Using Machine Learning to Predict Factors Leading to Employee Attrition
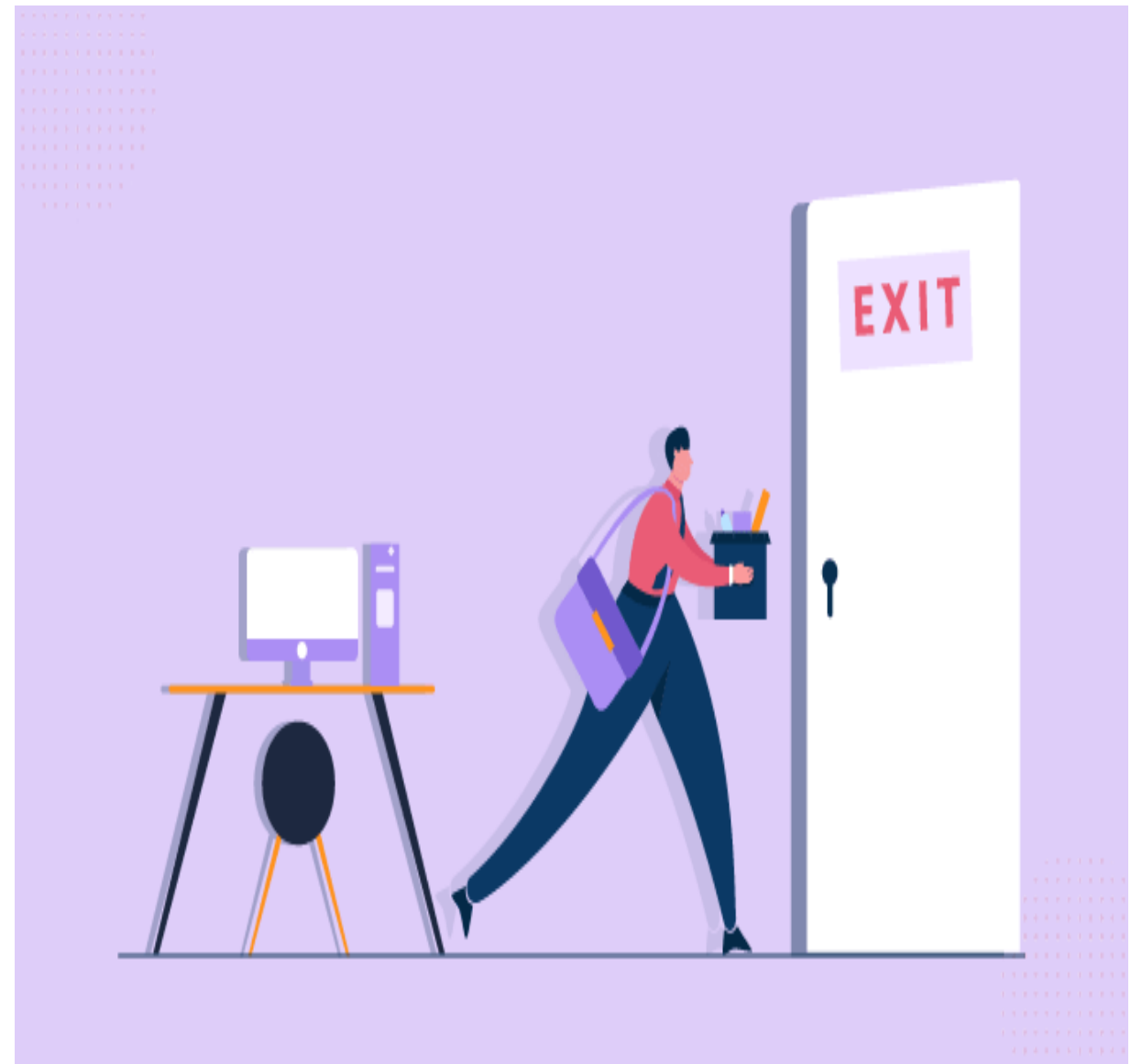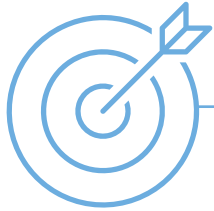
Tom Kim

May 10, 2023

Overarching goal and analysis approach

Our Goal

Identify Canterra's **probability of attrition**, understand the **factors impacting turnover** and **recommended changes** to encourage employees to stay in the workplace.

