



PREDICTING EMPLOYEE ATTRITION

Group B1

Lee Xin Qi

Leona Ang Qiao En

Leong Jia Wei, Marcus

Ng Mei Ting

Motivation & Problem Statement

Definition



Employee attrition is a process in which a number of people retire or resign and are not replaced

Motivation



Employee Attrition is the biggest talent-related challenge facing Singapore Sci & Tech industry



Alleviate economic and social costs

Problem Statement



Minimise employee attrition






Better understand the important variables a company looks at for valuable employee retention



Goal & Hypothesis

Goal

-  Predict attrition, with accuracy, sensitivity, specificity and area under curve (AUC) values more than 0.80
-  Identify the significant variables influencing attrition
-  Provide recommendations on how a company can remedy the situation

Hypothesis



JobSatisfaction is significant in the prediction of attrition



Data Cleaning & Data Exploration

Data Cleaning



No null values



Removal of variables:

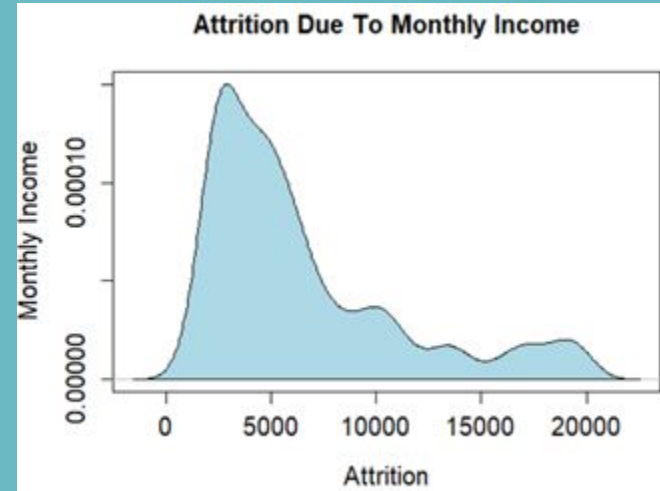
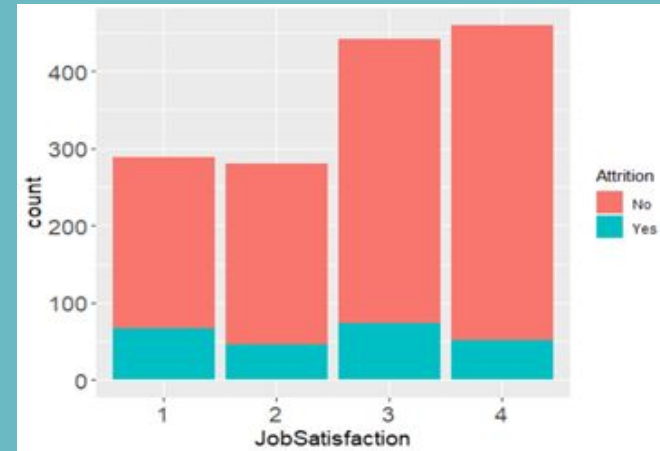
- Lack of variability - EmployeeCount, Over18, StandardHours
- Trivial identification - Employee Number

Data Exploration



Cleaned Dataset: 1470 observations, 31 variables

- About 16.1% Attrition (Imbalanced data)
- Higher level of job satisfaction, Lower Attrition
- Higher monthly income, Lower Attrition



Synthetic Minority Oversampling Technique (SMOTE)

- Dataset was split equally into training and test sets (735 observations each)
- Imbalance in training dataset - 606 Non-Attrition and 129 Attrition
- Performed SMOTE and up-sampled the training set
- Final training dataset - 903 observations with 516 Non-Attrition and 387 Attrition



Logistic Regression Model

Logistic Model was fitted onto training dataset (with all variables) after SMOTE

Logistic Model obtained was used to predict Attrition in test set



Not all the variables in the model were significant - Variable Selection and more needs to be done



Logistic Regression with Lasso

- 📌 Lasso performs both variable selection and regularization (shrinks coefficients of non-significant predictors to 0)
- 📌 The best lambda parameter for Lasso was tuned using cross-validation
- 📌 Logistic Model with Lasso was used to predict Attrition in test set

