Deloitte.

Predicting Employee Churn

Presented By:

Aidan Ryan, Shamarcus Dixon, Walter Friedrich, Justin Brown, and Karl Durant

MAKING AN IMPACT THAT MATTERS since 1845

AUGUST 2023

Meet The Team

Aidan Ryan
Core Business Operations

Shamarcus Dixon
Cyber and Strategic Risk

Walter Friedrich
Cyber and Strategic Risk

Justin Brown
Cyber and Strategic Risk

Karl Durant
Cyber and Strategic Risk

Business Understanding

01

PROBLEM: THE JOB MARKET IS EVOLVING RAPIDLY, AND COMPANIES ARE STARTING TO SEE AN INCREASE IN EMPLOYEE TURNOVER.

02

MOTIVATION: WE WANT TO BETTER UNDERSTAND THE VARIABLES THAT ARE ASSOCIATED WITH EMPLOYEE TURNOVER AND HOW WE CAN REDUCE THAT OUTCOME IN AN ORGANIZATION.

03

EXPLORE: HOW MACHINE LEARNING CAN IDENTIFY TRENDS IN THE WORKPLACE.

Hypothesis

If our team can identify trends that correlate to a decrease in churn, we can make recommendations for an organization to implement

If our team can identify trends that correlate to a decrease in churn, we can make recommendations for an organization to implement and reduce employee turnover. reduce employee turnover.

To confirm these hypotheses, we'll mainly look for higher model precision as an indicator of how successful our predictions are, given that we want a low false positive rate.

The Data

- Turnover.csv
- Data collected from Kaggle.com
- This dataset represents employee turnover based on variables that impact the workforce

stag	event	gender	age	industry	profession	traffic	coach	head_gender	greywage	way	extraversion	independ	selfc
7.030801	1	m	35.0	Banks	HR	rabrecNErab	no	f	white	bus	6.2	4.1	
22.965092	1	m	33.0	Banks	HR	empjs	no	m	white	bus	6.2	4.1	
15.934292	1	f	35.0	PowerGeneration	HR	rabrecNErab	no	m	white	bus	6.2	6.2	
15.934292	1	f	35.0	PowerGeneration	HR	rabrecNErab	no	m	white	bus	5.4	7.6	
8.410678	1	m	32.0	Retail	Commercial	youjs	yes	f	white	bus	3.0	4.1	
...	...	...	...	...	...	...	...	...	...	...	...	...	
10.611910	0	f	41.0	Banks	HR	rabrecNErab	yes	m	white	bus	8.6	3.4	
10.611910	0	f	41.0	Banks	HR	rabrecNErab	yes	m	white	bus	8.6	3.4	
118.800821	0	f	34.0	Telecom	Accounting	KA	no	f	white	bus	4.6	5.5	
49.412731	0	f	51.0	Consult	HR	empjs	no	m	grey	bus	3.8	7.6	
24.837782	0	f	29.0	Retail	HR	youjs	no	f	white	car	9.4	1.2	

Understanding The Data

Variables	Understanding
Event	Did the employee resign or not?
Gender	Employees gender
Age	In years, ranging from 18 to 58
Industry	Industry in which the employee works
Profession	The respondent's exact profession
Traffic	From what pipeline did candidate come to the company
Coach	Presence of a coach during probation
Head_gender	The supervisor's gender
Greywage	The salary isn't reported to the tax authorities
Way	How an employee gets to workplace
Extraversion	independence, self-control, anxiety,

Modeling The Data

Modeling Techniques

- Exploratory & statistical analysis
- Data visualization to identify trends and patterns.
- Predictive modeling or forecasting techniques
  - Linear regression and random forest.

Variables

- Mainly based on categorical data from turnover.csv

  'industry', 'profession', 'traffic', 'coach', 'wage'

Expected Results

"If our team can identify trends that correlate to a decrease in churn, we can make recommendations for an organization to implement and reduce employee turnover."

