Broad College of Business
MICHIGAN STATE UNIVERSITY

Employee Retention HR Data

GROUP 9

WHO WILL MAKE BUSINESS HAPPEN?
SPARTANS WILL.

HR Analytics Job Prediction

- **Aim of data set:** To understand the factors contributing to employee behavior and predict the likelihood of an employee leaving the company
- Data set compiled by the HR department of a company to predict employees' behavior.
- This dataset contains information about employees who worked in a company.

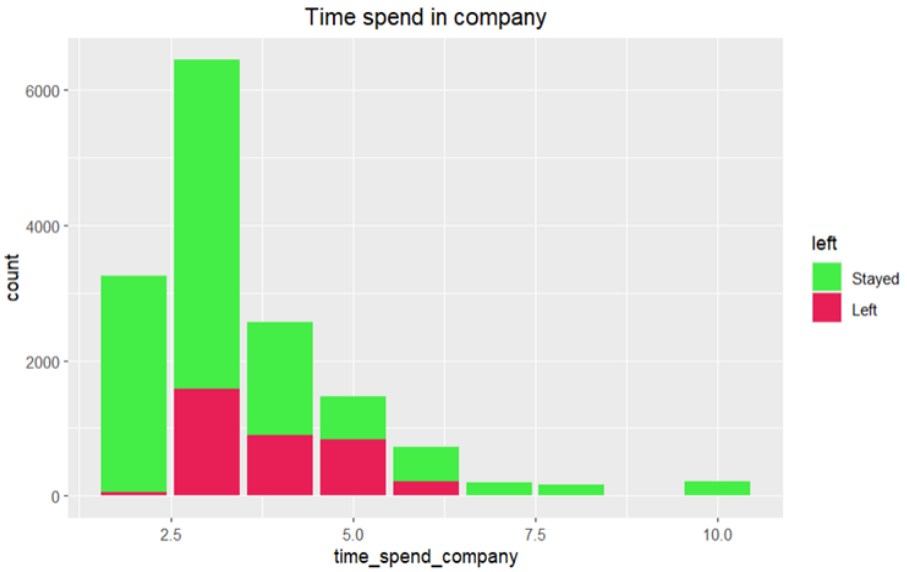
Data Description

Name	Type	Sample	Description
satisfaction_level	Quantitative	Continuous	Satisfaction level based on survey
last_evaluation	Quantitative	Continuous	Last evaluation score
number_project	Quantitative	Continuous	Number of projects they have worked on during the last 1 year
average_montly_hours	Quantitative	Continuous	Average monthly hours worked
time_spend_company	Quantitative	Continuous	Time spent in the company
Work_accident	Categorical / Dummy	0 or 1	Any accidents or mistakes done at work
promotion_last_5years	Categorical / Dummy	0 or 1	Promotion received in the last 5 years
Department	Categorical	Sales, HR, IT etc.	Department the employee belongs too
salary	Categorical	“low”,” medium” or “high”	Salary binned into low, medium and high brackets
left	Categorical / Dummy	0 or 1	Employee left or stayed in the company

