

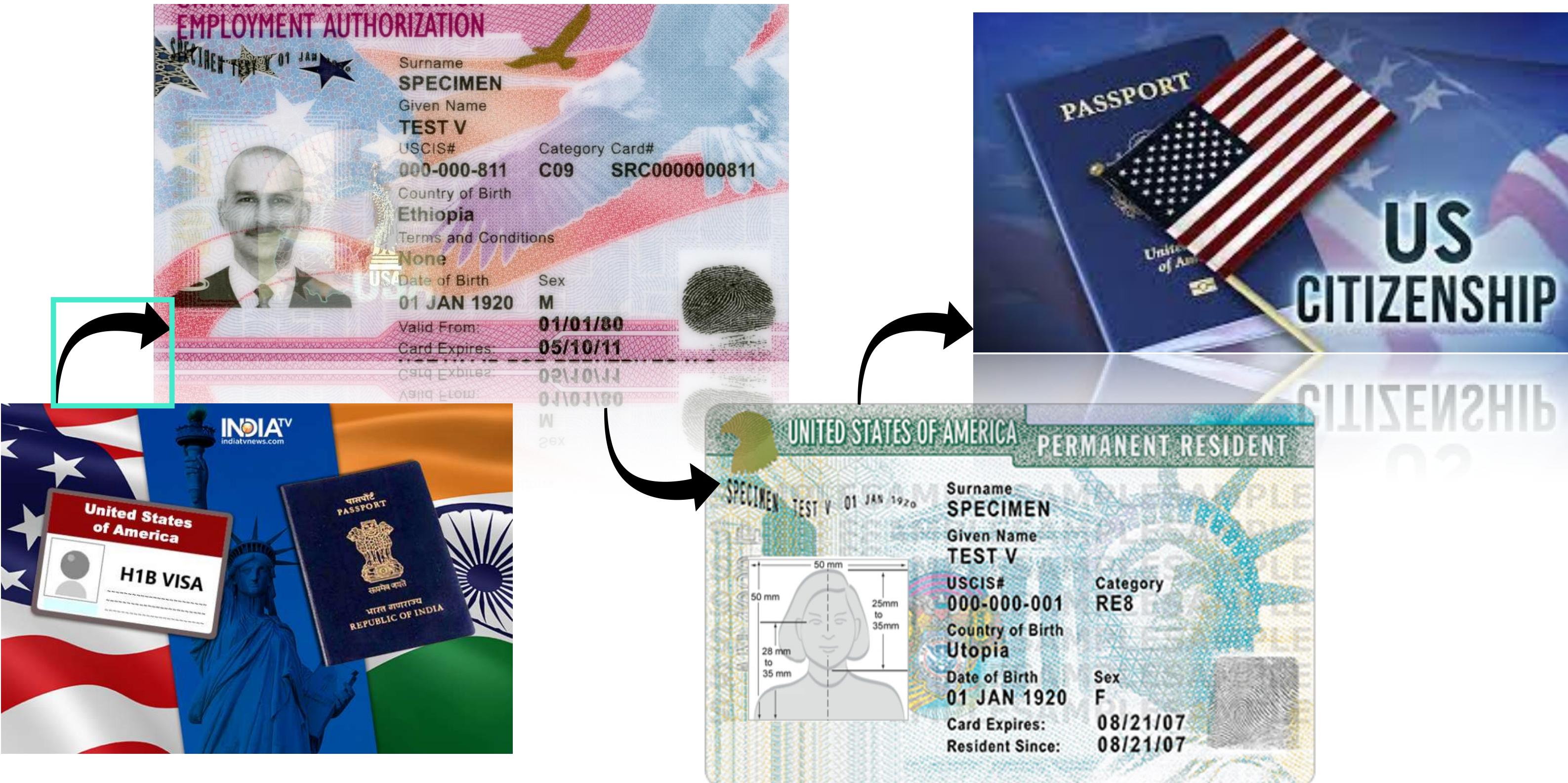
# U.S. PERMANENT LABOR CERTIFICATION

Predicting application results through  
Machine Learning Classification

JON YU - AUGUST 6, 2021

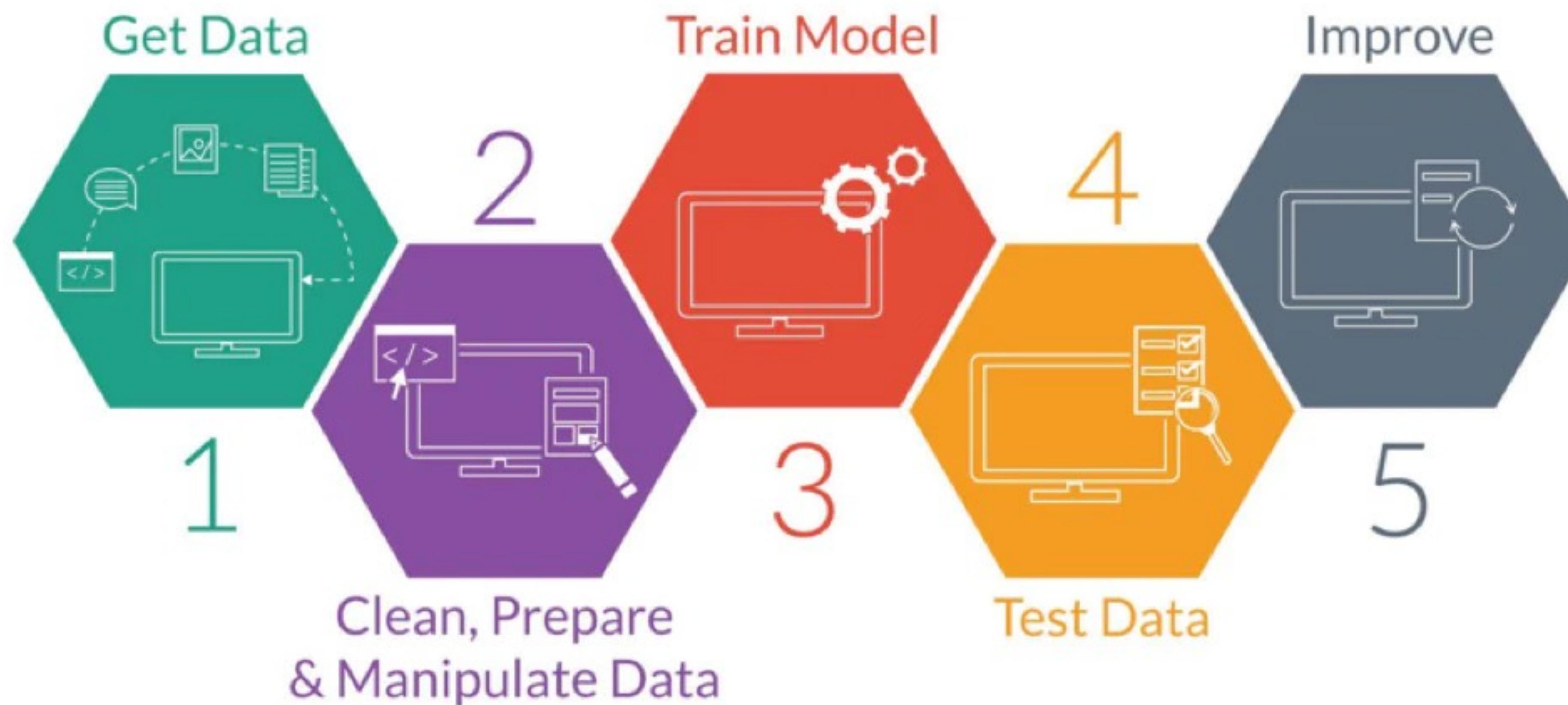


# INTRODUCTION



- Permanent Labor Certification serves as a stepping stone
- Eliminates recurring anxiety with temporary (H1B, TN, etc.) visa renewal for skilled foreign workers
- Motivation: provide employers and employees alike a predictive metric to determine risk of Perm Labor Cert application being declined

