfaction, lower job involvement, and high travel frequency are considerably linked to attrition in these groups, the opposite is true for the older one. Lower job level can be associated to attrition for mature employees, but not significant for junior ones as they are limited to junior positions, as discussed above. Interestingly, older employees left job with high satisfaction – this suggest that more dominating factor(s) could determine attrition in this group. A reasonable suggestion is health issue and retirement intention (Morrow, 1982).

Given the difference in patterns across age groups, the necessity of fitting separate models is again reinforced.

Table 1: *Correlation between each feature and attrition explained & examples of supporting studies*

Features	Correlation	Explanation	Reference
Job level	Negative	Higher job level also means stronger commitment to the organisation; thus, employees are less likely to leave the company as they reach more senior positions.	Michaels and Spector, 1982
Job and environment satisfaction	Positive	Satisfied employees are less likely to actively look for other opportunities. Thus, the more satisfied an employee, the more he/she is willing to stay.	Hom and Kinicki, 2001.
Marital status	Married negative	Married employees are less likely to change job given their intention to settle down. Marriage also indicates a stable and mature career which is negatively related to turnover.	Arthur, 1981
Job Involvement	Negative	Describes individuals' ego involvement with their work and indicates the importance of work to them. This positively influences job satisfaction and organisational commitment.	Parasuraman, 1982
Travel Frequency, overtime	Positive	These two features indicate work intensity and stress, which are indicators of leaving intention of employees.	Parasuraman, 1982

Figure 4 - Association between employee characteristics and attrition by age group



3.4. *Develop explanatory models*

The previous sections only covered features that were either considered related to attrition by empirical studies. As such, Logistic Regression was selected to take into account of other features such as, *Department, DistanceFromHome, Gender, StockOptionLevel, NumCompaniesWorked, EducationField*, etc. Categorical features were binary encoded, and class imbalance was addressed with SMOTE as oversampler. Feature selection was supported by visual inspection of the correlation heatmap (Figure 5 – Appendix 1), after which 7 features were removed, as detailed in Table 2. Then, pair plot was utilised to check for non-linear pattern, (Figure 6 – Appendix 2) and automated feature selection was implemented with LASSO. 3 models were fitted to the 3 age groups and 1 for the whole population.

Table 2: *Removed features and corresponding reasons*

Removed features	Reasons
EmployeeCount, Over18, StandardHours	Contains only "1", "Y" and "80" respectively;
Employee Number	No meaning as suggested by correlation heatmap
Job Role	Can be derived from "Job Level" and "Department"
Monthly Income	Highly correlated with "Job Level"
PerformanceRating	Highly correlated with "PercentSalaryHike", value ranged from 3 to 4 only

3.5. *Evaluate models, assess findings and unconsidered features*

As shown is Table 3, fitting a separate model for each age group significantly increase the models' performance. The AUPRC of 0.84, 0.91, and 0.87 arguably provide grounds for inference from these models.

The output coefficients are represented in Table 4, with colour transiting from Blue to Red representing low to high coefficient. In multivariate regression, the coefficient means the change in the result caused by a feature, provided that the others held constant. For instance, working overtime is associated with a 1.11-unit increase in the log-odds of attrition in the 25-55 age group. The coefficients verify that explanation of attrition does varies by age. The result suggests the negative association of job level, job involvement; the positive association of being single, working overtime to attrition, and how the explanation given by job satisfaction and environment satisfaction varies across age groups. These are consistent with the findings in section 3.3 and also suggests that empirical results are still applicable after at most 20 years. Furthermore, additional patterns are highlighted by unconsidered features. *DailyRate,* demonstrates negative association while *NumCompaniesWorked* relate positively to attrition. Interestingly*, DistanceFromHome, YearsSinceLastPromotion, YearsInCurrentRole* shows positive correlation with attrition in older employees, which is opposite to what observed in the rest.

