Machine Learning

Module 3 project
Car Insurance Cold Calling

LaShanni Butler

Flatiron School

5/28/19

Agenda

- Overview
- Machine Learning Workflow
- Background
- Problem Statement
- Machine Learning models
- Conclusion

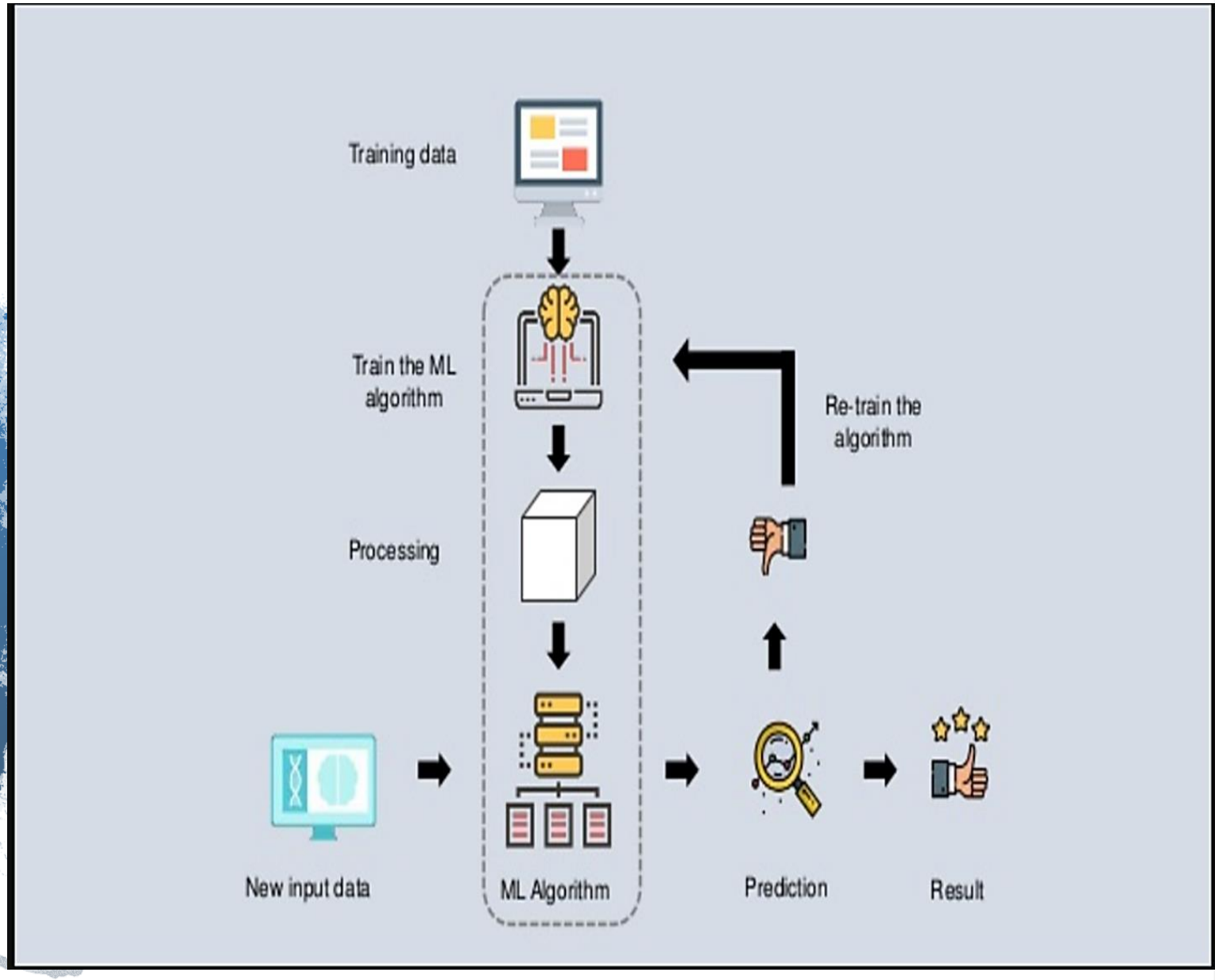
Overview

- What is Machine Learning?
  - ML works on the development of computer programs that can access data and use it to automatically learn and improve from experience. (I.e. – it's a technique that uses statistics to help machines learn from past data).
  - Examples include:
    - Amazon Echo (Alexa) functions off of ML, and develops its accuracy based off of the user(s) interactions
    - Google search algorithms
    - Email spam filters