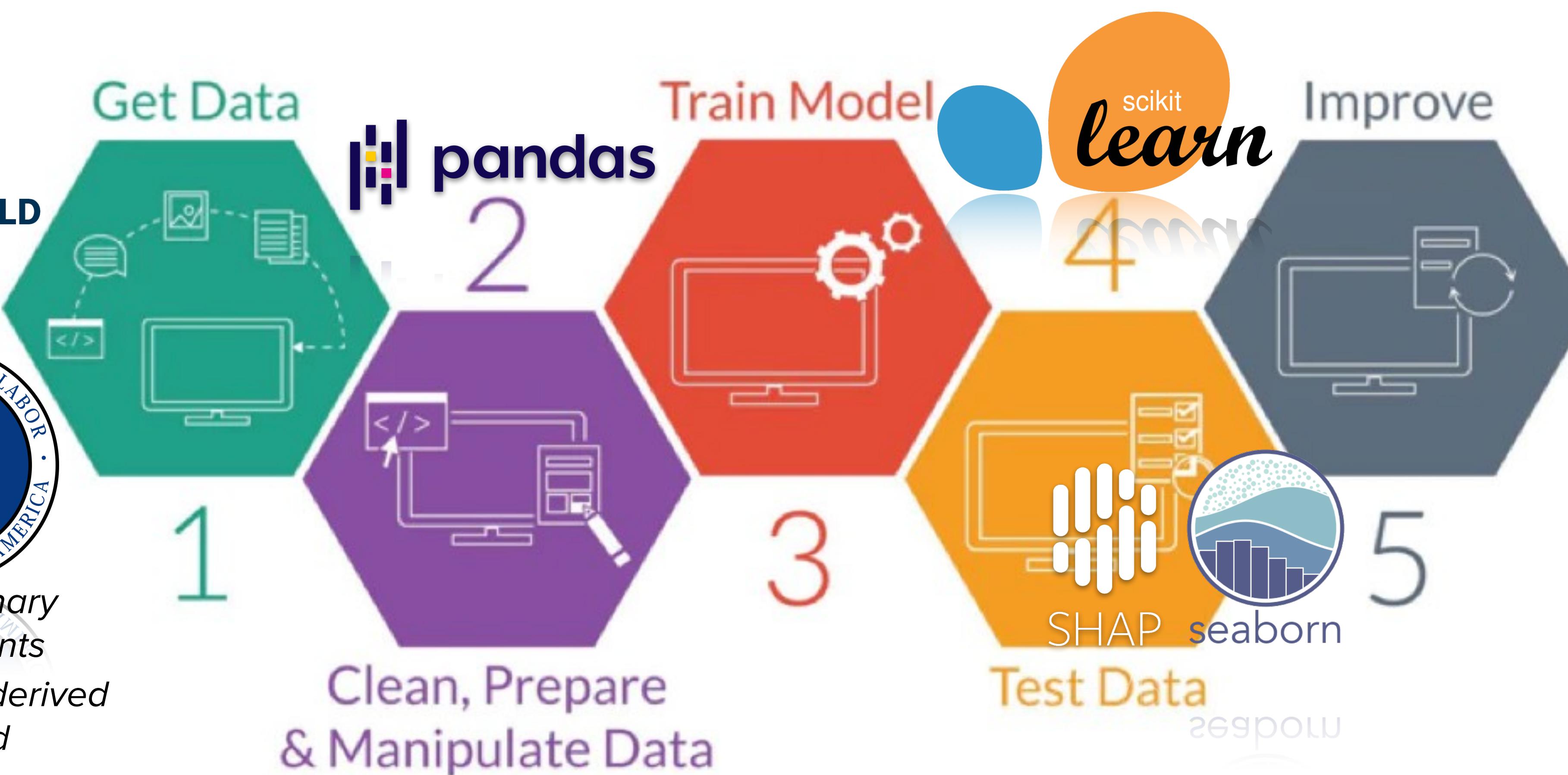
# DATA & METHODOLOGY



# DATA & METHODOLOGY



130k primary  
data points  
19 features derived  
by end