Accuracy

0.808

AUC

0.819

Sensitivity

0.704

Specificity

0.826

- 📌 Variables such as OverTime are deemed significant - large coefficients

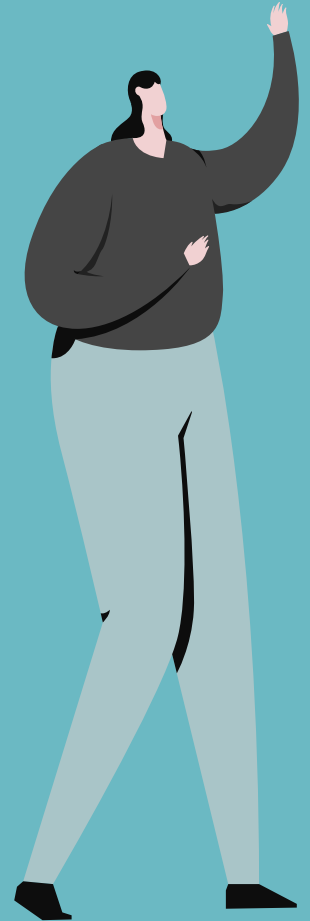
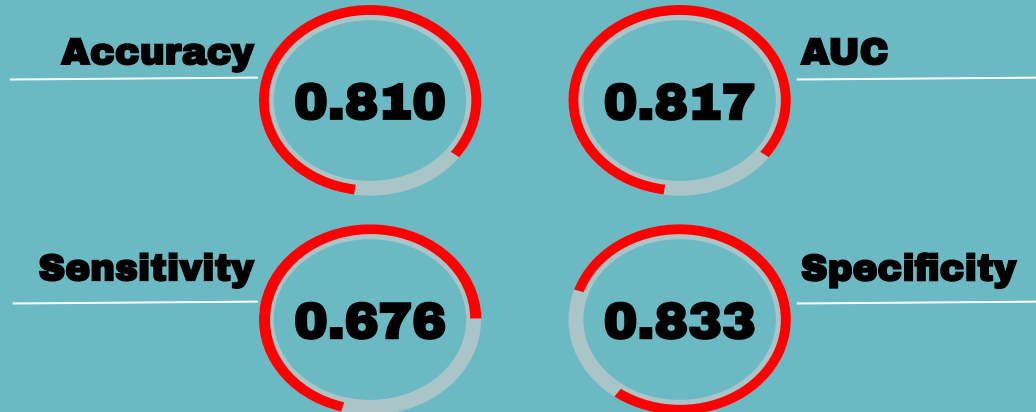


SVM Linear Kernel

Support Vector Machine (SVM) was fitted onto training dataset (with all variables) after SMOTE

Linear, Radial and Polynomial kernels were used and the best parameters for each were tuned using cross-validation

The metrics of radial and polynomial kernels are poor relative to SVM with linear kernel. The metrics for the linear kernel are as shown below.



