Mid Project

SHARK TANK AN ENTREPRENEUR - INVESTORS RESEARCH

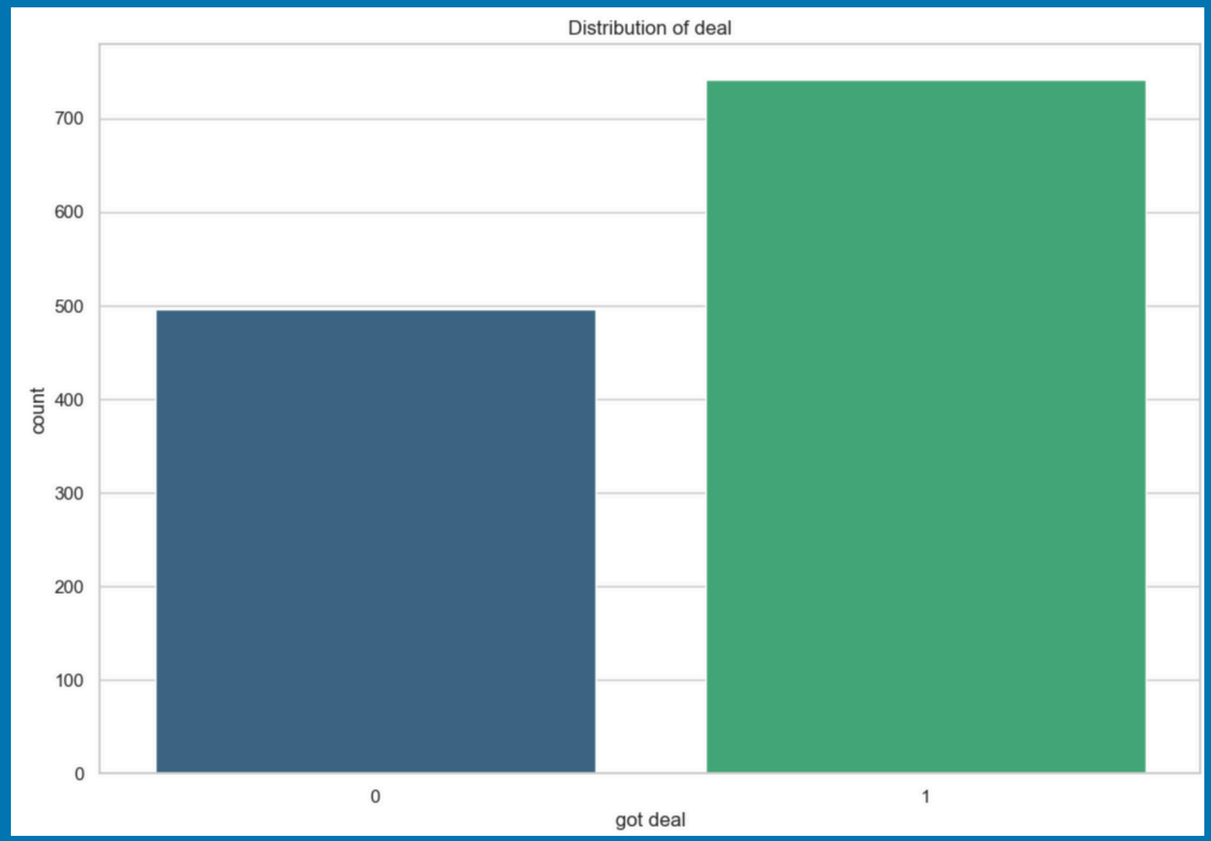
Elham Allahbakhshi

Project Overview

- industry
- gender
- pitchers_city
- pitchers_state
- original_ask_amount
- original_offered_equity
- valuation_requested
- total_deal_amount
- total_deal_equity
- deal_valuation
- number_of_sharks_in_deal
- investment_amount_per_shark
- equity_per_shark
- got_deal

Project goal:
Identifying factors that increases the probability of getting deals

