

Employee Attrition Analysis And Visualization:

How well we can retain our talent and better control the employee turnover?

GROUP 14

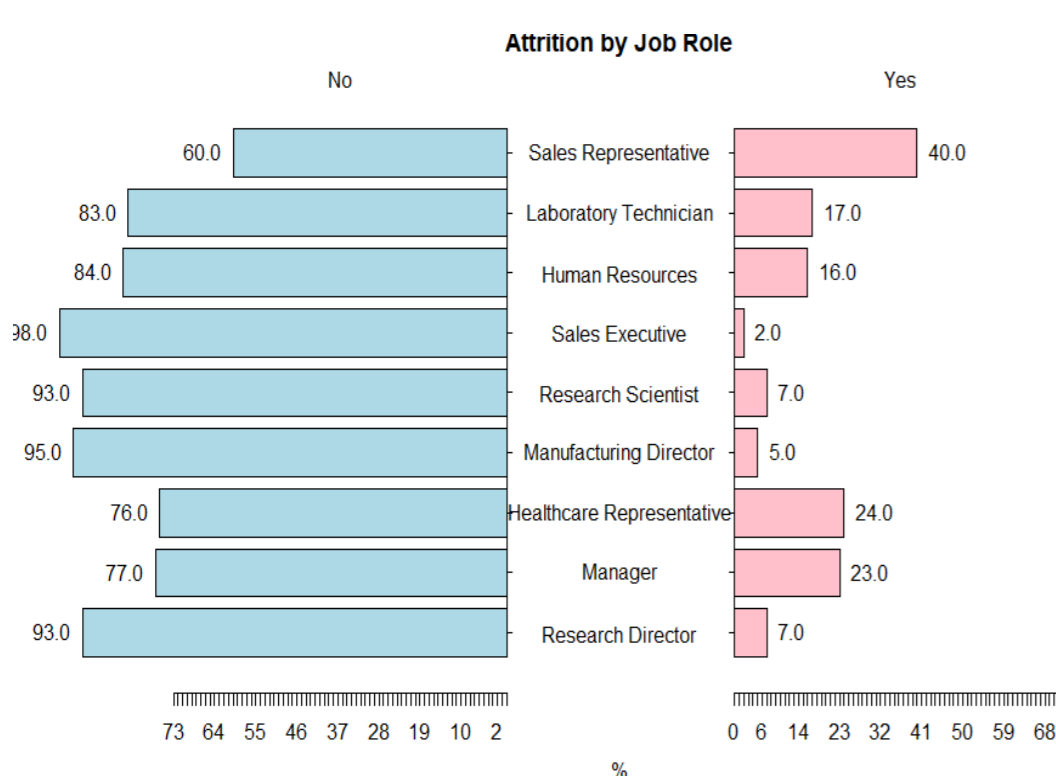
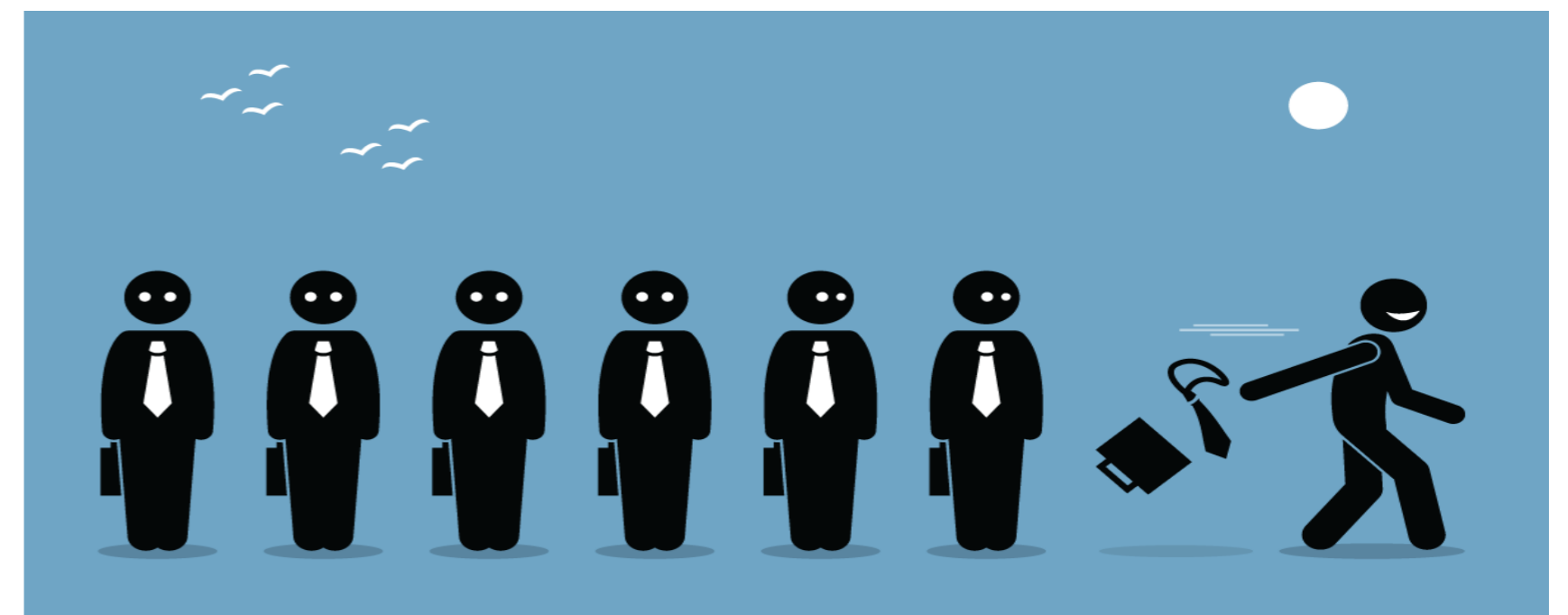
ZHU KEYU
SHI YAWEN
QIU YANG