Sources:

- Simplilearn: What Is Machine Learning: https://www.youtube.com/watch?v=HgBpFaATdoA&list=PLEiEAq2VkUULYYgj13YHUWmRePqiu8Ddy&index=3

Machine Learning workflow



Sources:
Simplilearn: What Is Machine Learning:
https://www.youtube.com/watch?v=HgBpFaATdoA&list=PLEiEAq2VkUULYYgj13YHUWmRePqiu8Ddy&index=3

Background

Types of Machine Learning Models:

- **Supervised**: Is a method used to enable machines to classify and predict objects, problems or situations based on labeled data fed to the machine. Data is labeled, direct feedback is given and machine predicts output.
  - Examples: Logistic Regression, Decision Trees, K-Nearest Neighbors, Random Forest

- **Unsupervised**: Systems are able to identify hidden data patterns from input given. Data is unlabeled, no feedback is given and machine finds hidden structure in data.
  - Examples: K-means Clustering, Partial Least Squares, Fuzzy Means, Principal Component Analysis

- **Reinforcement**: Systems are given no training and is rewarded or punished based on its last action. It helps increase efficiency.

Sources:

- Simpliearn: Machine Learning vs Deep Learning vs Artificial Intelligence: https://www.youtube.com/watch?v=9dFhZFUkzuQ&list=PLEiEAq2VkUULYYgj13YHUWmRePqiu8Ddy&index=4

Problem Statement

- The dataset: Contains consumer information from a company conducting cold calling. The dataset contains general info (i.e. – age, education, job, etc.) and cold calling info (i.e. – communication, last contact, previous contact attempts, etc.).

- The Ask: The client would like to know the most important factor that determines cold calling success. So we'll use predictive models (i.e. – Machine Learning techniques) to see which factor(s) are successful.

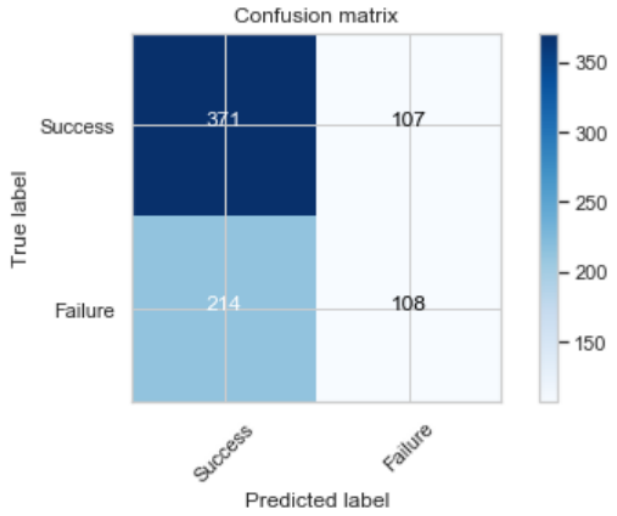
Sources:
Kaggle dataset: https://www.kaggle.com/kondla/carinsurance

K-Nearest Neighbors

- KNN is generally used to predict categorical values based on the nearest datapoints of interests

- Confusion matrix: a summary of prediction results on a classification problem. The number of correct and incorrect predictions are summarized with count values and broken down by each class.

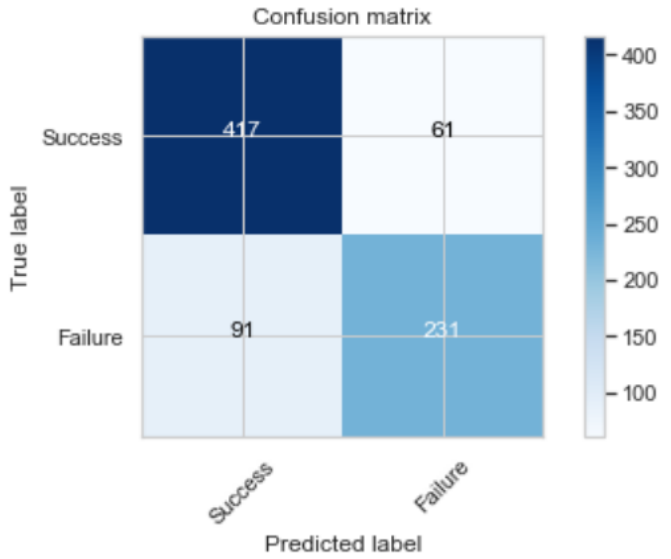
```
kNN Accuracy is 0.60
Cross Validation Score = 0.62
              precision    recall  f1-score   support

           0       0.63      0.78      0.70       478
           1       0.50      0.34      0.40       322

   micro avg       0.60      0.60      0.60       800
   macro avg       0.57      0.56      0.55       800
weighted avg       0.58      0.60      0.58       800
```



