HUMAN RESOURCE ANALYTICS

UNDERSTANDING THE EMPLOYEE DATA FOR HIGH ATTRITION

Name	Mail Id
Bharat Gurunathan Haridoss	bgurunathanharidoss01@qub.ac.uk

Queen's University group assignment

Table of Contents

1.0 Introduction 4
2.0 Methodology 5
3.0 Result & Discussion 11
3.1 Tableau Analyses 11
3.2 Tree Structure 14
4.0 Conclusion 17
5.0 References 18
Appendix 1 22
Appendix 2 22
Appendix 3 23
Appendix 4 23
Appendix 5 24
Appendix 6 24
Appendix 7 25

Table of Figures

Figure 1: Eliminating Age Outliers 5
Figure 2: Eliminating Total Working Years Outliers 5
Figure 3: Cleansing the data 6
Figure 4: Column Rosorter 7
Figure 5: Partitioning 8
Figure 6: Decision Tree Learner **Error! Bookmark not defined.**
Figure 7: Tableau Dashboard View 9
Figure 8: Tableau Visualization Attrition Vs Total Working Years & Monthly Income 10
Figure 9: Tableau Visualization Attrition Vs Percent Salary Hike & Performance Rating & Work Life Balance 11
Figure 10: Tableau Visualization Attrition Vs Job Role & Job Satisfaction 12
Figure 11: Tableau Visualization Figure 11: Attrition Vs Marital Status & Monthly Income 12
Figure 12: Scorer output of Decision Tree 13
Figure 13: Scorer output of Gradient booster 13
Figure 14: Scorer output of Random Forest Tree 13
Figure 15: Confusion Matrix of Decision tree, Random Forest & Gradient Boosting 14

1.0 Introduction

Data is known to be the new oil. As more data is collected and stored across industries, society and business practices are changing (Porter and Heppelmann 2015). A company incurs higher recruiting, hiring, and training expenditures. Employee attrition is the proportion of employees that leave a firm and are replaced by new hires (Apgar IV 2002). Employee Attrition can be voluntary when the person leaves on their initiative or involuntary when the employer makes the decision (Lazzari et al., 2022). Personal reasons, lack of job satisfaction, reduced pay or uncomfortable working environment, and high remuneration from other organisations lead to voluntary attrition (Krishna & Sidharth, 2022). When an employee leaves an organization, it takes a considerable amount of time to fill the gap and replacing them costs the company an average of 33% of their annual income. It costs time and money to hire a new employee (Darapaneni et al., 2022). Since employee attrition results in the loss of business prospects in addition to the loss of skills, experiences, and personnel (Raman et al., 2018). High-performing employees are the investment of organisations (P et al., 2022).

In addition, statistics after the pandemic outbreak showed a 14.7% increase in the unemployment rate in march 2020, when 21.8 million individuals were unemployed by that time, according to the Bureau of Labour Statistics (Seelam et al., 2022). So, organisations adopted analytical methods to prevent and predict employee turnover to retain existing employees and attract new talents in organisations (Yahia et al., 2021). IBM dataset used a correlation matrix to eliminate extraneous features. Moreover, variables including monthly income, the number of employers and age had a significant impact on employee attrition (Yang & Islam, 2021). The accuracy of the Random Forest model, which used 80% of the samples for training and 20% of the samples for testing, was 0.8456 and they also found that the attrition rate was 2.4 times higher for frequent travellers than for infrequent ones. (Yadav et al., 2018) in his study mentioned that the Random forest produced 98% accuracy and outperformed all ML models in terms of performance.

2.0 Methodology

All the data mining projects undergo the CRISP-DM model to prepare for both supervised and unsupervised learning. This model leads way to more successful and well-defined outcomes of machine learning (Purbasari, et al., 2021). CRISP-DM model comprises six elements of the procedural cycle and the following steps expose how this assignment is spread out into different steps involved in CRISP-DM:

The **business understanding** signifies that the business is witnessing a higher turnover rate which has compelled the board to take a deeper look into analysing the reason for such high employee attrition. The business has to find reasons and work accordingly to avoid a high amount of employee turnover. High turnover would mean extra expense to the business as they spend a lot on employee recruitment, training and retaining (Ongori, 2007). Therefore, to diminish the cost the business is looking for suitable modelling.

Data understanding is gathering data from different sources, and examining and describing it. After the data is collected, data quality issues are identified (Schröer, et al., 2021). In this case, the data is given in an excel sheet by the name of 'Employee.xlsx'. This sheet consists of 35 variables which display sensitive information and employee data of 1467 employees in total. Data is classified into nominal, ordinal, discrete and continuous data.

Data preparation is a very crucial step which includes removing or amending data quality issues. Data quality issues often lay disturbing effects on the final output so these data quality issues must be fixed primarily (Osborne & Overbay, 2004). The data quality issues in our data set were founded in 'Age' and 'Total working years'. According to UNICEF, the age of a person to work is 12-16 years (UNICEF, n.d.). The minimum age of 4 was an outlier which was removed by using Knime and the appended column was named 'Clean_Age' as shown in Figure 1. Another outlier identified was in total working years exceeding employee age, there were two outliers with total working years of 74 and 83, and the ages defined against these observations were 58, as portrayed in Figure 2. After removing these outliers the clean total working years were saved in the appended column as 'Clean TWY'. Moreover, when the clean data was imported into Tableau, the DOB category was changed from alphabetical form to time and date format.

