



EMPLOYEE ATTRITION

*Rani Misra and
Cheyenne Airington*





01. **INTRODUCTION**
Overview + research question

02. **FEATURE SELECTION**
Cleaning process + feature selection

03. **MODEL DEVELOPMENT**
Training + model selection

04. **CONCLUSION**
Results + recommendation



TABLE OF CONTENTS

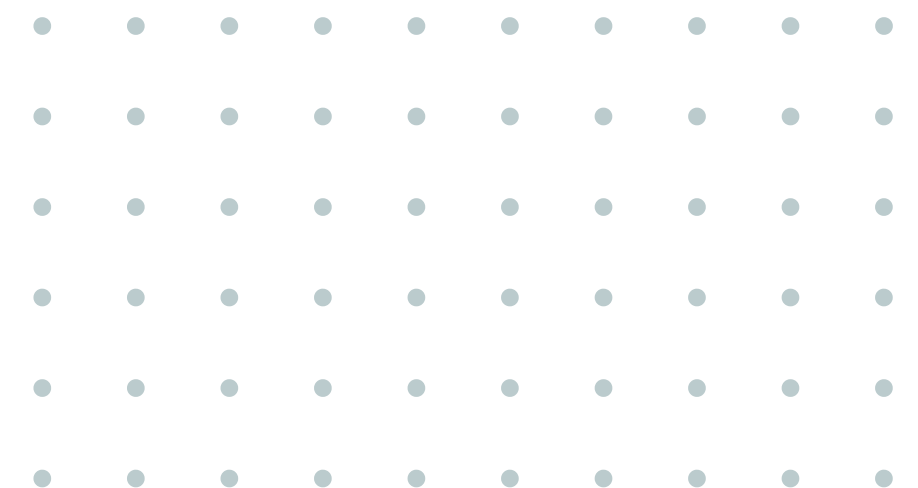


QUESTION

- Can we accurately predict if an employee is likely to terminate?

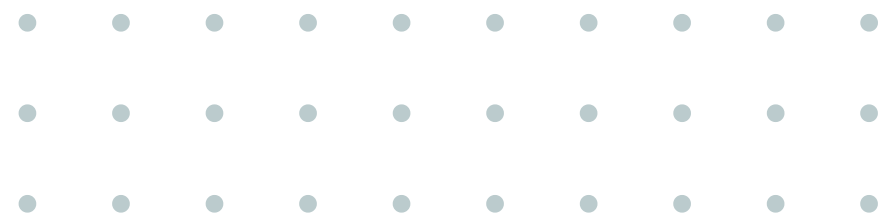
SOLUTION

- Train various classification models to predict a “Yes” and “No” attrition class
- Choose model based on accuracy and other evaluation metrics