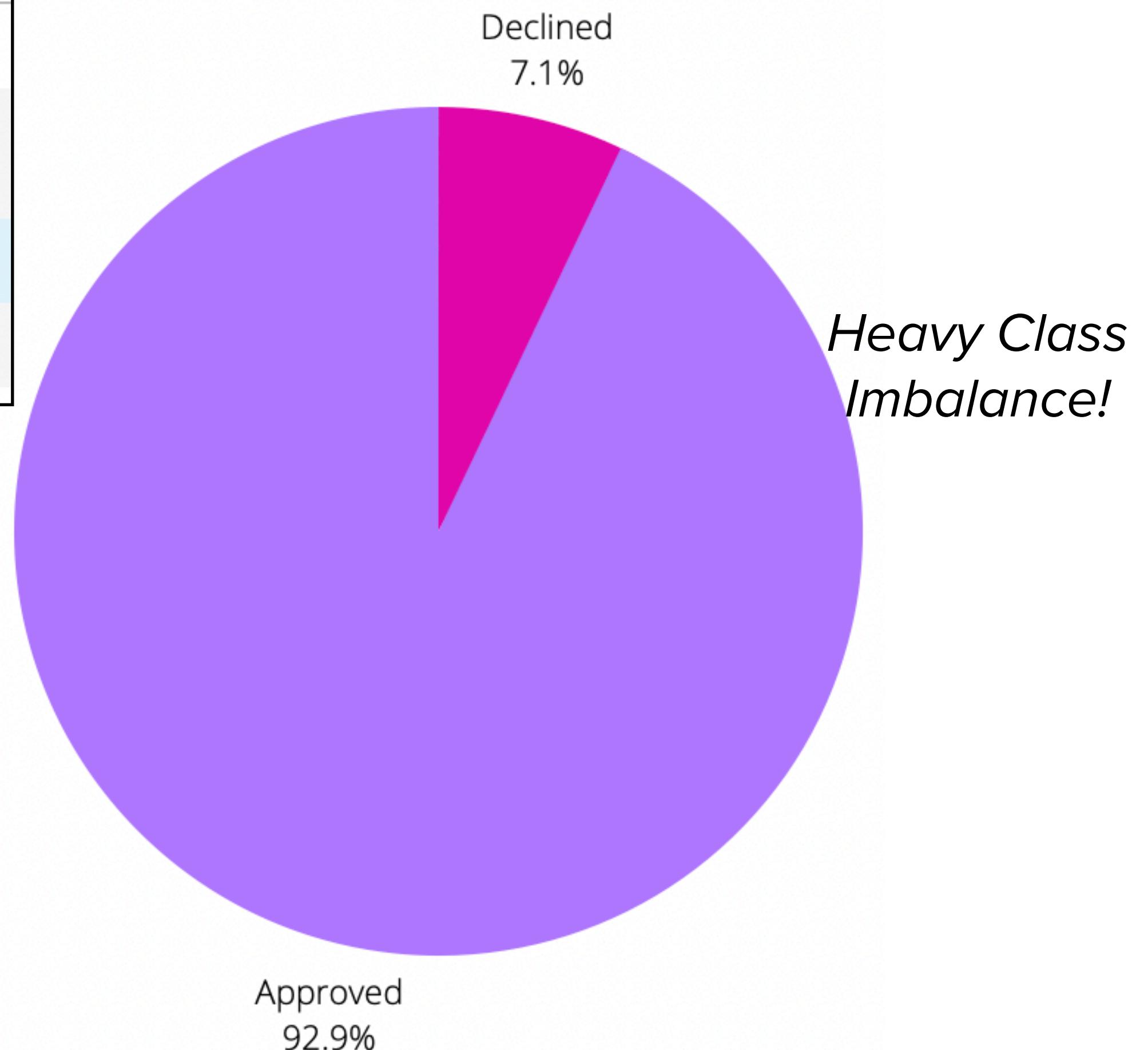
# DATA & METHODOLOGY

- True Class = application denied
- False Class = application approved
- Goal: Minimize FALSE NEGATIVE  
(Maximize Recall)