Table 3: *Logistic regression coefficient*

Features	Age Group				Features	Age Group			
	All	18-24	25-55	56-60		All	18-24	25-55	56-60
Age	-0.95	0.34	-0.44	-1.30	MaritalStatus_Married	0.61	0.56	-0.07	0.10
DailyRate	-0.37	-0.20	-0.07	-4.10	MaritalStatus_Single	1.10	1.03	0.46	1.93
Department_R&D	0.02	0.00	-0.35	0.81	MonthlyRate	0.51	0.38	0.08	-0.60
Department_Sales	0.34	0.35	0.13	-1.13	NumCompaniesWorked	0.74	0.49	0.36	2.21
DistanceFromHome	0.78	-0.86	0.68	0.48	OverTime_Yes	1.36	1.59	1.11	0.41
Education	0.39	0.13	0.03	1.12	PercentSalaryHike	-0.37	0.65	-0.10	0.54
EducationField_Life Sciences	-0.14	0.31	-0.30	-1.39	RelationshipSatisfaction	-0.19	-0.59	-0.37	-0.01
EducationField_Marketing	0.50	0.52	0.07	-0.30	StockOptionLevel	-0.11	-0.56	-0.43	0.20
EducationField_Medical	-0.13	-0.53	-0.15	-0.20	TotalWorkingYears	-0.56	-0.08	-0.47	-0.23
EducationField_Other	-0.05	-0.23	-0.11	0.00	TrainingTimesLastYear	-0.11	-0.12	-0.30	-0.94
EducationField_Technical Degree	0.35	0.18	0.25	1.27	TravelFreq	0.26	0.58	0.25	-1.25
EnvironmentSatisfaction	-0.66	-0.24	-0.66	1.31	WorkLifeBalance	-0.37	-0.17	-0.34	-0.05
Gender_Male	0.38	-0.40	0.05	-1.34	YearsAtCompany	0.20	-0.45	-0.23	1.13
HourlyRate	0.19	0.10	-0.15	-1.04	YearsInCurrentRole	-0.92	-0.94	-0.45	1.02
JobInvolvement	-0.62	-0.54	-0.60	-0.26	YearsSinceLastPromotion	0.75	-0.58	0.08	1.57
JobLevel	-1.16	-0.45	-0.66	-1.96	YearsWithCurrManager	-0.71	-1.13	-0.43	0.33
JobSatisfaction	-0.70	0.09	-0.55	-0.51					

Table 4: *Area under the precision-recall curve for each model*

Age Group	Area under the precision-recall curve
All	0.77
18 – 24	0.84
25 – 55	0.91
56 – 60	0.87

4. Findings

This analysis assessed the explanatory power of various factors in employee attrition, whether and how the explanation varies across age groups, and whether previous research findings were still applicable in the case of 1,470 employees working in the pharmaceutical industry. Figure 1 shows that attrition dropped from its peak at age of 18, then slowly decreases from age of 25 before rising from age of 55 onwards. Also, further analyses suggest that high attrition in junior employees can be related to not having graduated while the unusual fact that older employees left job with high satisfaction can be justified by health and retirement.

The visualisation method implemented in Section 3.3 and 3.4 confirms that overall, attrition can be explained by low job level and satisfaction, being single, high travel frequency, working overtime and low job involvement. By contrasting average profiles of leaving and staying employees across age groups, it can be noticed that the explanation for attrition does indeed vary by age. In specific, job satisfaction and involvement with high travel frequency are strongly related to attrition in junior and senior employees, while older workers demonstrated an opposite trend. To further assess

these findings, logistic regression was implemented which gave supporting results. Hence, it is concluded that the results of previous empirical studies on attrition is applicable in this case.

Furthermore, group-wise logistic regression helped highlighted patterns in unconsidered factors: *DailyRate* correlate positively with attrition while *NumCompaniesWorked* shows the contrary, and *DistanceFromHome, YearsSinceLastPromotion, YearsInCurrentRole* shows positive association with attrition in older employees, which is opposite to what observed in the rest.

5. Critical reflection

5.1. Implications of findings for domain

For all of the records, the most useful factors in explaining attrition are *MaritalStatus_Single, Overtime,* and *Job Level.* Other factors such as *DistanceFromHome, YearsSinceLastPromotion, YearsInCurrentRole* helped describe attrition differently across age groups. However, for junior and older workers, it may exist other factors that are not included in the dataset which may mainly drive attrition in these two age groups, for example not having graduated for the 18-24 age group or being close to retirement for the 55-60 age group. As such, one may conclude that these factors explain the unusual pattern in attrition, such as satisfied older workers still left job.

