Predicitve Analysis for **XYZ's Marketing Campaign**

Ssu Hsien Lee, Alan Chen, Kasturi Pal, Deepanshu Kataria

Business Problem

Who is likely to become a premium subscriber?

Business Model

The company provides a host of basic services for **free** while giving additional **premium features** for an additional monthly **subscription fee**.

ISSUE

XYZ needs **new subscribers** to take up the monthly subscription plan **to increase revenue.**
However, for the next marketing campaign, we don't know **which current non-subscribers** are likely to respond to the marketing campaign and **convert to premium subscribers**

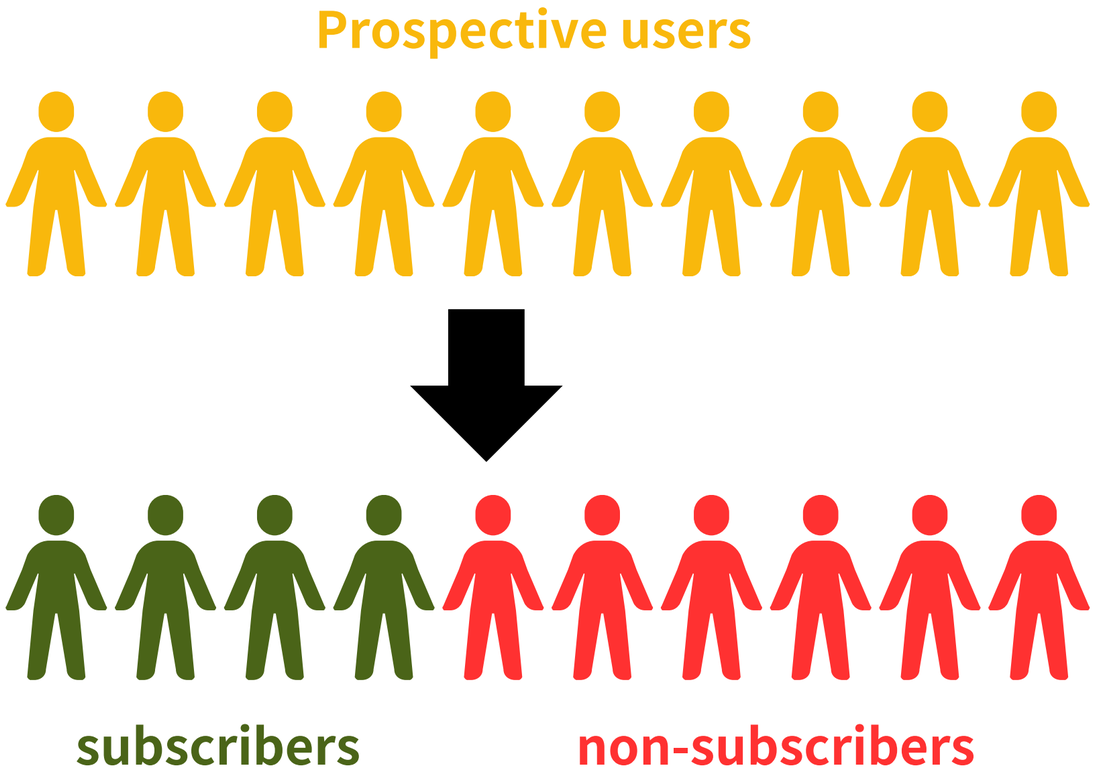
Solution

We analyze previous marketing campaign data and **develop a predictive model** to help the executives of XYZ **better predict** which existing users are most likely to take up their subscriptions.