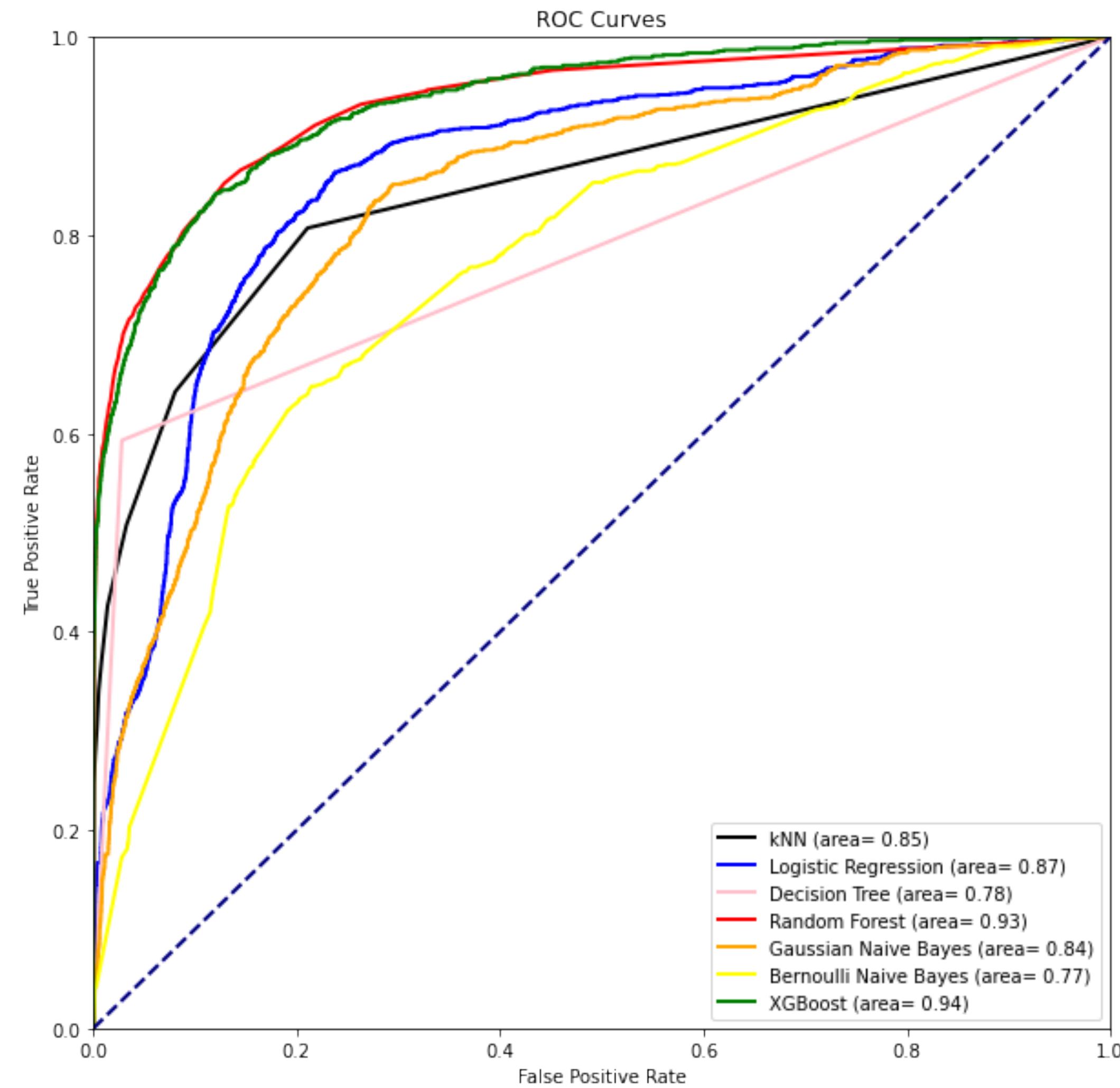
$$\text{Recall} = \frac{tp}{tp + fn}$$

CASE_STATUS	CASE_NUMBER
Certified	85218
Certified-Expired	32953
Denied	6517
Withdrawn	5308