This analysis has only considered attrition by grouping of ages, on top of which further grouping may also be considered. Previous research has pointed out attrition may vary across different job level, e.g managers vs junior employees; or job characteristics which can be related to the department of an employee (Cotton et al., 1986; Ekkehart, S., 2006).

The following paragraph is dedicated to point out the limitations of this analysis. First, the grouping of age was decided based on the trend in attrition rate and later assessed by the visualisation of observations with multidimensional scaling, which resulted in the 25-55 age group being quite large. As such, local pattern which the visualisation method fails to demonstrate may be ignored. As a result, a more popular age grouping, as in government's census data, can be implemented to address this issue. Secondly, concerning logistic regression, similar to other regression methods, this assumes no interactions between variables and group of variables. While such interactions have been addressed by correlation heatmap and feature pair plot, VIF as an evaluation metric could be implemented as a more formal approach. However, given that this dataset contains mostly categorical features, considering VIF does not necessarily provide significant change to the outcome. Thirdly, data veracity should not be overlooked: The measure of some features such as *RelationshipSatisfaction, JobSatisfaction, WorkLifeBalance,* etc. may vary by respondents' characteristics and is unclear that pre-normalisation of such features has been done by the data provider. Finally, the employees' average profiles presented in section 3.2 and 3.3 were presented using the average of each feature, which means that extreme outliers were not considered.

5.2. How well the analytical approaches enabled answers to research questions

Visual analytics approaches

The aim of this analysis was to characterise attrition by age. As a result, efforts have been dedicated to find a reasonable grouping. Stacked bar plot was utilised to verify that the variation in attrition rate was preserved while the visualisation of multidimensional scaled data confirmed that employees in the same age group were not separable. This preserved the change in pattern while minimising the number of groups to facilitate subsequent analysis. To contrast each age groups and to verify the applicability of empirical studies at the same time, radar plot was implemented, where the

average value of each of the available features that were considered to be related to attrition were visualised. Studying these plots has helped contrast the age groups which showed different pattern in attrition. Analysing the mean value allows discovery of the most common pattern in attrition by age groups. By combining the patterns shown in the radar plots and coefficient heatmaps, the most useful features in explaining attrition were quickly noticed. In addition, correlation heatmap and features pair plot, though do not directly answer the research questions, facilitated feature exploration and selection.

Computation methods

Logistic regression is one of a few interpretable machine learning algorithms, yet still provides decent result and does not rely on heavy assumption as much as its popular counterpart which is linear regression. To ensure the model's robustness and generalisability to new data, LASSO was selected as a regularisation method, which helped leave out insignificant features that do not necessarily explain attrition. The models were evaluated and selected using the area under the precision-recall curve which are considered suitable for class imbalance problems. Finally, data pre-processing steps such as feature selection, data normalisation, binary encoding of categorical variables boosted the models' performance while ensure that categorical variables are taken into account.

5.3. Applicability of approaches to other applications/domains

In general, this analysis seeks to identify factors that may contribute to an outcome using widely implemented analytical techniques. The analytical methods implemented do not require abundant computational resources and mostly utilised matplotlib, seaborn, and sklearn libraries, which are open-sourced. As such, the approaches should be applicable to a wide range of problems. However, the applicability of the implemented analytical approaches may be limited due to the following reasons.

First, the analysis clustered employees into 3 age groups, whose data were later analysed using 3 models. This approach may not be applicable to other domains where grouping may not be appropriate, for example geo-spatial or time-series data.

Secondly, the analytical methods may not be applicable for big data problems where the size of the data is one of the most immediate issue. Multidimensional scaling method computes the Euclidian distance between each data point, which would be extremely memory hungry, and its visual output would be hard to analyse given the huge number of points in the graph (Flexer, 1997). In addition, while radar plot proved to be useful in displaying less than 10 features, interpreting 30 to 40 features at the same time would be questionable.

Finally, the topic analysed is employee attrition, which has been extensively researched and knowledge can also be drawn from studies on employee turnover. The availability of research findings has helped refined the analytical process as well as detect abnormality in the results. However, for other topics such as gun violence, the availability of research findings may be scarce thus would not necessarily support the analytical approaches implemented in this analysis.