How we'll get there
A multi-phased statistical approach to uncovering the richest insight
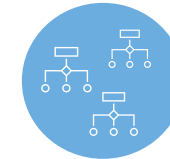
Logistic Regression

Explains the relationship between one dependent variable (attrition) and several independent variables in the dataset

Classification Decision Tree

Assesses structurally the binary decisions that lead to employee attrition

Bootstrap Aggregation "Bagging"

Curtails variance by creating multiple trees and taking a majority vote for the most common class among predictors of attrition

Random Forest

Implements diversity in the bootstrap aggregated model to reduce and remove correlation in attrition trees

Identifying the best model and understanding its performance power in predicting attrition

Measures of Model Performance
(Based on all Models)

Model	AUC-ROC	Sensitivity	Accuracy	Specificity
Decision Tree Model	68%	57%	72%	75%
Logistic Regression Model	73%	69%	65%	64%
Extended Bagging Model	52%	89%	80%	78%
Random Forest Model	**98%**	**96%**	**91%**	**90%**

Less Optimal Model

Most Optimal Model

Interpretation of the Random Forest Model
(Based on Random Forest Model)

98% chance the random forest model can **distinguish between positive versus negative attrition** classes

96% the random forest model has a high probability of identifying employees who actually left Canterra

9 in 10 the random forest model can **successfully identify whether an employee is predicted to resign or stay**

Professional experience, tenure with the company and years with current manager impact attrition in decision tree model

Top Predictors of Attrition
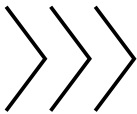*(Based on Decision Tree Model)*

Decision Tree Factors of Attrition
(Based on Decision Tree Model)

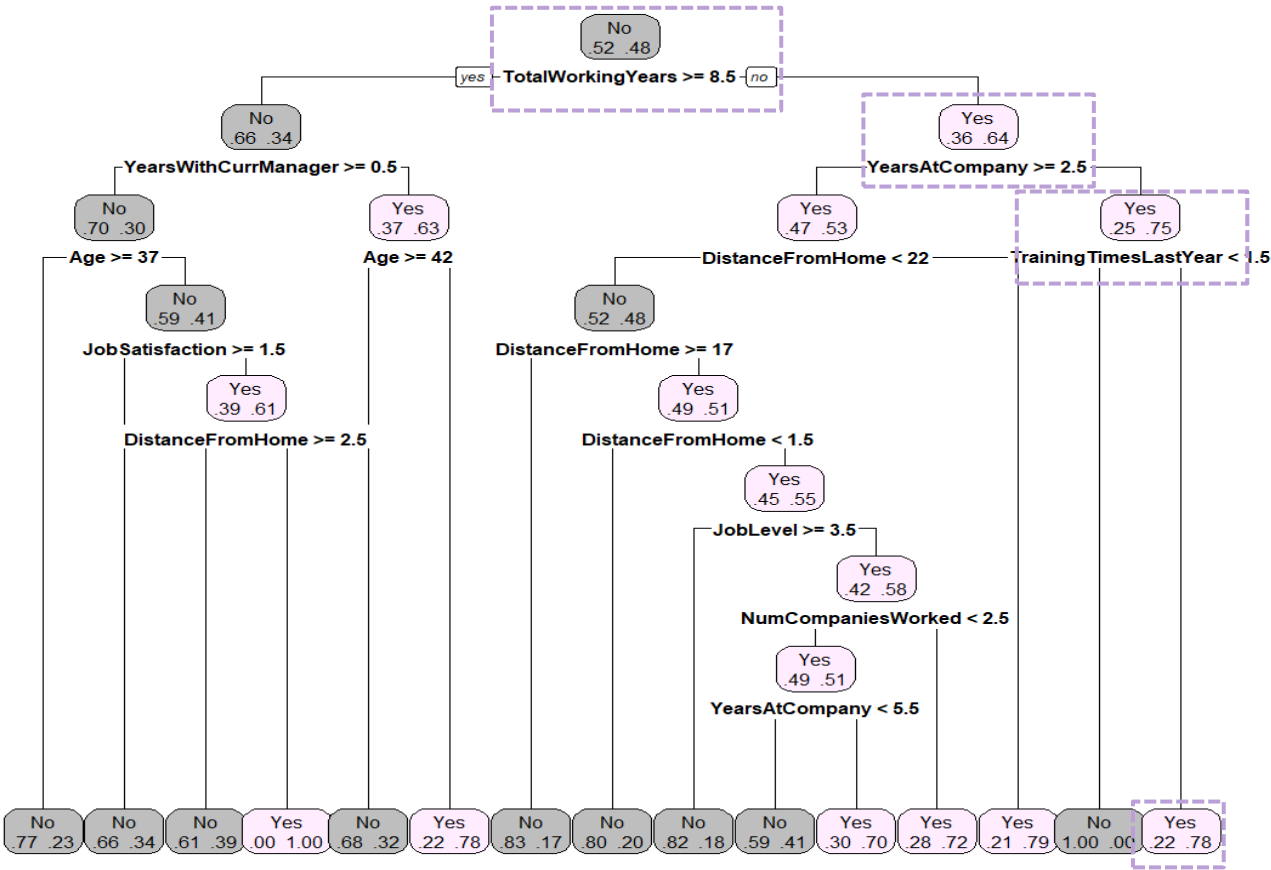**Total Working Years**
(based on more/less than 8.5 years)

