PREDICTING UBER AND LYFT SURGES TO INCREASE TAXI-CAB COMPETITION

BOSTON SURGE PRICE DETECTION

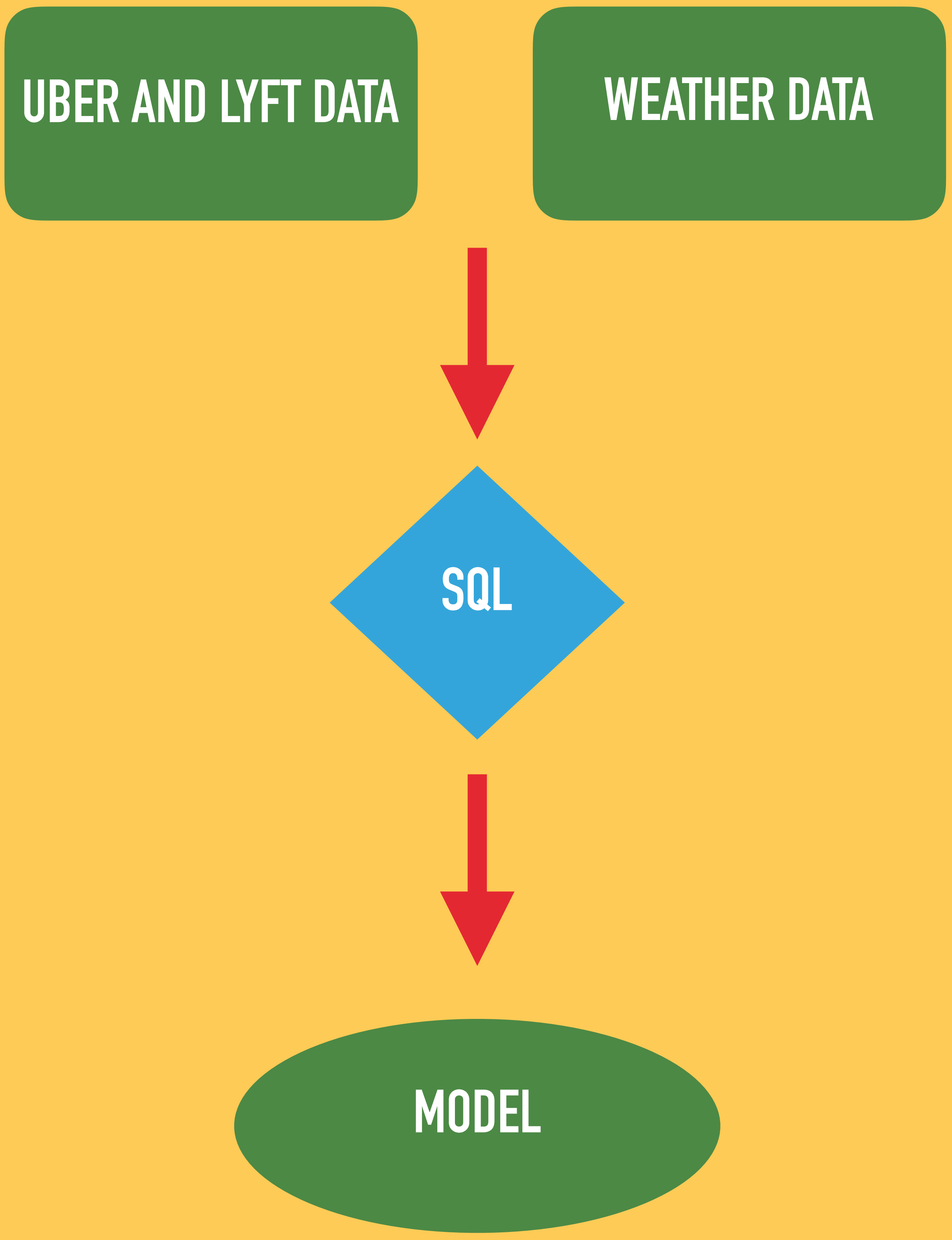
Samad Memon

PREMISE

- Traditional cabs have difficulty competing against Uber/Lyft
- One advantage of cabs: no “surge” pricing
- Goal: Create service to alert cabs when/where surge pricing is expected

DATA

- Features: time, location, and weather
- Target: Surge or No Surge
- 18 days' worth of Uber and Lyft data in Boston
  - ~700k rides
- Hourly weather forecasts



