Analysis of Employee Performance and Job Satisfaction

Aisha Binti Muhammad Fakhar Iqbal
School of Computing
Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur, Malaysia
tp050393@mail.apu.edu.my

Dewi Octaviani
School of Computing
Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur, Malaysia
dewi.octaviani@apu.edu.my

Halimaton Saadiah Hakimi
School of Computing
Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur, Malaysia
halimaton.saadiah@apu.edu.my

***Abstract*—Human Resource (HR) Managers devote long hours to generating descriptive reports that provide insight into employees' performance and job satisfaction. However, they were unable to answer a question as to why employee performance and job satisfaction are low. HR managers tend to rely on intuition to apply HR strategies in their companies. HR managers would be able to better understand their employees using HR analytics that can then be turned into interactive dashboards to further answer the question. This paper applied exploratory data analysis and predictive modelling to gain insights into a company's employee performance and job satisfaction data and suggest potential HR management strategies. SAS Enterprise Miner and Power BI were used to complete the research. The results show that Decision Tree was concluded to be the most optimal model with a cumulative lift of 2.258 for the performance rating and 1.307 for job satisfaction.**

***Keywords*— exploratory data analysis, predictive modelling, HR analytics, employee performance, job satisfaction.**

I. Introduction

Analytics is a revolutionary technique that can help us understand data patterns that help decision-making at an industrial or business level. Analytics can help HR Managers shed light on hidden patterns and challenges within the organization that may be affecting their business and company and, importantly, the employees. Edwards & Edwards (2019) stated that HR management teams spend a lot of time generating descriptive reports and comparing them to investigate what is happening in the organization over time. There is no doubt that such descriptive reports help the organisation understand what is happening in their organisation, such as low employee performance or high attrition rate. However, it is unable to address the question of 'why' it is happening.

Moreover, many HR leaders continue to use trends, intuitive instinct, industry standards, and benchmark tests as the foundation for decision-making. However, only certain HR leaders use statistics based on past data. Failure to integrate data from different sources restricts the utility of methodologies for human resources managers to provide findings to improve employee conditions such as performance and satisfaction. Lack of identification of employee needs, therefore, also results in a high number of employee attrition. (Etukudo, 2019)

The importance of knowing which factors in the workplace affect job satisfaction can be demonstrated by the research findings of the 2012 Human Resource Management Society Study Survey, where 6 out of 10 employees pointed out that wages were very significant for overall worker satisfaction. This reason was ranked only 3% less than talent and upskilling opportunities being provided at the workplace and just 1% less than job security (Yamoah, 2014).

However, HR managers have had difficulty bringing data together from different sources. This can result in data being available in different formats. This may be difficult to bring together under a single dataset. Therefore, HR needs the initiative to collect data from various resources. This is where predictive and exploratory data analysis plays a role. Higher recovery opportunities can therefore be achieved by businesses that assess their employee's existing abilities and offer incentives for employees to learn additional skills. This paper aimed to provide insights that result in actionable results thus HR managers can refer to analytical data rather than their instinctive feeling when it comes to facilitating their current employees. HR managers can, therefore, facilitate their existing work staff better with solid analytical data to support their next step in improving employee satisfaction and performance within the organization.

II. Literature Review

A. Human Resource Analytics

HR Analytics is defined as a systematic recognition and quantification of the generators of corporate results to make smarter choices (Keerthi & Reddy, 2018). Five types of HR analytics applications can be categorised, including identification and management of talent, management of critical workforce, prediction of employee preferences and needs, determining staffing requirements and business goals, and scaling recruiting source networks and targets. Like all analytical techniques, HR analytics has to go through all the stages of data analysis such as data collection, application of methodologies as well as challenges that need to be overcome. It is crucial to allow the organization to collect data consistently and be usable and of high quality as a key prerequisite for effective HR analytics. It is difficult to produce accurate results without the right data to evaluate, anticipate, forecasting, and maximize HR processes and capacities (Joerik van Dooren, 2012).

HR analytics can be utilised to discover employee performance and job satisfaction. As high employee performance and job satisfaction are key to completive and good business performance. Here is where HR analytics comes in as it allows company managers to correlate the survey data, gather significant business findings, and target initiatives on the main results-orienting areas (Mondore et al., 2011). Job Satisfaction can be defined as an effective response to a job arising from an employee's contentment with the job they are performing. Through the application of HR analytic-based work systems, the performance of the company can be increased. Concerning this, the role of high-performance work systems (HWS) in Job Satisfaction and greater involvement were investigated by Fabi, et al (2015). The results of the study concluded that HWS had a positive effect on job satisfaction based on 730 employees' data, which contributed to the growth, encouragement, and potential of HR analytical practices. In addition, employee performance is also one of the most important factors when it comes to boosting company performance. When employee performance is elevated, business or company performance can then be placed at a competitive level with the rest of the booming businesses. Positive employee performance reflects on employee job satisfaction.