Years with Current Manager
(based on more/less than 0.5 years)

Years at Company
(based on more/less than 2.5 years)

No .52 .48 — TotalWorkingYears >= 8.5
No .66 .34
YearsWithCurrManager >= 0.5
No .70 .30
Age >= 37
No .59 .41
JobSatisfaction >= 1.5
Yes .39 .61
DistanceFromHome >= 2.5
Yes .37 .63
Age >= 42
Yes .36 .64
YearsAtCompany >= 2.5
Yes .47 .53
DistanceFromHome < 22
No .52 .48
DistanceFromHome >= 17
Yes .49 .51
DistanceFromHome < 1.5
Yes .45 .55
JobLevel >= 3.5
Yes .42 .58
NumCompaniesWorked < 2.5
Yes .49 .51
YearsAtCompany < 5.5
Yes .25 .75
TrainingTimesLastYear < 1.5

No .77 .23
No .66 .34
No .61 .39
Yes .00 1.00
No .68 .32
Yes .22 .78
No .83 .17
No .80 .20
No .82 .18
No .59 .41
Yes .30 .70
Yes .28 .72
Yes .21 .79
No 1.00 .00
Yes .22 .78

Logit highlights categorical factors such as frequent business travel, single status, and tenure predicting attrition

Logistic Regression

Top Predictors of Attrition
(Based on Logistic Regression Model)

Business Travel Frequently*
(73:1 odd ratio)

Marital Status Single*
(68:1 odd ratio)

Years at Company
(84:1 odd ratio)

524% **Increase** in the probability of attrition among employees who frequently travel for business*
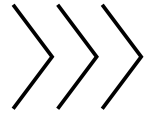
270% **Increase** in the probability of attrition among employees who are of single status*

12% **Increase** in the probability of attrition for each year that an employee works at Canterra