Business Problem

Analytics- enabled Marketing Campaign Stategy

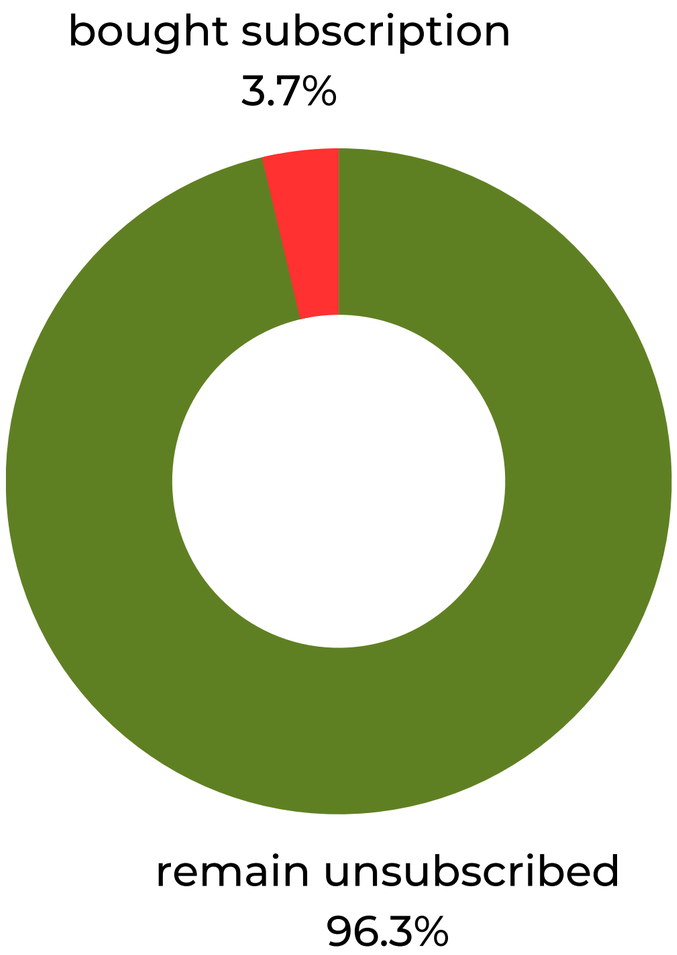
Predictive Model



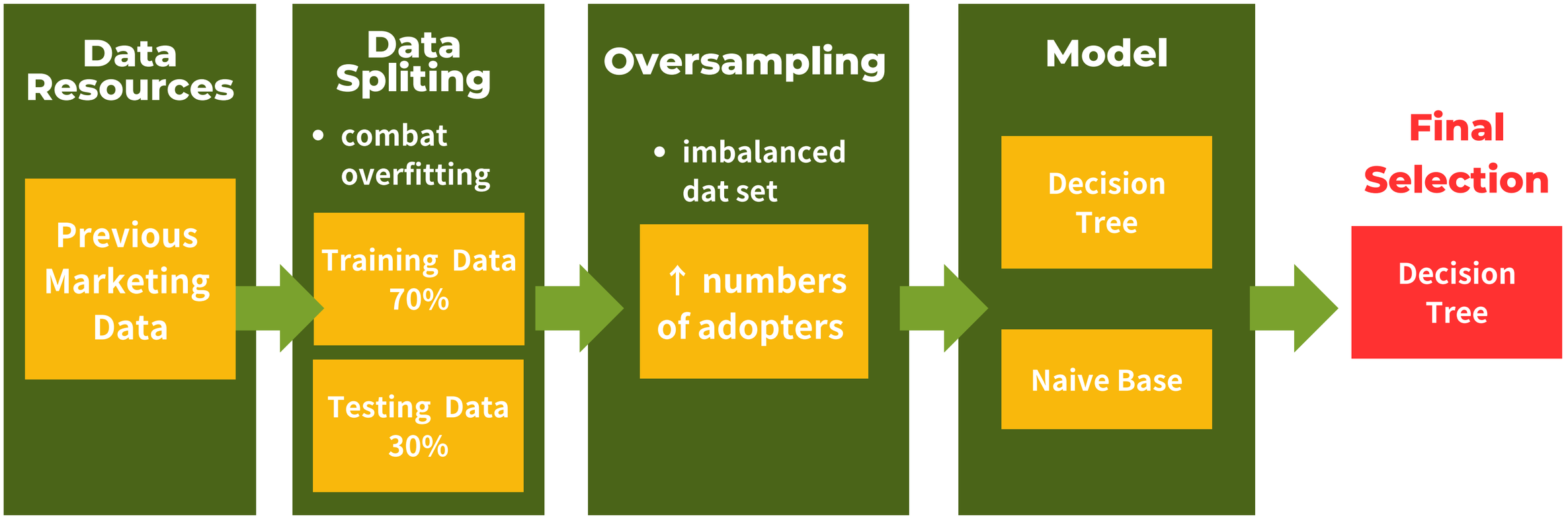
Data Description

Overview of Data:

- Total of 41,540 records
- 25 attributes
- Subscribers proportions after previous marketing campaign



Solution Map - Model Building Process



How do we know if our model works?