Broad College of Business
MICHIGAN STATE UNIVERSITY

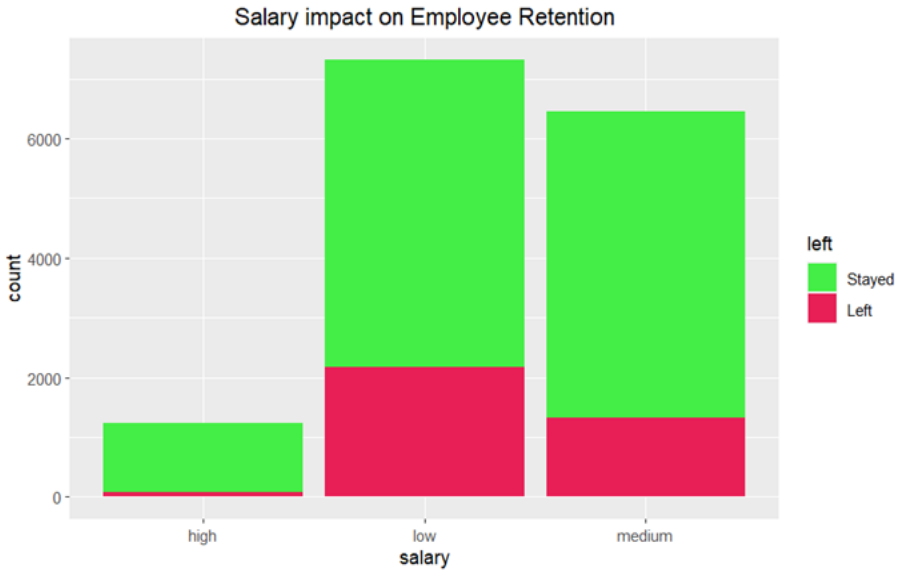
Exploratory Data Analysis

Comparing Time Spent by the employees and attrition rates



- The majority of those who left were employed for between 2.5 and 6 years

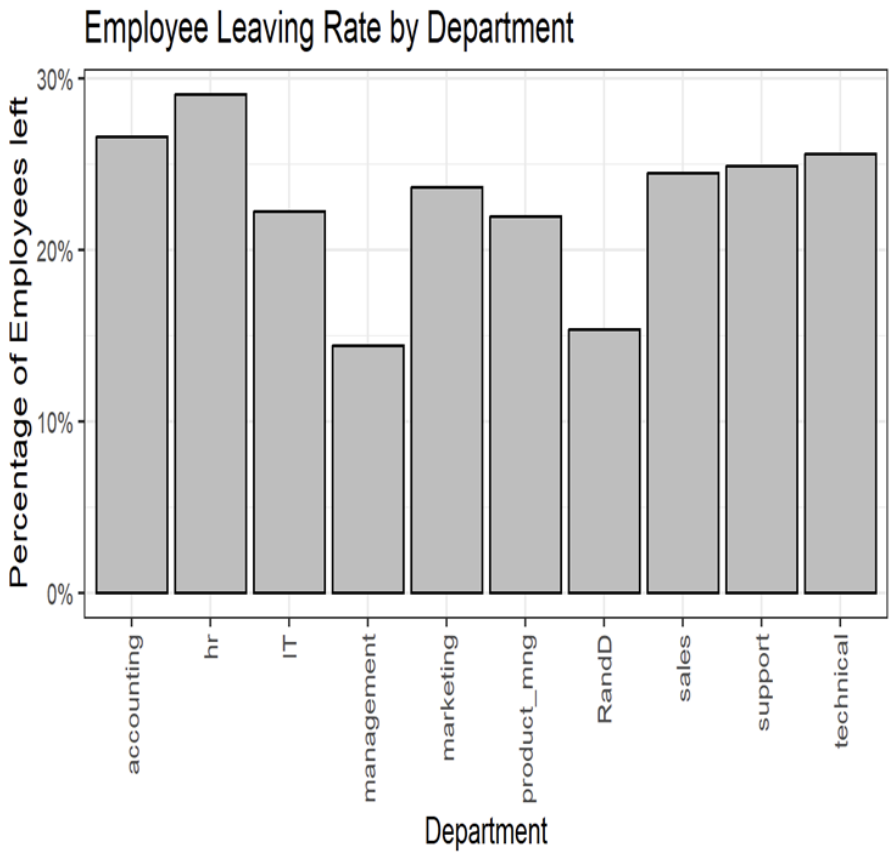
Comparing Salary of the employees and attrition rates.



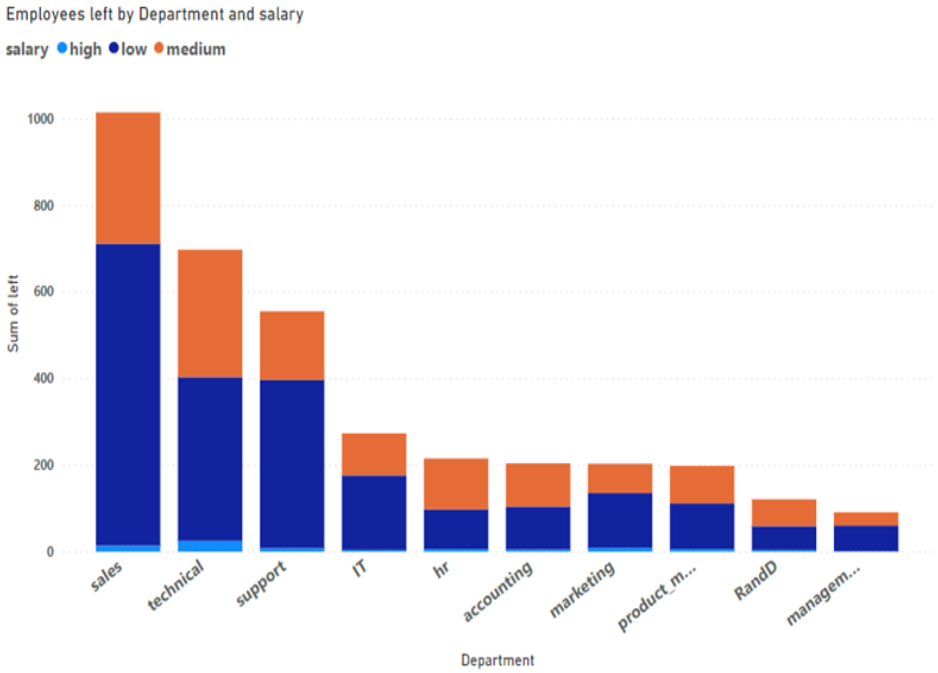
- Compared to the few of employees who were paid high salaries, the majority of those who left the company were paid relatively low salaries.

