DATA SCIENCE CAREER CHANGE LIKELIHOOD

Finding Future Data Scientists for RADS, Inc.

January 2022

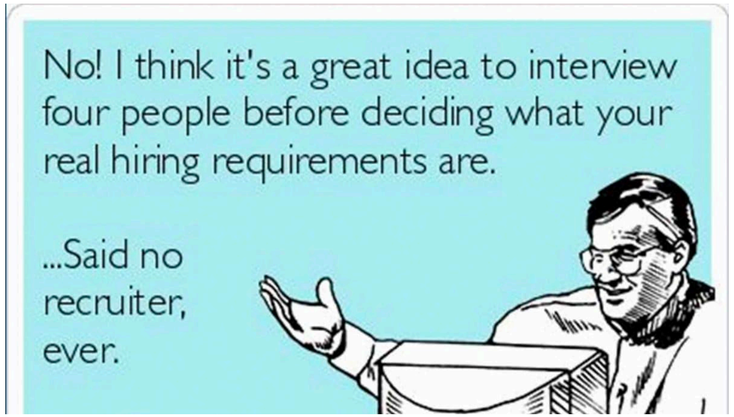
CONTEXT & OBJECTIVE

- **CONTEXT:** **After conducting a Data Science training program, RADs would like to know which participants are likely to change their jobs**

- **OBJECTIVE:** **Find potential candidates for Recruiting Awesome Data Scientists Incorporation (RADS, Inc.) - a Data Scientist recruiting firm looking for potential future Data Scientists**

WHY IT MATTERS // BUSINESS VALUE

- **Identify potential talent using a data-driven approach**
- **Save recruiters and hiring managers time**
- **Refine training program to recruit and train more talent internally**



DATA SOURCE

The data is from a company that is active in Big Data and Data Science and ran a training program with the intention to hire data scientists among people who successfully passed courses they conducted

This dataset includes attributes that will help us predict candidates that are likely to be looking for a job change

Attributes:

- **City Code (actual cities not provided)**
- **City Development Index**
- **Gender**
- **Relevant Experience**
- **Education Level**
- **Major (Discipline)**
- **Years of Experience**
- **Company Size**
- **Company Type**
- **Years Since Last Job**
- **Training Hours**

METHODOLOGY

DATA

- 14,283 employees/ training program participants
- Data includes: city code, training hours, gender, relevant experience

ANALYZE

25% of employees changed careers

MODEL

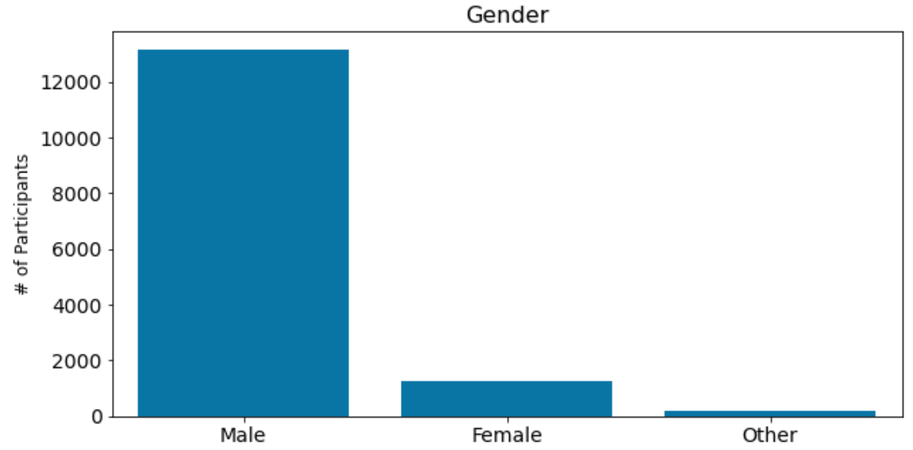
- Find the best model to predict likelihood of career change
- Determine the most important factors that influence likelihood of career change

OBSERVATIONS ABOUT THE DATA: THE MAJORITY OF PARTICIPANTS WERE MALE AND STEM MAJORS

The dataset is skewed:

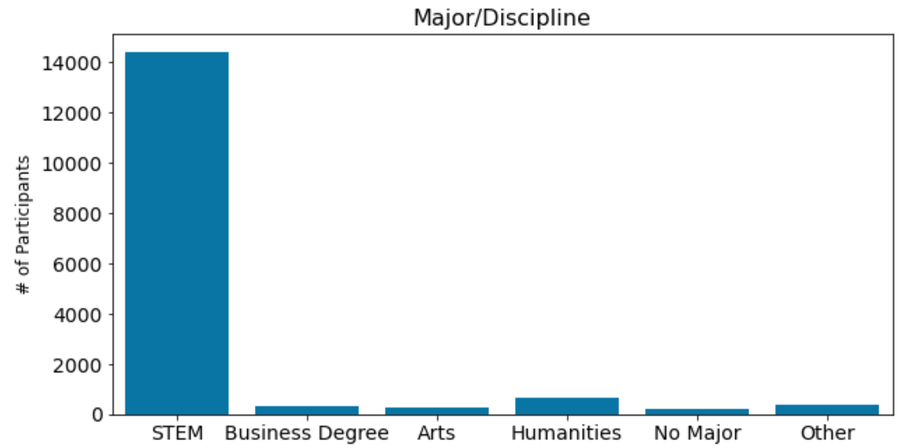
69% Male

7% Female



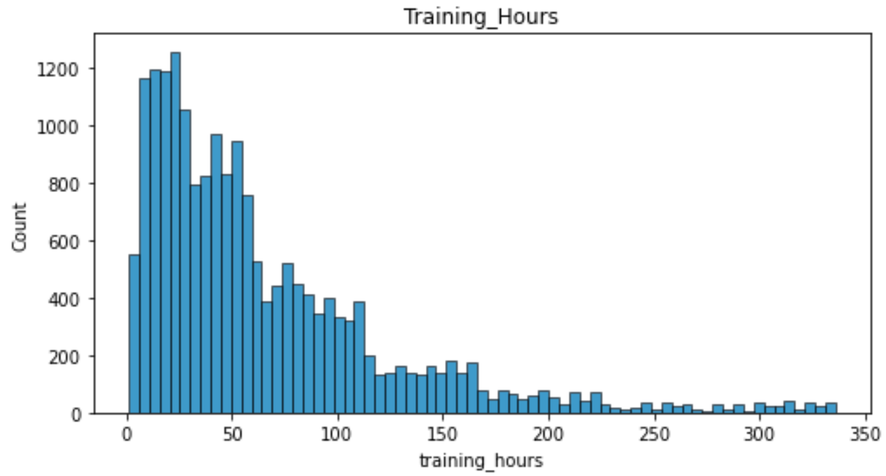
76% STEM Majors

(Science, Technology, Engineering, Math)

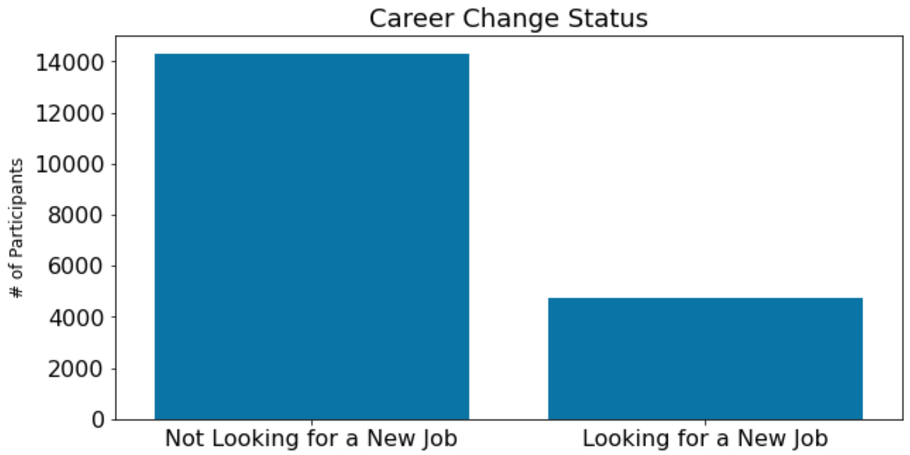


MOST TRAINED FOR 50 HOURS OR LESS AND ABOUT 25% CHANGED JOBS

For those that participated in the training program, most trained for 50 hours or less