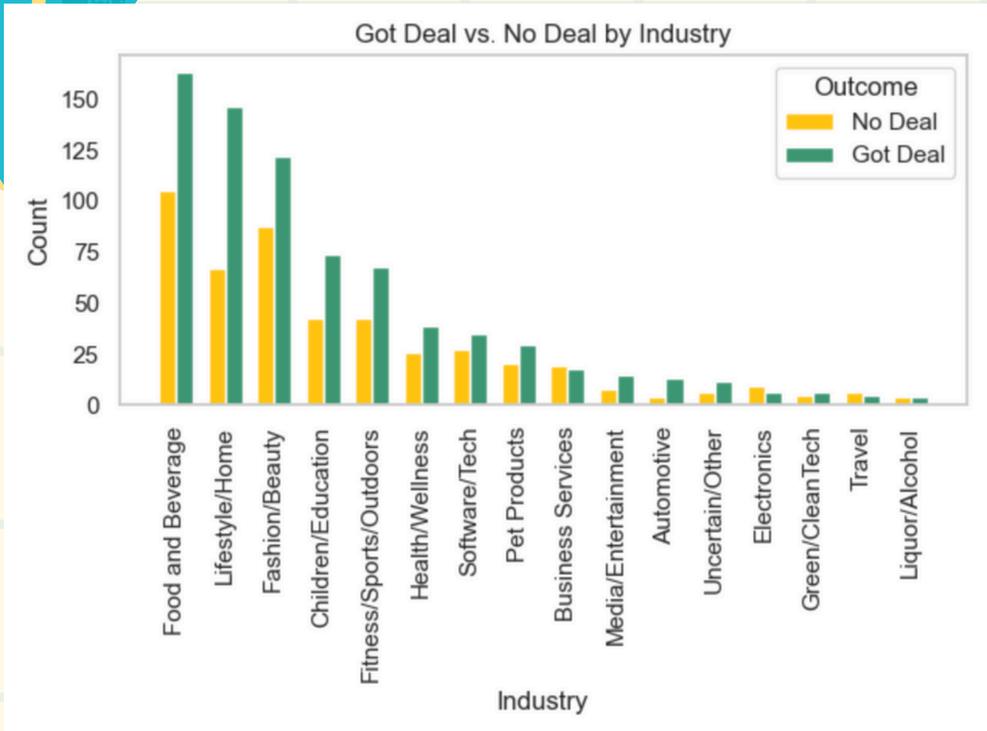
Distribution of number of deals

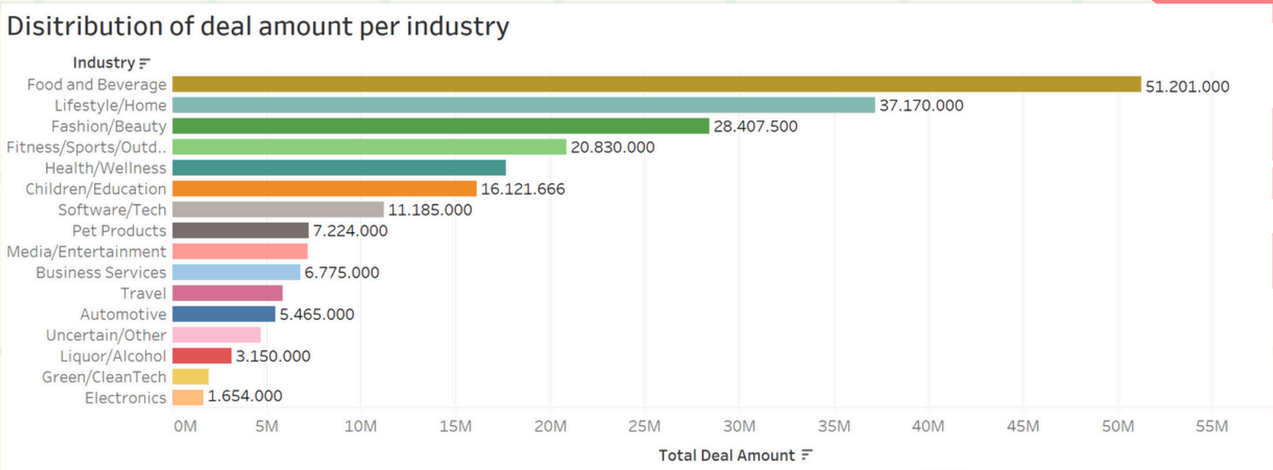


Out of 1238:

Got deal 756
Not got deal 482

Industry





Total deal amount / industry

Food and Beverage $51 million deal amount
Lifestyle/Home $37 million deal amount
Fashion/Beauty $28 million deal amount

Got deal/ industry

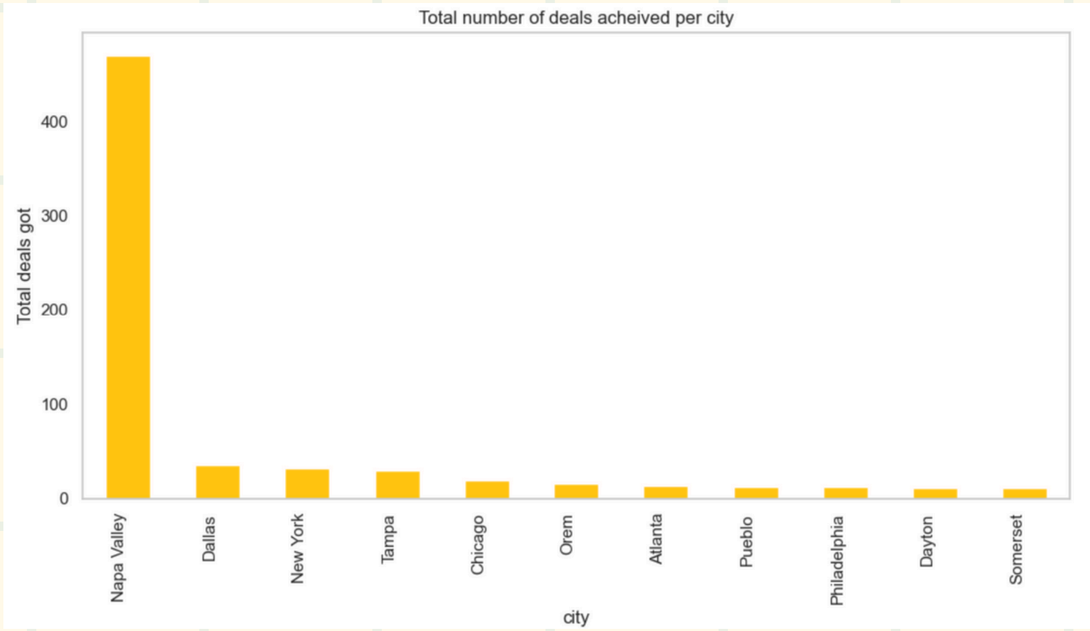
- Food and Beverage 163 deal, 105 no deal
- Lifestyle/Home 146 deal, 67 no deal
- Fashion/Beauty 122 deal, 88 no deal

HYpothesis testing

H0: There is no significant association between got deal and Industry.
H1: There is significant association between got deal and Industry.