Broad College of Business
MICHIGAN STATE UNIVERSITY

Attrition rates differ within various departments

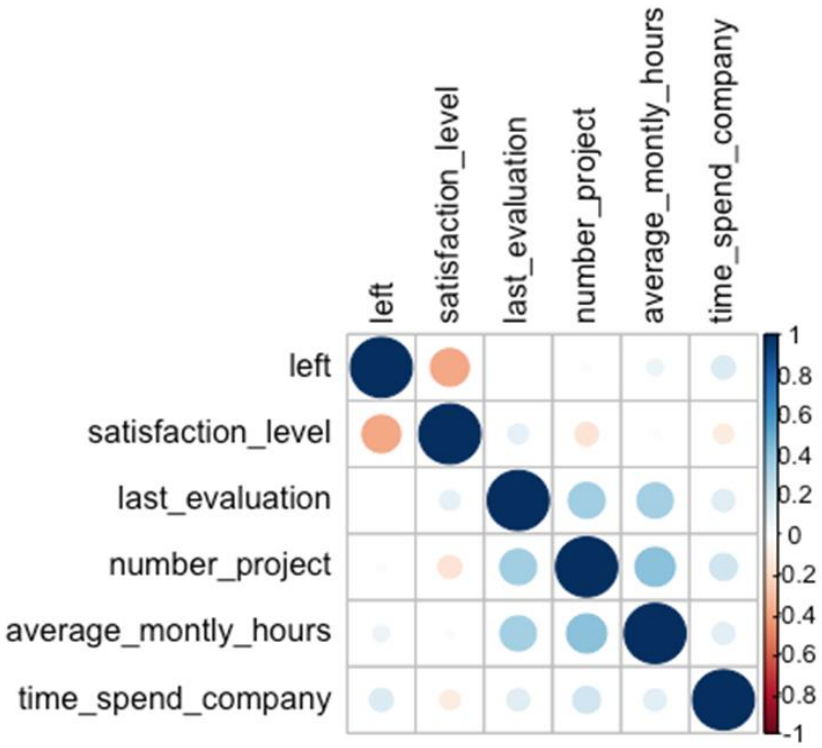


Attrition rates within various departments and salary levels

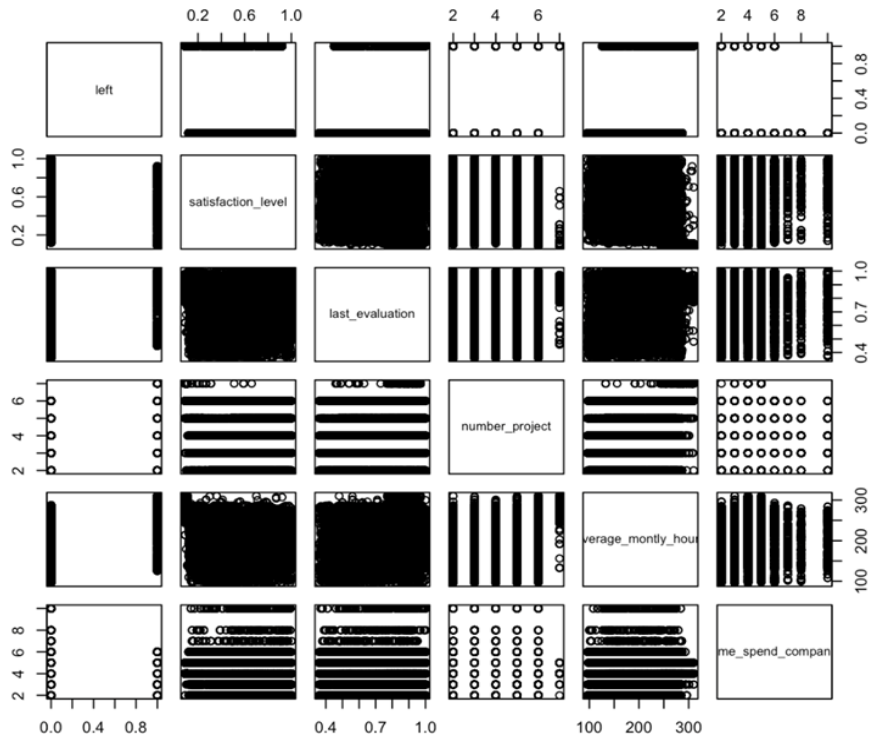


Broad College of Business
MICHIGAN STATE UNIVERSITY

Correlogram



Scatter Plot Matrix



Models

- **Logistic Regression**
- Attrition rates vs Salary
- Attrition rates vs Department, Time Spent
- Attrition rates vs Satisfaction level, Salary
- Attrition rates vs Department, Salary
- Attrition rates vs Promotion
- Attrition rates vs All Features

- **Decision Trees**
- Decision Tree model with 5 variables
- Decision Tree Full model

Models for Interpretation

Logistic Regression

Model 1
glm(left ~ promotion_last_5year)

INTERPRETATION

Smaller p-value (<0.05)

Odds of leaving the company	
Promoted	X
Not Promoted	5X

Model 2
glm(left ~ salary)

INTERPRETATION

Smaller p-value (<0.05)

Reference Group	Odds of leaving the company	
High salary	Low salary	5X
	Medium salary	3X

Model 3
glm(left ~ Department + salary)

INTERPRETATION

