Towards Understanding Employee Attrition using a Decision Tree Approach

Saadat M Alhashmi
College of Computing and Informatics
University of Sharjah
Sharjah, UAE
salhashmi@sharjah.ac.ae

Abstract—**Employee attrition is a serious problem, and this has been a focus of research for the last few decades. Various approaches from exit interviews to psychological studies have addressed this issue. The idea is to avoid or minimise people leaving an organisation before an employer finds a replacement. With the abundance of data, lately, researchers from the Artificial Intelligence community have also addressed this issue. This research addressed the employee attrition issue using a decision tree approach to publically available data. The results are promising from this work in progress study, and future work-study will include more parameters and test the model on a local supermarket data**

Keywords— Employee Attrition, Decision Tree, Turnover Rate, Churn rate

I.INTRODUCTION

Employee turnover is a problem that affects all businesses, regardless of industry and size of the company. Employee attrition leads to high costs for businesses. Tangible costs include training expenses and the time it takes from when an employee starts to when they start contributing. Intangible costs involve what is lost when an efficient employee quits: new product ideas, excellent project management, or customer relationships. It is essential to understand the key variables that have an impact on turnover. When an existing employee leaves the organisation and a new employee hired in their place, this is known as employee attrition. The churn rate is the percentage of employees leaving the company over a specific period. Factors such as age, salary, pay and job satisfaction can affect employee churn. Voluntary and involuntary are two types of employee churn. Voluntary churn is when employees leave their organisation for personal reasons. These reasons divided into two types: positive and negative reasons. Positive reasons are when an employee leaves to have a better chance or opportunity in terms of salary, benefits and work environment. Negative reasons, on the other hand, are when an employee leaves the organisation due to lack of promotion and exciting work. The second type of employee churn is involuntary which represents those who are released from their services by the organisation [1].The bottom line is that businesses of all sizes are affected by high employee turnover rate. This paper using the data mining technique shows that it is possible to predict what are the chances of someone leaving an organisation. Such a prediction would help senior management to take pre-emptive measures, either by exploring ways to retain their existing staff, or hiring and training in advance. The structure of this paper is as follows: Section 2 talks about the background and significance of employee retention problem; Section 3 talks about the data used and how it can help organisations to learn and address employee attrition issues based on the data. Section 4 talks about results and Section 5 conclude with conclusion and discussion and future work.

II. LITERATURE REVIEW

Employee attrition is a problem that has been a focus of research for the last few decades. Much work mainly in the field of psychology, has tried to address this issue [2], [3]. Some industries like call centres tend to have a high attrition issue compared to others, but in general, this affects all industries [4]. Besides costs, it affects employers in other ways such as the loss of experience and knowledge that employees take away with them. According to Mobley [3], there is a strong relationship between job satisfaction and turnover rate. However, they are also of the view that the employee withdrawal decision process also has a few intermediate linkages. While the literature review reveals that age, overall satisfaction and commitment are consistently and negatively related to attrition, less than 20 % of the variance is explained [3]. According to Jain [5], the common causes of attrition are Poor training, Poor management, Lack of growth and advancement opportunities, Inaccurate job profiles, The feeling of not being appreciated.

Every organisation has attrition of employees. People either resign or retire. This issue can have severe implications for the viability of an organisation if this does not happen in an orderly fashion and if employees leave unexpectedly. Frierson and Si used machine learning technique to show which employee(s) are high-risk to leave and also if the department they work in increases the probability of attrition. Another research by Shankar et al. [6] used various classification methods, for example, Decision tree, Logistic Regression, SVM, KNN, Random Forest, Naive Bayes methods on human resource data to help companies predict employee attrition.

Shahnawaz and Jafri [7] collected data from 80 managers from Information Technology Enabled Services employees and classified it into two broad categories of stayers and leavers. The focus was to identify how job satisfaction and organisational commitment predict employee turnover intentions for stayers and leavers. Findings were interesting as it merely showed that job attitudes were highly related to intention to leave the organisation in both categories of employees. Overall, the

978-1-7281-3841-1/19/$31.00 ©2019 IEEE

attrition issue is crucial. While there are many reasons for this, but the three reasons mentioned below are significant.

– *Cost implications*: Employee resignation causes the company loss of productivity. There are other economic costs related to employee attrition, such as the company must pay the workers who are handling the leaving employee's work until the organisation hires new ones. Companies need to spend money on job advertising, interviewing potential substitutes, in addition to the fees related to the actual recruiting and hiring an employee [8].