What HRs are facing is how well they can retain their talent and better control the employee turnover. In the past, employee attrition was like the weather – those paying attention could more or less see it coming. However, it really matters that what will affect the employee and analysis what we can do to reduce the unsatisfying attrition rate in order to achieve higher work efficiency, and therefore, to improve the control of attrition.

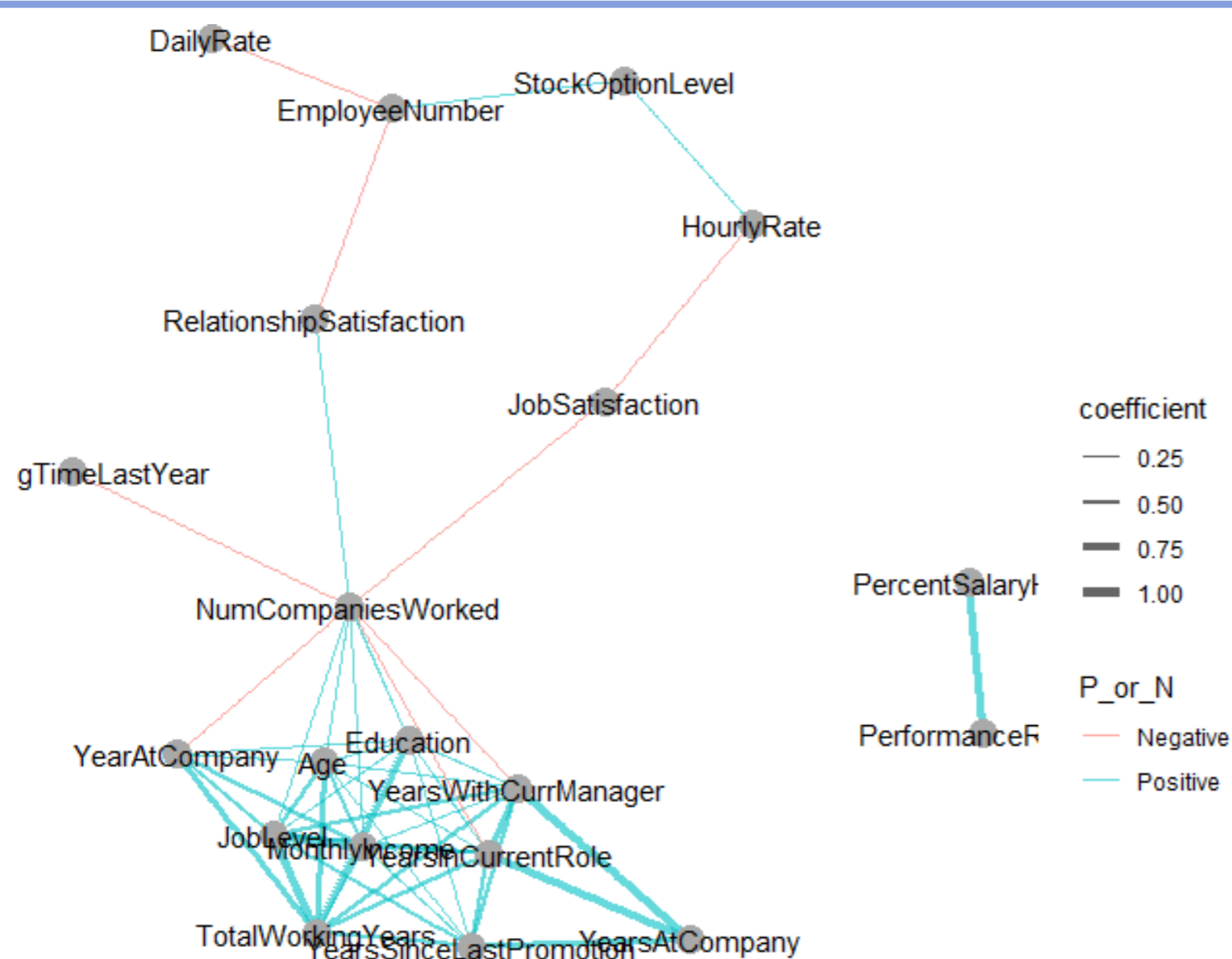
Motivation: We intend to design an application which can provide an interactive, user-friendly platform for visualizing network charts and radar charts, which will enable users to perform the analysis with different selection to see different results. From which can help not only to predict unwanted attrition, but to have proven action plans at your fingertips to help you reduce it, based on organization's unique attributes.

Design Framework: We aim to provide an application for Human Resource Department to help them to manage employee attrition. Our application consists of five major modules namely: Data Search, About Employee, Explore Features, User Portrait and Model Results, depicting different steps of the Attrition analysis to the our users.