Single Decision Tree

 Serves as a benchmark for Decision Tree Ensembles

Accuracy

0.661

AUC

0.693

Sensitivity

0.704

Specificity

0.654

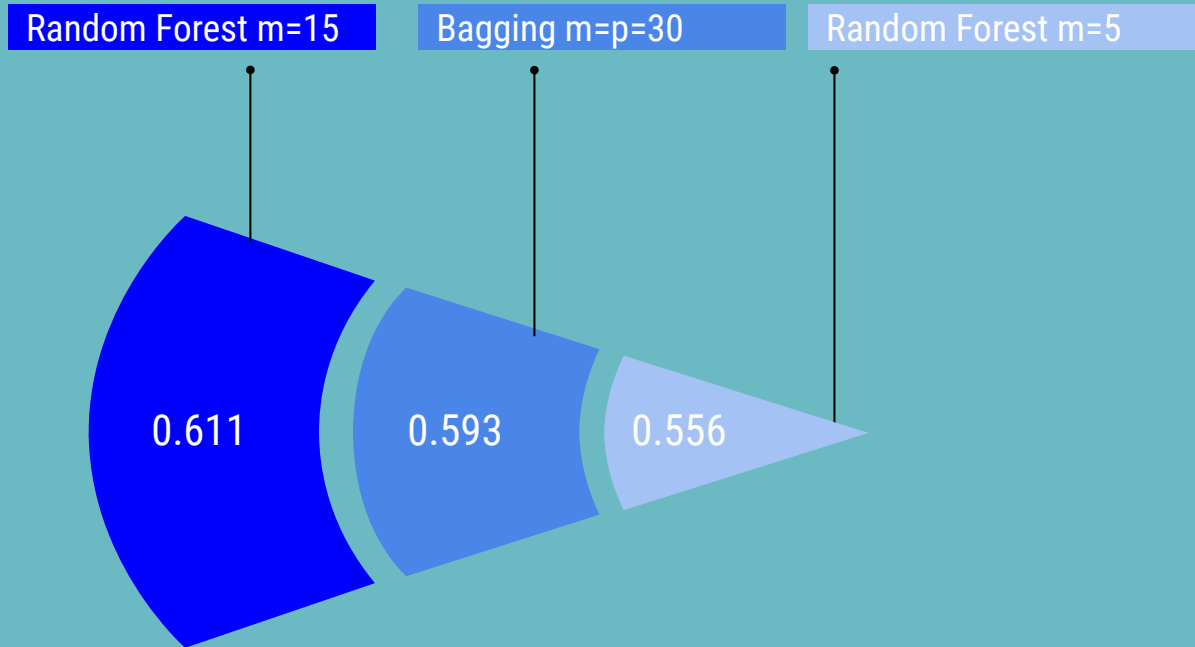
 Top 5 variables

1	Overtime
2	JobRole
3	TrainingTimesLastYear
4	JobSatisfaction
5	YearsAtCompany

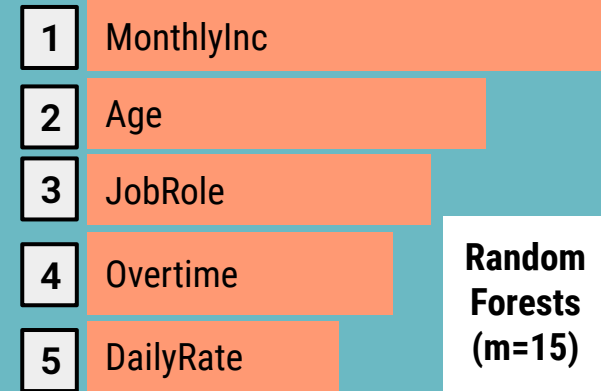


Bagging & Random Forest

- Performs better than Single Decision Tree, except on sensitivity
- Random Forests models perform slightly better than Bagging model only on sensitivity metric as shown below



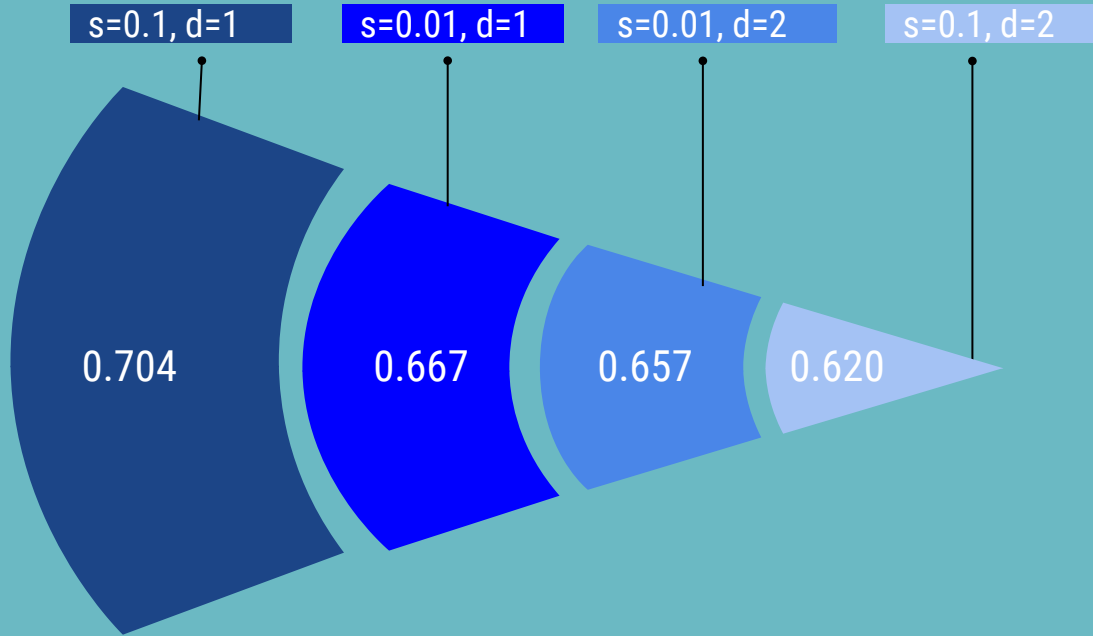
Variable Importance



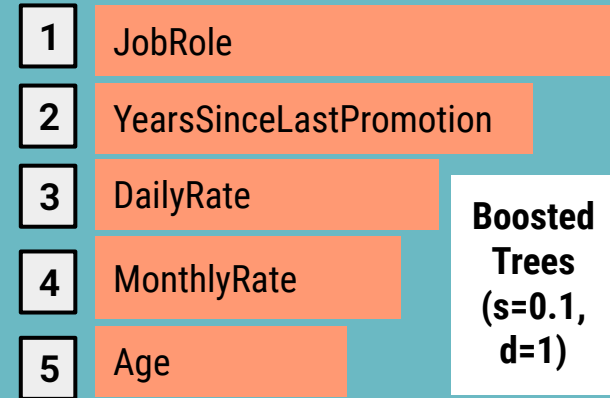
Boosting

Performs better than Single Decision Tree for all metrics

Model with $s=0.1$ and $d=1$ performs the best



Variable Importance



Comparison of models

 None of the statistical models achieved 0.800 and above for all evaluation metrics

 Logistic Regression with Lasso & Gradient Boosted Trees have the highest metric values

Comparison	Logistic Regression With Lasso	Gradient Boosted decision tree
Works well for simple decision boundary	✓	✓
Less prone to overfitting	✓	✗
Better interpretability	✓	✗



Conclusion



FINAL PREDICTION MODEL :



Discussion & Recommendations

Based on Logistic Regression with Lasso Model

Significant Variables



OverTime: More OverTime, More attrition

→ Provide more support - Increase manpower to lower need to overtime



Job Involvement: Less JobInvolvement, More attrition

→ Profiling - draw out strengths and needs of employees

→ Engage employees with suitable work activities - help them take more ownership

Hypothesis Rejected



→ JobSatisfaction is not significant in predicting attrition

→ Contrary to existing subject knowledge

→ Participants may not being truthful in the survey





THANK YOU!

That's it from GROUP B1!



Questions?