About 25% of employees changed jobs

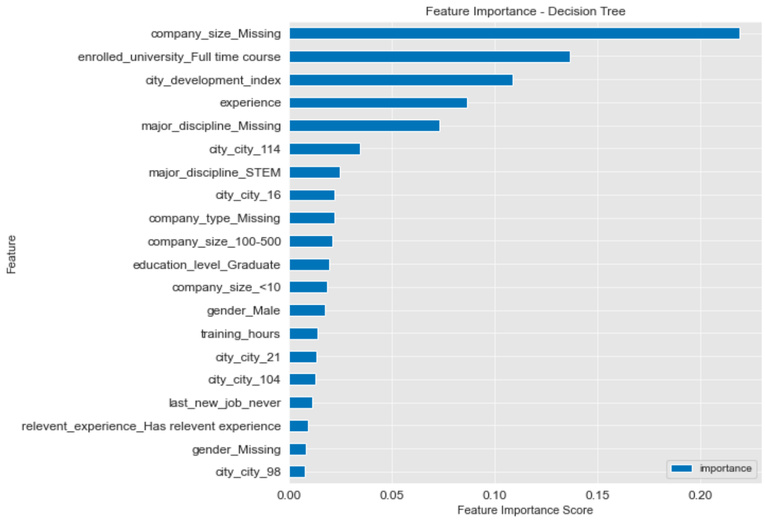


MODELS & METRICS: WE RAN VARIOUS MODELS AND CHOSE THE MODEL WITH THE HIGHEST RECALL RATE

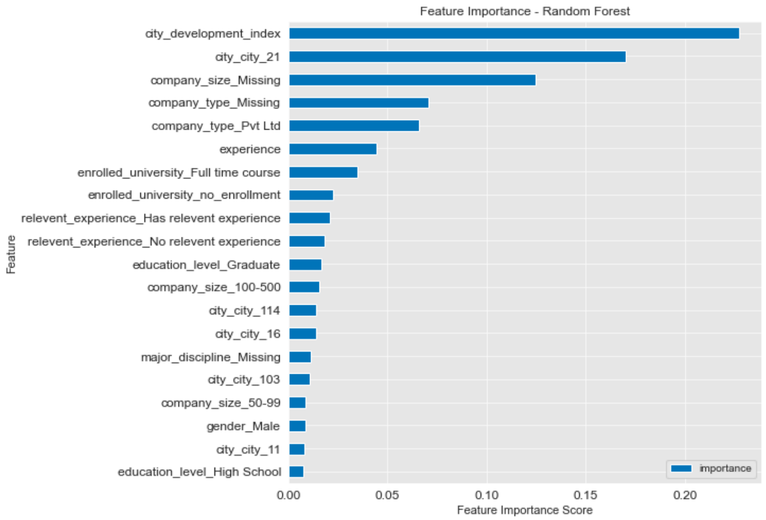
- Because there is a smaller portion of participants who changed jobs (25%), the metric we used to measure our models' performances was Recall
- Recall measures the % a model predicts True Positives (employees "Looking for a Job Change") that actually are looking for a job change
- **Our model is expected to predict True Positives (i.e. predict employees "Looking for a Job Change" that are actually looking for a job change) 77% of the time**