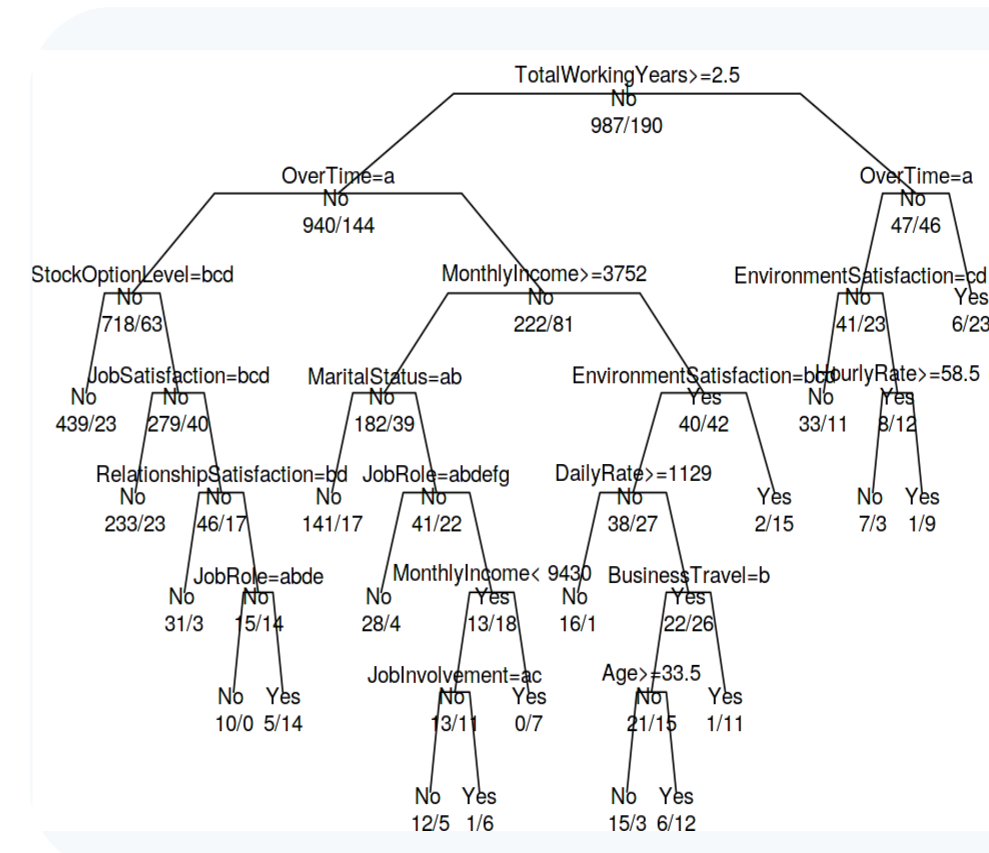
The Data Search Function is developed to help filter the raw employee data according to the specific features. Secondly, the EDA Analytics is used to preliminarily understand the pairwise relationship between two features and their effects on attrition. Network Graph is applied to understand the structural characteristic of the features as a whole and to select features.

Three methods: Decision Tree, Random Forest and XGB are utilized for building the attrition analytics model and predict the probabilities of employee attrition. Lastly, the User Portrait Analysis is applied to understand the differences among groups more intuitively based on the most important features that are generated from the model result. In short, users will have a clearer understanding of the distribution of employee attrition in different dimensions and get hints of how to improve their attrition management.