* Please note these findings are based on comparable results; 1) compared to employees who travel rarely; 2) compared to employees who are divorced
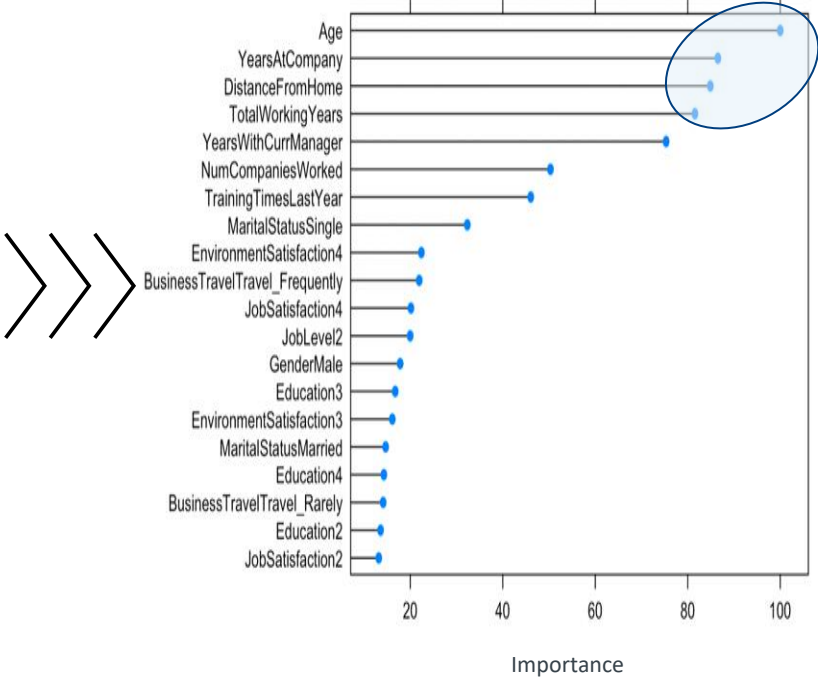
Age, working years, company tenure and commute play a role in retaining talent with bagging and random forest modeling
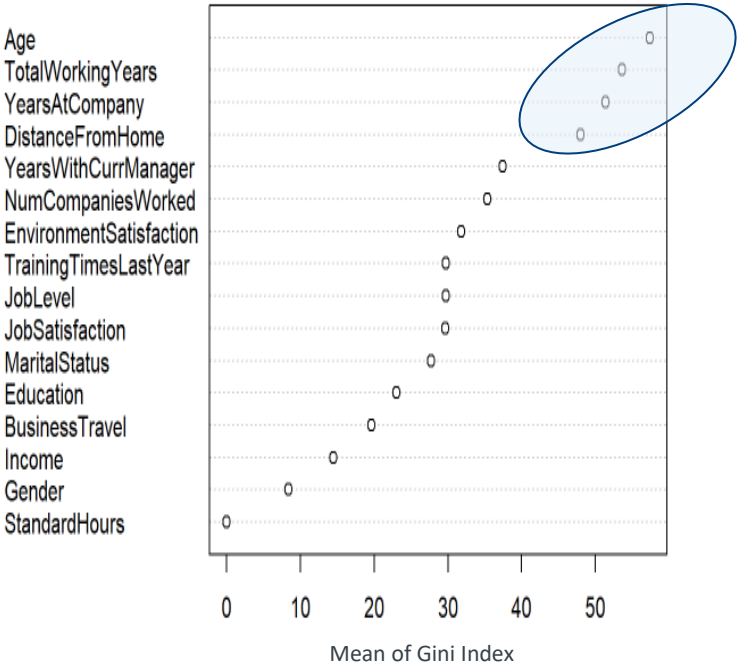

Top Predictors of Attrition
(Based on Bagging and Random Forest Models)

- Age
- Total Working Years
- Years at Company
- Distance from Home

Top Variables of Attrition
(Based on Bagging Model)