Breakdown of Processed Applications



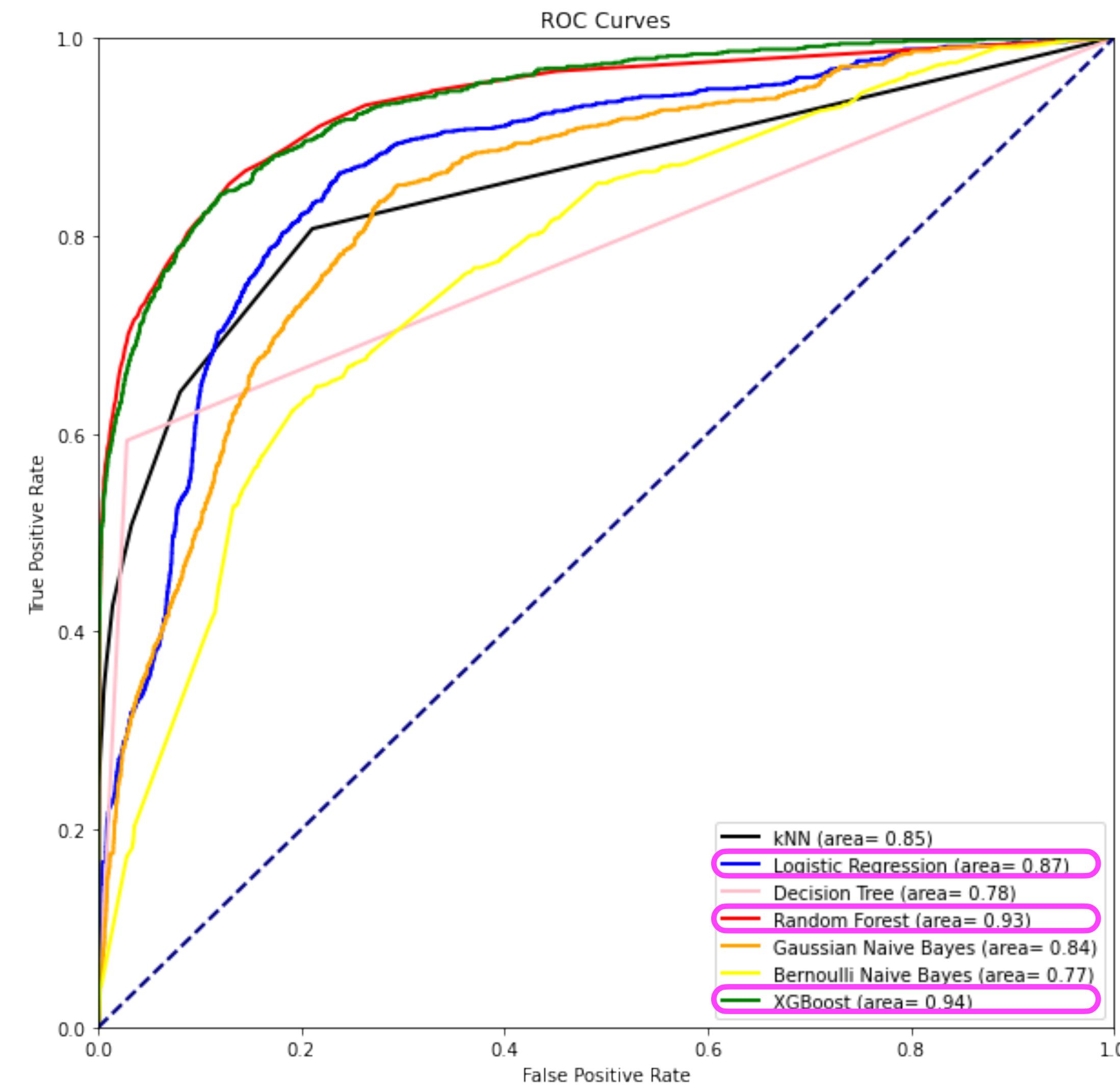
# RESULTS



# RESULTS

## Top 3 Models (by AUC)

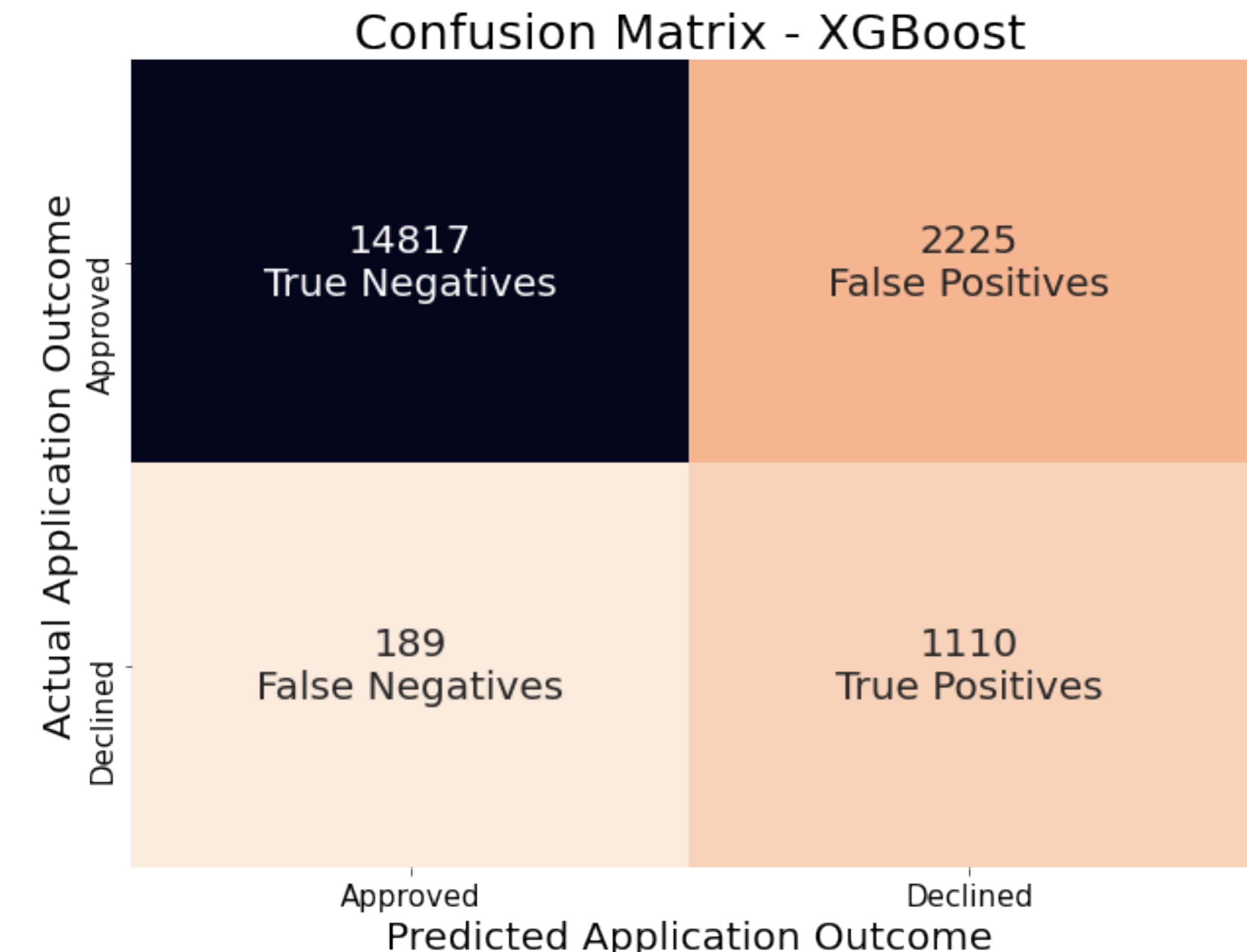
- 1) XGBoost (0.94)
- 2) Random Forest (0.93)
- 3) Logistic Regression (0.87)