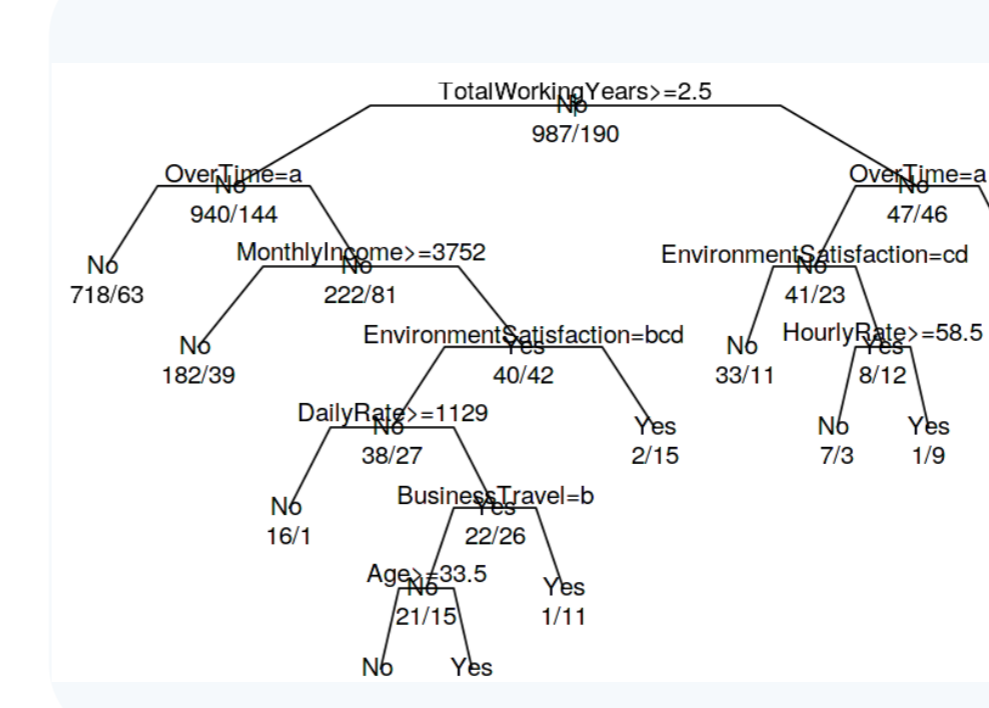


Network Analytics was generated by all features as nodes and their pairwise coefficient as link weight to get further interpretation of the pairwise relationship between two objects at a time and structural characteristic of the features as a whole. Library "ggraph" is applied to present the network graph of features. What can be interpreted is that "Monthly Income", "Age", "Job Level" and "Total Working Years" are the most "correlated" elements that affect whether or not an employee would leave the company. In the contrast, we can also learn that "Performance Rating" and "Percent Salary Hike" are in the isolated pair that only related with each other, but with relatively high correlation coefficient.

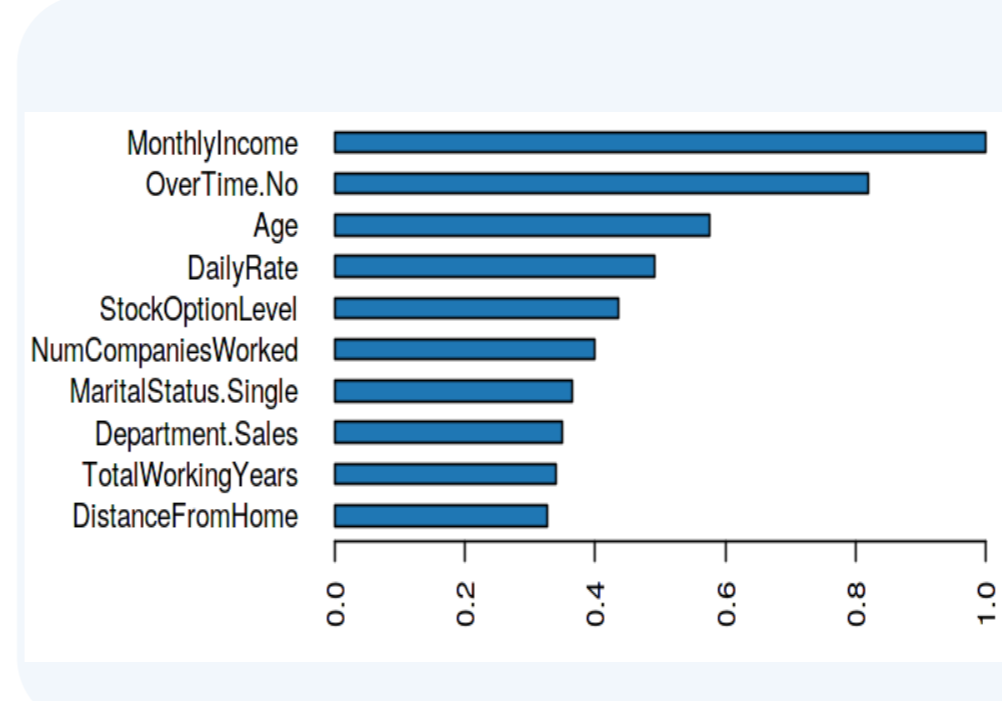
Decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences. The more often the features are chosen to split the tree, the more important the features are. From the feature importance result, two features MonthlyIncome and OverTime are far significantly more important than the others.



Random Forest is built by a collection of decision trees. However, the performance of Random Forest model is worse than the Decision and the top ten features importance of RF are strongly different from decision tree. It might be because of too many features to cause the overfitting.