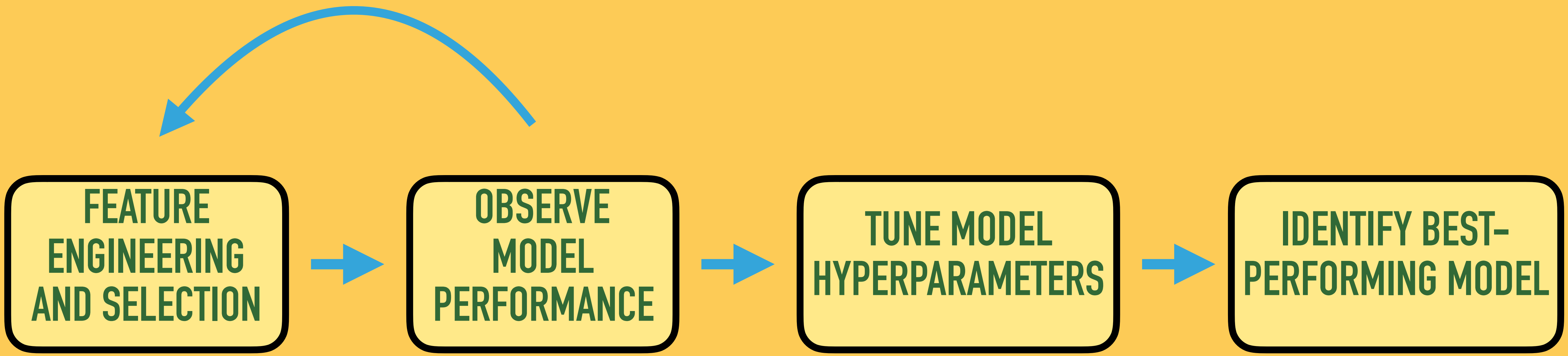
MODEL SELECTION

- Interpretability is a high priority
- Data is linearly inseparable
- Candidate models:
  - Decision Tree
  - Random Forest
  - Naive Bayes

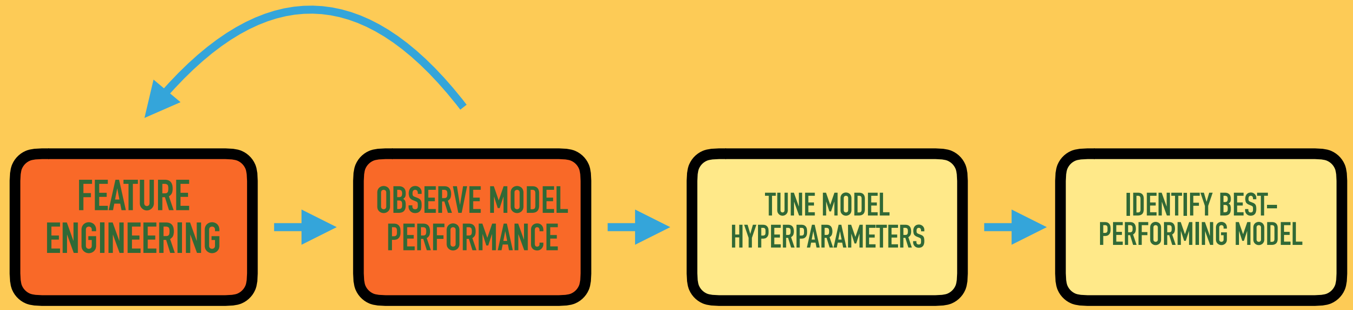
UNDERSTANDING PERFORMANCE METRICS

- Accuracy is a poor metric
    - Only 0.03% of rides had surge pricing
- $Precision = \dfrac{Detected\,Surge}{Detected\,Surge + Falsely\,Detected\,Surge}$
- $Recall = \dfrac{Detected\,Surge}{Detected\,Surge + Undetected\,Surge}$
- Maximizing precision makes the most business-sense