Control Variable	Reference Group	Odds of leaving the company	
Department	High salary	Low salary	6X
		Medium salary	3.5X

Control Variable	Reference Group	Department	p-value
Salary	Accounting Department	Management	<0.05
		R&D	<0.05

Model 4
glm(left ~ Department + time_spend_company)

INTERPRETATION

Control Variable	Odds of leaving the company	
Department	Year	X
	Year + 1	1.27X

Control Variable	Reference Group	Department	p-value
Salary	Accounting Department	Management	<0.05
		R&D	<0.05
		Product_mgmt	<0.05
		IT	<0.05

Models for Prediction

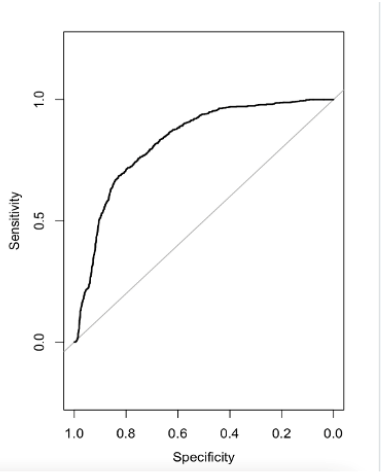
Logistic Regression Model with all variables

Part-1

- Logistic model with all independent variables

Part-2

- Used threshold probability value for prediction



AUC	0.827
Accuracy	0.795
Sensitivity	0.685
Specificity	0.829

Logistic Regression Model with selected variables

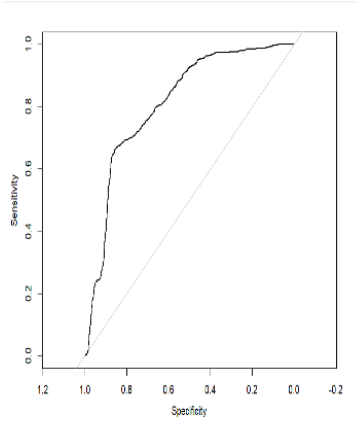
Part-1

- Logistic model with selected variables using “glmulti" package in R (8 selected variables)

Part-2

- Used threshold probability value for prediction (value = 0.34)