DATA OVERVIEW

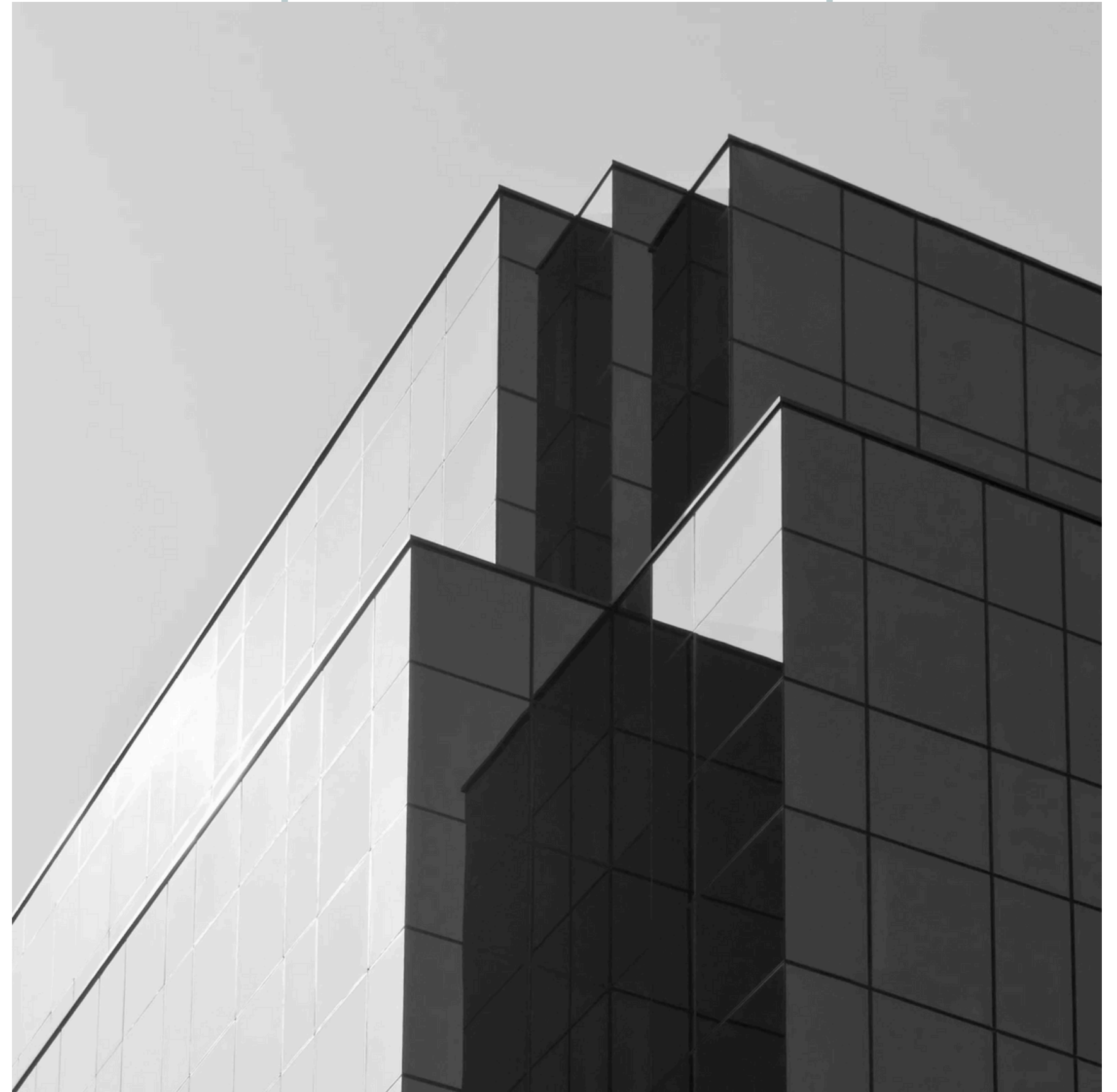
- Employee Attrition dataset (Kaggle)
- 35 fields with a combination of categorical, continuous, and distinct data
- Key data regarding employees like time in role, education, department, etc.