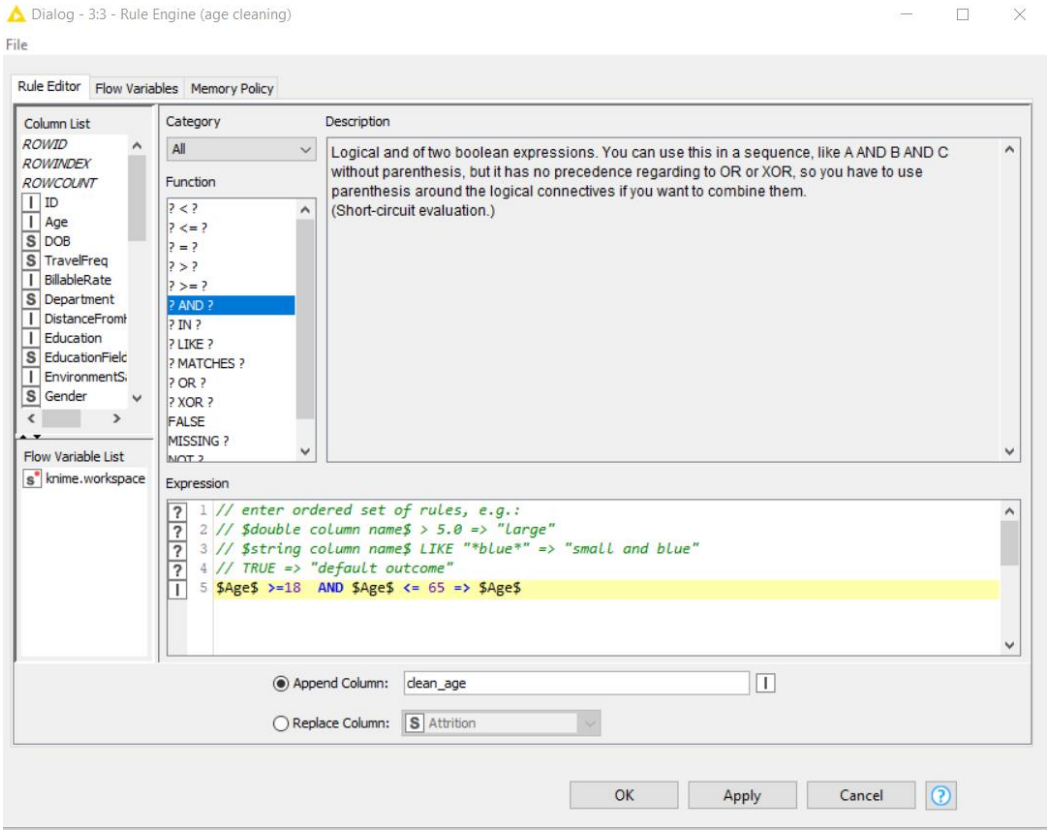


Figure 1: Eliminating Age Outliers

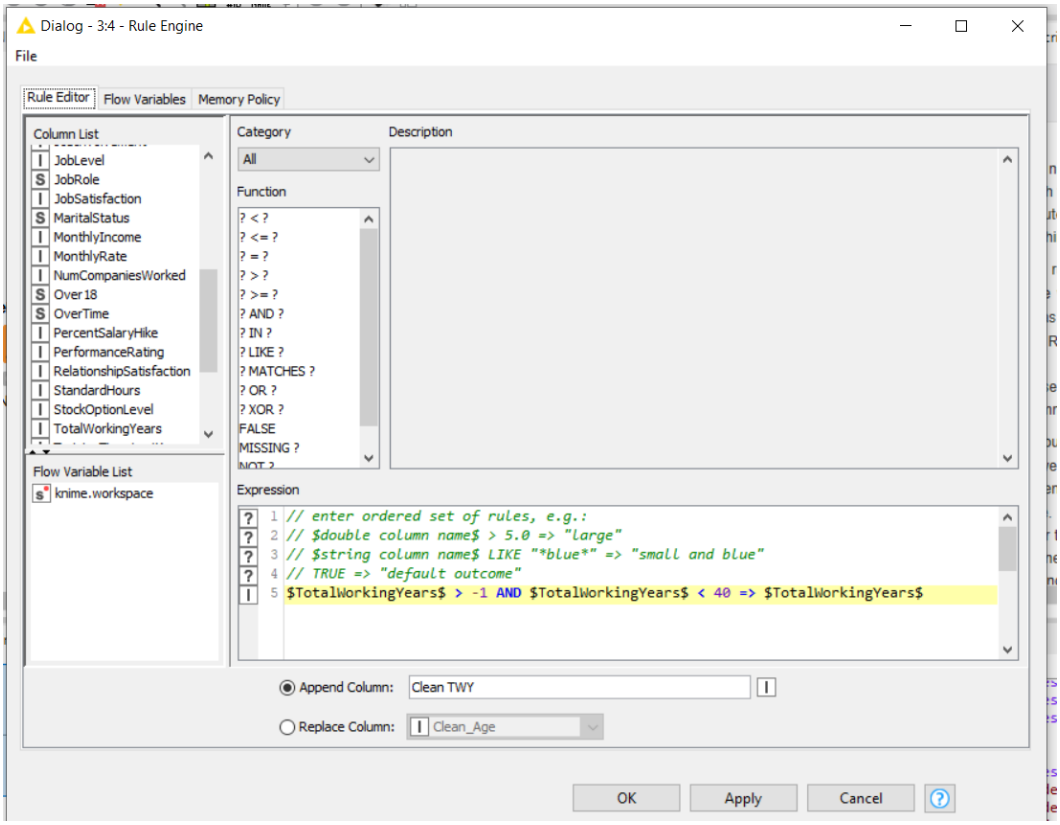


Figure 2: Eliminating Total Working Years Outliers

The Excel file is imported into KNIME using an Excel Reader node, and then the Statistics node is placed by dragging and dropping (box with built-in processing action). Right-click any node to see the results. There are no missing values in the dataset, which is great news. Variables are also shown in the Statistics view. Create a Column Filter node to exclude the variables. Class imbalance affects a learning algorithm during training by biassing decision rules towards the majority class and maximising predictions based on the predominant class in the dataset when training models on such datasets(Nguyen, Bouzerdoum et al. 2009).

Creating a new workflow in KNIME, the data set can be imported using the excel reader node. The data can be checked for accuracy by connecting the excel reader node to the table display node. This allowed us to verify that the column headers were correctly imported by visually inspecting the data. Before the raw data was cleaned, the bar chart node was linked, providing visualisations of any relevant elements that would be useful in the prediction stage.

The statistics node is linked to the excel reader node to visually indicate any problems or data that may be an exception. This includes a histogram (Appendix 1) showing the distribution of each variable and summary statistics for all variables in the data. The statistical node makes it easy to see how numbers and names are related. Many challenges emerged when they were examined. Inaccurate data in the age range had to be fixed. To identify outliers using a scatter plot (Appendix 3), a column connector is used to compare the cleaned age column and the total working hours column.

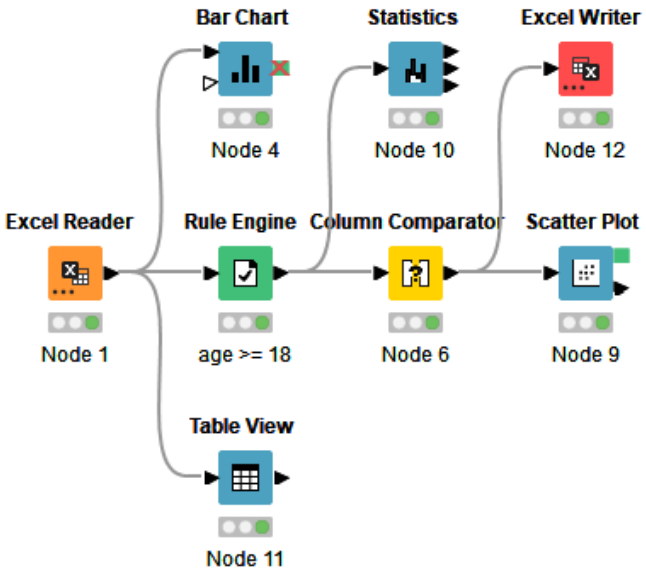


Figure 3: Cleansing the data

Modelling is a phase in which specific techniques are used to build the test case and model. Data mining techniques are used according to the data and business problem. Nevertheless, it is crucial to evaluate the model against the criteria that have been set and opt for the best model (Bueno, et al., 2022). In this data, we have used the Gradient booster, the random Forest tree, and the decision tree. Start by creating training and test sets from the dataset using an 80/20 split (Lever, Krzywinski et al. 2016); the model will be trained using 80% of the data, and its performance will be evaluated using the remaining 20%. A new node called

Partitioning is used. After splitting, our data is now ready to be used as input for machine learning models. Among the models we will train are decision trees, random forests, and gradient boosting. These models may be trained by dragging and dropping the learner nodes, which can then be connected to and customised with the predictor nodes. To improve model performance, begin by modifying model parameters, doing feature engineering, and balancing data approaches. we concluded that the best model was the Random Forest (RF).

A decision tree is a three-part structure with roots, branches, and leaf nodes. The result of a test is represented by a branch, each internal node represents a test on an attribute, and a class label is demonstrated by the Leaf node. The root node is the topmost node in the tree (Naik, A. and Samant, L., 2016).

The decision tree technique considers each conceivable variable to determine which approach provides the most accurate results when dividing data into two categories. This continues to be done until a predetermined goal is accomplished. After that, Column Resorter is used to move the Attrition dependent variable to the most recent data column.

.

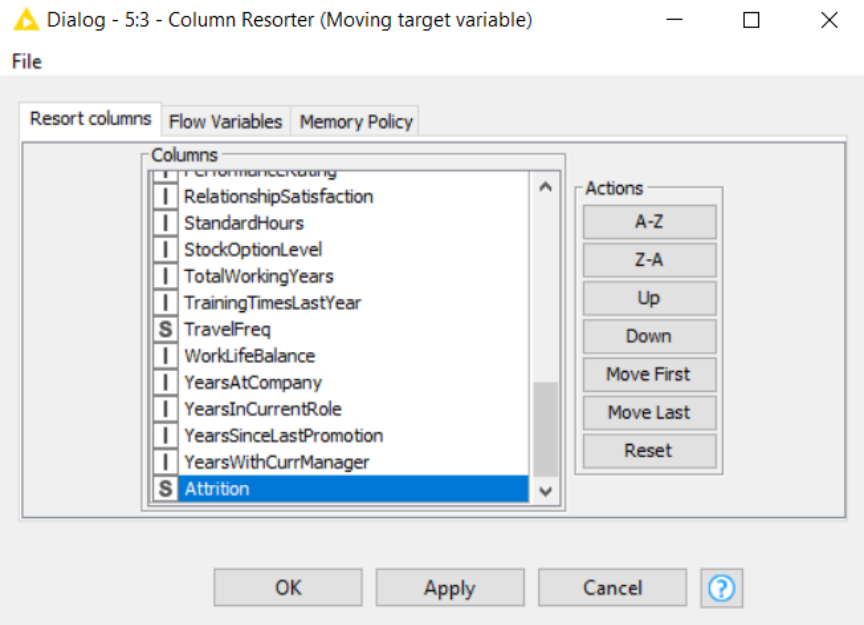
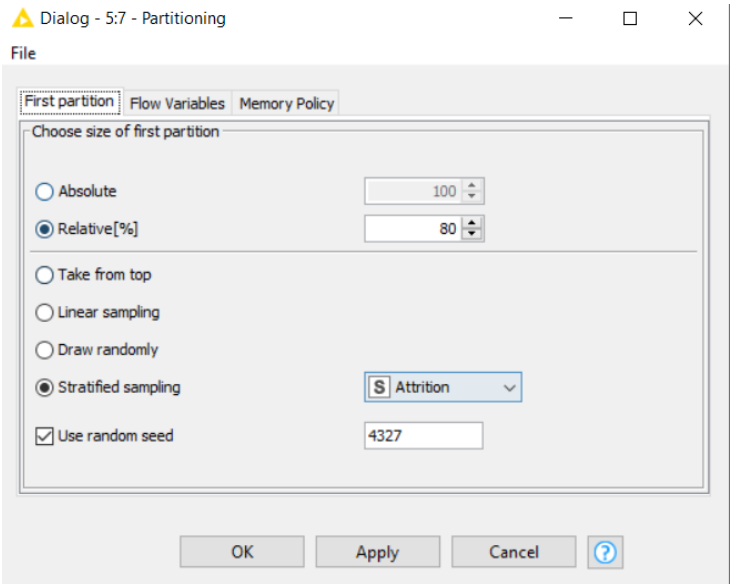