B. Exploratory Data Analysis and Data Visualization

Exploratory Data Analytics (EDA) helps to find the right way to modify data sources to have the appropriate answers. It enables the exploration of patterns, evaluate anomaly, testing theories, and verifying conclusions by data experts. An example of the demonstration of data visualization through EDA was completed by Liu, et al., (2020) where the researchers used bar charts to demonstrate factors affecting employee turnover. These bar charts are shown below. The bar chart was also used to visualise the Entity Sentiment analysis conducted by the researchers through Google Cloud NLP. The generation of dashboards is also a great technique when it comes to demonstrating data to the HR management, information management team, and other managers at the company in an understandable and visually-capturing way. The main purpose of data demonstration through dashboards is the ease of data understanding for those who are not comfortable with interpreting numbers or statistical data in a table. Davidescu et al., (2020) investigated the implications of Sustainable Human Resource Management and demonstrated their finding using a dashboard. This is a great example of data visualization that caters to HR management. The results above show statistics understandably and clearly for HR managers to make informed decisions. Dashboards enable decision-makers to understand the new HR key indicators easily. The dashboard helps HR professionals to explain their findings and expertise based on comprehensive data analysis.

C. Predictive Model

Predictive modelling helps to show what are the factors that actively affect employee attrition, job satisfaction, workplace environment satisfaction, and more. One of the techniques is Logistic Regression which helps to predict the likelihood that one or more independent inputs can be inferred by a binary or common goal. Liu, et al., (2020) describe Logistic Regression as being a very important model for representing the outcome of each employee turnover which could mean less job satisfaction. To predict the results of more than one level, logistic regression is the most suitable type of regression that can be used. Moreover, Decision Trees are widely used for predictive and descriptive analysis. A decision tree exhibits a hierarchical segmentation of data. Parts of a decision tree include node rules that assign each part or record of the data to a leaf node. It includes subsequent probabilities of each leaf node with the target variable being assigned to each level of the leaf node as well (Sarma, 2017). The summary of similar works in human resources analytics can be seen in Table I.

III. METHODOLOGY

This research adopted the CRISP-DM methodology which consists of business understanding, data understanding, data preparation, modelling, evaluation, and deployment as seen in Fig.1.

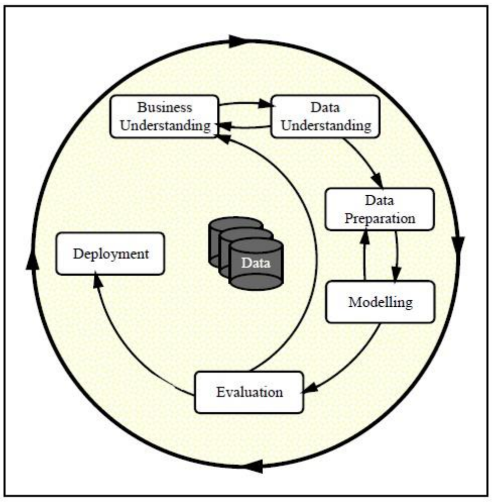

Fig. 1. CRISP-DM Methodology (Wirth & Hipp, 2000)

Business Understanding refers to understanding the company or business goals and objectives that may facilitate the project. Understanding where the company may be lacking in their ways of handling employees for example, through the overview of company background and employee data available. The data will be viewed and attributes that may give meaningful insights can be chosen moreover certain questions about the project (to be answered) can be drawn from them. E.g.: do working hours affect job satisfaction in employees, etc. data quality will be checked by going through every data table. Any attributes that do not pose a clear purpose can be eliminated in this step. Data will then be cleaned, and any outliers can be identified, all irregular or missing data will be identified as mentioned before. The data preparation phase further includes data selection, data partition, data transformation, etc. Modelling for this research used Regression and Decision Trees to build the models. Evaluation is where it is made sure that the results are correct and tested for any irregularities. Evaluation of the project is important, to determine if the information matches Deployment. The results of the research are evaluated using MAE, MSE, and R2 or a Chi-Square test. The review stage would consist of making sure all areas of the evaluation phase have been covered and fully executed, with no area being left out or overlooked.

TABLE I. SIMILAR WORKS IN HUMAN RESOURCE ANALYTICS.