- Employees that have a coach are less likely to churn
- Employees that have a higher wage are less likely to churn
- Employees that have a shorter commute are less likely to churn
- *Employees that don't have a coach are more likely to churn*
- *Employees that have a lower wage are more likely to churn*

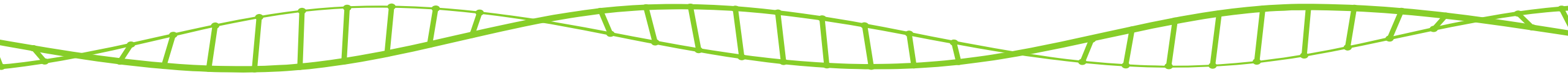
Candidate Models

We initially trained the data using logistic regression, leveraging its simplicity and statistical significance.

To improve prediction accuracy and handle complex nonlinear relationships in the data, we experimented with decision trees.

Ultimately, we settled on a random forest model because of its powerful ensemble-based approach and its ability to minimize overfitting.

Logistic Regression Model



Client Service excellence

Logistic Regression is a statistical and machine learning algorithm used for binary classification tasks, where the goal is to predict one of two possible classes (e.g., yes/no, true/false, 0/1).

Key Features

- It is simple, interpretable, and easy to implement.
- It can handle both linear and nonlinear relationships between features and the target.
- It is sensitive to feature scaling, so it is essential to preprocess the data properly.

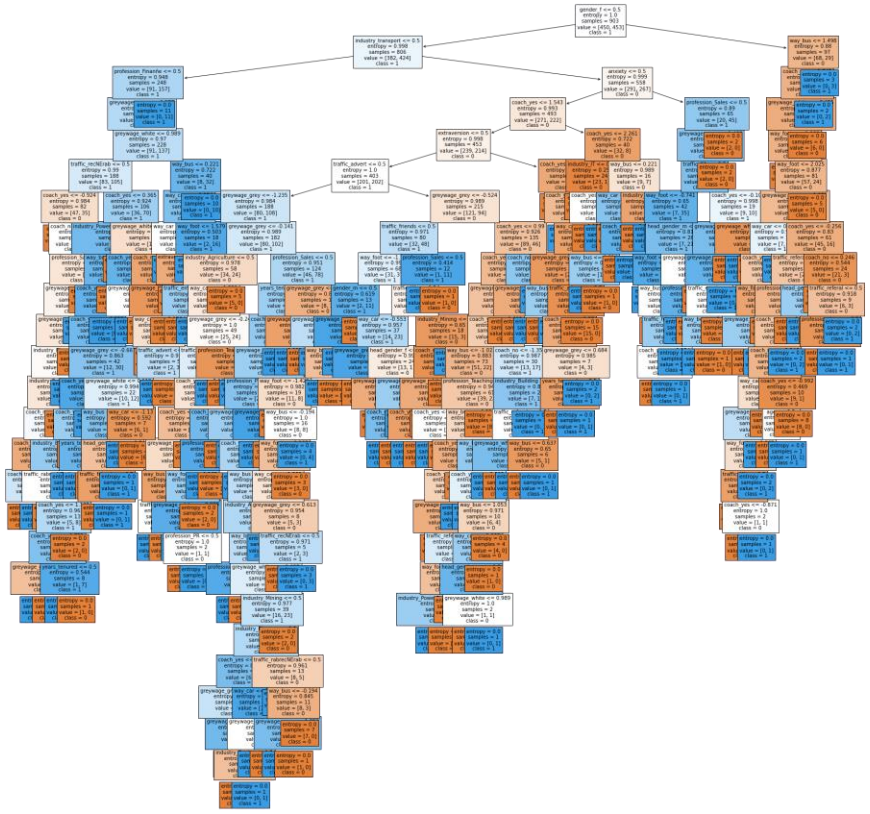
The model accuracy was 66%, so we decided to try out more complex models.