Figure 4: Column Resorter

For the training models to be effective, they had to be designed with 80% of the data considered. After that, the model's performance was evaluated by utilising the last 20% of the available data, which was performed by splitting the node.



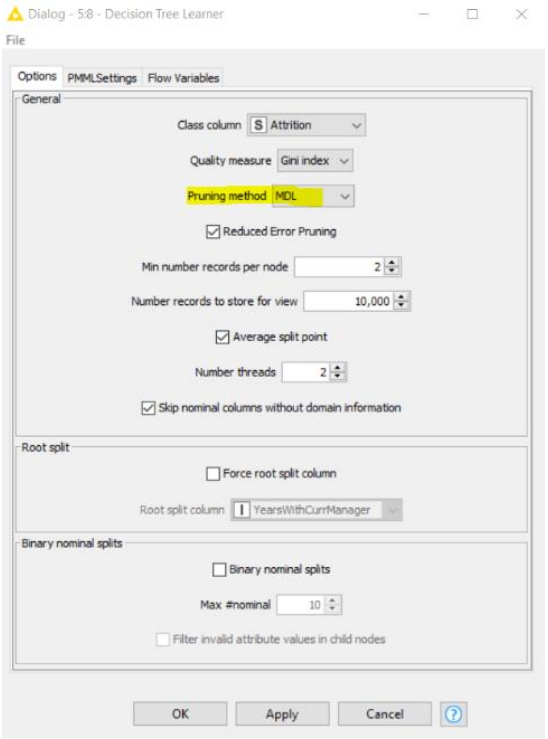


Figure 16: Decision Tree Learner

Figure 5: Partitioning

To prevent overfitting in general, the minimum description length (MDL) concept appears to be a natural, non-ad hoc method. A subtree should be trimmed if the description length of the classification of the training model is too extensive, according to the MDL-based method this is known as pruning (Kononenko, I., 1998). After that, the decision tree learner node (APPENDIX4) was inserted and determined to prune using the MDL pruning approach.

The decision tree predictor node is utilised as the next node to provide a more precise forecast. The decision tree illustrates the sequence of splits that provide the cleanest leaf nodes; the remaining 20% of the splits must be examined to assess the performance of the tree. To do this, the partitioning node and the decision tree learner are connected as the decision tree predictor. The predictions produced for each employee are displayed at the decision tree predictor node (Appendix 5).

Evaluation and Deployment: The models created by using gradient boosting, forest tree and decision tree are then evaluated to see which models best fit the data (detailed explanation in the Results section). And in the last deployment phase, a final report is generated which includes the creation of deployment, monitoring and maintenance strategy (Alogogianni & Virvou, 2021). The data can be used by the stakeholders easily through the dashboard available in Tableau which shows visualizations. These may show explicit relationships and variations in the variables or factors that need to be improved for a lower level of employee turnover as shown in the figure below:

Figure 7: Tableau Dashboard View

3.0 Result & Discussion

3.1 Tableau Analyses

Visualisations using Tableau are one of the most attractive forums for all the stakeholders to extract the maximum information regarding the high rate of employee attrition. Following are some visualisations that aid to draw insights and reasons for employee turnover, leading to appropriate measures to solve the business problem:

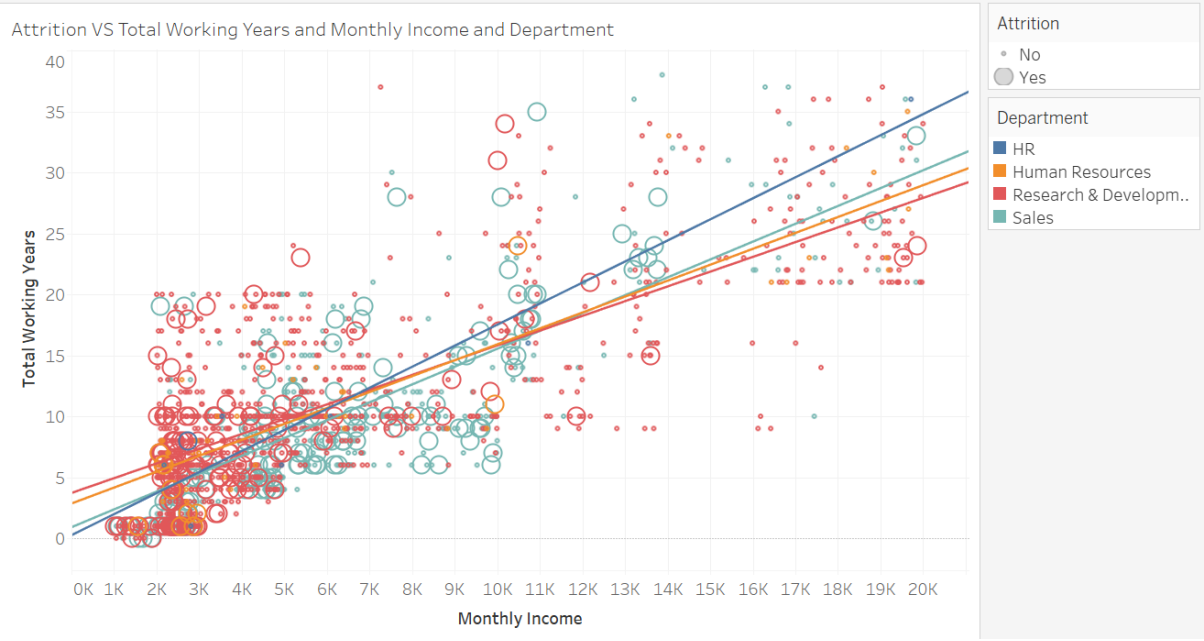


Figure 8: Tableau Visualization Attrition Vs Total Working Years & Monthly Income & Department

Figure 8 depicts a positive linear relationship between total working years and monthly income. It also demonstrates that the employee that has more work experience and higher monthly income is more likely to continue staying in the job. This is supported by (Fallucchi et al., 2020) as they state that employees with fewer years of experience are more likely to leave the company, both in absolute terms and as a percentage within their category. They further state that resignations reduce gradually as salary increases. Further, the graph also proves that the attrition rate for the sales department is highest as the total working years and the income is increasing. Also, as the total working years and monthly income increase, the attrition rate lessens.