Authors	**Description**	**Dataset Used**	**Analysis Techniques**	**Evaluation Techniques**	**Outcomes**	**Limitations**
Davidescu et al., (2020)	Study the relationship between working flexibility and employee satisfaction and job performance.	Data from 16 districts, and 220 people.	Descriptive statistics and charts. logistic regression	Confusion Matrix	The outcome found that the higher the salary, the higher the job satisfaction. Motivational factors such as flexibility in time and the workplace were also uncovered.	Research conducted among Romanian employees may not apply to all types of employees.
Liu, et al., (2020)	Predictive analysis to build two machine learning models. Sentiment analysis by the use of Google Cloud Natural Language Processing.	Data from the aluminium company 1866 employees.	K-nearest Neighbours, Logistic Regression. Random Forest Gradient boosting Decision Tree. (CHAID),and (CART) Entropy Google Cloud NLP: Sentiment analysis.	ROC (Receiver Operating Characteristic) AUC (Area under the curve), Accuracy, recall, and precision.	The Logistic Regression results indicate - workers who were relocated or promoted and those with lower benefits payments, have a higher chance of leaving. Employees have the best attitudes towards progress, and growth activities. Employees have the poorest attitudes to leadership, managers; training, growth, opportunity; and work-life balance.	No limitations associated with the research
Qureshi et al., (2020)	This study examines the role of HR analysis, performance pay, and human resources participation in Job Satisfaction and firm performance in various	The dataset consists of 317/331 responses to a survey conducted among the various 27 different multinational firms.	Data mining was done with the help of two data mining software. SmartPLS Version 3.2.7 and SPSS, Statistical Package for Social Sciences (Version-21).	Pearson's Correlation test, Cronbach's alpha, reliability value estimations, AVE, as well as probability values and t-stats.	The effects of partial less square equation modelling have also shown the optimistic and important influence of HR analytics, HR engagement, and Work satisfaction on firm performance in Malaysian multinational companies.	The research was conducted among Malaysian Multinational companies only. The study may not apply to worldwide data.
Etukudo, R, (2019)	In this descriptive, qualitative case study, HRA's strategies were examined for improving HRA's business performance. The study was conducted with 5 HR managers in the USA, D.C. & Lagos, Nigeria, who successfully utilized HRA in strategic decision-making.	Data collection was done through interviews. observation and document reviews as well. 5 participants were interviewed.	Data triangulation analysis was used as the analysis technique. Using NVivo software.	Data evaluation methods included member checking, triangulation, engagement over some time, and reflective journaling.	All participants agreed that there is a need for HRA.	The small sample sizes Interviews were restricted. research is limited to the US and Nigeria and the findings may not be applicable in other countries.

In the deployment, data will be prepared for presentation to the end users, thus they can access the results of the analysis. Includes planning for the next steps. Here planning for the next stage should be done at the corporate or workplace level. The major change is that companies would concentrate on the processes/competencies/attitudes of workers, which have proven to have a direct influence on the business results that the company seeks to accomplish, not just an implied consequence or the impression that the right thing is to happen. An anticipated return can now be used to direct the HR plan and creation, or modifications must be adapted to each particular organization (Mondore et al., 2011).

IV. Results and Discussion

This section describes the results of the analysis and discussions the findings.

A. *Data Understanding and Preprocessing*

The data for this project was collected via Kaggle and is consisting of employee survey data of up to 4000 records. The data dictionary is shown in Table II. The next one is to determine the roles of variables, there are 2 target variables "Performance Rating" and "Job Satisfaction". Setting their role as target variables means that these variables will be the ones that will be predicted by the input and other variables. The target variables are Ordinal data. Ordinal data is data that has naturally occurring levels. The variables that have a numeric distance between them such as Years Since the Last Promotion are called interval variables. The variables that have data in only two levels are known as binary variables such as gender (M/F) and Attrition (Y/N) Similarly, data that has only one level is known as unary (Sarma, 2017). The rest of the variables will be treated as input variables that will be used to predict the target variables.

Mean imputation is used to treat the missing values as it is the most common and unbiased way for data analysts to treat missing interval values, it also helps measure the central tendency of the values. According to Sarma (2017), it is important to handle missing values as it can cause the predictive model and insights to be impacted and affect the results. Missing values are often ignored by predictive models such as Regression and Decision Trees, however, completely avoiding missing values can cause us to miss out on the important information

TABLE II. DATA DICTIONARY

Variable	Description of Variable	Data_type	Ranking
Age	The Age of every employee	Integer	
Attrition	If the employee has left or not	Boolean (Yes/No)	
BusinessTravel	How frequently employees go for a business trip	Character	
Department	Company Departments (Sales, Reasearch and Development, etc.)	Character	
DistanceFromHome	Distance between work and home in Kms	Integer	
Education	Education Level	Integer	1 'Below College' 2 'College' 3 'Bachelor' 4 'Master' 5 'Doctor'
EducationField	Field of education (Medical, Life Science etc.)	Character	
EmployeeCount	Employee count	Integer	
EmployeeID	Unique ID for each employee	Integer	
EnvironmentSatisfaction	Satisfaction Level of each employee in the Work Environment	Integer	1 'Low' 2 'Medium' 3 'High' 4 'Very High'
Gender	Gender of employee	Character	
JobInvolvement	Level of Job Involvement for each employee	Integer	1 'Low' 2 'Medium' 3 'High' 4 'Very High'
JobLevel	Level of responsibility at the company on a scale of 1 (Basic) to 5 (Expert)	Integer	
JobRole	Job role name in the company	Character	
JobSatisfaction	Levels of Job Satisfaction at the company	Integer	1 'Low' 2 'Medium' 3 'High' 4 'Very High'
MaritalStatus	Employee Marital Status (Married, Divorced, Single etc.)	Character	
MonthlyIncome	Employee Monthly income in INR (Indian Rupees)	String	
NumCompaniesWorked	Employee company count (number of the companies the employee has worked for)	Integer	
Over18	Employee age over 18 or not	Boolean (Yes/No)	
PerformanceRating	Employee's rate of performance at the company measured from 1 (low) to 4 (outstanding)	Integer	1 'Low' 2 'Medium' 3 'Good' 4 'Outstanding'
StandardHours	Hours worked by employee on average	Integer	
StockOptionLevel	Right of buying stock given to the employee (usually high level employees have this option)	Integer	
TotalWorkingYears	Duration of years employee has worked at the company	Integer	
TrainingTimesLastYear	The amount of time the employee has attended training in the year	Integer	
WorkLifeBalance	Balance of personal life and work life for this employee, measured in levels.	Integer	1 'Bad' 2 'Good' 3 'Better' 4 'Best'
YearsAtCompany	Duration of years the employee has spent at the current company	Integer	
YearsSinceLastPromotion	Duration of years passed since the employee has gotten a promotion	Integer	
YearsWithCurrManager	Duration in years spent with the current manager	Integer	