REFERENCES

1. Ahituv, A. and Lerman, R.I., 2005. Job Turnover, Wage Rates, and Marital Stability: How Are They Related?.
2. Armstrong, M. and Taylor, S., 2014. Armstrong's handbook of human resource management practice. Kogan Page Publishers.
3. Arthur, M., 1981. Turnover and the occupational career. *Industrial Relations Journal*, *12*(3), pp.28-39.
4. Bureau of Labor Statistics. (2017). *Employed persons by detailed occupation and age, including median age*. [online] Available at: https://www.bls.gov/opub/ted/2017/employment-population-ratio-and-labor-force-participation-rate-by-age.htm [Accessed 24 Dec. 2017].
5. Bureau of Labor Statistics. (2017). *Employment–population ratio and labor force participation rate by age*. [online] Available at: https://www.bls.gov/opub/ted/2017/employment-population-ratio-and-labor-force-participation-rate-by-age.htm [Accessed 24 Dec. 2017].
6. Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, pp.321-357.
7. Cohen, A. and Golan, R., 2007. Predicting absenteeism and turnover intentions by past absenteeism and work attitudes: An empirical examination of female employees in long term nursing care facilities. *Career Development International*, 12(5), pp.416-432.
8. Cotton, John & M. Tuttle, Jeffrey. (1986). Employee Turnover: A Meta-Analysis and Review With Implications for Research. *Academy of Management Review*. 11. 55-70. 10.5465/AMR.1986.4282625.
9. Davis, J. and Goadrich, M., 2006, June. The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning* (pp. 233-240). ACM.
10. Ekkehart, S., 2006. Labour Turnover, Wage Structure, and Natural Unemployment. *Journal of Institutional and Theoretical Economics*, 134(2), pp.337-364.
11. Hom, P.W. and Kinicki, A.J., 2001. Toward a greater understanding of how dissatisfaction drives employee turnover. *Academy of Management journal*, 44(5), pp.975-987.
12. Flexer, A. (1997). Limitations of self-organizing maps for vector quantization and multidimensional scaling. In *Advances in neural information processing systems* (pp. 445-451).
13. Ioffe, S. and Szegedy, C., 2015, June. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning (pp. 448-456).
14. Iqbal, A., 2010. Employee turnover: Causes, consequences and retention strategies in the Saudi organizations. *The Business Review*, Cambridge, 16(2), pp.275-281.
15. Kotsiantis, S.B., Zaharakis, I. and Pintelas, P., 2007. Supervised machine learning: A review of classification techniques.
16. Michaels, C.E. and Spector, P.E., 1982. Causes of employee turnover: A test of the Mobley, Griffeth, Hand, and Meglino model. *Journal of applied psychology*, 67(1), p.53.

17. Morrell, K.M., Loan-Clarke, J. and Wilkinson, A.J., 2004. Organisational change and employee turnover. *Personnel Review*, 33(2), pp.161-173.

18. Morrow, P.C., 1982. Human resource planning and the older worker: Developing a retirement intentions model. *Journal of Organizational Behavior*, *3*(3), pp.253-261.

19. Ongori, H., 2007. A review of the literature on employee turnover.

20. Oxford Economics (2017). The cost of Brain Drain: Understanding the financial impact of staff turnover. [online] Oxford: Oxford Economics, pp.8 - 27. Available at: https://www.oxfordeconomics.com/publication/open/246524 [Accessed 14 Dec. 2017].#

21. Palei, S.K. and Das, S.K., 2009. Logistic regression model for prediction of roof fall risks in bord and pillar workings in coal mines: An approach. Safety Science, 47(1), pp.88-96.

22. Parasuraman, S., 1982. Predicting turnover intentions and turnover behavior: A multivariate analysis. *Journal of Vocational Behavior*, *21*(1), pp.111-121.

23. Saito, T. and Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, *10*(3), p.e0118432.

24. Stark, D.E. and Shah, N.H., 2017. Funding and publication of research on gun violence and other leading causes of death. *Jama*, *317*(1), pp.84-85.