MODELING PROCESS



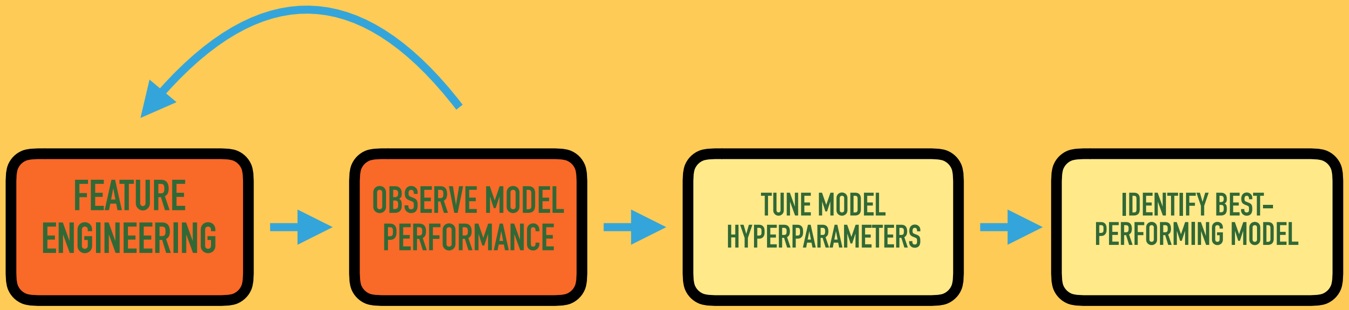
BASELINE MODEL



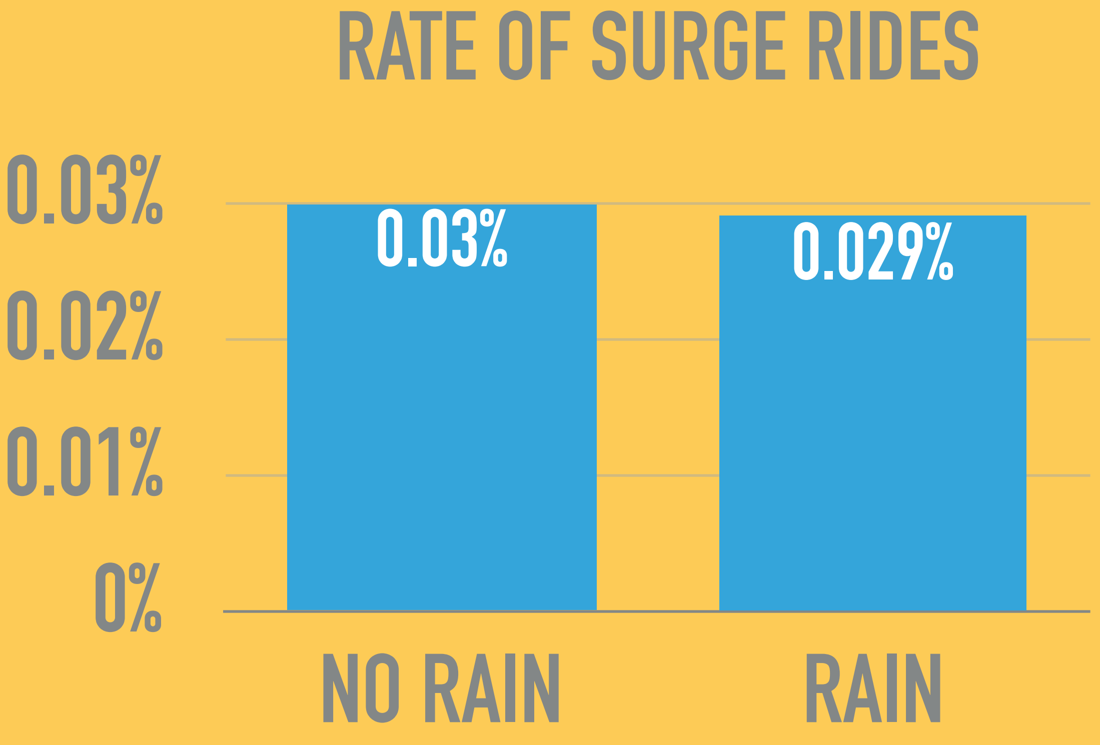
- Baseline model results:

	Accuracy	Recall	Precision	F1
Decision Tree	95.16%	15.61%	15.08%	15.34%
Random Forest	96.02%	5.10%	9.88%	6.72%
Naive Bayes	66.07%	100%	7.65%	14.22%