Table III shows the results of missing values, which can be seen which impute method was used, the imputed value, and the number of missing values that were imputed. The missing values for Environment Satisfaction and Work-Life Balance were imputed by level 3 using the method count. Job Satisfaction was also imputed using level 4. The number of Companies worked, and Total Working Years were imputed using the mean value to imputed values of 2.6948 and 11.2799 respectively. The dataset was then divided into 70 for training and 30 for validation (Sarma, 2017). The training and validation set is then used for developing the prediction models such as Regression and Decision Trees, and the validation is used to evaluate these models.

TABLE III. MISSING VALUES SUMMARY

Imputation Summary
Number Of Observations

Variable Name	Impute Method	Imputed Variable	Impute Value	Role	Measurement Level	Label	Number of Missing for TRAIN
EnvironmentSatisfaction	COUNT	IMP_EnvironmentSatisfaction	3.00000	INPUT	ORDINAL		25
JobSatisfaction	COUNT	IMP_JobSatisfaction	4.00000	TARGET	ORDINAL		20
LOG_NumCompaniesWorked	MEAN	IMP_LOG_NumCompaniesWorked	1.08626	INPUT	INTERVAL	Transformed NumCompaniesWorked	19
LOG_TotalWorkingYears	MEAN	IMP_LOG_TotalWorkingYears	2.29418	INPUT	INTERVAL	Transformed TotalWorkingYears	9
WorkLifeBalance	COUNT	IMP_WorkLifeBalance	3.00000	INPUT	ORDINAL		38

The next phases are variable selections for employee performance and job satisfaction ratings using Chi-Squared and R-Squared Selection. These techniques analyse the strength of the relationship the variables have to the target variable and select the variables based on this relationship. (Sarma, 2017) . Fig 2. shows that Job Role, Age, Gender, Job Involvement, and Total Working Years are the top 5 variables that have the most contribution to the prediction of Employee Performance due to a high relative importance score of 1, 0.873154, 0.665522, 0.59983 and 0.591808 respectively.

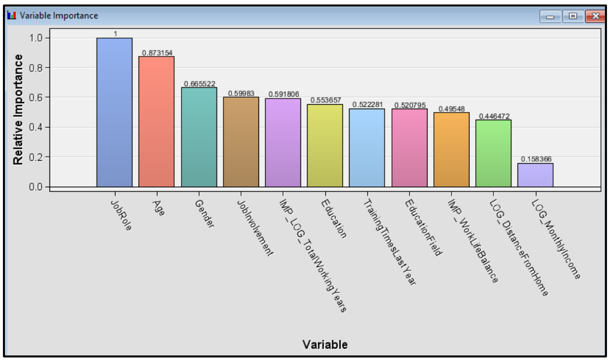

Fig. 2. Variables Selection for Employee Performance Rating

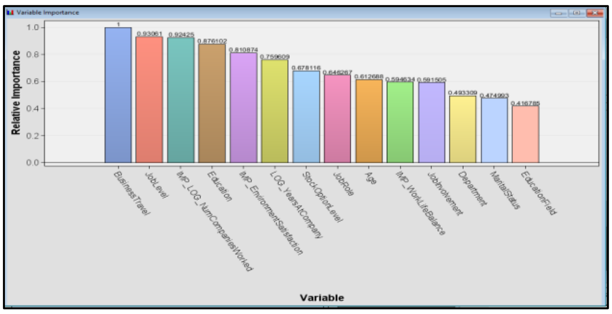

Fig. 3. Variables Selection for Job Satisfaction Rating

Whilst, the variables that can contribute the most to the prediction of Job Satisfaction are Business travel, Job Level, Number Of Companies Worked, Education, and Environment Satisfaction, with relative importance values being 1, 0.93061, 0.92425, 0.876102, and 0.810874 respectively as shown in Fig 3.

B. *Model Buildings and Evaluations*

Two models were constructed for the target Performance Rating to evaluate employee performance. The logistic regression and decision tree models were constructed with RMSE (Root Mean Squared Error), ASE (Root Mean Squared Error), and MISC (Misclassification Rate) used to evaluate the quality of the model.