# RESULTS

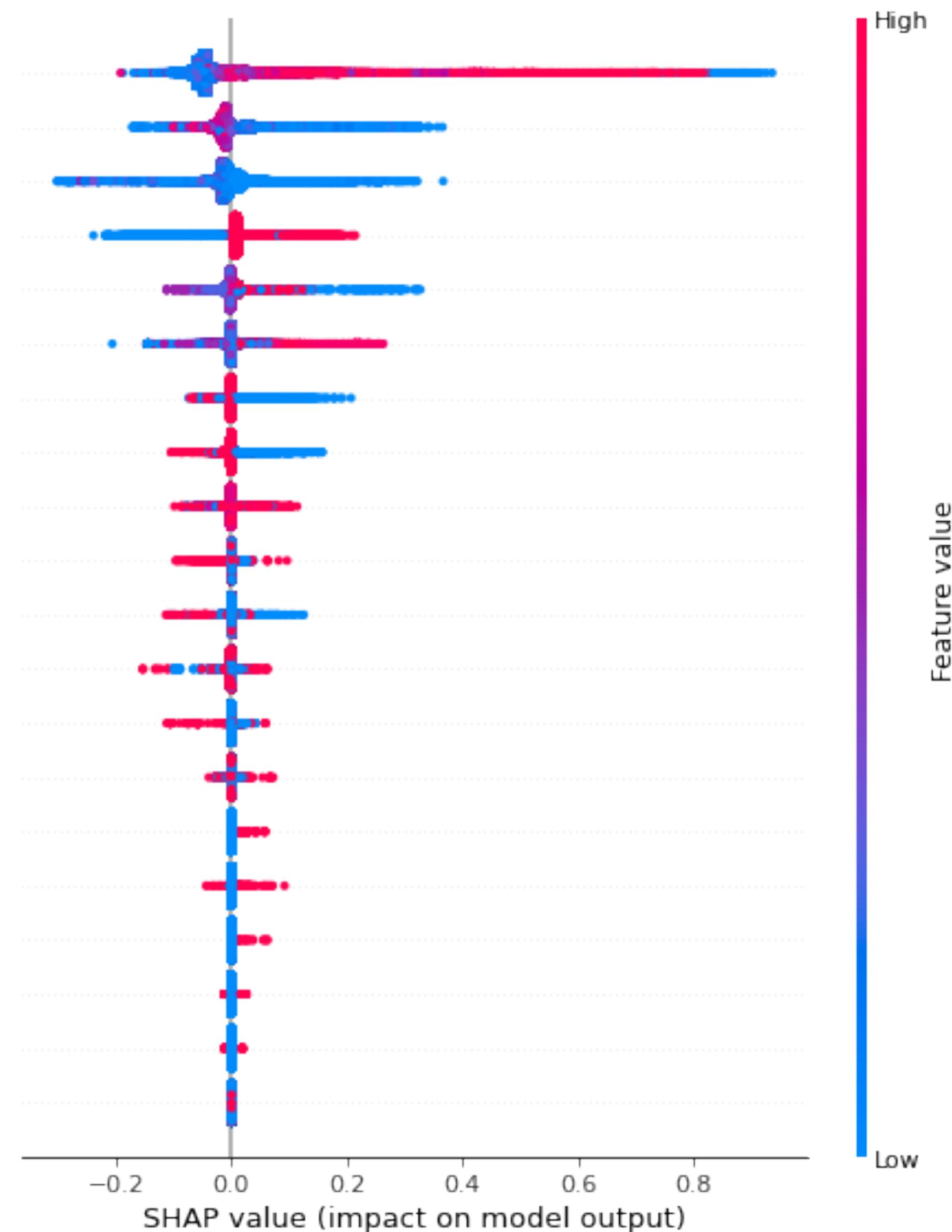
	Baseline	Tuned XGBoost
AUC	0.87	0.93
F1	0.39	0.48
Recall	0.79	0.85