DATA CLEANING

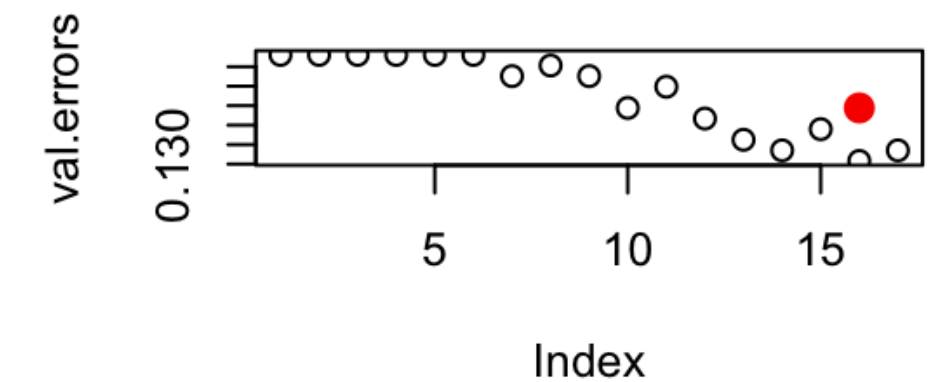
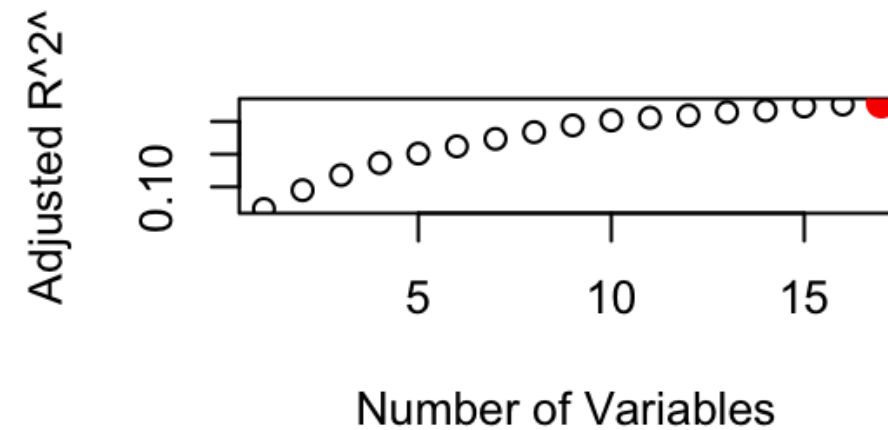
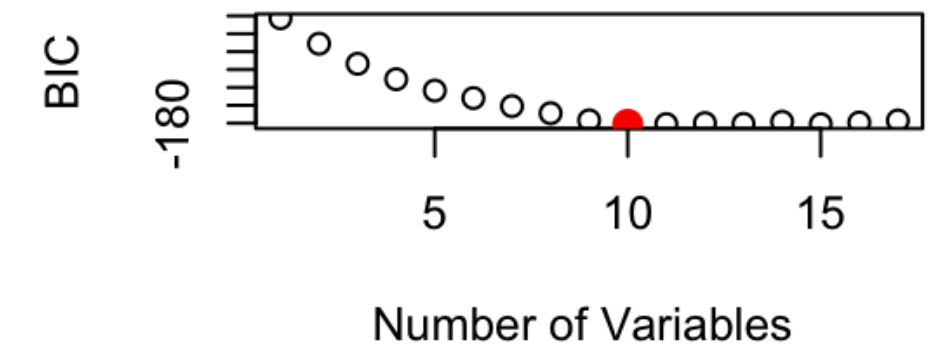
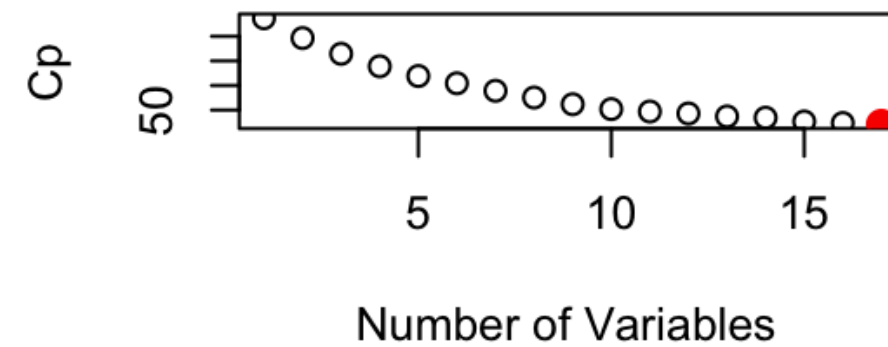
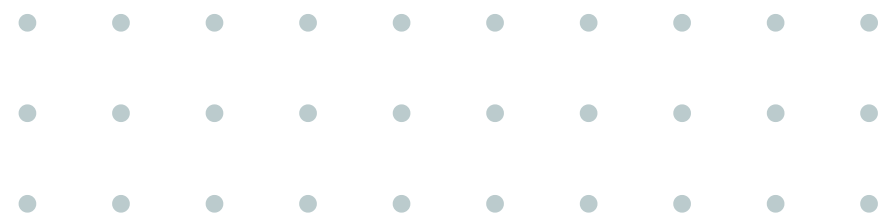
To reduce redundancy we:

- Removed constant fields, "EmployeeCount", "Over18", and "StandardHours"
- Removed "EmployeeNumber"
- Checked for nulls (none were found)



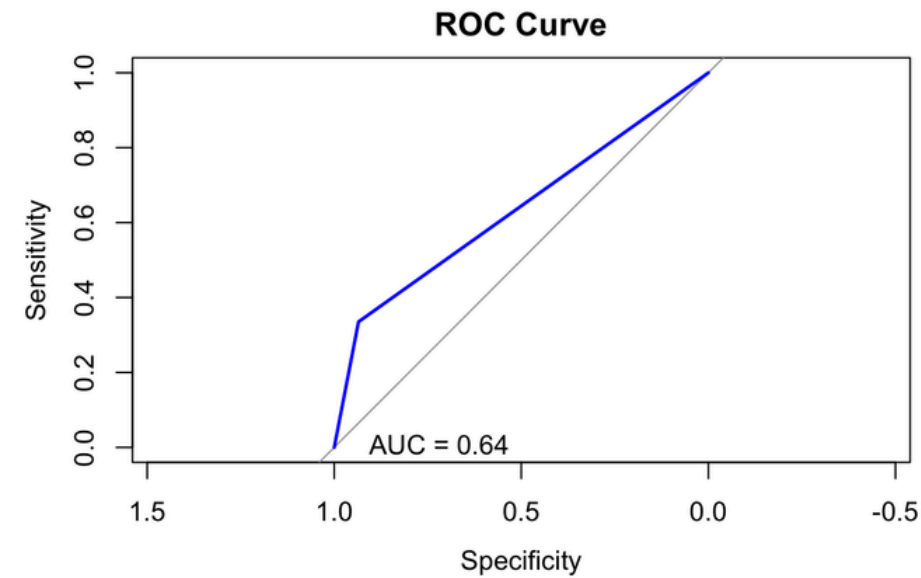
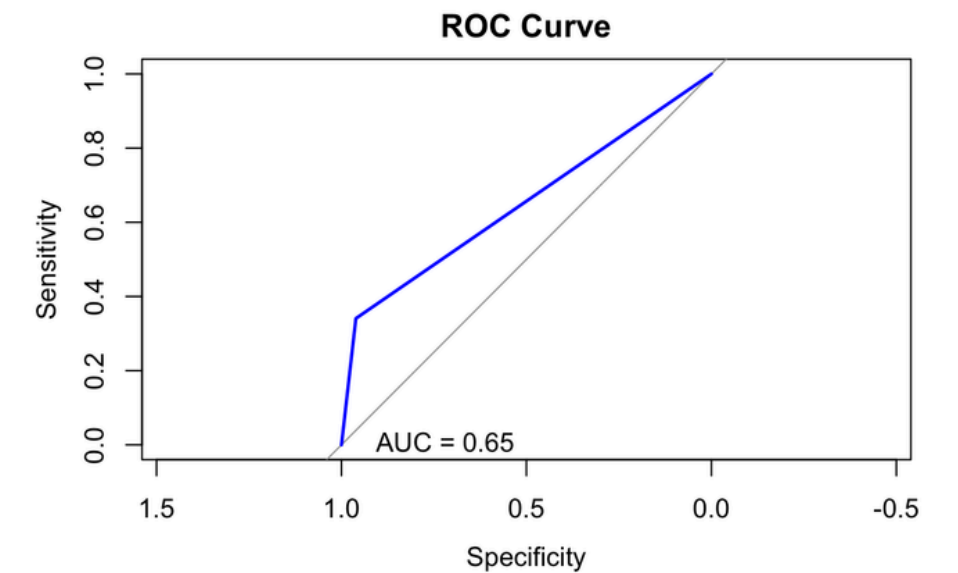
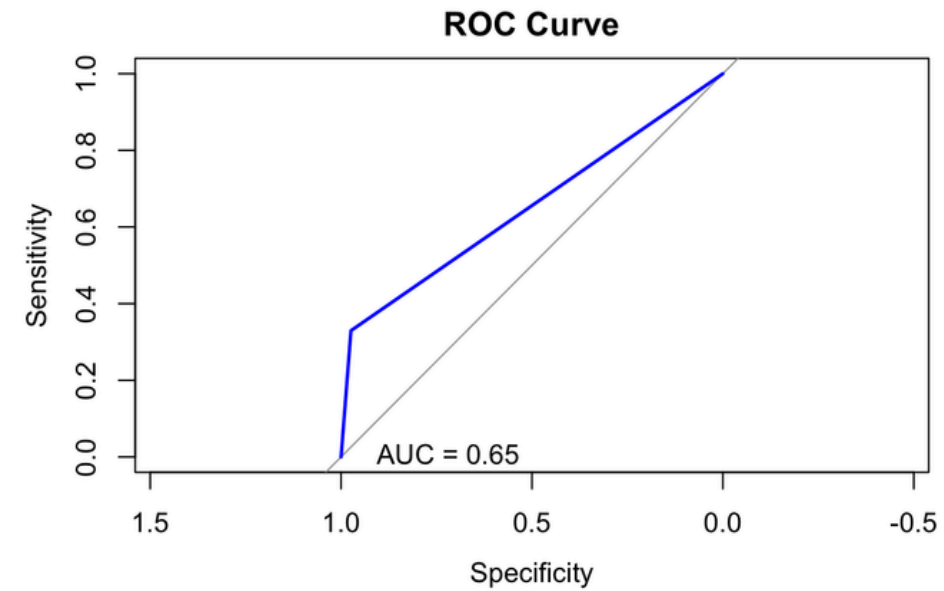
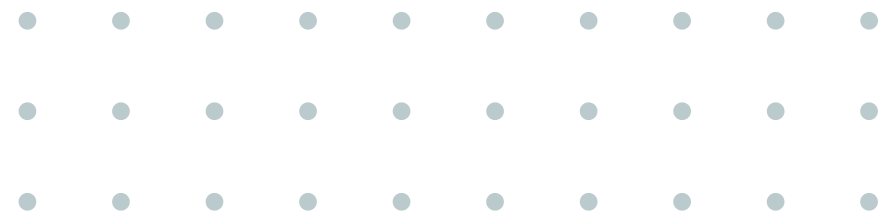
FEATURE SELECTION