CITY

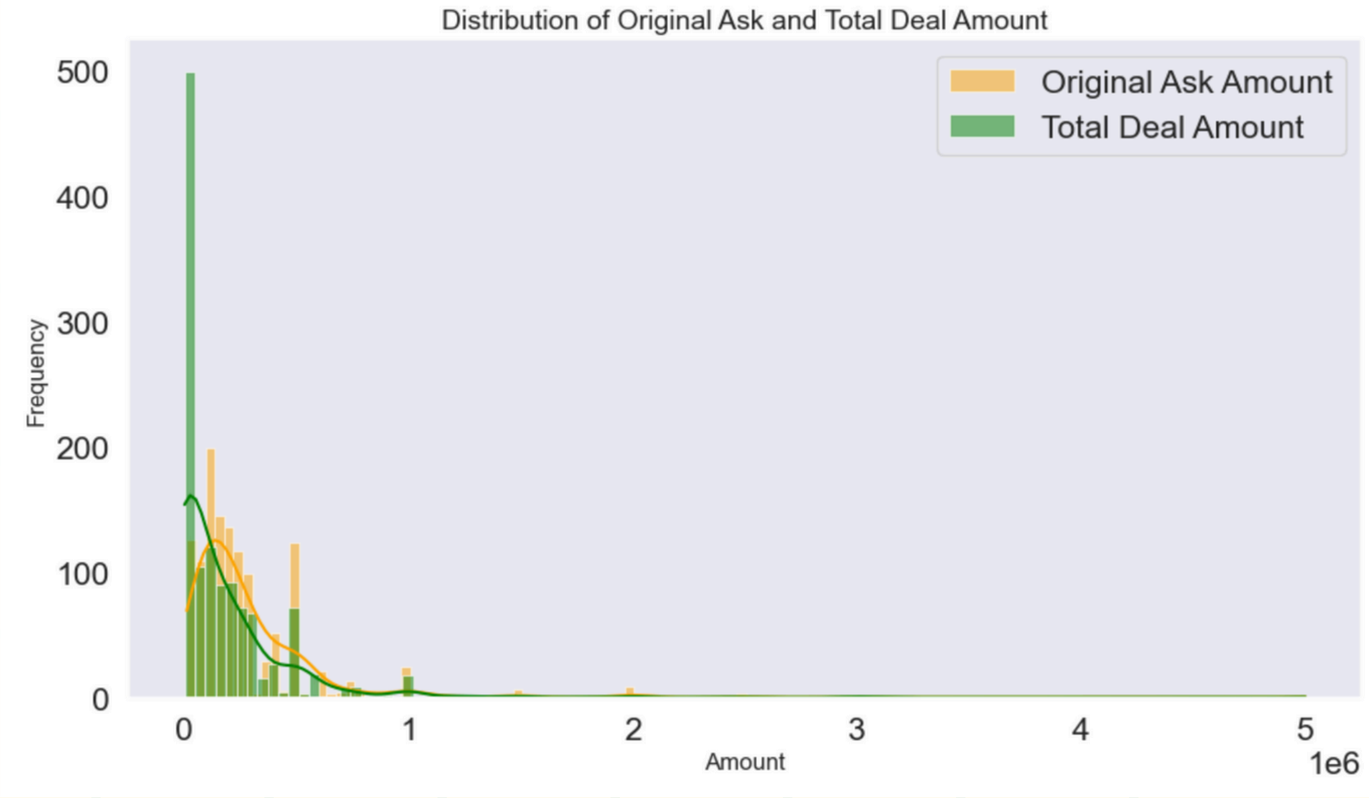
Got deal / CitY	
Napa Valley	470
Dallas	36
New York	32



HYpothesis testing

- Null hypothesis (H0): The mean of original amount asked by entreprenuer is equal to the total_deal_amount.
- Alternative hypothesis (H1): The mean of original amount asked by entreprenuer is not equal to the total_deal_amount.