Logistic Regression

- The simplest classification algorithm used for binary or multiclassification problems (datasets where y = 0 or 1, where 1 denotes the default class).

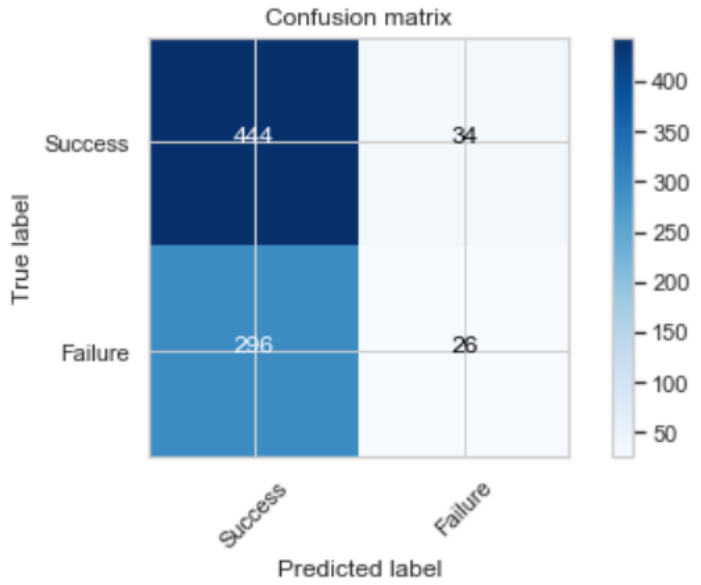
```
Logistic Accuracy is 0.81
Cross Validation Score = 0.81
              precision    recall  f1-score   support

           0       0.82      0.87      0.85       478
           1       0.79      0.72      0.75       322

   micro avg       0.81      0.81      0.81       800
   macro avg       0.81      0.79      0.80       800
weighted avg       0.81      0.81      0.81       800
```



Support Vector Machine (SVM)

- Is widely used classification algorithm. SVM creates a separation line which divides the classes in the best possible manner. Ex - dog or cat, disease or no disease.

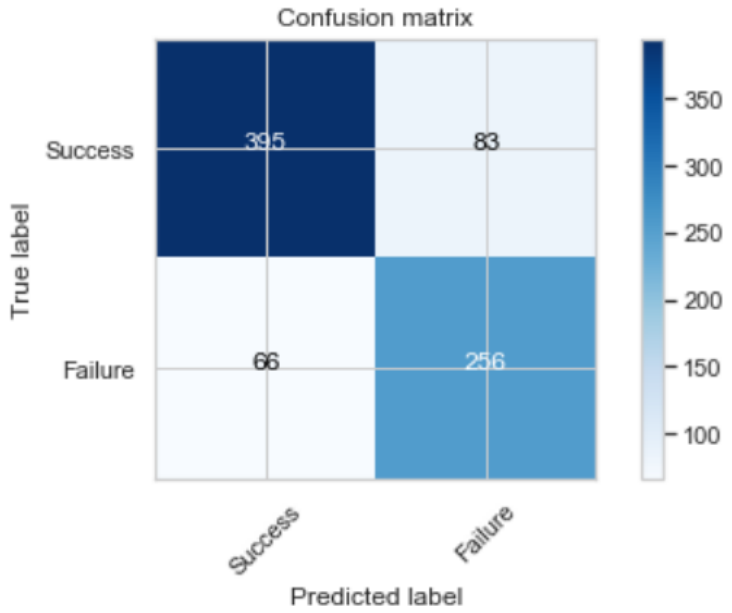
```
SVM Accuracy is 0.59
Cross Validation Score = 0.59
              precision    recall  f1-score   support

           0       0.60      0.93      0.73       478
           1       0.43      0.08      0.14       322

   micro avg       0.59      0.59      0.59       800
   macro avg       0.52      0.50      0.43       800
weighted avg       0.53      0.59      0.49       800
```



