Sixth Sense Bank

Jiayan Han
Kasandra Woo
Ryan Chen
Sean Jung

Agenda

1. Meet the Team
2. Data Mining Process Overview
3. Who are we?
4. Our Task
5. Data Understanding and Preparation
6. Our Model
7. Model Deployment
8. Q&A

The Team

Sean Jung

President

Ryan Chen

COO

Jiayan Han

Head of Data Analytics

Kasandra Woo

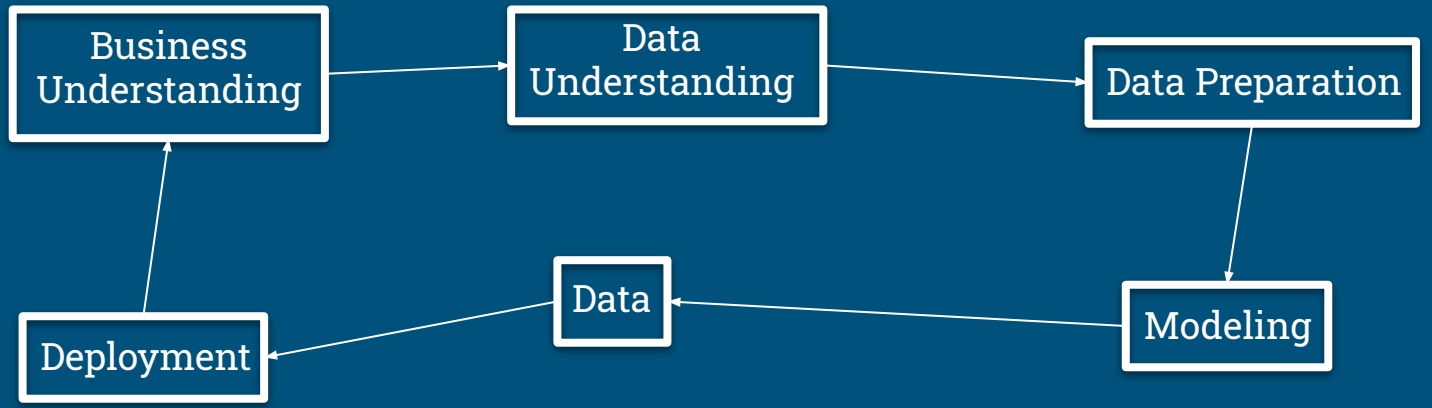
Data Analyst

Data Mining Process Overview

What actions can our bank take to impact churn?
What attributes impact churn the most?

What are the informative variables to answer our question?
What improvement can be done to current features?

Identify target variables
Identify predictor variables
Perform feature engineering



How should we implement the model?
What are some potential risks with the implementation?