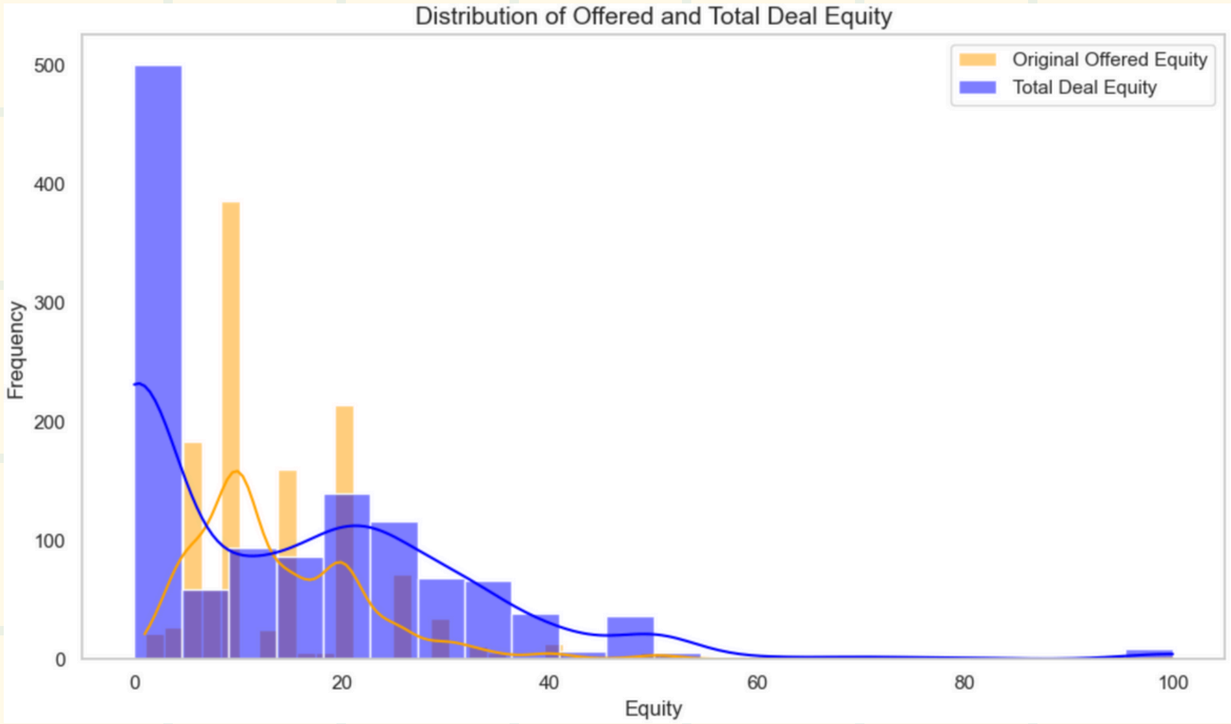
Result:
Reject null hypothesis
Effect size:
Cohen's d 0.03 medium