Table IV shows the evaluation results for the regression models. It can be seen that all the values of ASE, and MISC are closer to 0 than to 1. A MISC of 0.15 as per validation data means that 15% of the data only was misclassified whereas, 0.84 or 84% of the data has been classified correctly. Further in this evaluation, we can calculate the accuracy and precision of the model as well. The RMSE value can be seen to be nearer to 0, this means that the data is concentrated near the Regression line of best fit instead of being spread out. The result of the job satisfaction model is depicted in Table V. It can be seen that ASE is closer to 0 than to 1, meaning it has a low value, this is desirable as it shows that the accuracy of the model is 1-0.23= 0.76 which means that the model is 76% accurate with a 23% error rate. The MISC of 0.38 shows that 38% of the model outcome was misclassified wherewith 0.60833 or 61% of the data was classified correctly. RMSE value of 0.48 can be seen to be 48% which means that data is a little spread out from the line of best fit.

TABLE IV. REGRESSION MODELS EVALUATION

Evaluation Technique	Train	Validation
Performance Rating		
ASE	0.127262	0.130547
RMSE	0.35761	0.361312
MISC	0.152823	0.155873
Job Satisfaction		
ASE	0.23523	0.238452
RMSE	0.485635	0.488316
MISC	0.384241	0.389894

As for the decision tree models, the fit statistics in Table V show the evaluation statistics chosen to evaluate this model. The misclassification rate and average squared error for the performance rating can be seen to be nearer to 0 than to 1. The misclassification rate for validation data can be seen as 0. 156509 whereas the average squared error is 0.131302. For the misclassification rate, it can be said that 15% of the data was misclassified Whereas 84% of the data has been classified correctly. The main evaluation technique that will be used to analyse this model in more detail is the Average Squared Error rate. Whilst for the job satisfaction tree, the misclassification rate of the validation sample at 0.38 means that 38% of the data was misclassified which is an optimized misclassification rate as compared to the 68% misclassification rate the Decision Tree had before the recording of the Job Satisfaction target level from 4 levels to 2 levels.

TABLE V. FIT STATISTICS FOR DECISION TREES

Evaluation Technique	Train	Validation
Performance Rating		
ASE	0.123874	0.131302
MISC	0.1506	0.156509
Job Satisfaction		
ASE	0.23241	0.24006
MISC	0.38035	0.38009

The assessment plot in Fig 4. shows the data sequentially being split and the Average squared error that is corresponding with each subtree. As we have chosen the average squared error assessment measures the subtree assessment plot shows that the most optimal Decision Tree is at leaf number 7 with the average squared error all 0.1313 Even though the initial maximal tree was built up to more than 20 leaves. This means that the accuracy of the predictions after leaf number 7 starts to decrease as the optimality of the leaves starts to decrease after leaf 7. The Subtree assessment plot in Fig 5. shows that the tree was built up to more than 30 leaves. However, the ASE shows that the most optimal predictions were made up to leaf number 12 as it has the lowest Average Square Error at the value of 0.2401. After this point, the validation line can be seen going upwards, which means that the accuracy of the model worsened after this point. Therefore, the leave selection in the settings of this Decision Tree node was 12 leaves for the most accurate predictions.