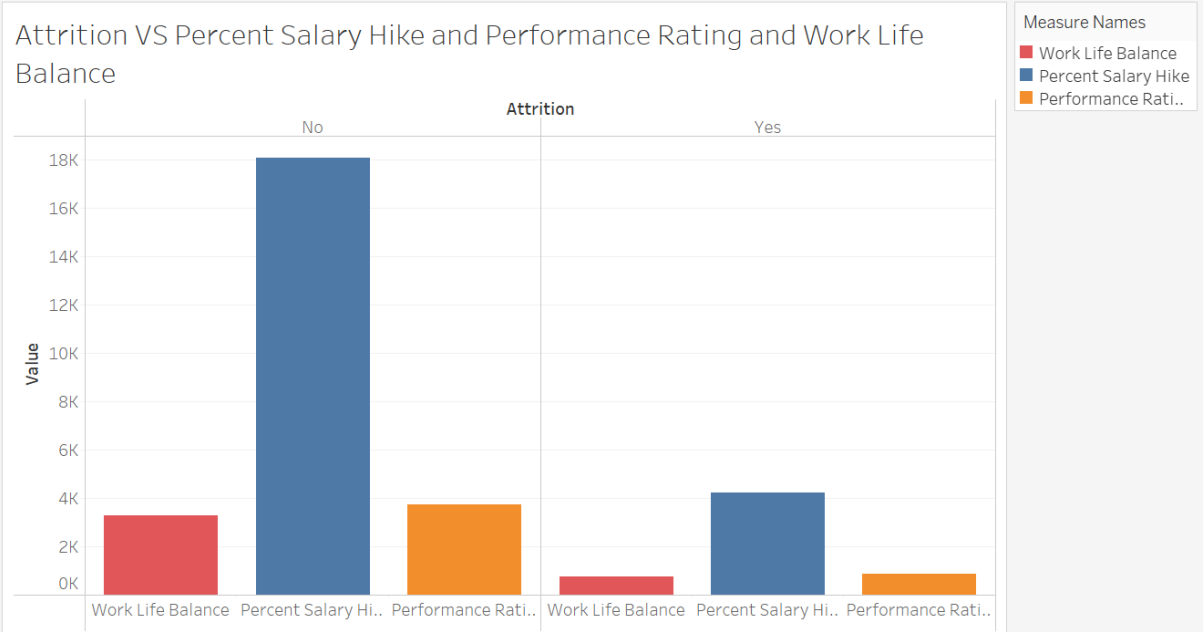


Figure 9: Tableau Visualization Attrition Vs Percent Salary Hike & Performance Rating & Work-Life Balance

Figure 9 shows that the employees who received higher salary hikes were more likely to stay than the employees who received lower salary hikes. This is supported by (Jain and Nayyar, 2018) who state that companies can freeze pay increases and promotions, and employees can experience a lack of professional and financial advancement, forcing them to seek new opportunities and increasing turnover. Similarly, the employees whose performance rating was good were more likely to stay than the employees who received a lower performance rating. According to (Zimmerman and Darnold, 2009) the supervisor's assessment of work performance has a far greater influence on an employee's desire to quit than objective performance assessments, but only slightly more than self-assessment of performance. The graph shows that the higher the performance rating, the more likely the employee will stay in the organisation. According to (Jaharuddin & Zainol, 2019) there is a connection between work-life balance and the desire to quit. They further state that the companies that offer work-life balance inculcate engaged employees that have improved levels of productivity, profitability, growth, customer satisfaction, and employee retention due to lower employee turnover and reduced intention to leave the company. This can be seen in the graph; as the work-life balance is higher, there are lesser chances of employee turnover.

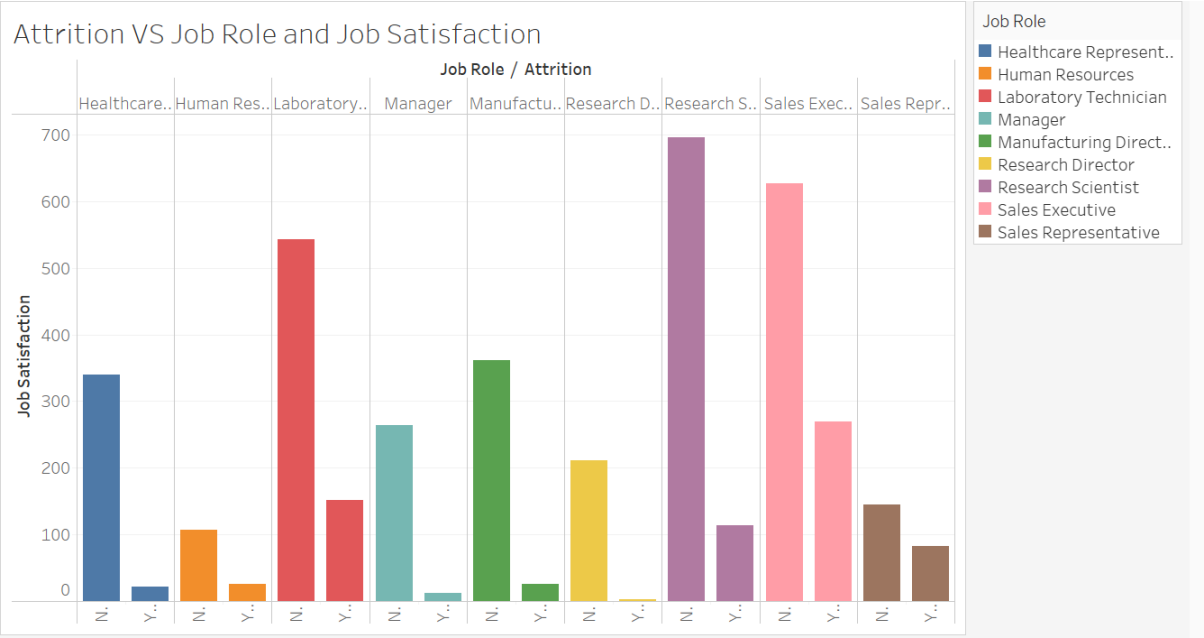


Figure 10: Tableau Visualization Attrition Vs Job Role & Job Satisfaction

Figure 10 shows that the higher the job satisfaction for a job role, the less likely the employee is willing to leave the company. This is supported by (Medina, 2012) who states that job satisfaction and propensity to quit are negatively correlated, and research has shown that minimal turnover improves Organisational productivity and performance. Likewise, in the decision tree, it was revealed that workers with the job title Sales Executive, with over time are more likely to leave (Appendix 5).

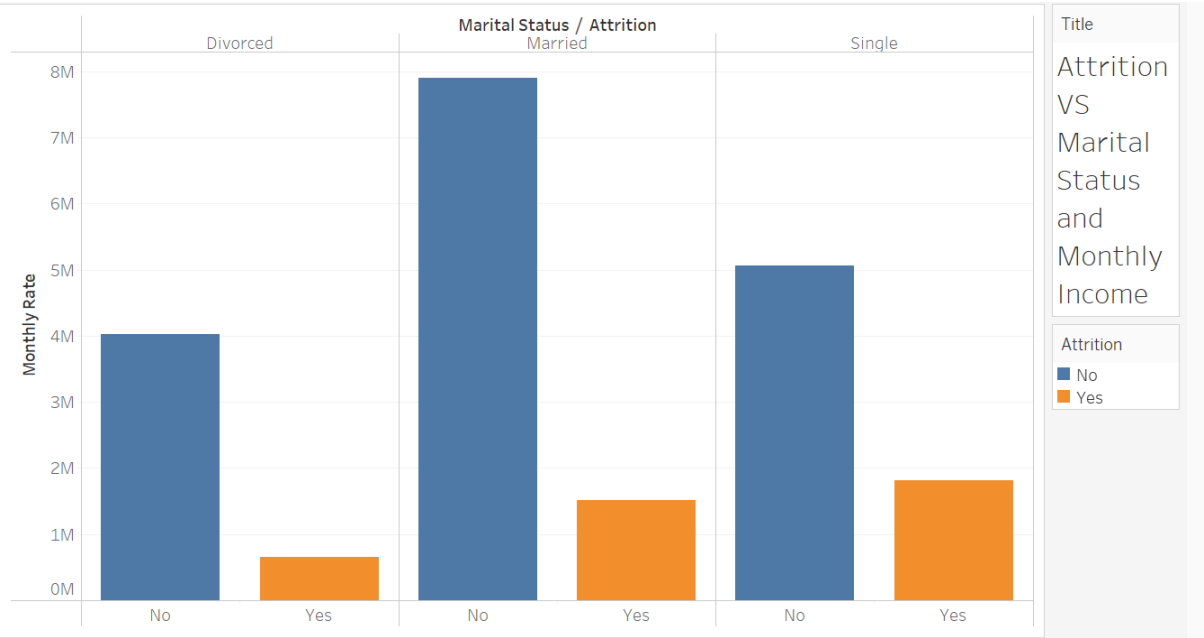


Figure 11: Tableau Visualization Figure 11: Attrition Vs Marital Status & Monthly Income

Figure 11 illustrates that people who are married and have high incomes are more likely to stay in organisations. Similar is the case of divorced and single employees. This is supported by (Jain and Nayyar, 2018)as they claim that employee churn was highest among unmarried employees and lowest among divorced employees.