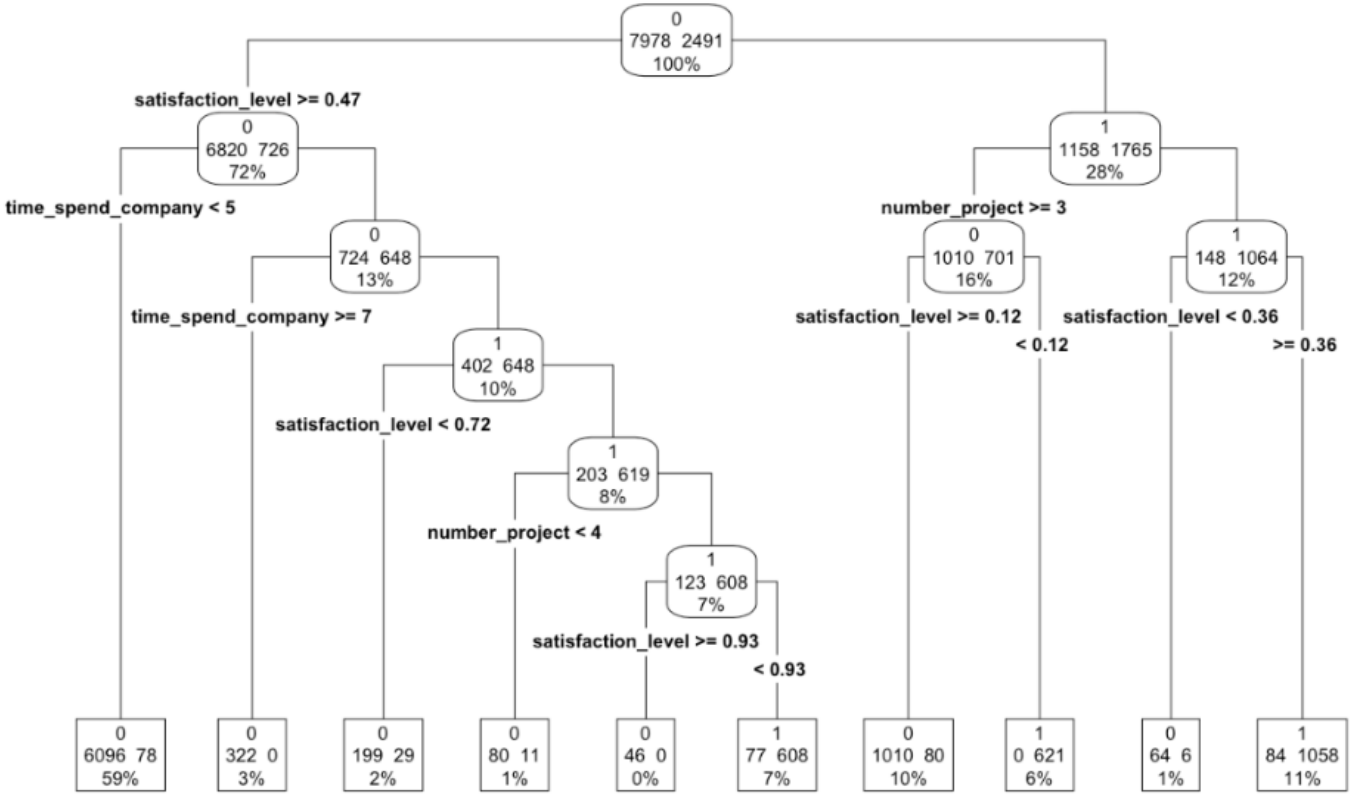
AUC	0.810
Accuracy	0.765
Sensitivity	0.663
Specificity	0.856



Decision Trees

Tree model with 5 variables:

satisfaction_level, number_project, time_spend_company, promotion_last_5years, salary