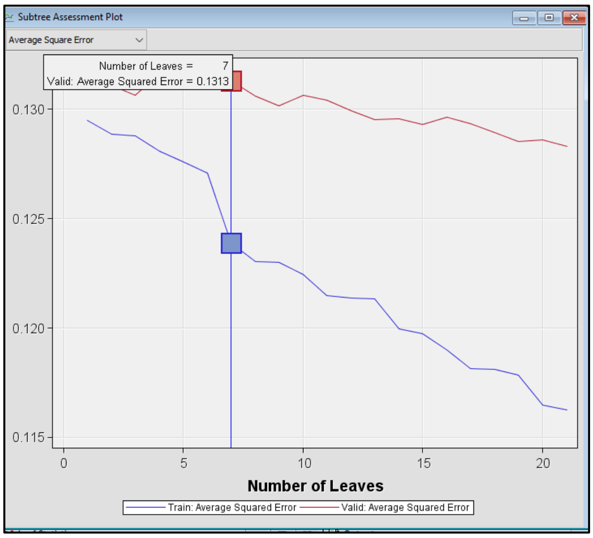

Fig. 4. Subtree Assessment Plot - Performance Rating

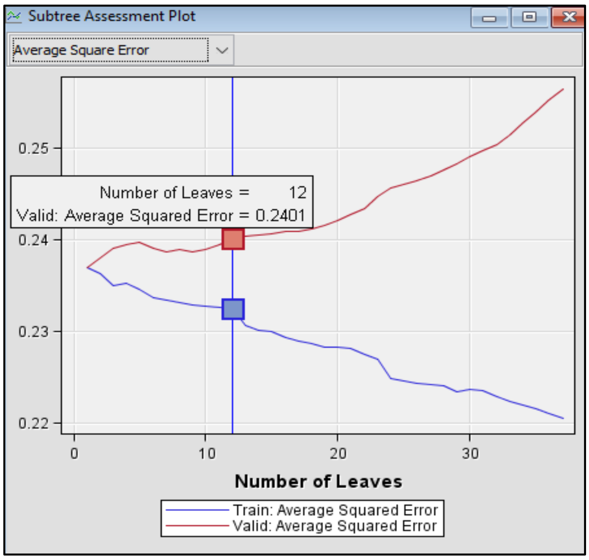

Fig. 5. Subtree Assessment Plot – Job Satisfaction

Model comparisons were shown that the higher the cumulative lift is meaning the model is performing well. It can be seen that the Decision Tree denoted by the blue line has a better cumulative lift of 2.258 as compared to Regression 1.784. In line with job satisfaction that shows the decision tree has a better cumulative lift of 1.307 as compared to Regression 1.155 as shown in Fig 6 and 7.

Subsequently, the results from the Decision Tree model as well as general HR employee data were visualised using Power BI and three dashboards were created. The first dashboard contains an overall view of the employee demographic. The Decision Tree data was exported using the save data node in SAS Enterprise Miner, as a CSV file to be visualised in Power BI . Minor pre-processing steps were taken using Power Query to visualise the data better such as grouping job levels, age, income, and formatting probabilities to percentages. The employee demographic data is seen in Fig 8.

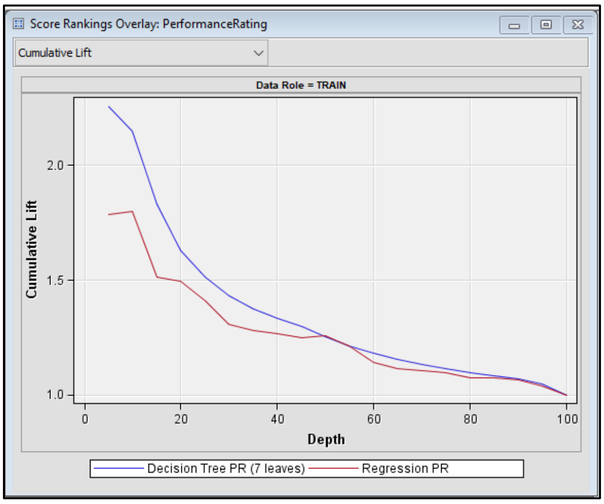

Fig. 6. Cumulative Lift - Performance Rating

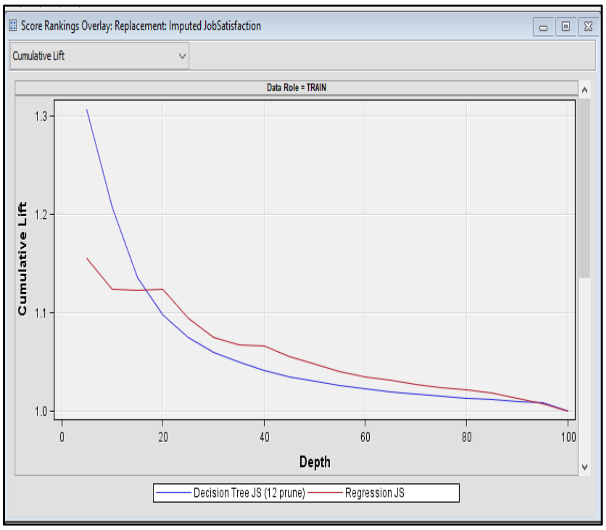

Fig. 7. Cumulative Lift – Job Satisfaction

Whilst, Fig 9. shows the Job Satisfaction dashboard. A decomposition tree was constructed for high Job Satisfaction as well as low Job Satisfaction to see which factors of a variable affect the Job Satisfaction level the most according to what the Decision Tree model had predicted. The vertical bar showing the nodes on the right can be used as buttons to see what the actual Decision Tree had predicted moreover important variables such as Total working years, Department, and Marital Status have been shown as clickable buttons to make the tree interactive and see different outcomes. Below is a demonstration of 2 Node results in the dashboard: The decomposition tree in green shows how important variables react to “4 – High Job Satisfaction” showing the probabilities of high Job Satisfaction among variables and the decomposition tree in red shows “1 – Low Job Satisfaction” showing the probability of low Job Satisfaction among variables. A score Card has been used to show the probability that the nodes have predicted as percentages for each level of the target.

Similarly, a dashboard was created showing the Decision Tree modelling results for factors that affect employee performance ratings according to the Decision Tree predictions (Fig 10.). The vertical bar on the right, similarly, shows the node rules that are clickable to view the outcome of the results. Important variables such as Age, Department, Education Level, and Gender have been shown as clickable buttons to make the tree interactive and view how the variables react in the Decision Tree and view the change in percentages. Monthly Income has been shown as a funnel chart. Decomposition tree in green shows how important variables react to “4 – High - PR” showing the probabilities of high-performance rating and “3 – Good/Average - PR” showing the probability of low-Performance Rating among variables. A score Card has been used to show the probability that the nodes have predicted as percentages for each level of the target. Demonstration of the scorecards and Decision Trees can be seen below.