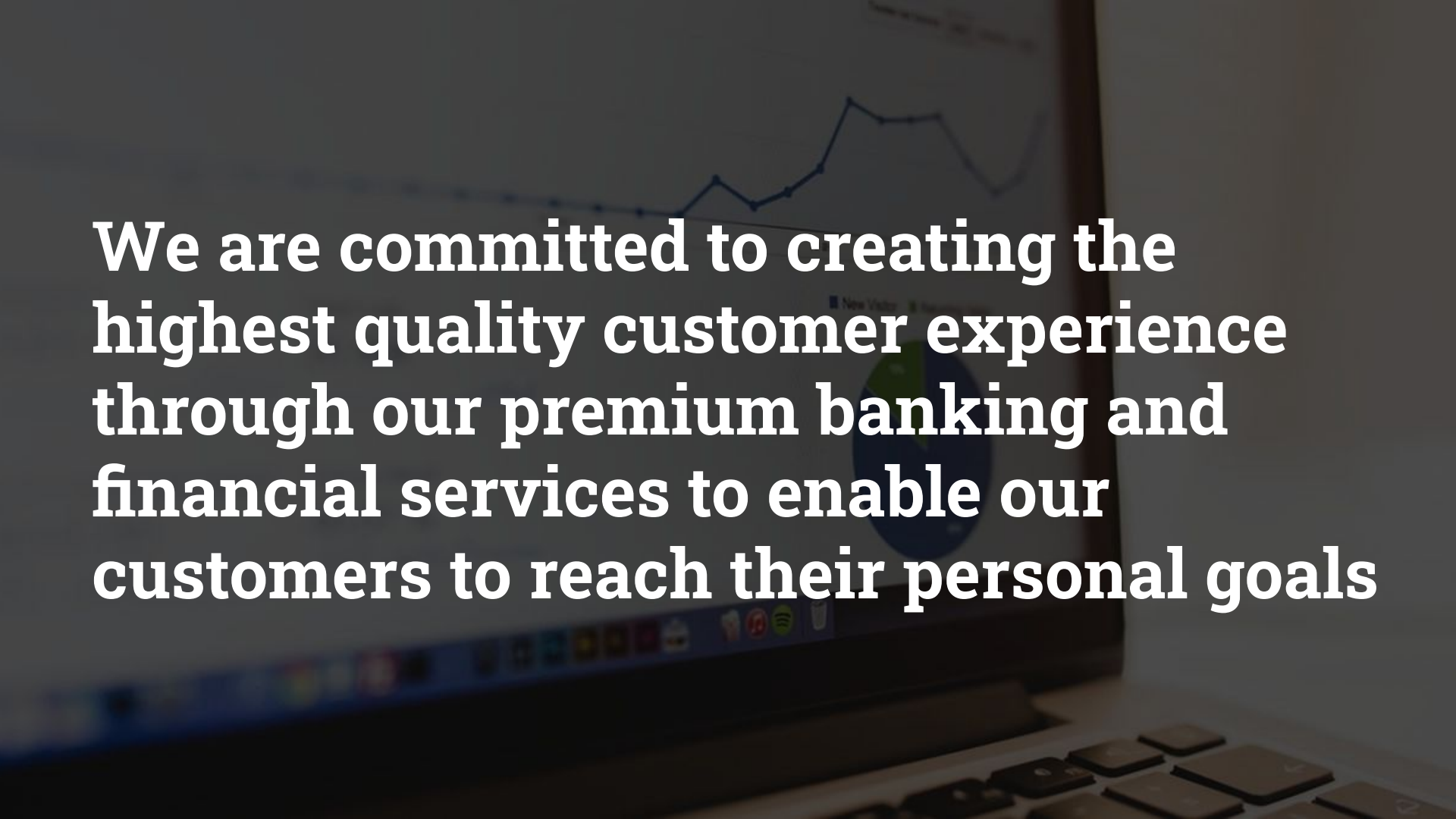
Evaluation

Confusion Matrix
Accuracy, Precision, Recall, F-1

kNN
Logistic Regression
Decision Tree

Who are we?

- Internationally recognized and trusted bank
- Located in France, Spain and Germany
- Provide world class banking and financial products and services to our customers


We are committed to creating the highest quality customer experience through our premium banking and financial services to enable our customers to reach their personal goals