?

We build different models. However, which model is the best?

Identify evaluation performance metrics

Receiver Operating Characteristic (ROC)

+

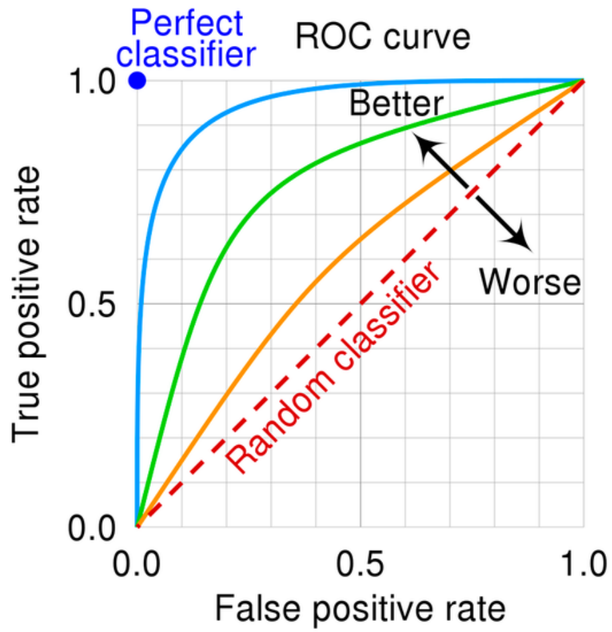
AUC

- **scale-invariant**
- **classification-threshold -invariant**

resoruses:https://commons.wikimedia.org/wiki/File:Roc_curve.svg#filelinks

What does AUC-ROC mean to the problem in context?

The most suitable metric for this markegint campaign analysis

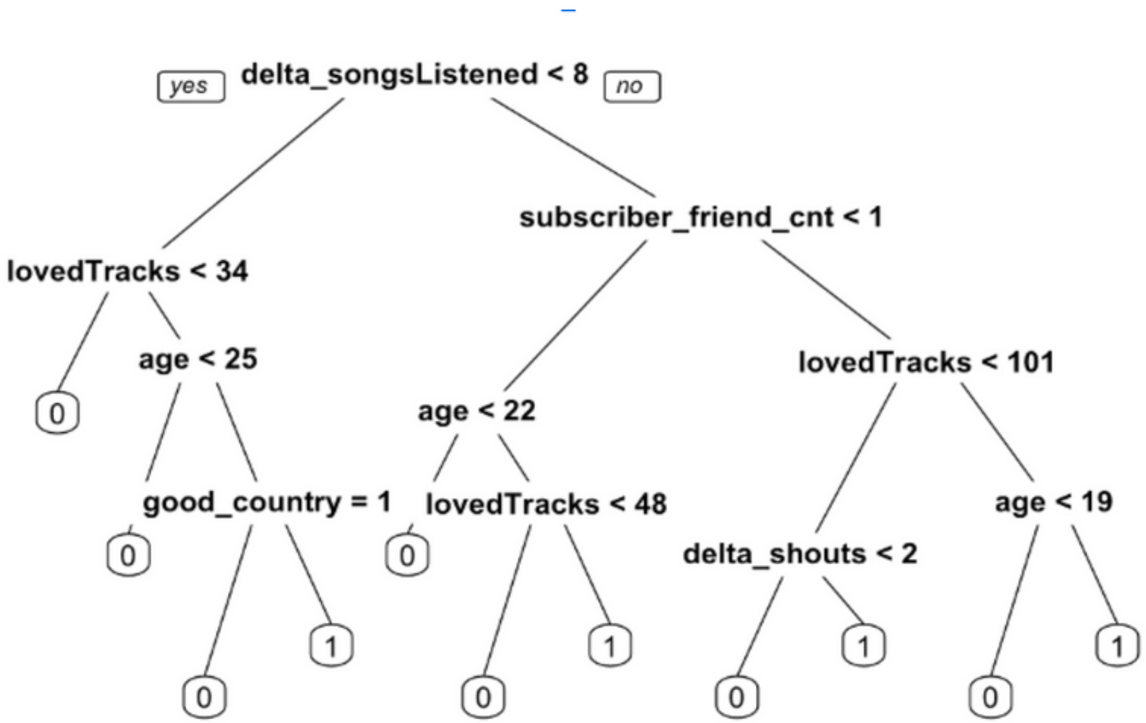


- ROC curve **ranks** the **predictive probabilities** of all the data points.
- AUC is used to **correctly rank the probability** of converting to subscribers ahead of the maintain as free users.
- AUC value should > 0.5(random model performance)
- AUC is used to predict what percentage of users has a higher predictive probability of being likely to subscribe over the not subscribers records.