– *Overall business performance*: Employees in companies with high turnover rates are less productive and much less efficient than they might have been in a lower turnover environment. This also makes a company with low retention rate less competitive and productive as they do not have a stable workforce. According to Huselid [9], high-performance work practices and firm performance have an economically and statistically significant impact on both employee turnover and short and long-term measures of corporate financial performance.

– *Challenging to control company environment:* Research consistently displays that employees change jobs more often because of the work environment or inter-employee relations rather than because of the difficulty of the job. This is something that organisations have little or no control over as they cannot interfere with the employees' relationships or feelings. Shalley et al. [10]conducted an extensive study and found that there is a direct link between job satisfaction and fewer instances of intentions to leave.

III. METHODS

In this project, we aim to analyse the factors that lead to employee turnover in companies and to identify the factors that have the most influence over employee churn. We used IBM's HR employee attrition dataset [11] to make a model that would help us in predicting employee attrition. The dataset contains information about employees such as age, gender, nature of work, position, education, salary. Overall, there are 34 features recorded in the dataset. However, for our analysis, not all features were essential or useful. For example, standard weekly hours were not suitable for us because all records had the same values, so these features discarded from the analysis. After we select the features that we want to keep in the data set, the data selection and cleansing step will be finished.

A. The Relationships between Data

As explained above, we only selected a subset of the attributes to use in our analysis. The attributes selected are some common factors that can result in employee turnover. We used the RapidMiner program to create the correlation matrix. The attributes selected lists in the correlation matrix below.

The correlation matrix gives us insights about the relationship between attributes, by analysing the matrix, we can see that some attributes are not related to other attributes meaning they are independent of other features. Such attributes are environment satisfaction Job satisfaction over time, and work-life balance. While other attributes are highly associated with other attributes, for example, the job level is highly correlated with monthly income and years in the current role is correlated with years at the company. Some attributes have relationships with other attributes, but the association between them is not very strong. For example, age is correlated with job level and monthly income, while years since last promotion is correlated with years at company and years in the current role.

TABLE I Correlation Matrix

Attributes	Age	EnvironmentSatisfaction	JobLevel	JobSatisfaction	MonthlyIncome	OverTime	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion
Age	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
EnvironmentSatisfaction	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
JobLevel	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
JobSatisfaction	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MonthlyIncome	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
OverTime	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]
WorkLifeBalance	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]
YearsAtCompany	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
YearsInCurrentRole	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]
YearsSinceLastPromot...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1

B. Detection Trends

In the process that we created, we wanted to analyse the factors that impact employee churn, and we also wanted to create a model that will help us in predicting employee attrition by analysing previous employee records (our dataset).

The model that we used in our process is the Decision Tree. Decision Trees are one of the most popular predictive models because they help people understand the possible results based on a series of related choices. A decision tree shows us the outcome of each choice we make. When Decision Trees are used to build a predictive model, it is called decision tree learning because it takes the features (characteristics) in the data you insert and learns the relationships between them then tries to predict its outcome (value). For example, in our case, the decision tree model will take into account the features that we selected and the relationships between them to determine paths for which whether employees will leave or not [12]

Once we import the dataset, it contains all the features and attributes of the employees, including the attrition attribute. We set the role of the attrition attribute as a label, meaning that it is the target attribute whose value we want to predict based on the values of other attributes. Then, the second operator used is the select attributes, this operator allows to select the attributes that include in our assessment, because as we explained earlier. There are more than 30 attributes in the dataset, and we selected ten attributes(shown in the correlation matrix, Table 1) that we believe they would have a significant impact on attrition. The third step was to create a correlation matrix; the correlation matrix as we explained above shows the relationships or associations between each attribute. Then, we used the select by weight operator; this operator weights the impact of each attribute on employee attrition. Finally, we use the decision tree operator to create our model; in our process, the select by weight operator will be going into the decision tree. We specify the weights of the attributes that will be going into the decision tree because we only want attributes with higher weights to be in our decision tree. The decision tree is helpful because when we understand the behaviours behind attrition, we can minimise its risk of occurring and maximise the likelihood of an employee staying within the company [13]