	Validation	Test
AUC	0.93	0.93
F1	0.46	0.48
Recall	0.83	0.85

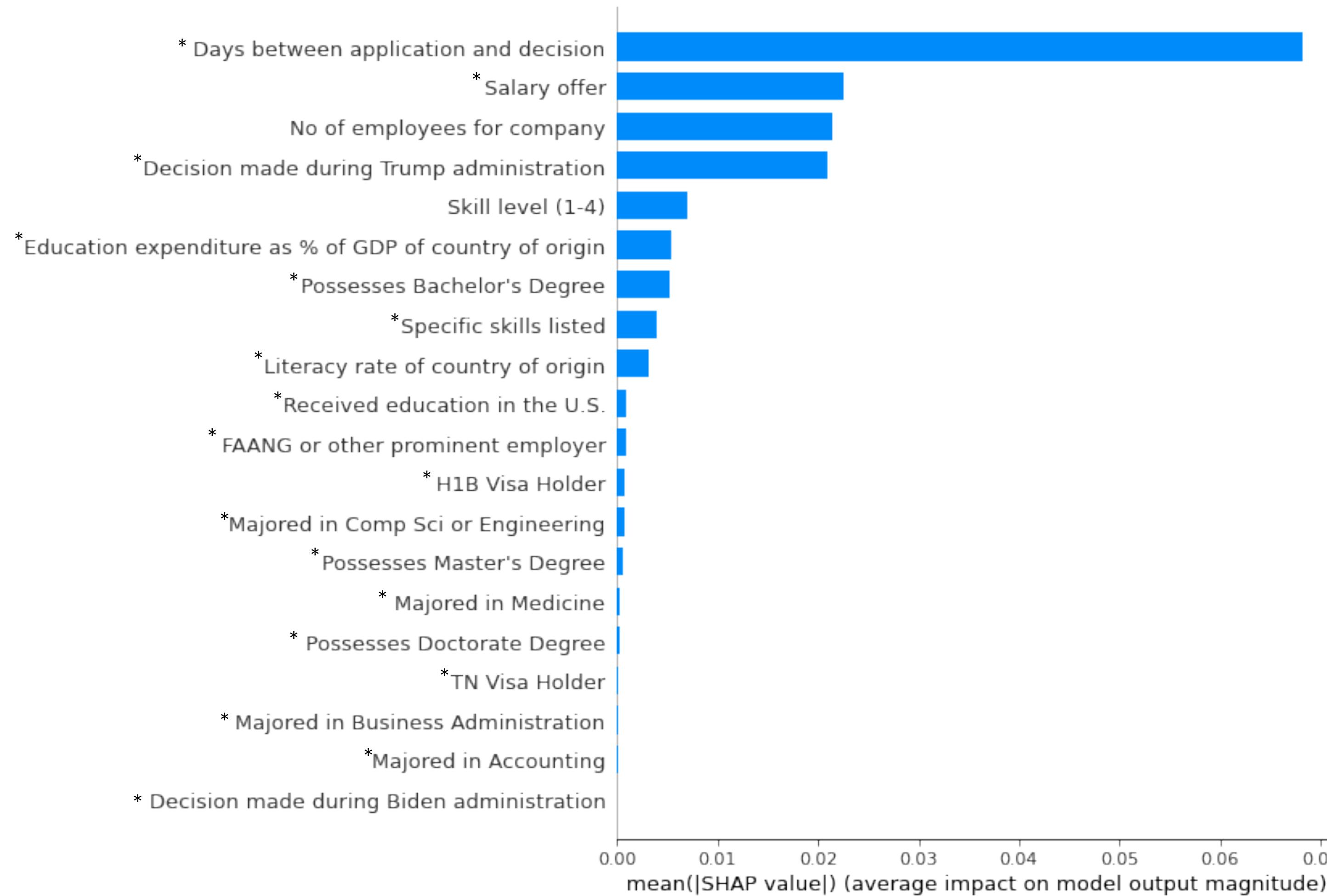


# RESULTS

Days between application and decision  
Salary offer  
No of employees for company  
Decision made during Trump administration  
Skill level (1-4)  
Education expenditure as % of GDP of country of origin  
Possesses Bachelor's Degree  
Specific skills listed  
Literacy rate of country of origin  
Received education in the U.S.  
FAANG or other prominent employer  
H1B Visa Holder  
Majored in Comp Sci or Engineering  
Possesses Master's Degree  
Majored in Medicine  
Possesses Doctorate Degree  
TN Visa Holder  
Majored in Business Administration  
Majored in Accounting  
Decision made during Biden administration



# CONCLUSIONS



- XGBoost offers best classification metrics with **hyper parameter tuning**
- Biggest contributors to application outcome are
  - 1) Days elapsed between application and decision (interpretive)
  - 2) **Salary offer (predictive):**  
Higher salary = increased likelihood of acceptance
  - 3) **Number of employees at company (predictive):**  
Larger company = increased likelihood of acceptance
  - 4) Whether or not decision was made during Trump administration (interpretive)
  - 5) **Skill level of applicant (predictive):**  
Higher skill level = increased likelihood of acceptance
  - 6) Education expenditure of applicant's country of origin (interpretive, out of control):  
Higher spending = decreased likelihood of acceptance (?)
  - 7) **Possesses Bachelor's Degree & specific skill set = increased likelihood of acceptance**

\* Derived through Feature Engineering

A detailed statue of the Statue of Liberty, showing her from the waist up, holding the tablet and torch.

# FUTURE WORK

- Seek out more continuous feature variables from World Bank Open Data
- Check models against application data from Obama / Bush administrations
- Deploy machine learning model to web app via Streamlit Share using only *predictive* features