3.2 Decision Tree Structure

Apart from the tableau, using knime and obtaining some insights are explained as follows using the prediction's accuracy, forecasting is evaluated. This is accomplished by including a scorer node that provides the accuracy as a percentage and a confusion matrix.

The gradient-boosting algorithm creates a series of decision trees from a series of training trees. It may be used for both categorising and forecasting future events. At each level, a new decision tree was constructed based on the failures of the prior one. This helps to reduce errors. With GBM, fitting strategies for analysing and training a dataset need greater time and storage capacity (Reddy, P.V. and Kumar, S.M., 2022). The accuracy of the decision tree is 81.633% (Figure 12).

Correct classified: 240	Wrong classified: 54
Accuracy: 81.633%	Error: 18.367%
Cohen's kappa (κ): 0.396%	

Figure 12: Scorer output of Decision Tree

The efficiency of the model can be evaluated by utilising the Gradient booster machine in addition to the random tree node in the same manner. This makes it possible to compare the model's results. 85.374% of precision may be attributed to the gradient booster machine(Figure 13).

Correct classified: 251	Wrong classified: 43
Accuracy: 85.374%	Error: 14.626%
Cohen's kappa (κ): 0.455%	

Figure 13: Scorer output of Gradient booster

Random Forest Tree provided the most accurate predictions (Yadav et al., 2018). The accuracy rate was 85.7% (Figure 14), which is considered to be a reasonable accuracy. The Random Forest Learner can be viewed in Appendix 6.

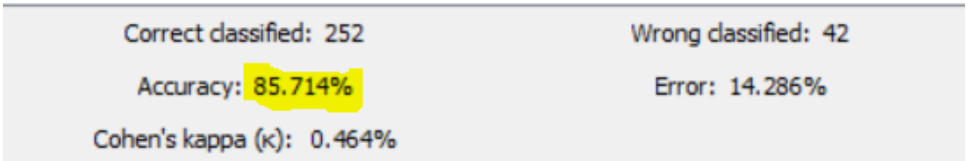


Figure 14: Scorer output of Random Forest Tree

Using matrices, it is possible to illustrate the derivation of several classification performance metrics from a multiclass confusion matrix. When the primary classes are further split, one method for calculating the likelihood of confusion is to utilise a confusion matrix that has been expanded (Visa, 2011). The use of confusion matrix analysis to validate. It performs a severe validity check and provides additional information on the kind and causes of mistakes since it is resilient to any data distribution and type of connection (Ruuska, et al 2018.). The confusion matrix for decision trees, random forests, and the gradient boosting model is shown in the figure below. It indicates that for decision trees, there is a chance that 26 out of 212 employers will leave the company, for random forests, there is a chance that 11 out of 227 employers of them will leave the company, and for gradient boosting, there is a chance that 12 out of 226 employers of them will leave the company.

The confusion matrix of decision trees:

Confusion Matrix - 5:10 - Scorer

File Hilite

Attrition \ ...	Yes	No
Yes	28	28
No	26	212

The confusion matrix of random forests tree:

Confusion Matrix - 5:19 - Scorer

File Hilite

Attrition \ ...	Yes	No
Yes	25	31
No	11	227

The confusion matrix of gradient boosting machine:

Confusion Matrix - 5:16 - Scorer

File Hilite

Attrition \ ...	Yes	No
Yes	25	31
No	12	226

Figure 15: Confusion Matrix of Decision tree, Random Forest & Gradient Boosting

Employees frequently leave their jobs or the company due to many reasons, including the belief that their work or workplace does not live up to their expectations or the perception that there is a mismatch between their skills and the responsibilities of their position (Zhang, 2016). One of the most common reasons why employees leave their jobs or the company is dissatisfaction with their work environment (Hammerberg, 2002). In addition to this, there is

an exceedingly inadequate quantity of both training and feedback, as well as very few opportunities for promotion or professional growth. Moreover, workers think they are underappreciated and acknowledged. They endure stress as a result of working too much, and they feel disconnected from their home, and professional lives. The last factor is a lack of confidence in those in leadership positions (Al-Suraihi, et.al., 2021).

Employee retention may be directly influenced by factors such as job satisfaction. Organisations must take the initiative to adopt the employee motivation process to improve overall employee performance by producing high-quality goods and providing first-rate services. Employees are the foundation of every business (Al Mamun, C.A. and Hasan, M.N., 2017).

Career growth affects turnover and company success. Employee Career Growth did not allow employees to upgrade their abilities, hence most left the company to find better positions. This limited employees' opportunities to further append their careers through training, delegating, the Mentoring programme, and job rotation. (Anzazi, N., 2018)

4.0 Conclusion

Conclusively, the modelling process, as analyzed by the algorithms, gives insights into which aspects are most predictive of employee attrition. The most accurate models can be predicted in the Random Forest Tree model that affects attrition in the firm by the tree structure. Improvement targets for the predictive model deployment can be established and continue to evaluate the work - monitor the model's performance and improve where possible.

From the visualizations, it is clear that the monthly income, per cent salary hike, total working years, job satisfaction level and work-life balance have positive relationships with attrition and are pertinent in determining the employee turnover rate as defined in the wider academic literature. Moreover, tableau visualisations can be used by both internal and external agents, working for the business by analysing and getting future insights. Therefore, companies can focus more on maintaining work-life balance and ensuring employees have good job satisfaction levels which will in turn help lower the attrition rate. Higher per cent salary hikes and performance ratings lead to lower attrition rates. Similarly, if a company has a good working culture, the employees are most likely to stay for a longer period. An employee having greater working experience and salary is more likely to stay. Moreover, employee turnover was highest among unmarried workers and lowest among divorced workers.

Organisations incur various types of expenses as a result of turnover (E, Mohammed et al. 2015). The cost of attrition is broken down into two categories: direct costs such as requisitioning, replacement, hiring and selection, temporary labour, and management time; and indirect costs such as costs related to employee morale, pressure on the workforce as a whole, cost of learning and quality of services ((E, Mohammed et al. 2015). Therefore, it is very important to find and solve the problems leading to high employee attrition in the company. Organisations that work hard to reduce employee churn typically have a competitive advantage because the ability of the organisation to succeed is directly impacted by the loss of highly talented individuals (Mehta & Modi, 2021). Therefore, business executives must comprehend the primary causes of staff turnover and take appropriate action to boost their company's performance by controlling overall workflow (Darapaneni et al., 2022).

5.0 References

Alogogianni, E. & Virvou, M., 2021. Data Mining for Targeted Inspections Against Undeclared Work by Applying the CRISP-DM Methodology. *12th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pp. 1-8.

Al-Suraihi, W.A., Samikon, S.A., Al-Suraihi, A.H.A. And Ibrahim, I., 2021. Employee Turnover: Causes, Importance And Retention Strategies. *European Journal Of Business And Management Research*, 6(3), Pp.1-10.

Al Mamun, C.A. And Hasan, M.N., 2017. Factors Affecting Employee Turnover And Sound Retention Strategies In Business Organization: A Conceptual View. *Problems And Perspectives In Management*, 15(1), P.63.

Anzazi, N. (2019) "Effects of total customer solutions strategic positioning on organizational performance in telecommunication industry, in Kenya," *European Journal of Business and Management* [Preprint]. Available at: https://doi.org/10.7176/ejbm/11-6-13.

Apgar IV, M., 2002. 21 The Alternative Workplace: Changing Where and How People Work. *Managing Innovation and Change*, p.266.

Augustine, M. O., Et Al. "The Impact Of Organisational Structure On Employee's Commitment In The Nigeria Manufacturing Sector. *Journal of Business and Organizational Development, 10(4)*, p. 35-47 "

Bueno, I., Carrasco, R., Ureña, R. & Herrera-Viedma, E., 2022. A business context aware decision-making approach for selecting the most appropriate sentiment analysis technique in e-marketing situations. *Information Sciences*, pp. 300-320.

Darapaneni, N. *et al.* (2022) "A detailed analysis of AI models for predicting employee attrition risk," *2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC)* [Preprint]. Available at: https://doi.org/10.1109/r10-htc54060.2022.9929893.

Fallucchi, F. et al. (2020) 'Predicting Employee Attrition Using Machine Learning Techniques', Computers, 9(4). Available at: https://doi.org/10.3390/computers9040086.

Hammerberg, J. H., 2002. Reasons given for employee turnover in a full priced department store.

Jaharuddin, N.S. and Zainol, L.N. (2019) 'The impact of work-life balance on job engagement and turnover intention', *The South East Asian Journal of Management*, 13(1), p. 7.