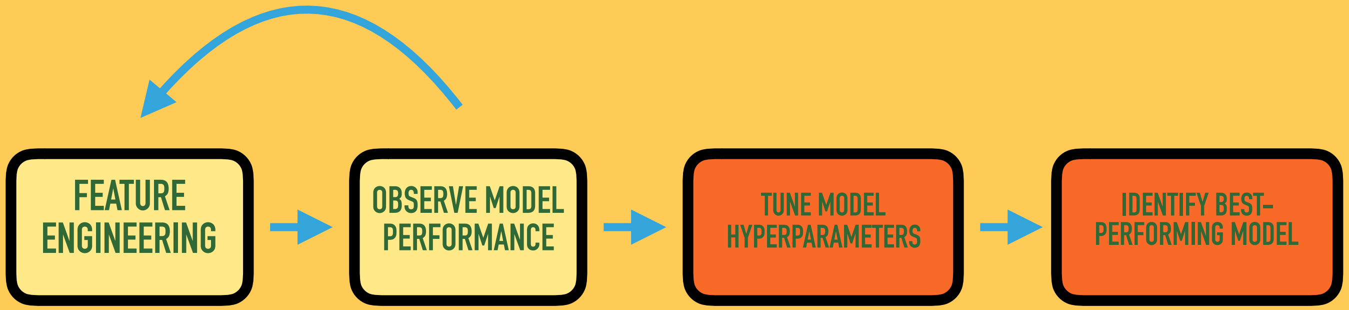
FEATURE ENGINEERING



- Recent rides
- Cyclical time
- Boston sports team game days
- Weather observation:

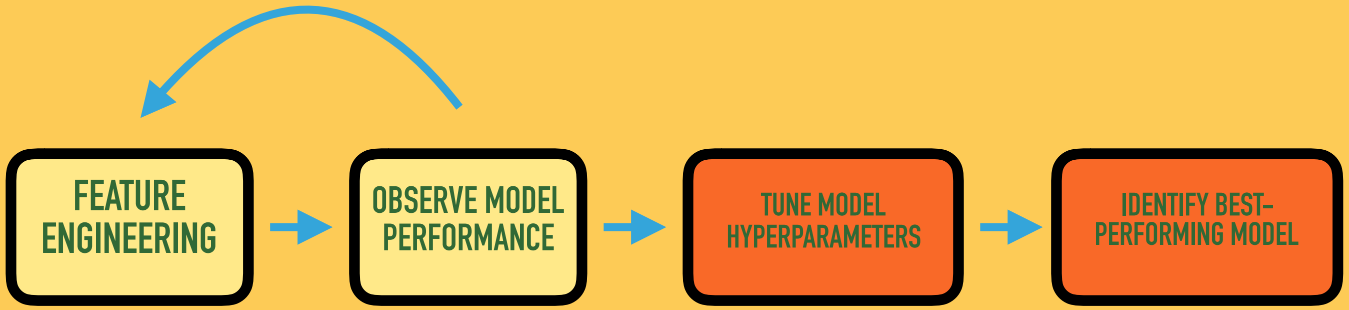


COMPARING MODEL PERFORMANCE

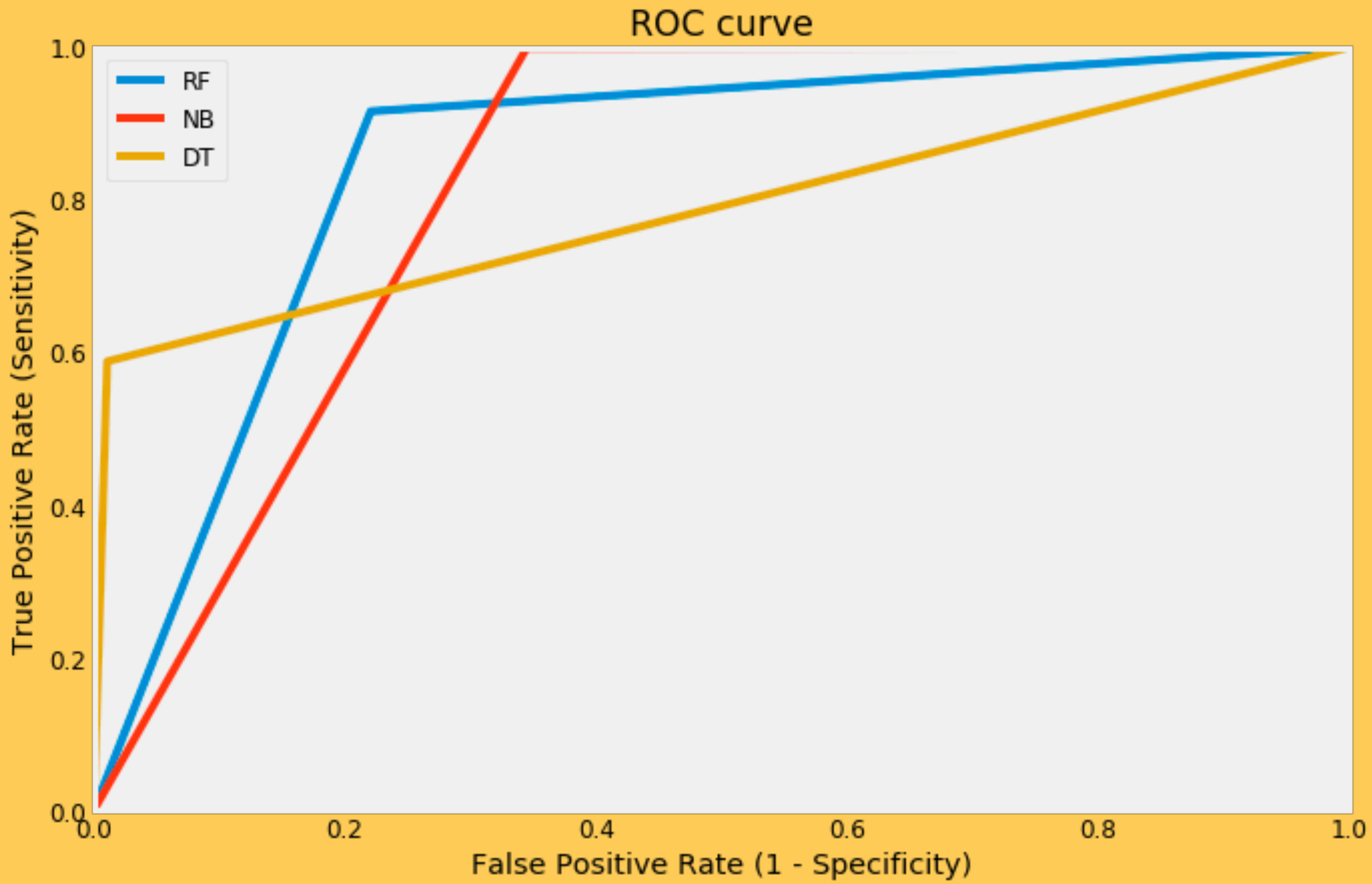


	Accuracy	Recall	Precision	F1
Decision Tree	97.70%	58.85%	63.17%	60.93%
Random Forest	78.35%	91.62%	11.53%	20.48%
Naive Bayes	66.57%	99.88%	8.33%	15.38%