Decision Tree

- Is an inverted tree shaped algorithm used to determine a course of action. Each tree branch represents a possible decision

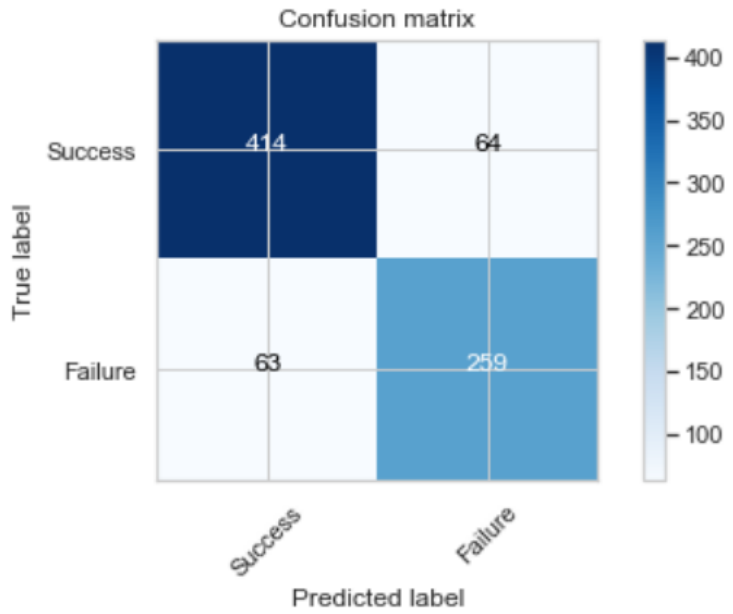
```
Decision Tree Accuracy is 0.81
Cross Validation Score = 0.81
              precision    recall  f1-score   support

           0       0.86      0.83      0.84       478
           1       0.76      0.80      0.77       322

   micro avg       0.81      0.81      0.81       800
   macro avg       0.81      0.81      0.81       800
weighted avg       0.82      0.81      0.81       800
```



Random Forest

- Similar to Decision Tree, but multiple trees are used. Each "tree" observation is classified. Usually increased the accuracy when DT is used as an algorithm.

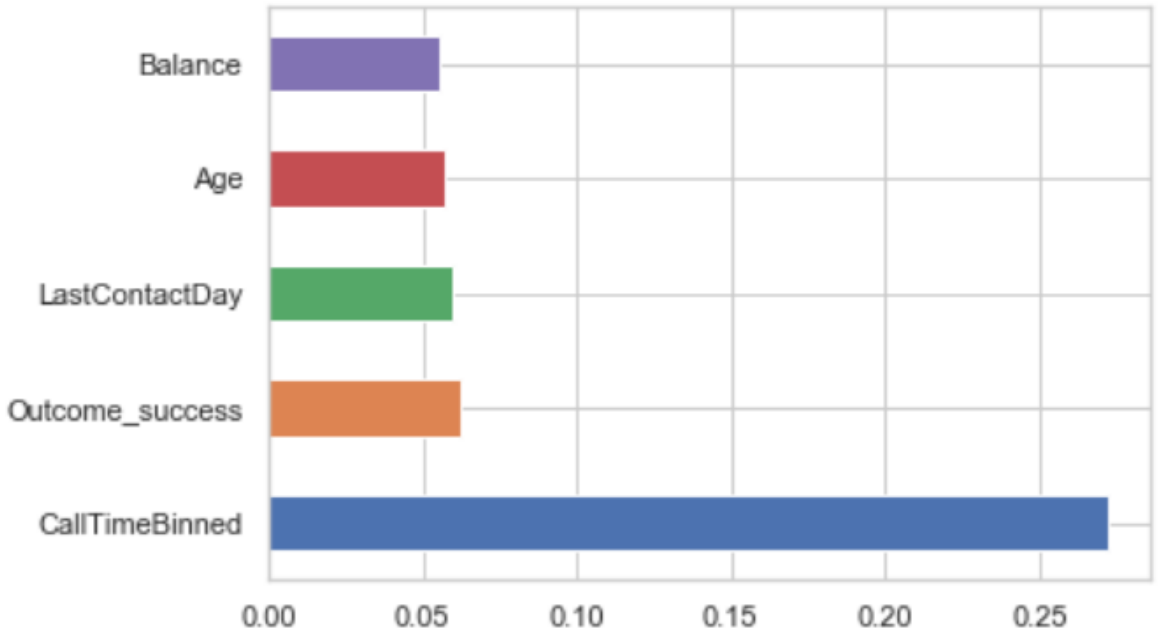
```
Random Forest Accuracy is 0.84
Cross Validation Score = 0.84
              precision    recall  f1-score   support

           0       0.87      0.87      0.87       478
           1       0.80      0.80      0.80       322

   micro avg       0.84      0.84      0.84       800
   macro avg       0.83      0.84      0.84       800
weighted avg       0.84      0.84      0.84       800
```