Random Forest

A Random Forest is an ensemble learning technique used in machine learning for both classification and regression tasks. It is based on the concept of creating multiple decision trees during the training phase and combining their predictions to make more accurate and robust predictions.

We worked through it in steps:

- Create an initial model
  – we saw a jump in most model metrics!
- Fine tune the model using a grid search



Comparing the Models

Logistic Regression

The goal is to predict one of three possible classes

(yes/no, true/false, 0/1)

Our accuracy was: 0.67

Decision Tree

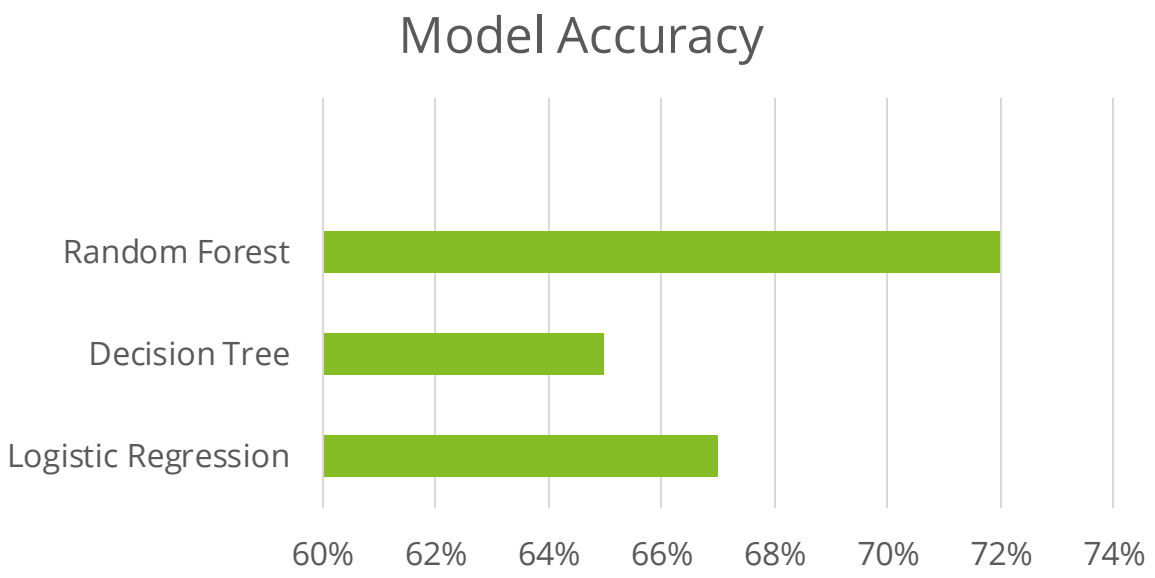
The goal of a decision tree is to create a predictive model that can make decisions or predictions based on a set of input features

Our accuracy was: 0.65

Random Forest

The goal is to create multiple decision trees during the training phase and combining their predictions to make more accurate and robust predictions.