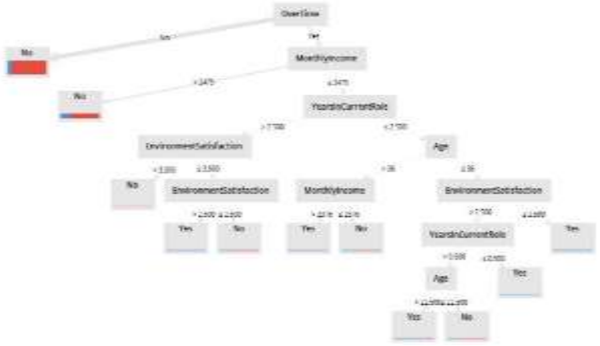

Fig. 1. Decision Tree showing the distribution of occurrences with the items in the dataset

The Decision Tree above shows different paths that lead to different decisions. The tree starts with the overtime attribute. If an employee works do NOT work overtime, then it is most likely that he/she will NOT leave the company. The distribution is 944 out of the 1054 employees who do not work overtime will not leave the company. While only 110 out of 1054 employees who work overtime will leave the company. This means that the overall average is NO Attrition as shown in the decision tree and the distribution table above, when we consider the overall average as no, then it is 71.70% of the total employees will not leave if they do not work overtime (1054/1470). Then if the employee works overtime (YES), then we have more branches, so we look at his/her income. If the monthly income is GREATER than 2500, then it is most likely that the employee will Stay at the company (NO attrition). The distribution is as follow; 268 employees out of 347 employees who work overtime and have income more than 2500 will NOT leave, while 79 employees out of 347 employees who work overtime and have income more than 2500 will leave (YES). So, the overall result is still NO attrition.This is an explanation of the first two scenarios, but as the decision tree gets more branches, it becomes easier to explain it with paths as follows:

Path 3: Overtime=Yes, monthly income < 2500, years in current role>2.5 years, environment satisfaction= very high.

In this case, only 5 employees conform to these attributes, and they did NOT leave the company. Although the negative attributes such as the fact that they work overtime, their income is low, and they did not get any promotions in the last 2 years, their environment satisfaction is very high and that is maybe the reason they did not leave the company.

Path 4: Overtime=Yes, monthly income < 2500, years in current role>2.5 years, en-vironment satisfaction= medium.

In this case, only 4 employees conform to these attributes, and they left the company (YES attrition). This makes sense because the factors that impact the employee's decisions are all negative, since they work overtime, have a low income, did not get promoted, and they do not like the environment very much.

Path 5: Overtime=Yes, monthly income < 2500, years in current role<2.5 years, Age>36, monthly income>2300.

In this case, only 4 employees conform to these attributes, and they left the company (YES attrition.

Path 6: Overtime=Yes, monthly income < 2500, years in current role<2.5 years, Age<36, Environment satisfaction=low-medium.

In this case, 13 employees conform to these attributes and all of them have left the company (YES attrition). This is expected as they are working overtime and not getting paid enough, they are not satisfied with the environment, and they are considerably young, so they may be motivated for new and better opportunities.

Path 7: Overtime=Yes, monthly income < 2500, years in current role<2.5 years, Age<36, Environment satisfaction=medium-high, years in current role> half a year, age>22.5

In this case, 9 out of 11 employees who conform with these attributes have left (YES attrition), while only 2 did not leave.

Path 8: Overtime=Yes, monthly income < 2500, years in current role<2.5 years, Age<36, Environment satisfaction=medium-high, years in current role> half a year, age<22.5

In this case, there are only 4 employees who have these characteristics, 3 of them did NOT leave the company (NO ATTRITION), while only one left. It is expected that they are students because their age is below 22, and that is why most of them did not leave because they do not have much power and they need a job to help them in covering their expenses.

IV.RESULTS

The results showed that the number of people who left the company is 237. However, the number of people who stayed in the company is 1233. So, we can see that people who stay in the company usually are more than people who leave. Another aspect we looked at is a relationship between overtime and employee attrition. 125 of the employee who works overtime don not stay in the company. However, around 950 of the employees who do not work overtime do not leave the company. So, we can say that people who work overtime leave more frequently than people who do not.The results showed how monthly income related to employee attrition. The most common group of people who leave the company are in the low-income bracket between $1000 and $6000. Moreover, most people with high income between $15000 and $21000 stay on in the company. This is reasonable as people who have low income will not be satisfied with their job so they will leave for better jobs, while people who get high income are satisfied with their job, so they do not leave. However, a few numbers of people in the high-income bracket left the company, and these are exceptions. Generally, we can understand that people with low-income leave the company more than people with high income [14]. The results showed how employee age relates to attrition. Most people who leave the company are between the ages of 27 and 35; this can be explained by the fact that young people are actively looking for more experience and better opportunities. On the other hand, with age, employees grow more attached to the firm and therefore prefer to remain in the company, rather than look for new opportunities.