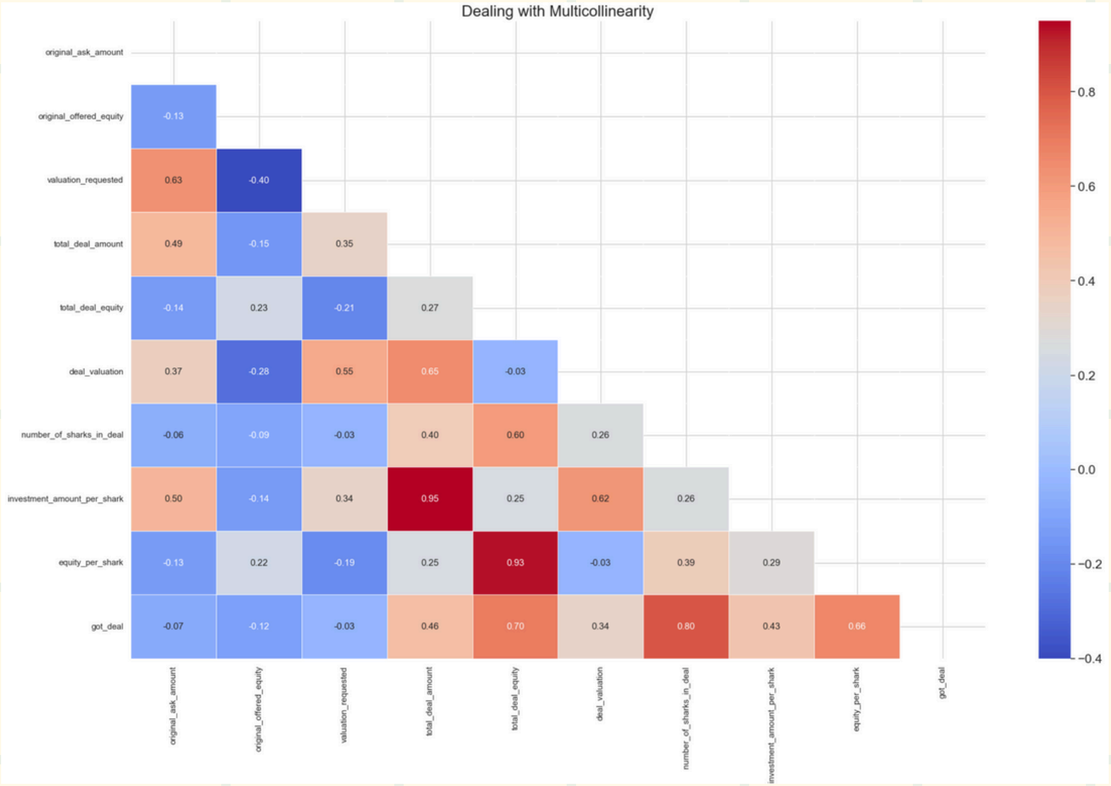
HYpothesis testing

- Null hypothesis (H0): The mean of original_offered_equity by entrepreneur is less than and equal the total_deal_equity.
- Alternative hypothesis (H1): The mean of original_offered_equity by entrepreneur is greater than the total_deal_equity.

Result:
Fail to reject null hypothesis
Effect size:
Cohen's d –0.06 small



Machine learning



Dealing with Multiculiniarity

- total_deal_amount and investment_amount_per_shark 0.95
- total_deal_equity and equity_per_shark 0.93
- deal_valuation and total_deal_amount 0.65
- original_offered_equity and valuation_requested –0.40
- original_offered_equity and deal_valuation –0.28

Features selected

- original_offered_equity
- investment_amount_per_shark
- equity_per_shark
- industry
- State
- City

Target

Got deal

Challenges and solutions

Technical Challenge

Overfitting models

Overcoming Challenges

- Cross validation
- Remove features
- Regularization
- Balancing data

Solution

Principal Component Analysis

Models

- Logistic model
- Decision tree
- Random Forest
- XGBoost
- Support Vector Classifiction

Machine learning results

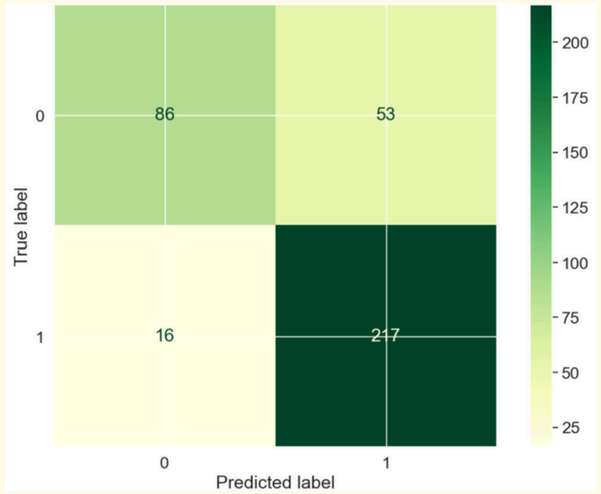
Models

- Logistic model
- Decision tree
- Random Forest highest score overall
- XGBoost
- Support Vector Classifiction

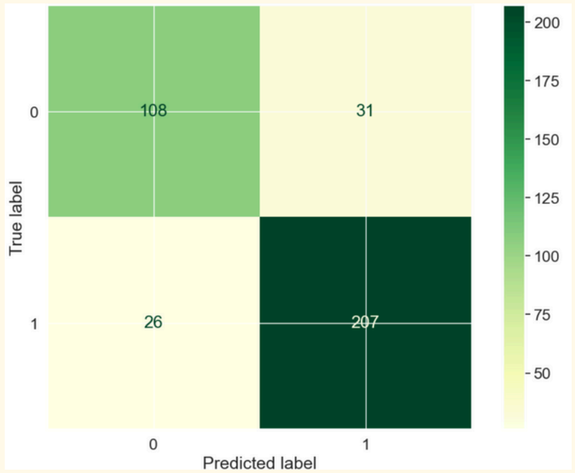
	Precision	Recall	F-score	Support	accuracy
Logistic model					
0	0.84	0.62	0.71	139	0.81
1	0.8	0.93	0.86	233	
Decision tree					
0	0.71	0.87	0.78	139	0.82
1	0.91	0.79	0.85	233	
Random Forest					
0	0.81	0.78	0.79	139	0.85
1	0.87	0.89	0.88	233	
XGBoost					
0	0.65	0.72	0.68	139	0.75
1	0.82	0.76	0.79	233	
SVC					
0	0.83	0.41	0.55	139	0.75
1	0.73	0.95	0.82	233	