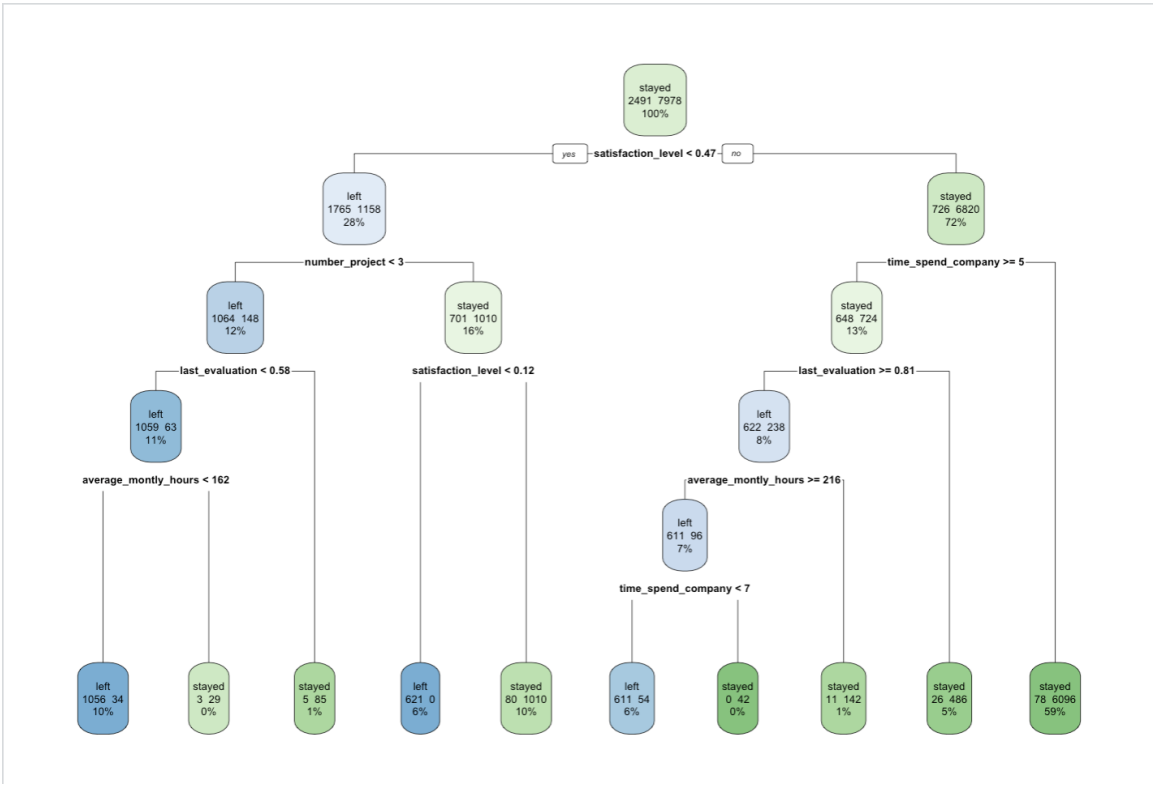
Strategies for Employee Retention

Firms can retain employees when,

1) The satisfaction levels of recently joined hires (<5 years) are greater than or equal to 0.47.
2) For Unsatisfied employees, if their number of projects are (>=3) and satisfaction levels should be at least 0.12.

Attrition rates are high when,

1) Unsatisfied employees work on less than 3 projects, and are last evaluated less than 0.58, with monthly average hours spent is less than 162 hours