Jain, R. and Nayyar, A. (2018) 'Predicting Employee Attrition using XGBoost Machine Learning Approach', in *2018 International Conference on System Modeling & Advancement in Research Trends (SMART)*, pp. 113–120. Available at: https://doi.org/10.1109/SYSMART.2018.8746940.

Krishna, S. and Sidharth, S. (2022) “Analyzing employee attrition using Machine Learning: The new AI approach,” *2022 IEEE 7th International conference for Convergence in Technology (I2CT)* [Preprint]. Available at: https://doi.org/10.1109/i2ct54291.2022.9825342.

Kononenko, I., 1998, November. The Minimum Description Length Based Decision Tree Pruning. *In Pacific Rim International Conference On Artificial Intelligence* (Pp. 228-237)..

Lever, J., Et Al. (2016). "Points Of Significance: Model Selection And Overfitting." *Nature Methods* 13(9), p. 703-705.

Lazzari, M., Alvarez, J.M. and Ruggieri, S. (2022) “Predicting and explaining employee turnover intention,” *International Journal of Data Science and Analytics*, 14(3), pp. 279–292. Available at: https://doi.org/10.1007/s41060-022-00329-w.

Medina, E. (2012) Job satisfaction and employee turnover intention: what does organizational culture have to do with it? Columbia university.

Mehta, V. and Modi, S. (2021) “Employee attrition system using tree based ensemble method,” *2021 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4)* [Preprint]. Available at: https://doi.org/10.1109/c2i454156.2021.9689398.

Mohammad Esmaieeli Sikaroudi, A. and EsmaieeliSikaroudi, A. (2015) ‘A data mining approach to employee turnover prediction (case study: Arak automotive parts manufacturing)’, Journal of Industrial and Systems Engineering, 8(4), pp. 106–121.

Nguyen, G. H., Et Al. (2009). "Learning Pattern Classification Tasks With Imbalanced Data Sets." *Pattern Recognition*: P. 193-208.

Naik, A. And Samant, L., 2016. Correlation Review Of Classification Algorithm Using Data Mining Tool: Weka, Rapidminer, Tanagra, Orange And Knime. *Procedia Computer Science*, 85, Pp.662-668.

Ongori, H., 2007. A review of the literature on employee turnover. *African Journal of Business Management*, pp. 49-54.

Osborne, J. W. & Overbay, A., 2004. The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation*, 9(1), p. 6.

P, J.P., D, V. and K. V, U. (2022) “Effective classification of IBM HR analytics employee attrition using sampling techniques,” *2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)* [Preprint]. Available at: https://doi.org/10.1109/icaect54875.2022.9808057.

Porter, M. E. And J. E. Heppelmann (2015). "How Smart, Connected Products Are Transforming Companies." *Harvard Business Review* 93(10): 96-114.

Purbasari, A. et al., 2021. CRISP-DM for Data Quality Improvement to Support Machine Learning of Stunting Prediction in Infants and Toddlers. *2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA*), pp. 1-6.

Raman, R., Bhattacharya, S. and Pramod, D. (2018) "Predict employee attrition by using predictive analytics," *Benchmarking: An International Journal*, 26(1), pp. 2–18. Available at: https://doi.org/10.1108/bij-03-2018-0083.

Reddy, P.V. And Kumar, S.M., 2022, October. A Method For Determining The Accuracy Of Stock Prices Using Gradient Boosting And The Support Vector Machines Algorithm. *In 2022 3rd International Conference On Smart Electronics And Communication (Icosec)* (Pp. 1596-1599).

Ruuska, S., Hämäläinen, W., Kajava, S., Mughal, M., Matilainen, P. and Mononen, J., 2018. *Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. Behavioural processes*, 148, pp.56-62.

Schröer, C., Kruse, F. & Gómez, J. M., 2021. A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, Volume 181, pp. 526-534.

Seelam, S.R. *et al.* (2022) "Comparative study of predictive models to estimate employee attrition," *2022 7th International Conference on Communication and Electronics Systems (ICCES)* [Preprint]. Available at: https://doi.org/10.1109/icces54183.2022.9835964.

UNICEF, n.d. Minimum age for admission to employment, s.l.: unicef.org.

Yadav, S., Jain, A. and Singh, D. (2018) "Early prediction of employee attrition using data mining techniques," *2018 IEEE 8th International Advance Computing Conference (IACC)* [Preprint]. Available at: https://doi.org/10.1109/iadcc.2018.8692137.

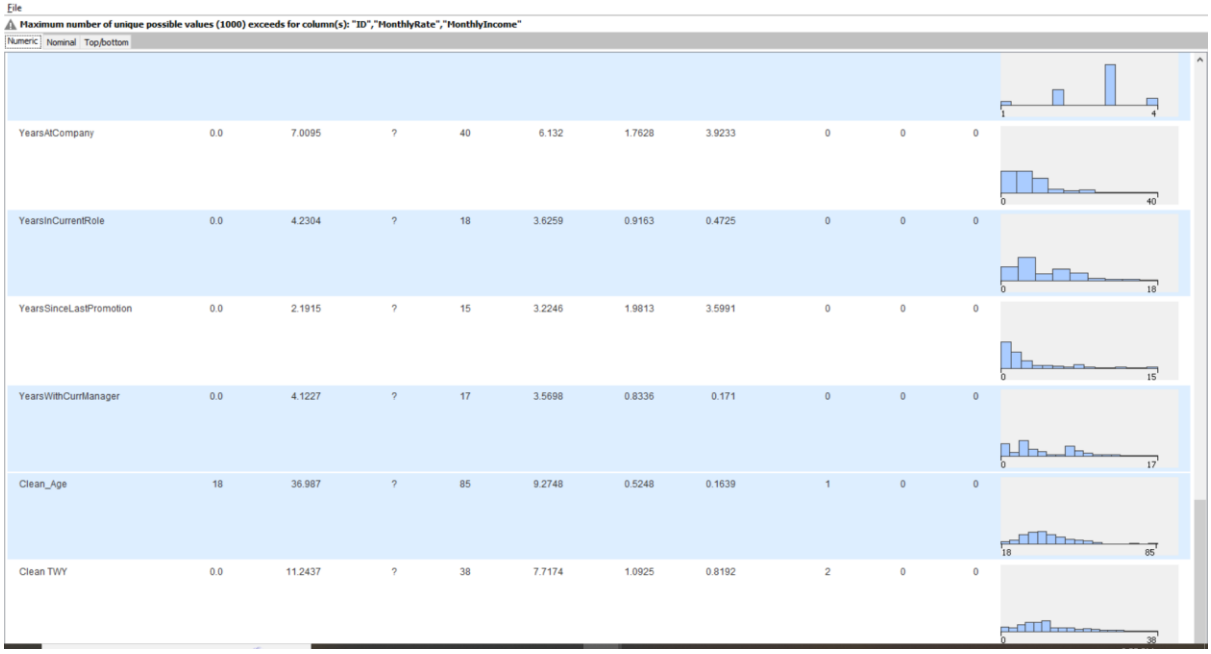
Yahia, N.B., Hlel, J. and Colomo-Palacios, R. (2021) "From big data to deep data to support people analytics for employee attrition prediction," *IEEE Access*, 9, pp. 60447–60458. Available at: https://doi.org/10.1109/access.2021.3074559.

Yang, S. and Islam, M.T. (2021) *IBM employee attrition analysis*, *arXiv.org*. Available at: https://arxiv.org/abs/2012.01286 (Accessed: December 15, 2022).