Our accuracy was: 0.72

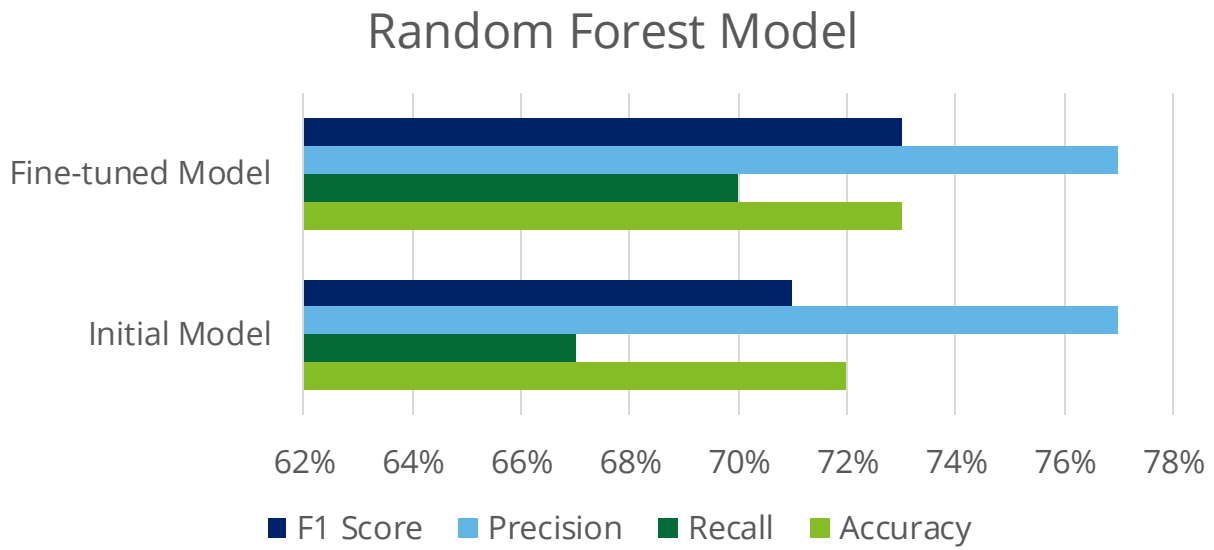


Chosen Model Metrics

The random forest model we decided on was more precise than our logistic regression model.

This means our model was more conservative and doesn't usually overpredict.

When it predicts a positive (a person churning) it is highly precise, however it may miss out on other employees that may be likely to churn.



Initial Model

Accuracy: 0.72

Recall: 0.67

Precision: 0.77

F1 Score: 0.71

Fine-tuned Model

Accuracy: 0.73

Recall: 0.7

Precision: 0.77

F1 Score: 0.73

Recommendations

Coaching

Integrate a coaching system to help foster a community and onboard new employees into positive work culture