Confusion Matrix

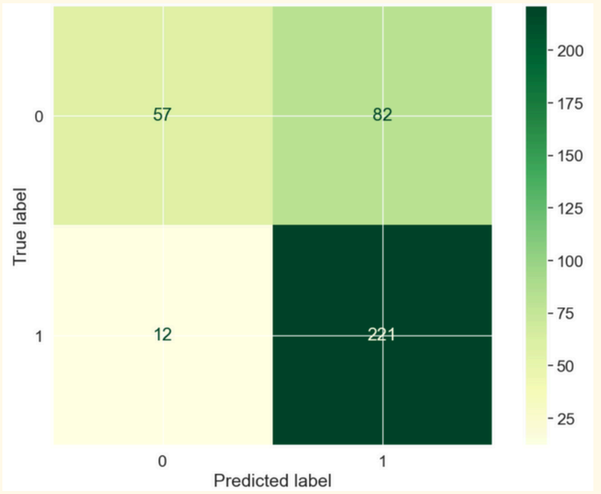
Logistic Model



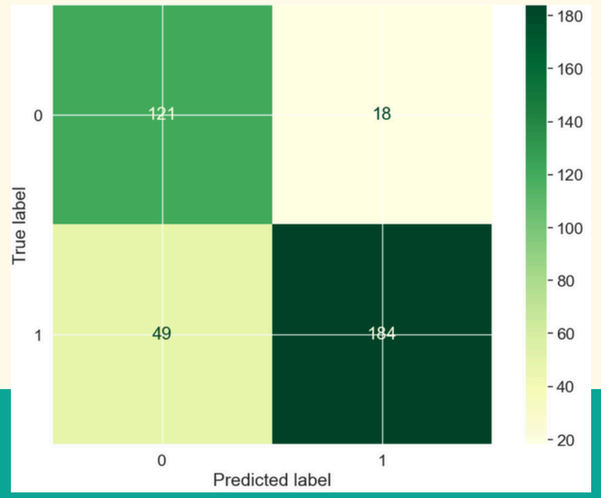
Random Forest



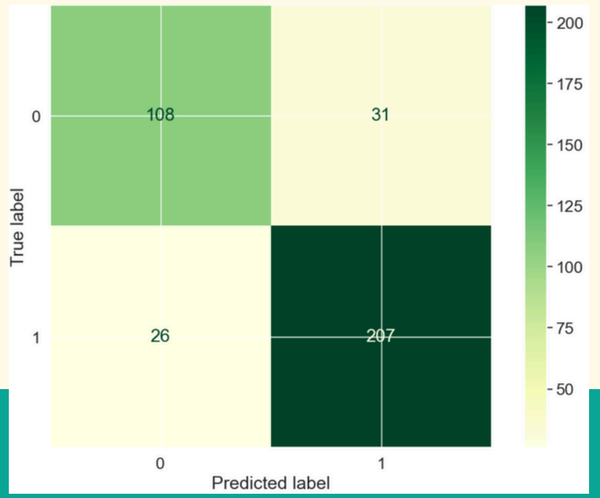
SVC



Decision tree



XGBoost



Planning

Mid project- start up ☆ Board Table

EDA

- Project planning
- Brainstorming
- Cleaning data — Dec 7
- EDA- t-test/chi2
- visualization -add Tableaue
- Finalization
- Streamlit

+ Add a card

ML

- correlation
- feature selection- Ensembled model: (random forest, xgboost...) --> Feature Importance to do feature selection
- classification model - metrics, confussion matrix
- logit model
- fine tuning

+ Add a card

Done

- Finalize notebook-check the result. and rerun all codes
- create Canva
- Github and readme file

+ Add a card


THANK
YOU