V. DISCUSSION AND CONCLUSIONS

The idea was to contribute in the field of employee attrition; Overall results were promising. The model developed in this project helped us to understand the problem better as information alone is not enough. Once we know, we can suggest solutions to the problem. For example, if most employees that leave the company have low to medium environment satisfaction, then the company must change its work environment and make it friendlier [15].

Finally, we can see how employee attrition is a severe problem that affects all organisations no matter what their size is, so it is imperative for employers to understand the cause behind attrition and the factors that impact employees' decisions. In order to make better-informed decisions, corporations can no longer rely on management experience; they must have hard evidence information based on valid data. The information can be generated by different tools that help in decision making such as we did in this project. This is a work in progress, and we are expecting to test the hypothesis on a local supermarket with a high churn rate the text edit has been completed.

REFERENCES

[1] A. Frye, C. Boomhower, M. Smith, L. Vitovsky, and S. Fabricant, "Employee Attrition: What Makes an Employee Quit?," SMU Data Sci. Rev., vol. 1, no. 1, p. 9, 2018.

[2] W. H. Mobley, R. W. Griffeth, H. H. Hand, and B. M. Meglino, "Review and conceptual analysis of the employee turnover process.," Psychol. Bull., vol. 86, no. 3, p. 493, 1979.

[3] W. H. Mobley, "Intermediate linkages in the relationship between job satisfaction and employee turnover.," J. Appl. Psychol., vol. 62, no. 2, p. 237, 1977.

[4] P. S. Budhwar, A. Varma, N. Malhotra, and A. Mukherjee, "Insights into the Indian call centre industry: can internal marketing help tackle high employee turnover?," J. Serv. Mark., vol. 23, no. 5, pp. 351–362, 2009.

[5] M. Jain, "Employee attrition-causes and remedies," J. Soc. Welf. Manag., vol. 5, no. 2, p. 69, 2013.

[6] R. S. Shankar, J. Rajanikanth, V. V Sivaramaraju, and K. V. Murthy, "Prediction of Employee Attrition Using Datamining," in 2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA), 2018, pp. 1–8.

[7] M. G. Shahnawaz and M. H. Jafri, "Job attitudes as predictor of employee turnover among stayers and leavers/hoppers," J. Manag. Res., vol. 9, no. 3, p. 159, 2009.

[8] J. Altman, "How Much Does Employee Turnover Really Cost," Huffpost.[Google Sch., 2017.

[9] M. A. Huselid, "The impact of human resource management practices on turnover, productivity, and corporate financial performance," Acad. Manag. J., vol. 38, no. 3, pp. 635–672, 1995.

[10] C. E. Shalley, L. L. Gilson, and T. C. Blum, "Matching creativity requirements and the work environment: Effects on satisfaction and intentions to leave," Acad. Manag. J., vol. 43, no. 2, pp. 215–223, 2000.

[11] McKinley Stacker IV, "SAMPLE DATA: HR Employee Attrition and Performance – IBM Analytics Communities,"2015. [Online].Available:https://www.ibm.com/communities/analytics/watson-analytics-blog/hr-employee-attrition/. [Accessed: 06-Jan-2019].

[12] S. B. Kotsiantis, "Decision trees: a recent overview," Artif. Intell. Rev., vol. 39, no. 4, pp. 261–283, 2013.

[13] D. Wilson, "Predicting Employee Churn with Data Mining - CDO Advisors,"2017.[Online].Available: https://www.cdoadvisors.com/predicting-employee-churn-data-mining/. [Accessed: 06-Jan-2019].

[14] J. Costa, "Employee Retention Ideas: How Important is Salary?," 2013.[Online].Available: https://www.furstperson.com/blog/employee-retention-ideas-it-takes-more-than-salary. [Accessed: 06-Jan-2019].

[15] D. K. Srivastava and P. Nair, "Employee Attrition Analysis Using Predictive Techniques," in International Conference on Information and Communication Technology for Intelligent Systems, 2017, pp. 293–300.