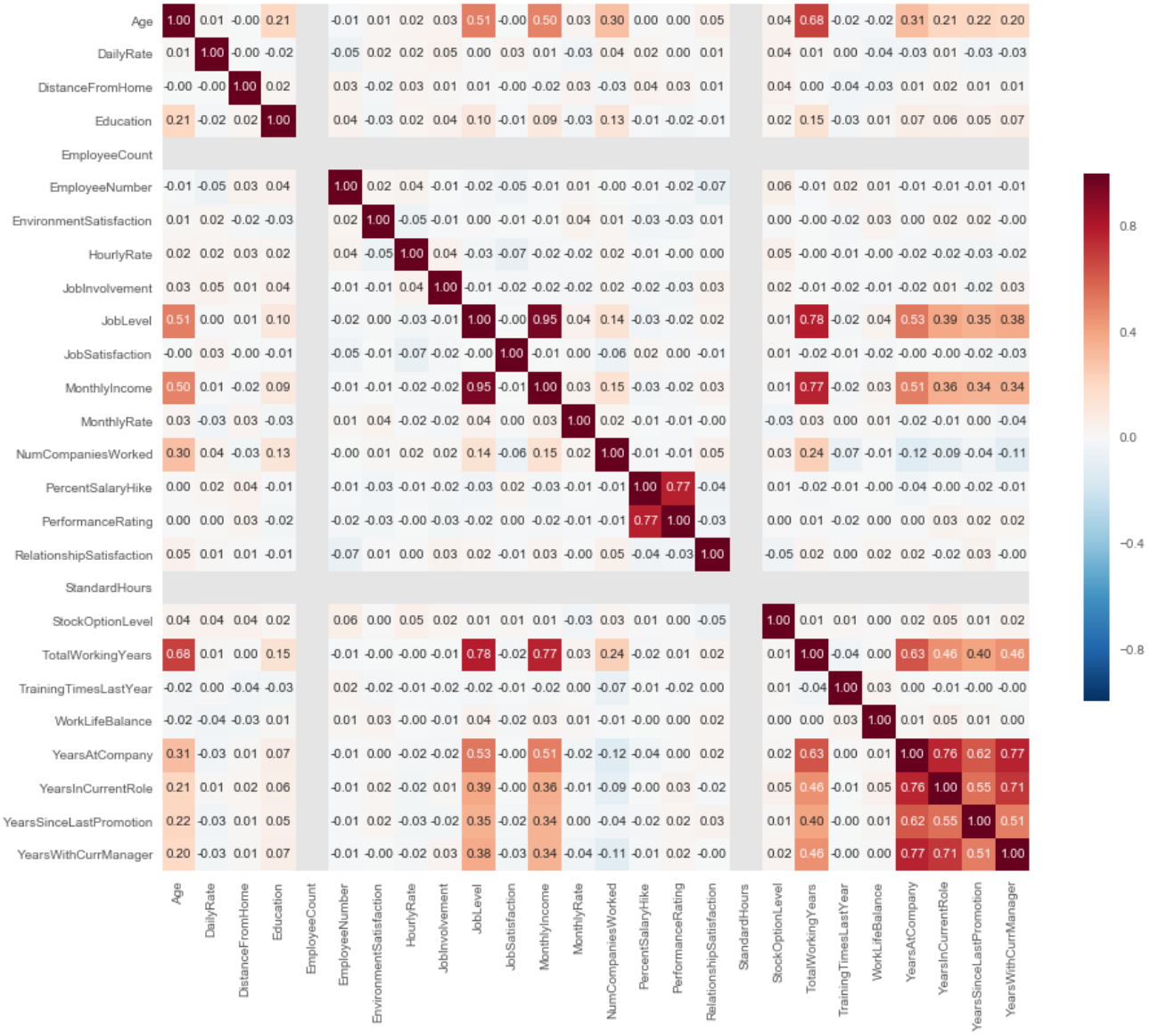
25. Ton, Z. and Huckman, R.S., 2008. Managing the impact of employee turnover on performance: The role of process conformance. *Organization Science*, 19(1), pp.56-68.

26. Watson Analytics Forum. (2017). HR Employee Attrition Sample Data Set Origin - Discussions. [online] Available at: https://community.watsonanalytics.com/discussions/questions/18014/hr-employee-attrition-sample-data-set-origin.html[Accessed 14 Dec. 2017]

APPENDICES

1. Appendix 1 – Correlation heatmap

Figure 5 - Correlation heatmap



2. *Appendix 2 – Pair plot*

Figure 6 - Pair plot