MODEL RESULTS

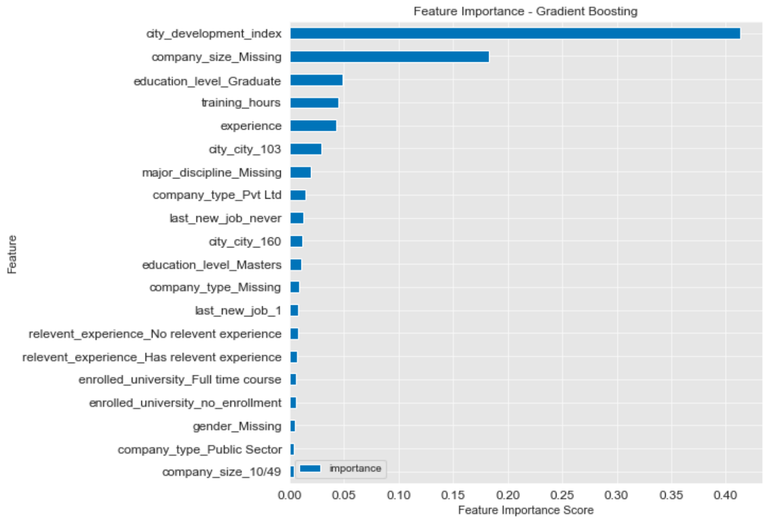
Decision Tree
66% Recall Rate



Random Forest
77% Recall Rate

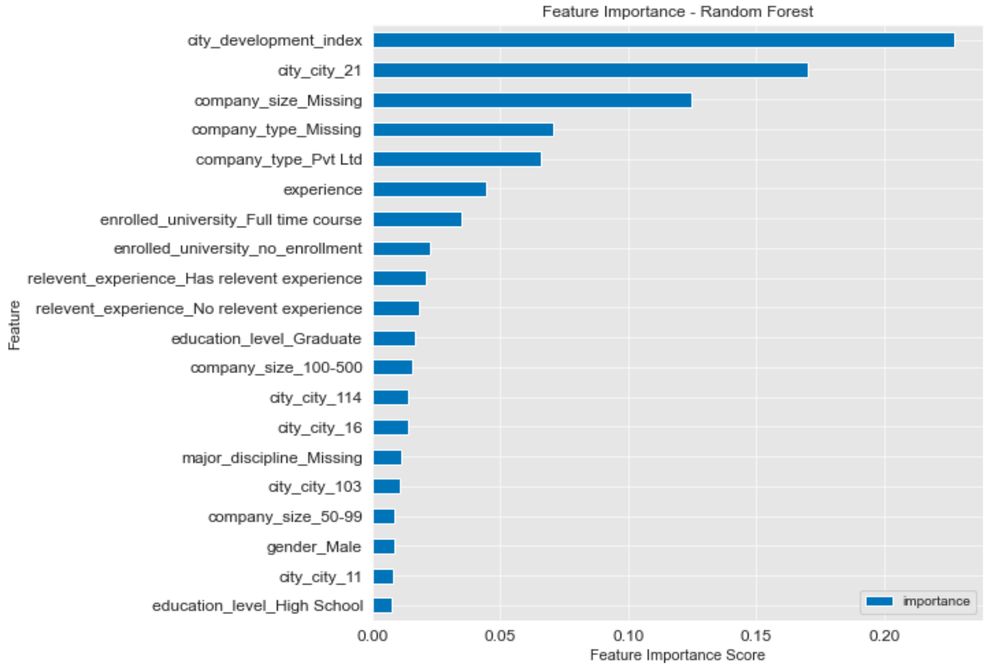


Gradient Boosting
56% Recall Rate



THE MOST IMPORTANT DETERMINANT OF JOB CHANGE IS CITY DEVELOPMENT INDEX

- Other important features (some which also ranked high in other models are):
  - Experience
  - Small company size (<10 to 500)
  - Education level: Graduate Degree



Classification Report for Training Data

	precision	recall	f1-score	support
Not Looking for Job Change	0.91	0.75	0.82	10726
Looking for a Job Change	0.51	0.77	0.61	3557
accuracy			0.76	14283
macro avg	0.71	0.76	0.72	14283
weighted avg	0.81	0.76	0.77	14283

Classification Report for Test Data

	precision	recall	f1-score	support
Not Looking for Job Change	0.91	0.75	0.82	3576
Looking for a Job Change	0.50	0.77	0.61	1185
accuracy			0.75	4761
macro avg	0.71	0.76	0.71	4761
weighted avg	0.81	0.75	0.77	4761

CONCLUSIONS & FUTURE WORK

- CONCLUSIONS: While this data gives us insight into the importance of location, company size, education, and experience in identifying those looking to change jobs, there are endless ways to use this data to help RADs build a future for Data Scientists

- FUTURE WORK: Identify recruitment opportunities and strategies for RADs' Data Science program (recruit more females; recruit 'unlikely' candidates that do not have a background in STEM; recruit in various cities)
- Offer another round of Data Science training curriculum to engage more employees
- Gather more data following Round 2 and re-run models
- Continue refining overall process and models
- Make sure all who want to explore Data Science have the opportunity!

THANK YOU!!
STAY RAD