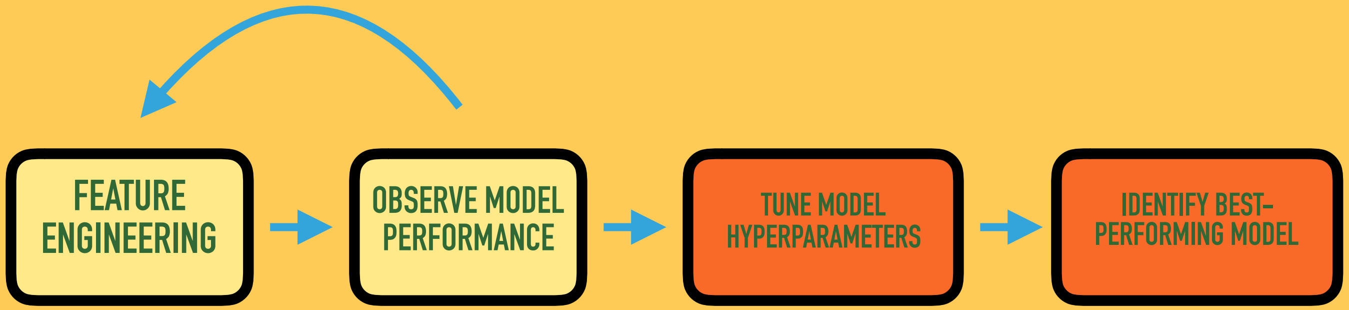
COMPARING MODEL PERFORMANCE

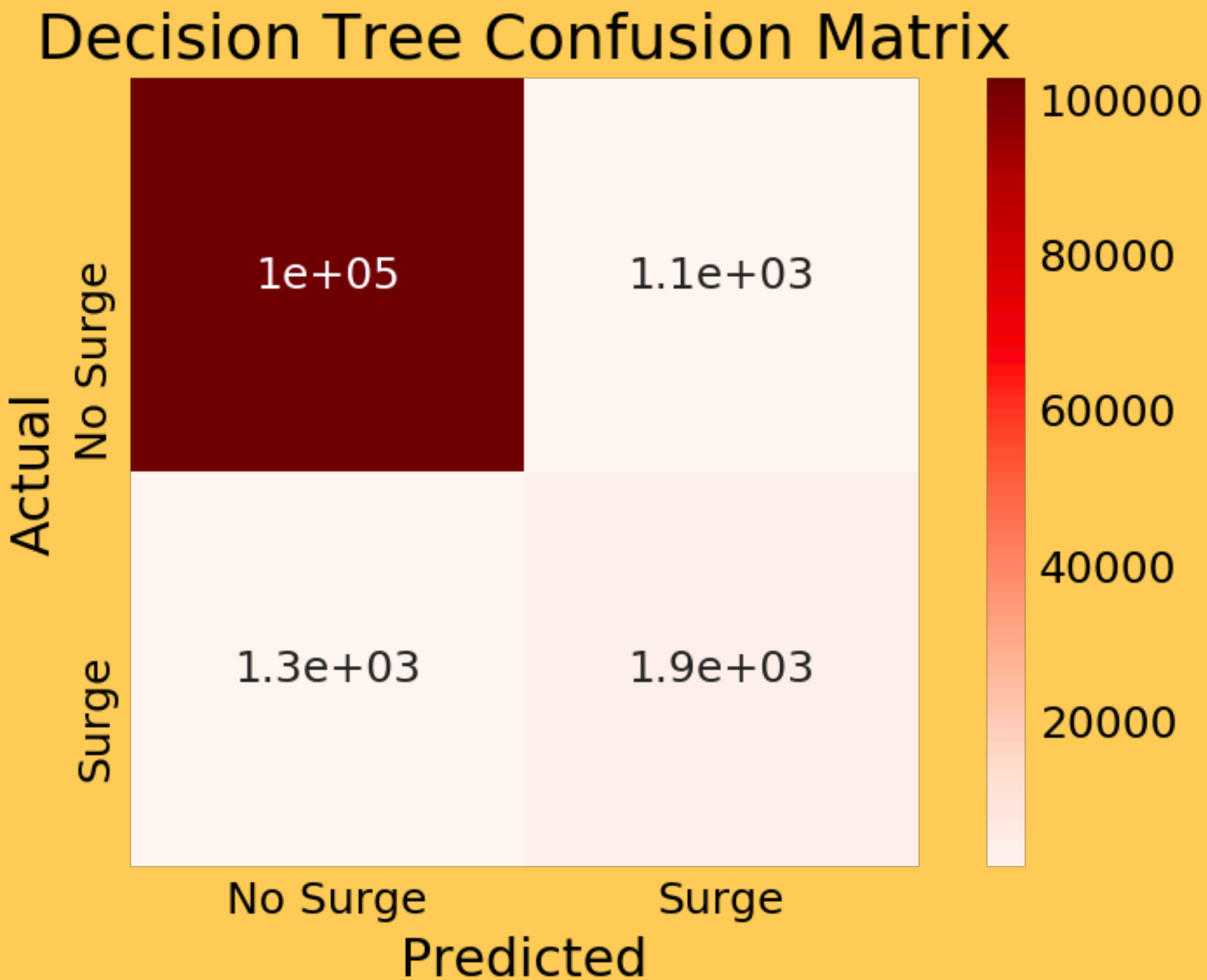


- For our business, it is essential to minimize FPR



DECISION TREE CONFUSION MATRIX





CONCLUSIONS AND FUTURE WORK

- Decision tree is the best model for the objective
  - Minimizes False Positive Rate, maximizes precision
- More data required
- Improve hyper-parameters
- Consider changing business model