Our Task

Continue to strengthen our internationally recognized brand

↓

Centered around increasing customer satisfaction and loyalty

↓

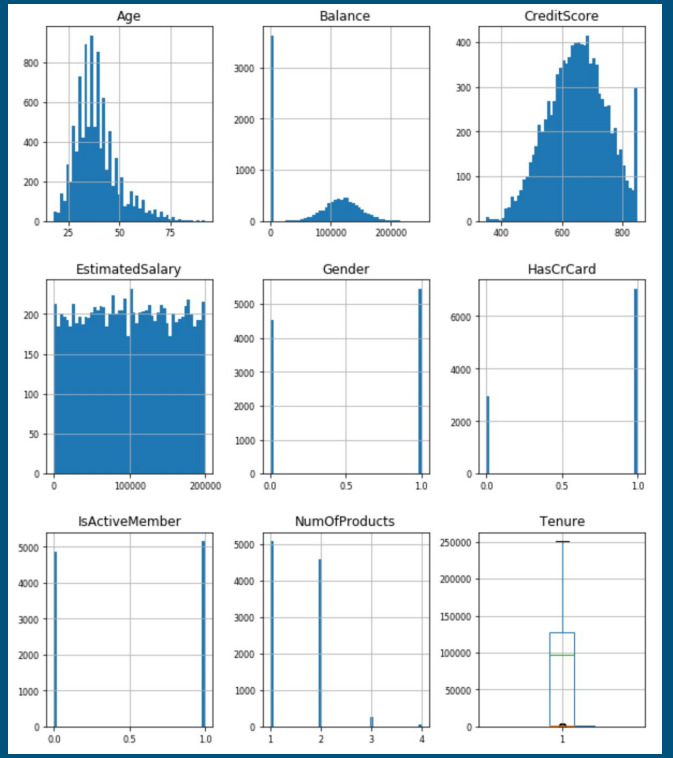
Decrease churn rate

↓

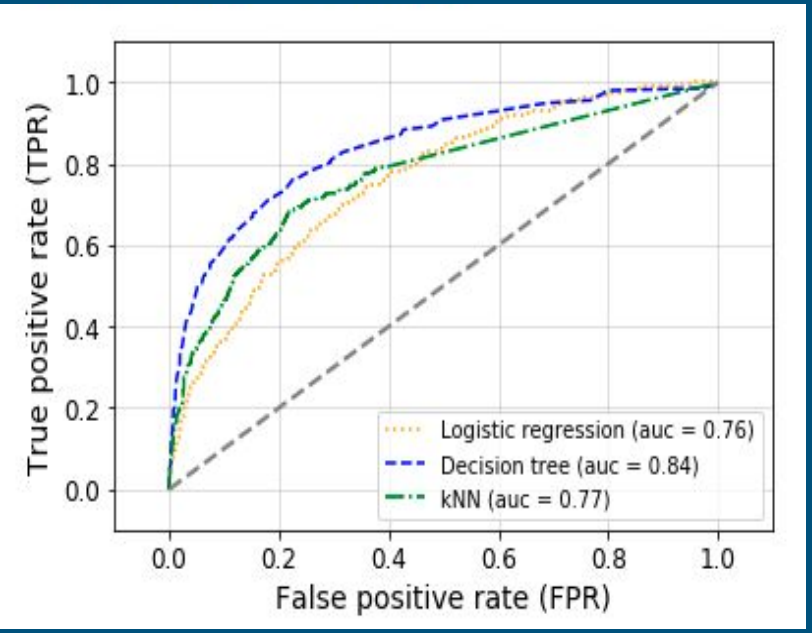
Create predictive models to predict churn and identify where we should focus our efforts

Data Understanding and Preparation

- **Transforming skewed variables**
  - Log(“*Age*”)
  - Log(“*Credit Salary*”)
- **Creating interaction terms**
  - “*Salary_balance*” = *"EstimatedSalary" * "Balance"*
  - *“tenure_product” = ‘Tenure’ * ‘NumOfProducts'*
  - *"Credit_salary" = "CreditScore" * "EstimatedSalary"*
  - *“credit_balance” = 'HasCrCard' * ‘”Balance'*
- **Encoding categorical variable**
  - *‘Geography’ column*
- **Standardizing data for kNN**



Model Comparison



	Non-nested CV Accuracy	Nested CV F1 Score
kNN	0.68	0.67
Logistic Regression	0.62	0.63
Decision Tree	**0.75**	**0.74**

AUC: 0.74 (+/- 0.02) [Logistic regression]
AUC: 0.83 (+/- 0.02) [Decision tree]
AUC: 0.76 (+/- 0.01) [kNN]

Decision Tree Model

Optimal Model:

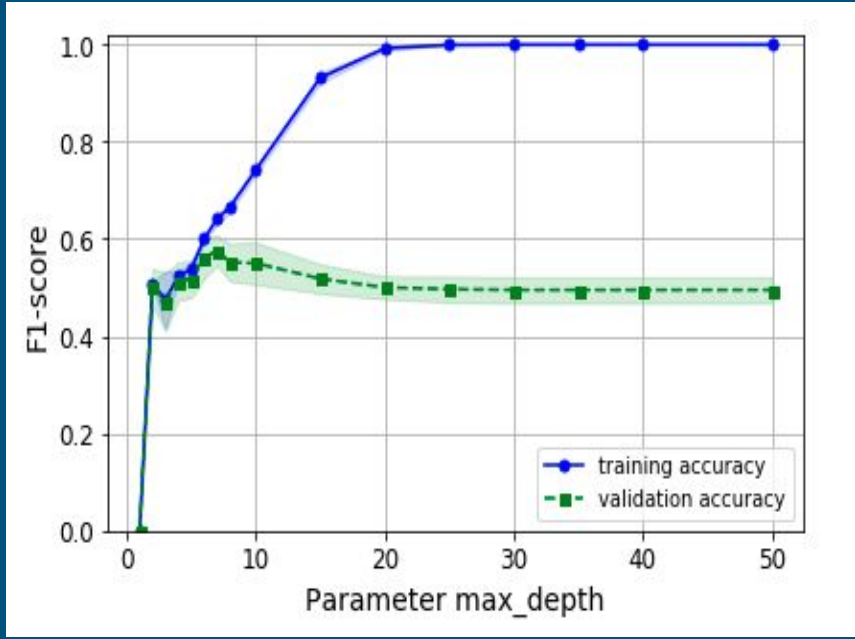
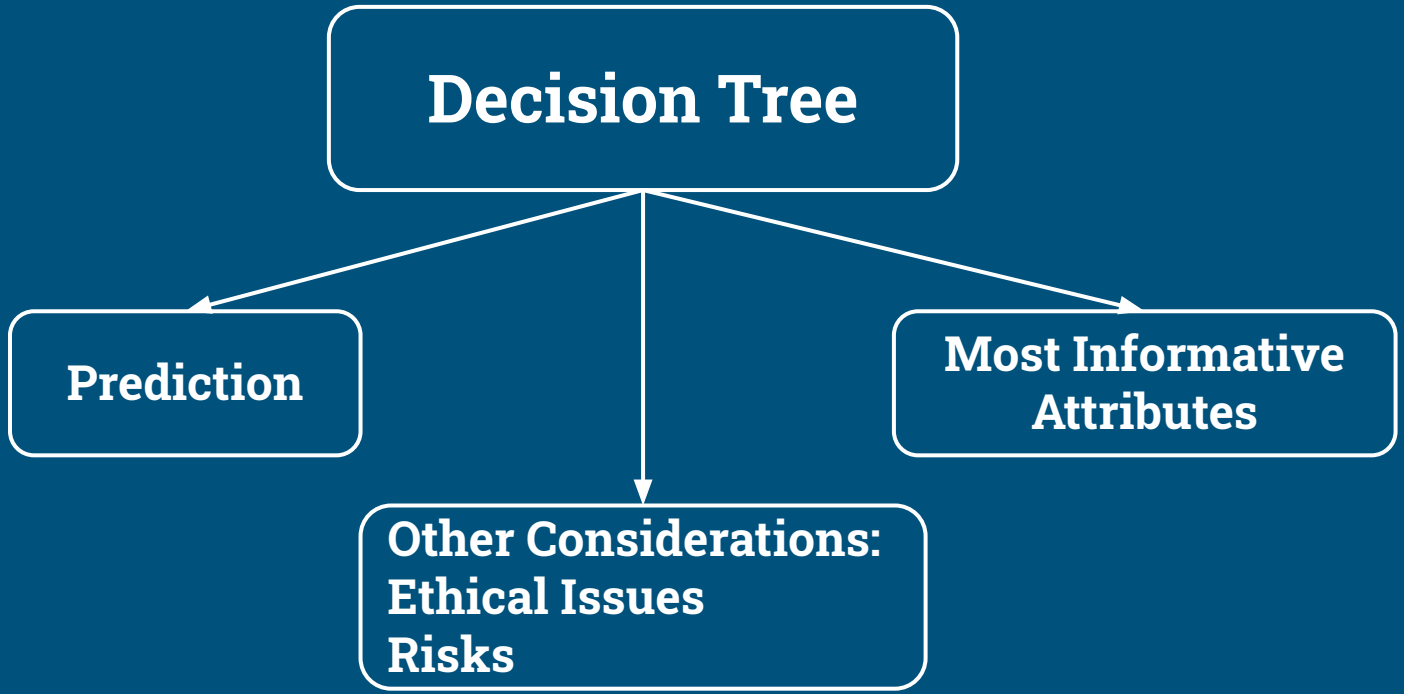
criterion='gini', max_depth=8,
min_samples_leaf=2, min_samples_split=25

- Accuracy: 0.85
- Precision: 0.71
- Recall: 0.47

Confusion Matrix:

	Actual Negative	Actual Positive
Actual Negative	6035	321
Actual Positive	875	769

Fitting Graph:



Model Deployment