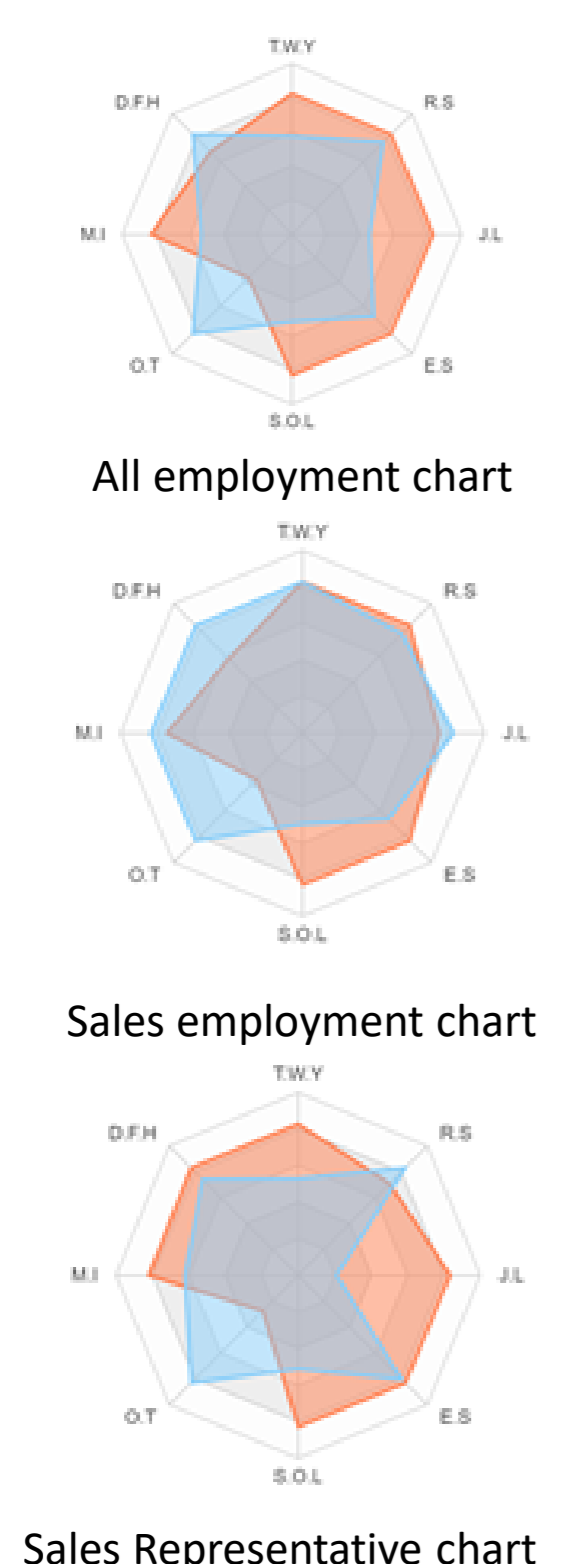


XGBoost is a scalable and improved version of the gradient boosting algorithm designed for efficacy, computational speed and model performance. Combining three performance results, the XGBoost model has the highest accuracy and AUC score. Therefore, this model is the best one.



Persona Profile analysis: A radar chart Library "rechart" was perfect for deploying with react components and presentation.

By selecting out three charts constituted by different employees, we can reveal that working overtime has a significant importance on total employees' attrition decision, and people with higher total working years, relation and environment satisfaction are less likely to leave the job. By contrasting the "yes" attrition of sales executives and sales representatives, average total working years for sales representatives are less than that of executives; moreover, executives have higher salaries and nearer distances from home still, cannot detain their attrition by comparing in the same group of "no" attrition.



Future Work: The application requires some advancements to improve its efficiency. The very first one is to allow users to load the dataset of his/her interest to enable different features. Also, we can develop a new function to predict the attrition, the most effective retention factors to retain valuable employees by typing in specific decision information. In the meantime, users will be granted to choose the prediction algorithm and not depend only on specific ones. Also, others can be provided such as changing the option to select the threshold values for the attributes while deciding on valuable employees and even retention factors. Further expansion of visualization can be added such as adding interactions for all the plots and graphs. To improve the performance of the dashboard would help our subscribers to improve user experience.