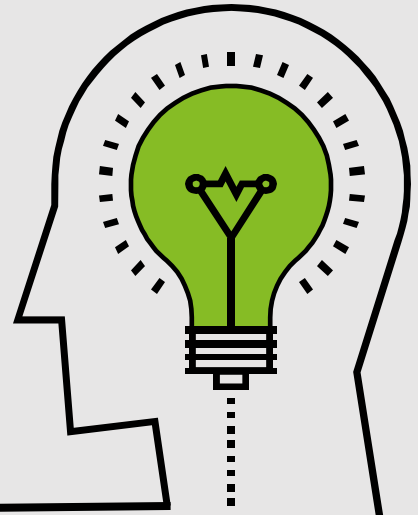
Commuting

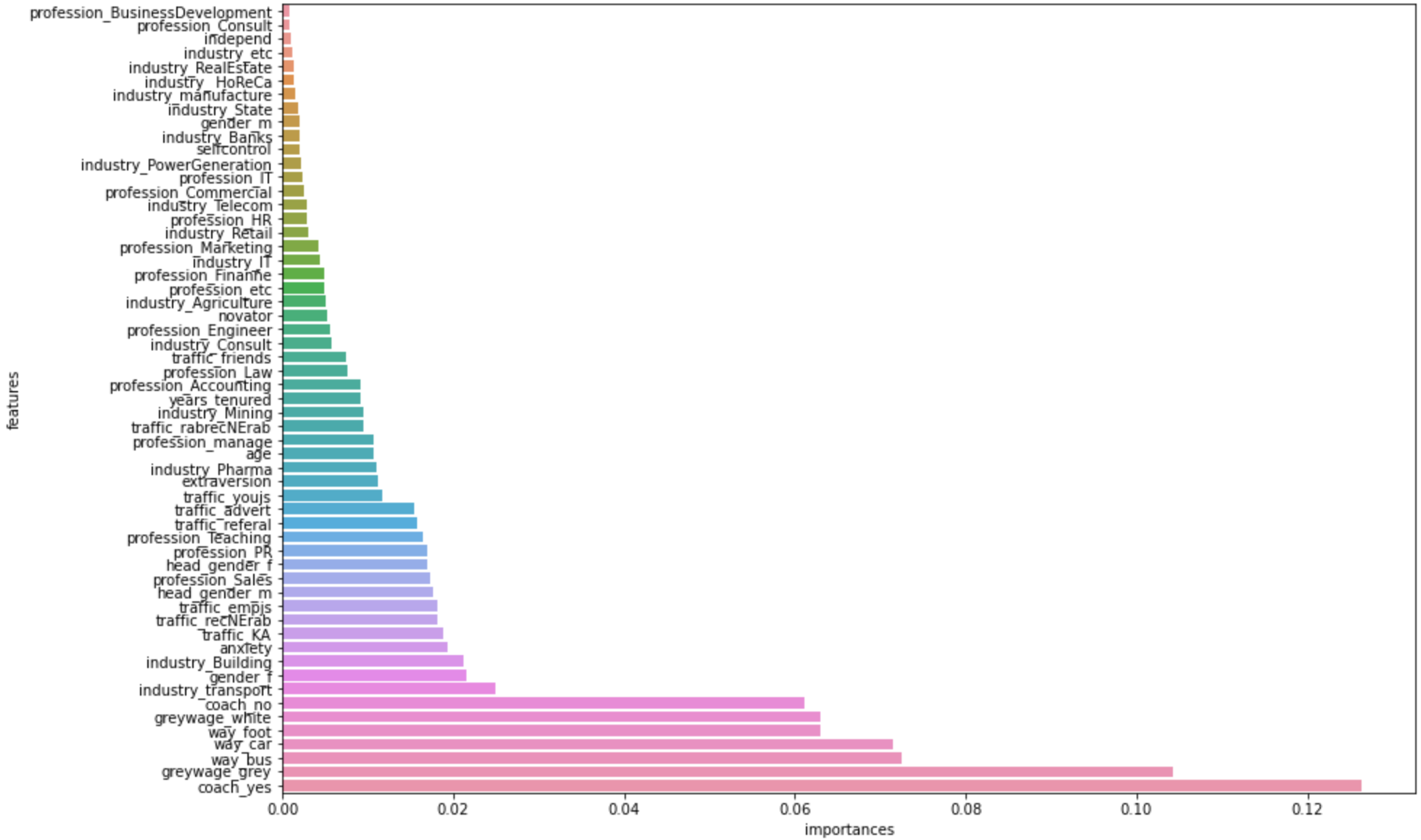
Introduce commuter benefits program for employees, a company bus, or a carpool program to make employees a little more excited to journey to the office

Wages

It costs more to hire a new employee than to pay a current one market wages. Make sure pay across the board is competitive to create satisfaction among your workforce

Thank You




profession_BusinessDevelopment
profession_Consult
independ
industry_etc
industry_RealEstate
industry_HoReCa
industry_manufacture
industry_State
gender_m
industry_Banks
selfcontrol
industry_PowerGeneration
profession_IT
profession_Commercial
industry_Telecom
profession_HR
industry_Retail
profession_Marketing
industry_IT
profession_Finance
profession_etc
industry_Agriculture
novator
profession_Engineer
industry_Consult
traffic_friends
profession_Law
profession_Accounting
years_tenured
industry_Mining
traffic_rabrecNErab
profession_manage
age
industry_Pharma
extraversion
traffic_youjs
traffic_advert
traffic_referal
profession_Teaching
profession_PR
head_gender_f
profession_Sales
head_gender_m
traffic_empjs
traffic_recNErab
traffic_KA
anxiety
industry_Building
gender_f
industry_transport
coach_no
greywage_white
way_foot
way_car
way_bus
greywage_grey
coach_yes
features
0.00
0.02
0.04
0.06
0.08
0.10
0.12
importances