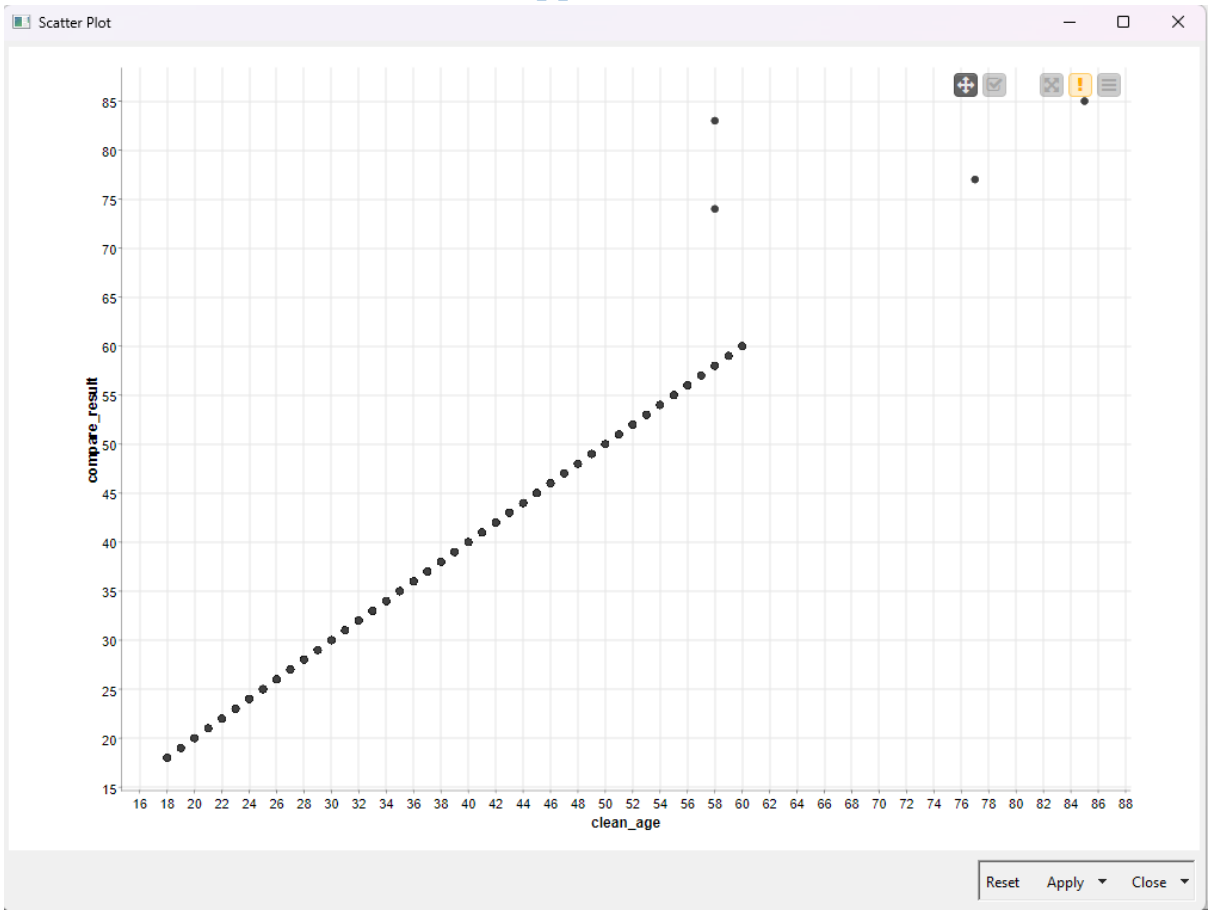
Zimmerman, R.D. and Darnold, T.C. (2009) 'The impact of job performance on employee turnover intentions and the voluntary turnover process: A meta-analysis and path model', *Personnel review* [Preprint].

Zhang, Y., 2016. A Review of Employee Turnover Influence Factor and Countermeasure. *Journal of Human Resource and Sustainability Studies*, 4(2), pp. 85-91.