Conclusions

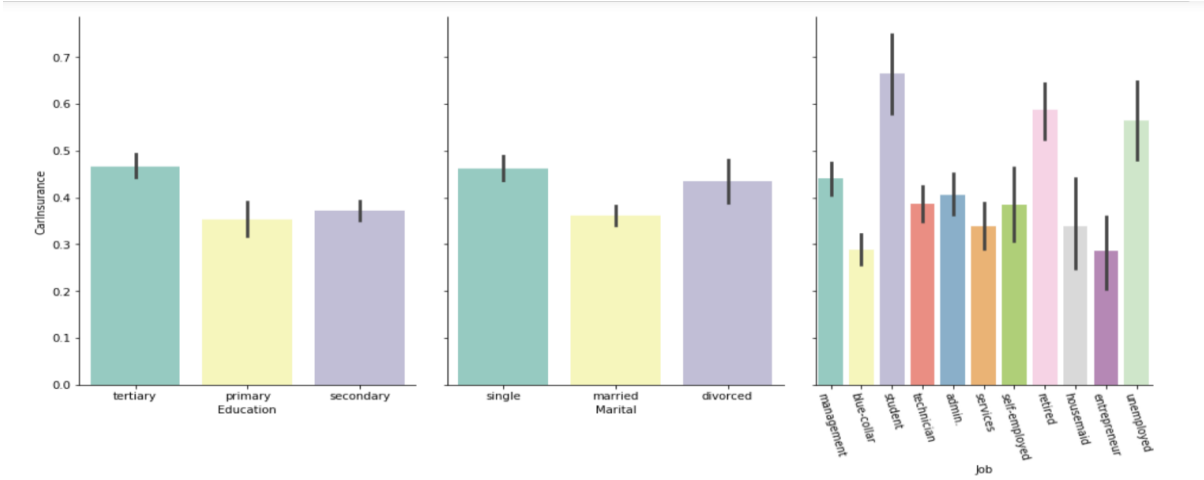
- Based on the predictive modeling, Random Forest had the highest percentage of 84%, and a cross validation score of 84%.

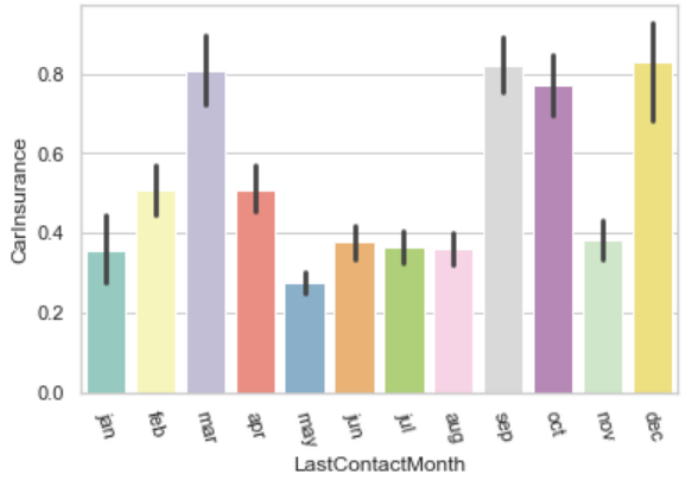
- Interestingly, Call Time had the most effect on customers who were cold called. Last Contact Day also showed would be more likely to purchase car insurance.



Conclusions

- Potential customers with tertiary (advanced) educations are more likely to purchase insurance
- Single people are more likely to purchase insurance
- Students, retired and unemployed people are purchasing the most car insurance polices
- Many people buy insurance in March, September, October and December





Questions?

- Link to recorded presentation: https://drive.google.com/file/d/19z0aYThNQTfa1epXunNkVkAjBTcQlF7F/view?usp=sharing