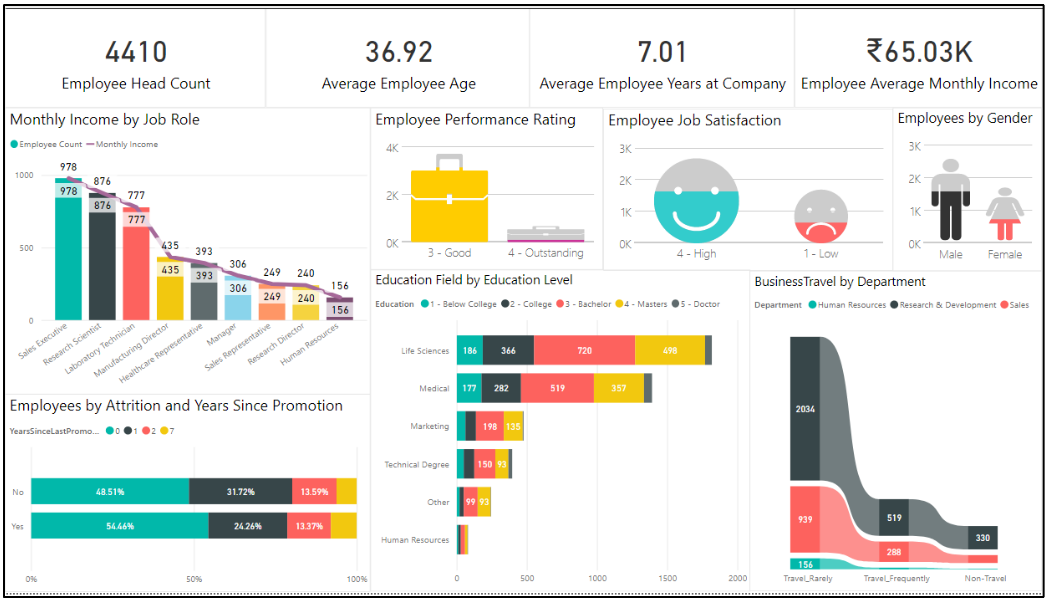

Fig. 8. Employee Demographic Dashboard

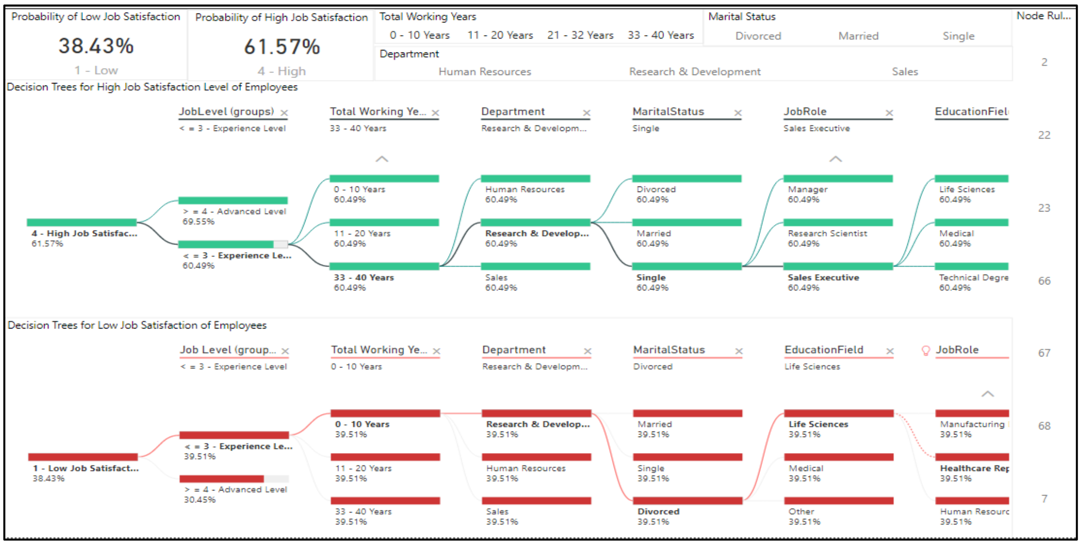

Fig. 9. Job Satisfaction Dashboard

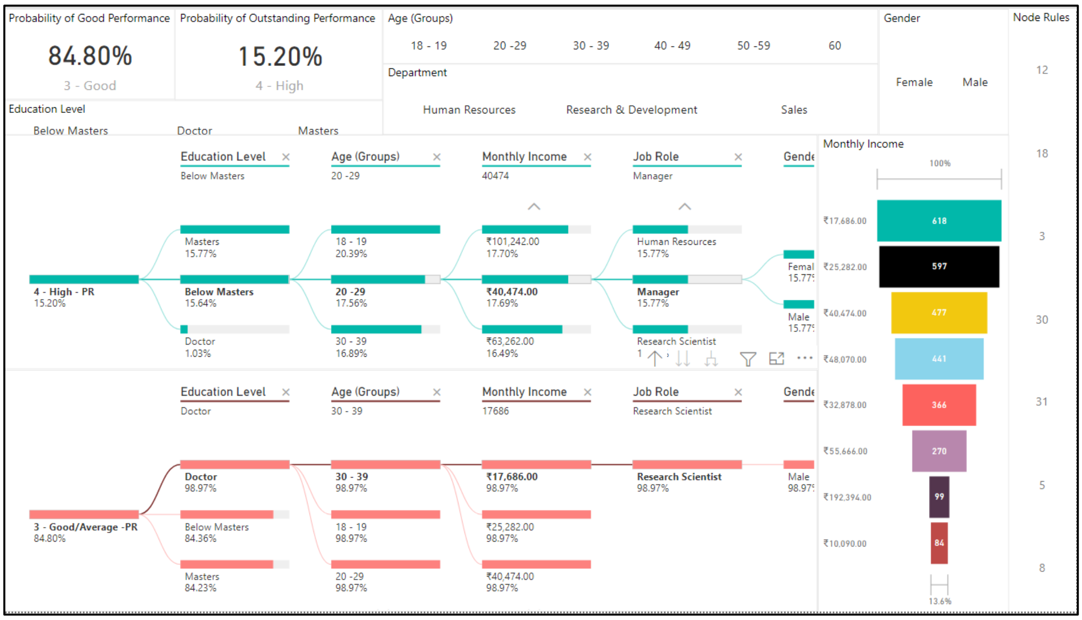

Fig. 10. Performance Rating Dashboard

V. Conclusion

In conclusion, extensive and detailed information has been obtained about the field of human resource management; the role analytics plays in driving insights over the years. HR departments have an extensive amount of responsibilities that need attention including staffing, employee benefits, medical care, training, and more. The application of HR analytics can help HR managers make informed decisions rather than making decisions based on intuition. Solid Analytical data may ensure the correct and authentic application of practices such as performance-paying schemes or provision of compensation that may boost employee performance and job satisfaction. Employee performance and Job Satisfaction hold a high value in the corporate world as these aspects are directly proportional to the well-being and performance of the company or business itself. Moreover, dashboards are an efficient way of displaying and explaining company HR data as they are interactive and less time-consuming. Improvements to the research can include looking into the different models that may fit a larger dataset the best based on the type of dataset and target variables.

Acknowledgment

The authors would like to thank to all School of Computing members who involved in this study. This study was conducted for the purpose of Employee Performance and Job Satisfaction project.

References

Batumalai, G., & Zolkipli, M. F. (2023). Review on the Enhancement of Cyber Security Policies That Impacts the Employment System. *Journal of Applied Technology and Innovation.*

Davidescu, A. A., Apostu, S. A., Paul, A., & Casuneanu, I. (2020). Work Flexibility, Job Satisfaction, and Job Performance Among Romanian Employees—Implications for Sustainable Human Resource Management. *Sustainability*, 12(15), 6086.