Importance

Top Variables of Attrition
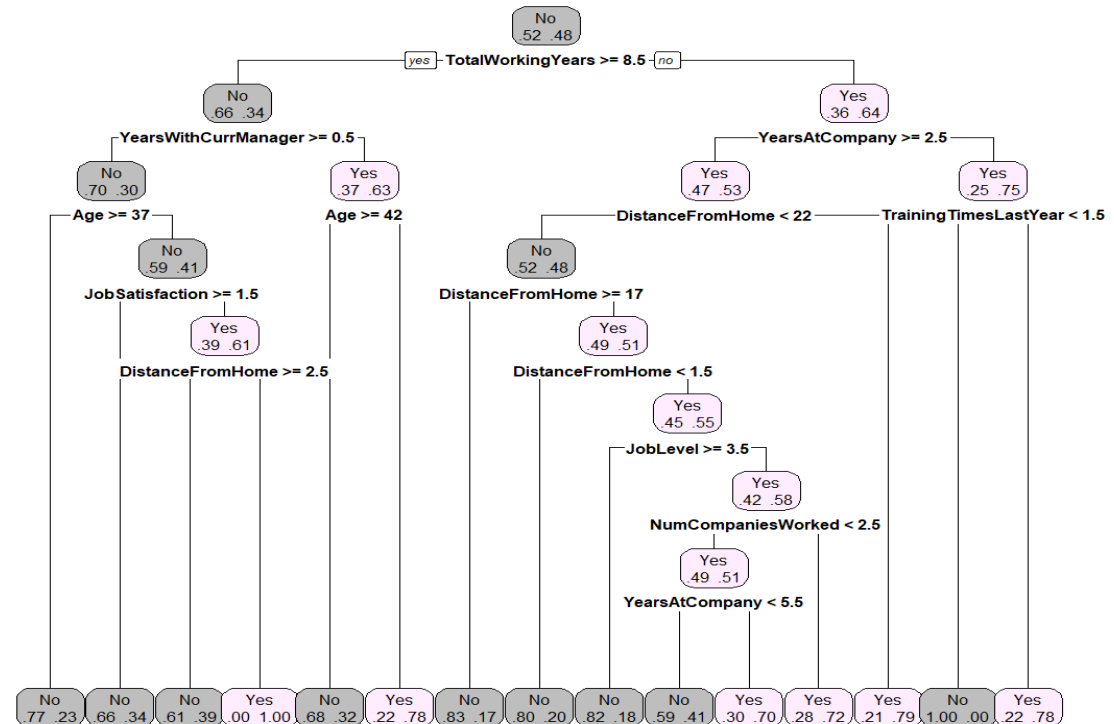*(Based on Random Forest Model)*

Mean of Gini Index

Overall, the random forest algorithm predicts attrition best, with directional insights from decision tree modeling

Top Predictors of Attrition
(Based Random Forest Model)

- Age
- Total Working Years
- Years at Company
- Distance from Home
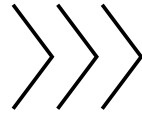
Decision Tree Factors of Attrition
(Based on Decision Tree Model)

Recommendations for improving retention at Canterra

	Finding	**Proposed Recommendations**	**Actions for the Initial 30-60 days**
1	Younger to middle-aged employees are the greatest at-risk group to leave Canterra.	Preemptively deter attrition by offering robust professional development opportunities, career growth maps and competitive pay.	• Implement quarterly check-ins for all staff • Share career roadmaps based on job level • Review salaries and develop a plan for meeting/exceeding industry standards
2	Staff with less professional working years tend to seek opportunities elsewhere.	Place greater rigor in vetting candidates by placing greater emphasis in understanding their long-term goals to ensure alignment with Canterra.	• Train interviewers/managers on asking goal-oriented questions in interview process • Align incoming employees to departments and disciplines that support long-term professional growth
3	Yet, tenure at the company also plays a role in driving attrition.	Institute a reward-and-recognition program for longstanding employees to incentivize commitment.	• Share monthly recognition emails for employee anniversaries • Develop a reward program such as sabbatical or additional PTO for seasoned employees
4	Distance from work matters – with greater commute times comes greater attrition.	Implement flexible work from home policy to mitigate stress and inconvenient travel for long-distance employees.	• Encourage employees to have 1-2 work from home days on a weekly basis • Provide a light stipend for employees to purchase items for a home office set-up
5	The random forest model can most accurately (91%) distinguish between employees who will retain versus leave.	Leverage the random forest model to identify current at-risk employees as well as to vet incoming, prospective candidates.	• Assess which current employees fit criteria and host discovery sessions to gauge overall satisfaction • Train human resources are the prominent factors to closely assess when vetting candidates

Thank You

GEORGETOWN UNIVERSITY McDonough
SCHOOL of BUSINESS