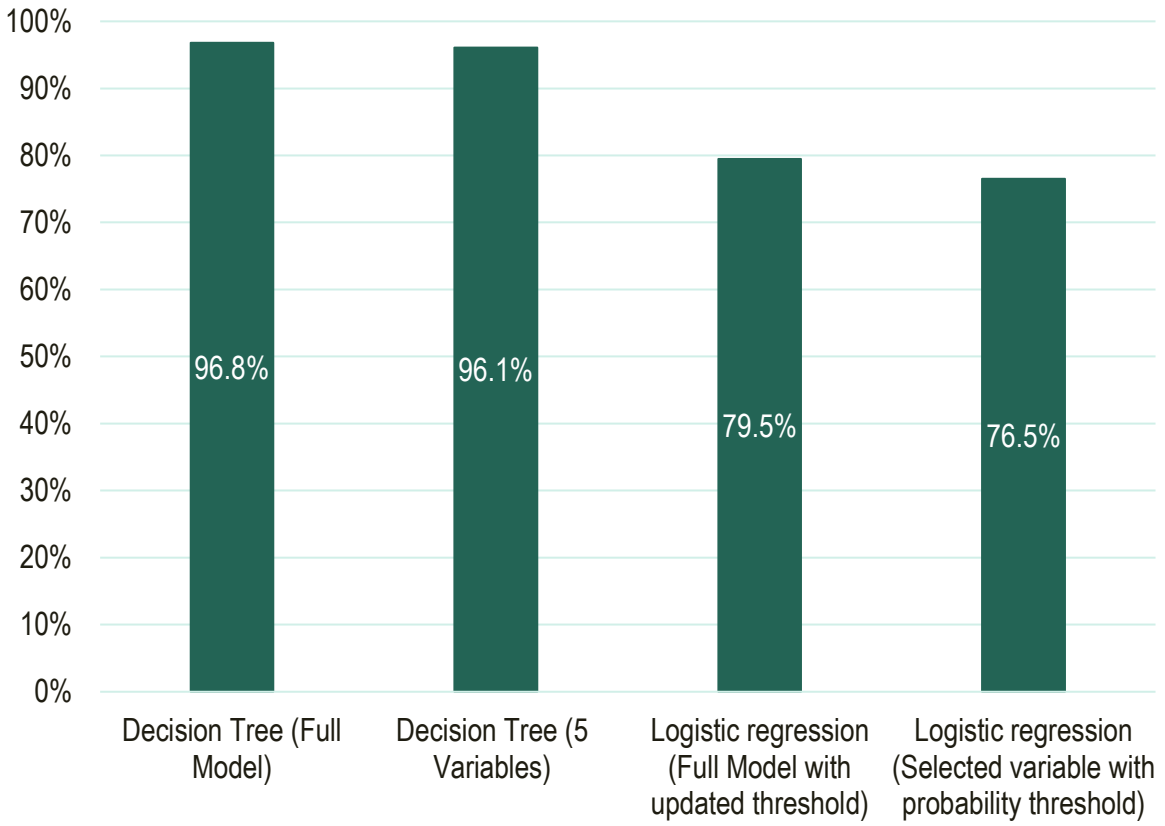


Accuracy : 96.8%

Model Accuracy

The chart illustrates that the decision tree model with all variables included has the highest accuracy rate of

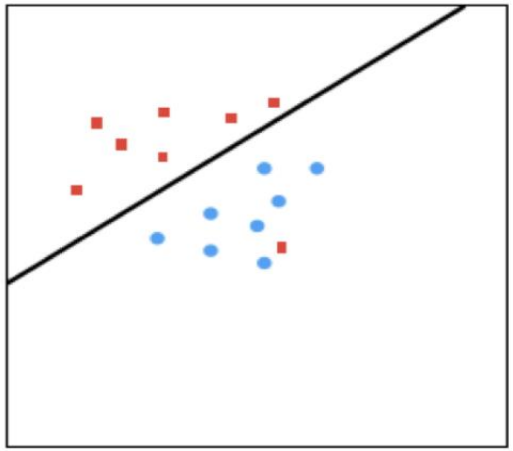
96.8%

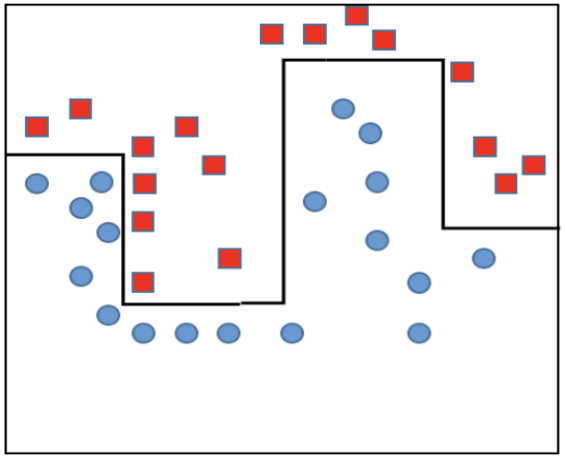


Broad College of Business
MICHIGAN STATE UNIVERSITY

Conclusion

- Choice between logistic regression and decision tree depends on nature of problem and data characteristics.
- Logistic regression is suitable for linearly separable data and binary classification problems.
- Decision tree is suitable for nonlinear data and can handle both categorical and numerical data.





- By comparing the models, we can see that the decision tree with all variables has the highest accuracy. Thus, can we can predict which employee will stay or leave the firm.

Broad College of Business
MICHIGAN STATE UNIVERSITY

THANK YOU

Barath Kurien Alapatt (alapattb@msu.edu)

Divisha Jain (jaindiv2@msu.edu)

Monika Sureshkumar (sureshk2@msu.edu)

Rohan Reddy Galipur (galipurr@msu.edu)

Sriya Kondabathula (kondabat@msu.edu)

Vamsi Gopala Krishna Kethepalli (kethepal@msu.edu)

WHO WILL MAKE
BUSINESS HAPPEN?
SPARTANS WILL.