Edwards, M. R., & Edwards, K. (2019). *Predictive HR analytics: Mastering the HR metric.* Kogan Page Publishers.

Etukudo, R. (2019). Strategies for using analytics to improve Human Resource Management (Doctoral dissertation, Walden University).

Fabi, B., Lacoursière, R., & Raymond, L. (2015). Impact of high-performance work systems on job satisfaction, organizational commitment, and intention to quit in Canadian organizations. *International Journal of Manpower.*

Joerik van Dooren (2012). HR Analytics in Practice (Master thesis). Retrieved from http://arno.uvt.nl/show.cgi?fid=127276

Lawler III, E. E., Levenson, A., & Boudreau, J. W. (2004). HR metrics and analytics–uses and impacts. *Human Resource Planning Journal*, 27(4), 27-35.

Liu, L., Akkineni, S., Story, P., & Davis, C. (2020, April). Using HR analytics to support managerial decisions: a case study. *Proceedings of the 2020 ACM Southeast Conference*, 168-175.

Keerthi, P. L., & Reddy, P. R. (2019). Adoption Issues of Hr Analytics. *Journal of Business and Management (IOSR-JBM)*, 21(2), 1-7.

Mondore, S., Douthitt, S., & Carson, M. (2011). Maximizing the impact and effectiveness of HR analytics to drive business outcomes. *People* and Strategy, 34(2), 20.

Qureshi, M. A., Thebo, J. A., Rehman, S., Shahbaz, M. S., & Sohu, S. (2020). The Role HR Analytics, Performance Pay and HR Involvement in influencing Job Satisfaction and Firm Performance. *International Journal of Advanced Science and Technology*, 29(11), 382-392.

Sarma, K. S. (2017). Predictive Modeling with SAS enterprise miner: Practical solutions for business applications. SAS Institute.

Tee Zhen Quan. (2022). Salary Prediction in Data Science Field Using Specialized Skills and Job Benefits – A Literature Review. *Journal of Applied Technology and Innovation*, *6*(3), 1–10.

Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards A Standard Process Model for Data Mining. *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, 1, 29-39.

Yamoah, E. E. (2014). Exploratory Analysis of Compensation and Employee Job Satisfaction. *Developing Country Studies*, 4(12), 27-35.