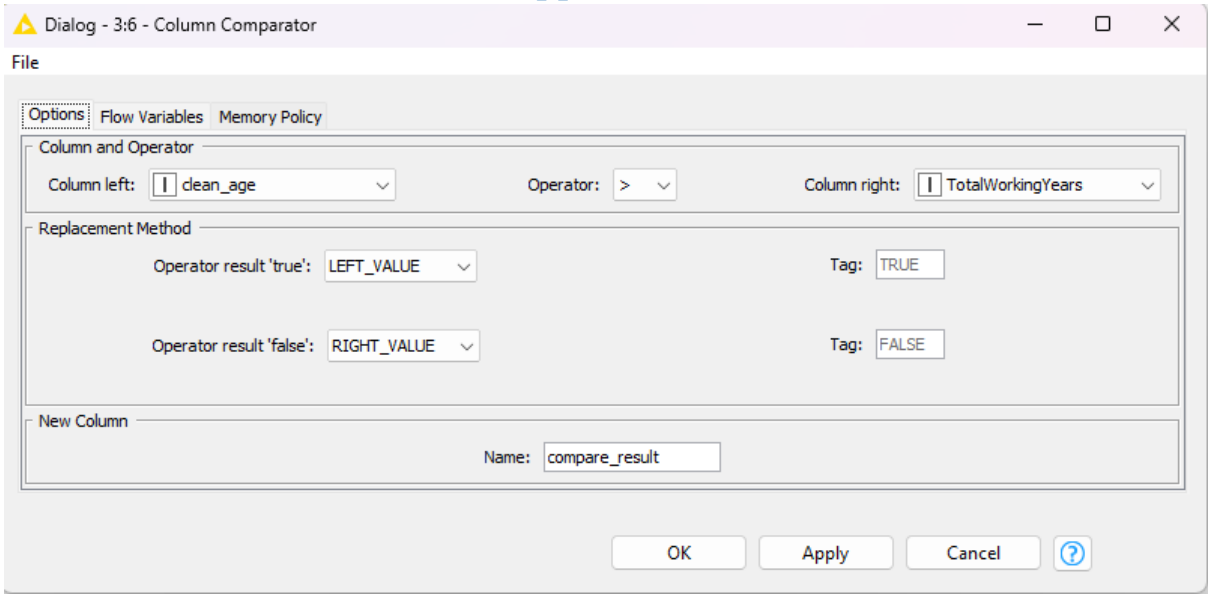
Appendix 1



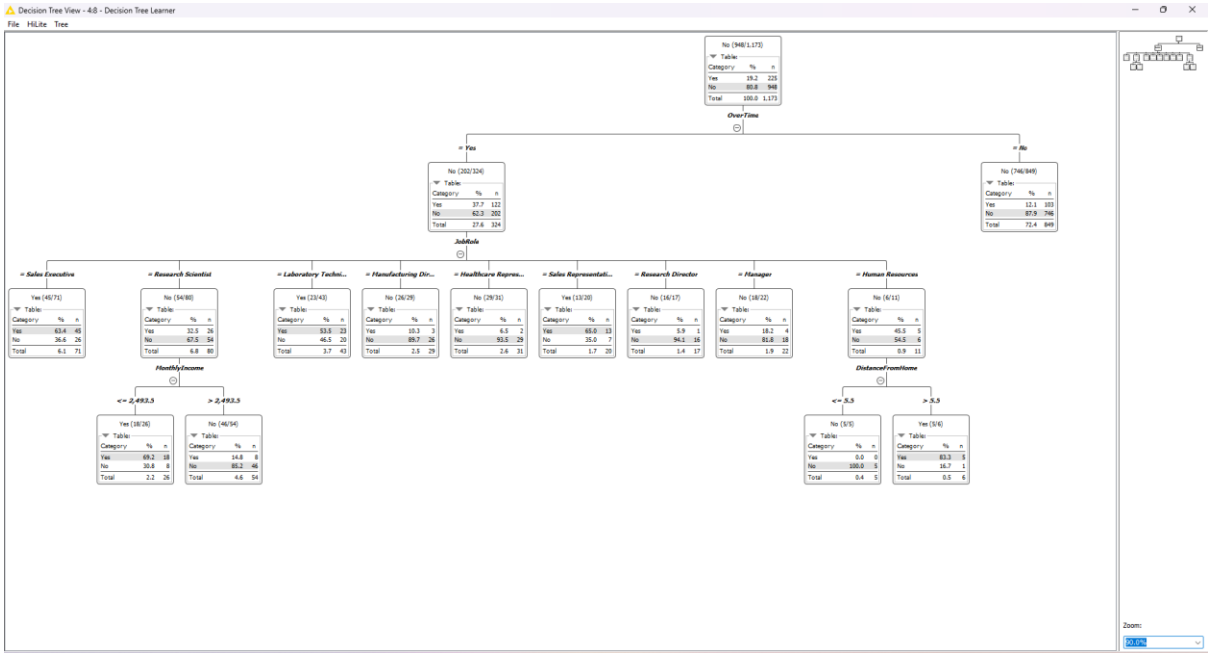
Appendix 2



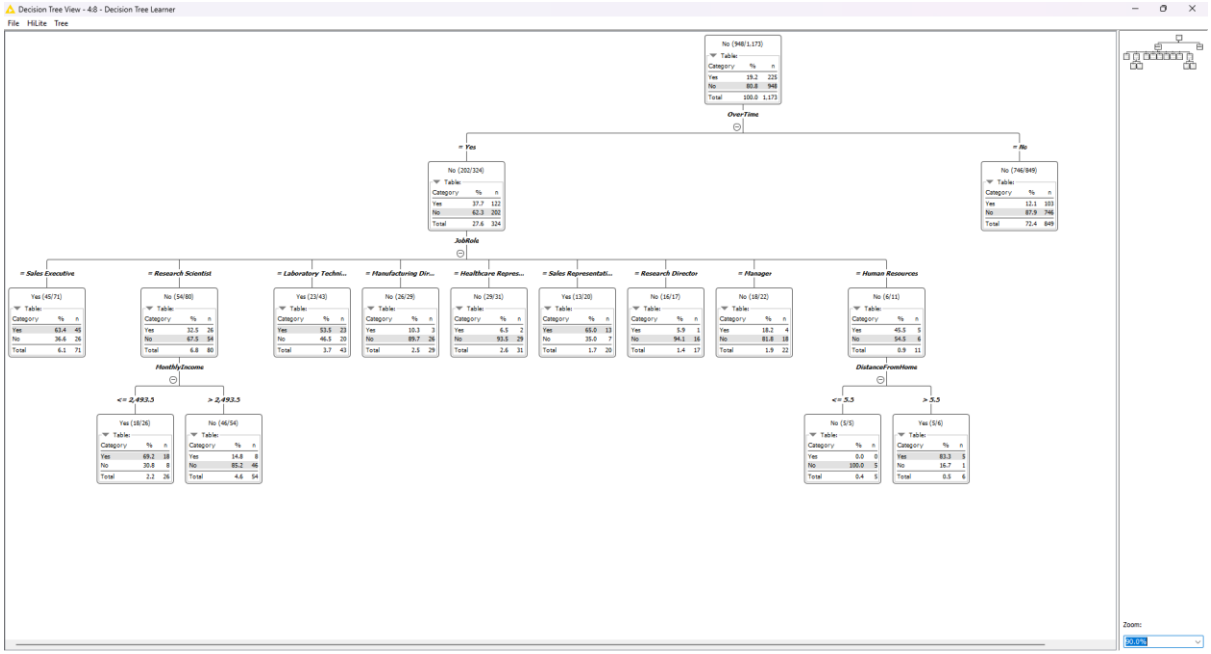
Appendix 3



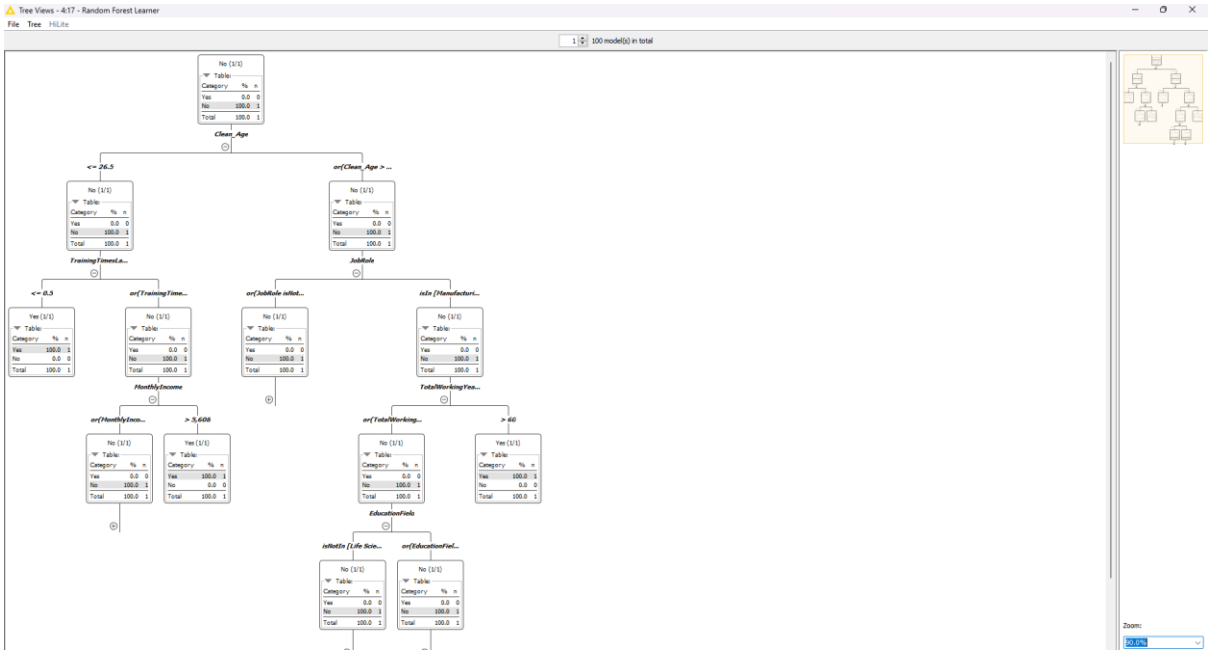
Appendix 4



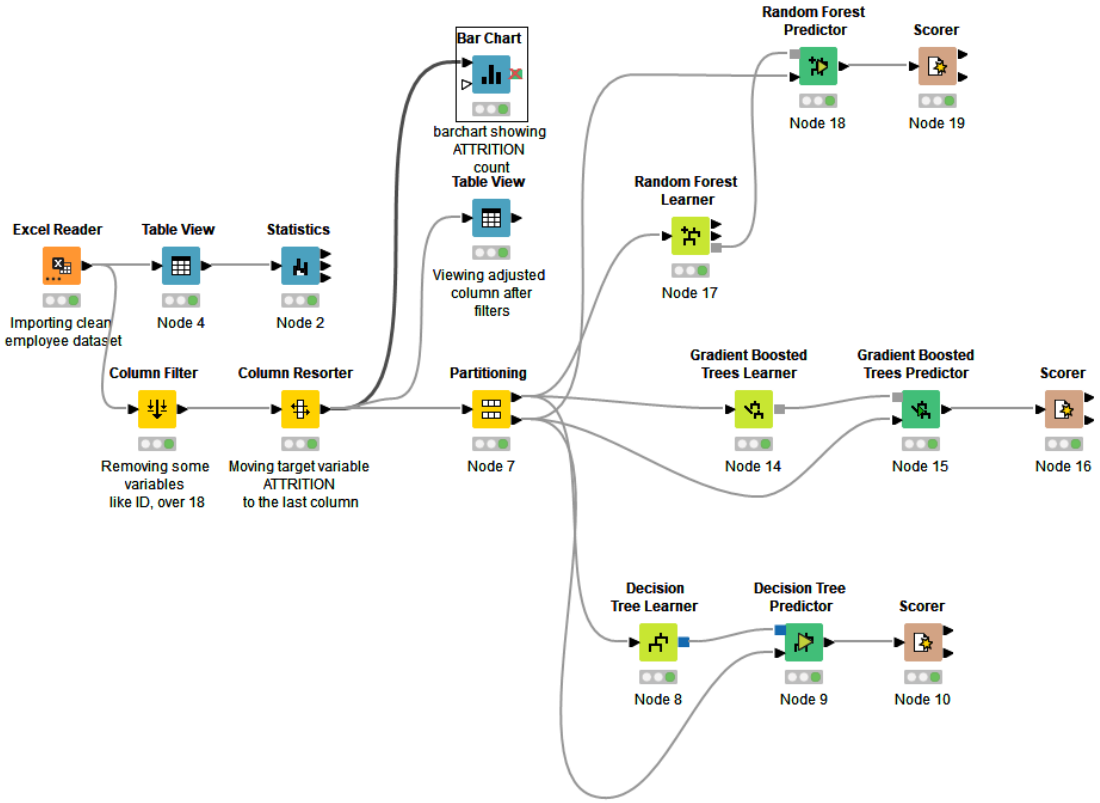
Appendix 5



Appendix 6



Appendix 7


Excel Reader
Importing clean employee dataset
Table View
Node 4
Statistics
Node 2
Bar Chart
barchart showing ATTRITION count
Table View
Viewing adjusted column after filters
Column Filter
Removing some variables like ID, over 18
Column Resorter
Moving target variable ATTRITION to the last column
Partitioning
Node 7
Random Forest Learner
Node 17
Random Forest Predictor
Node 18
Scorer
Node 19
Gradient Boosted Trees Learner
Node 14
Gradient Boosted Trees Predictor
Node 15
Scorer
Node 16
Decision Tree Learner
Node 8
Decision Tree Predictor
Node 9
Scorer
Node 10