Implication of the Model - Revelant features

Six most relevant features:

- Delta_songs_listened: changes in the number of listening songs
- Subscriber_friend_count: number of friends who are subscribers of the premium service
- lovedTracks: total number of different songs that the user "liked”
- Age: ages of users
- delta_shouts: changes in the number of wall posts received by the user
- good_country: countries where free usage is more limited, 1 – less limited



Implication of the Model - 4 Types of users are more likely to sbuscribe

- **Four types** of users are more likely to convert to premium subscribers

Delta_songs_listened < 8

age > 25
loved Tracks > 34
Less limited free usage coutnry

Delta_songs_listened > 8

Subscriber_friend_count < 1

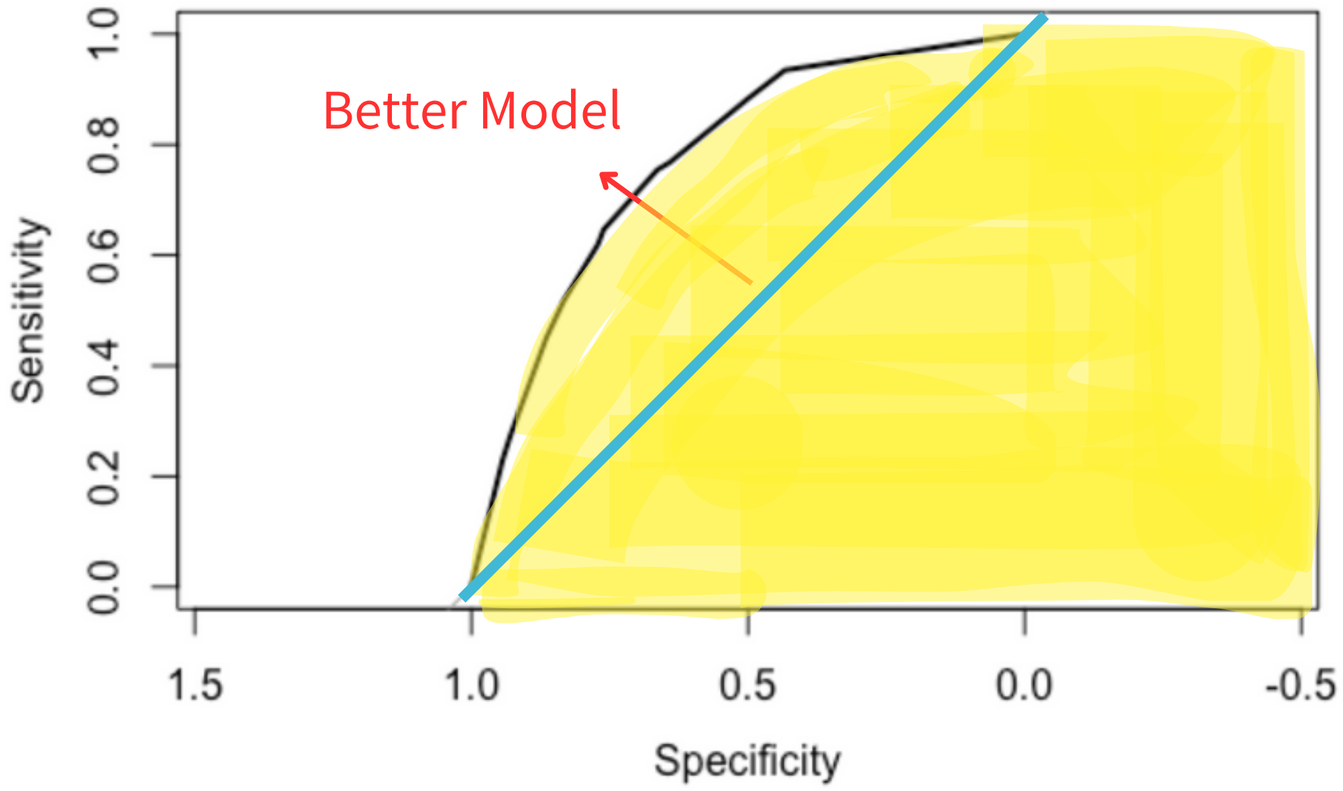
age > 22
loved Tracks > 48

Subscriber_friend_count >1

loved Tracks < 101
delta_shouts > 2

age > 19
loved Tracks > 101

Implication of the Model - Model Performance



	Decision TreeModel	BaseLine Model
AUC	**0.77**	0.5

The area under this ROC curve, **AUC**, equates to the model’ s **ability to predict classes correctly.** As a significant AUC would show that the model can achieve a high true positive rate with a correspondingly low false positive rate.

Application of the Model

Data Input

New Marketing Data

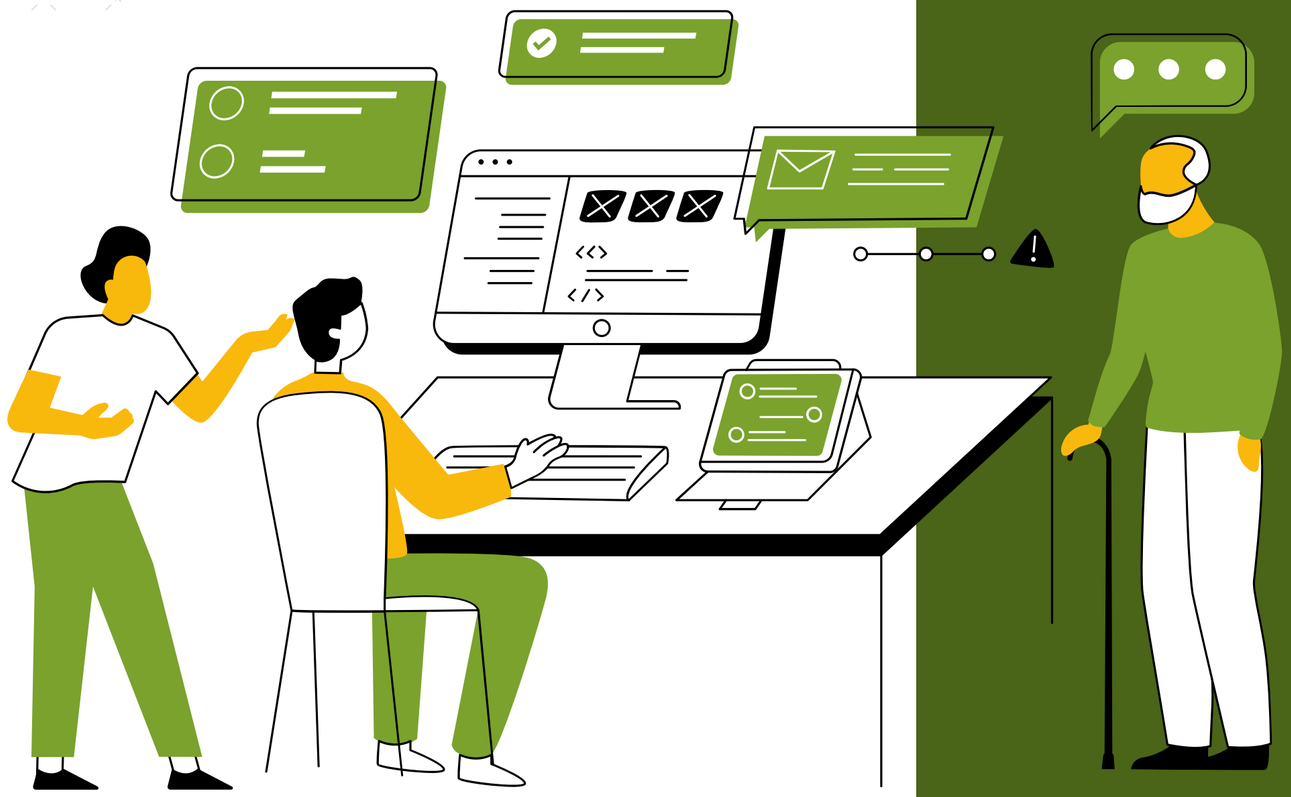
Decision Tree Model

Model Prediction

Outcome

predicted premium subscribers

+

important features of predicted premium subscribers

THANK YOU