- Used a forward selection process
- 10 features selected based on minimum BIC
 - Environment Satisfaction, Job Involvement, Job Satisfaction, # Companies Worked, Total Working Years, Work Life Balance
- Some features are dummy versions of original
 - Traveling Frequently, Being a Sales Representative, Single, Working Overtime



TRAINING

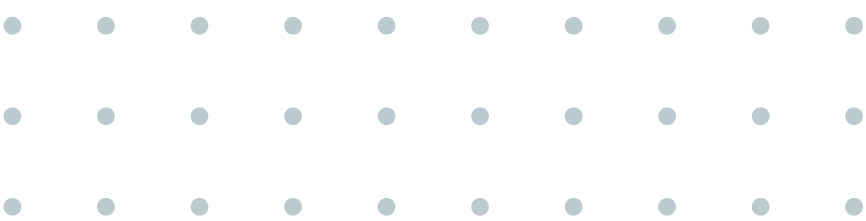
- Logistic Regression, LDA, and QDA training models tested.
- Accuracy, Recall, and ROC/AUC used as metrics for model selection.



MODEL SELECTION

Using the metrics and the overall significance of each model's ability to predict the data, the logistic model is chosen.

Model<chr>	Accuracy<dbl>	Recall<dbl>	AUC_Value<dbl>	P_Value<dbl>
LOG	0.8694470	0.9740260	0.6518175	0.001939485
LDA	0.8603808	0.9610390	0.6509105	0.021153476
QDA	0.8377153	0.9350649	0.6351302	0.519942645

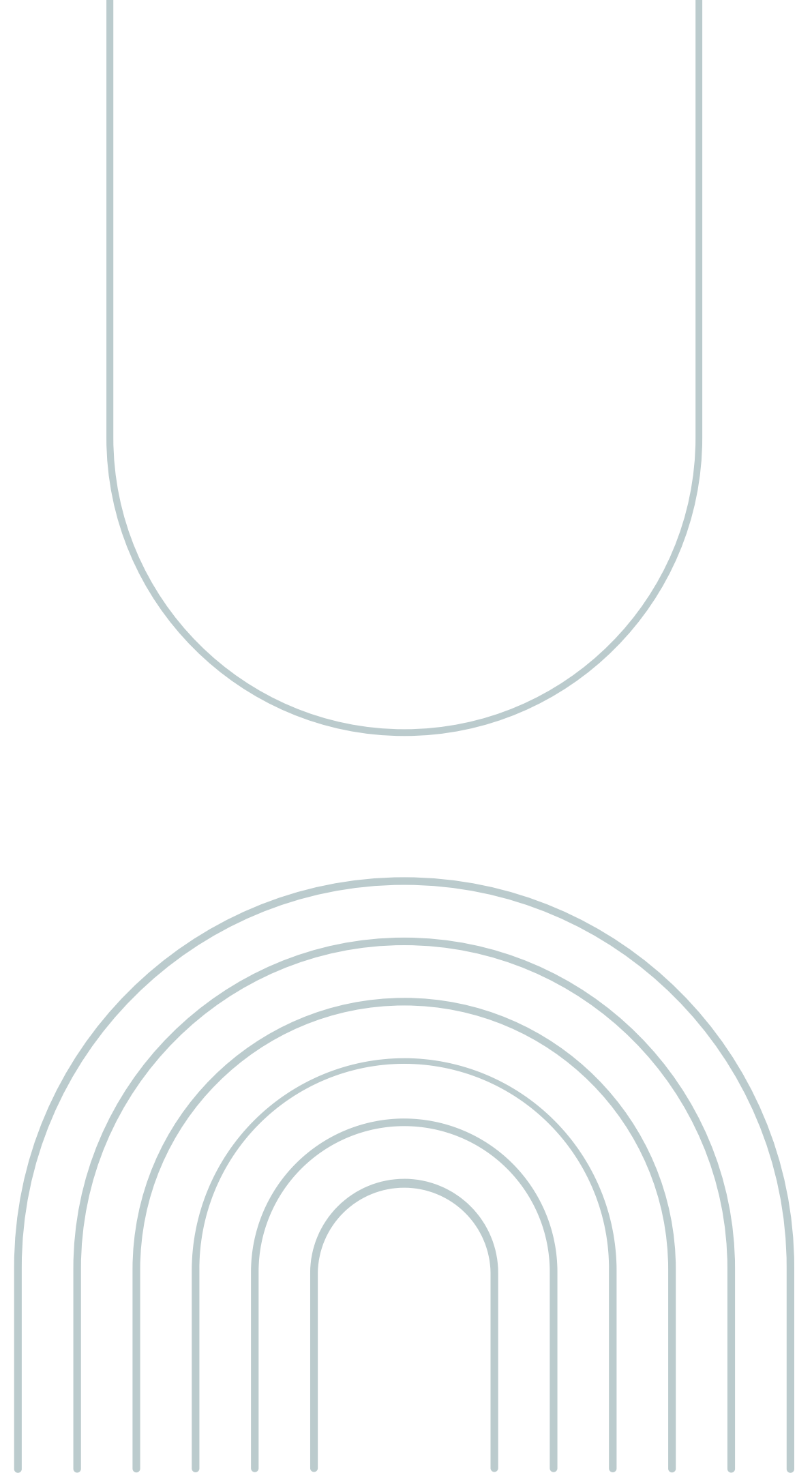


RESULTS

Logistic model provides almost 87% accuracy, a 97% recall rate, and 65% performance in the classification model.

RECOMMENDATION

- Conduct studies to find how to reduce significance of variables that are more likely to lead employees to attrition.
- Conduct further studies after implementation of new business practices to see if attrition is actually lowered.



THANK YOU