Slide Deck

Startup Acquisition Prediction


Eugenio Murillo

Problem Statement

The startup environment is considered one of the riskiest investments due to 90% of startups failing in the first year. This 90% can be reduced throughout the life of a startup depending on several factors and if the founders were able to create relationships and get funding to finance the startup operations.

The purpose of this project is to be able to predict if a startup is going to be acquired or fail based on several features. I also want to see which are the features that have the most weight on predicting a startup fail/success. This can lead to investors to use the model and decide whether to invest or not being one step ahead of the competitor investors.

Exploratory Data Analysis

We can see the detail of the states and the total number of startups Acquired or Not Acquired. The 3 top states with the most count of startups are 'CA', 'NY', and 'MA'. This tells us that these 3 states offer greater opportunity due to a more adequate environment for startups to grow and prosper.
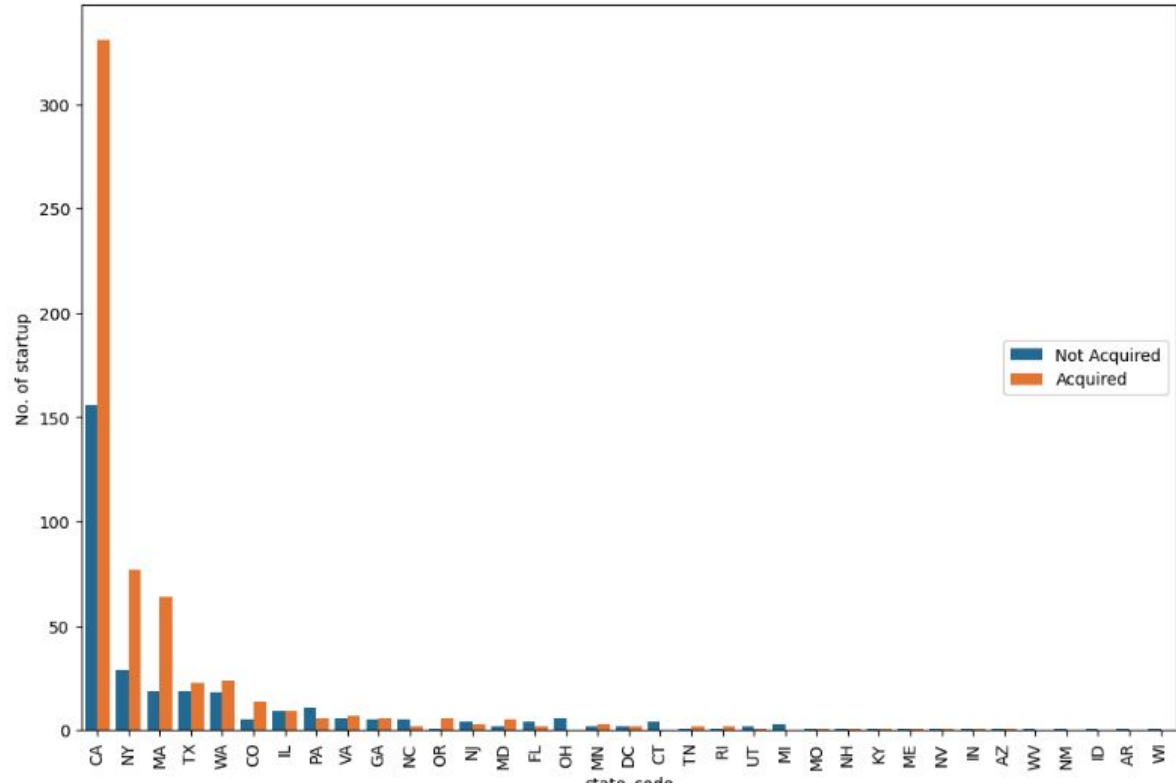
EDA state_code

We create a summary statistic named 'success_rate' which determines # startups successful in each state.
Even though there are startups with higher success rate their count of startups is just a small fraction compared to the top, which is why we focus on the count first.

state_code	Acquired	Not Acquired	total_count	success_rate
CA	331	156	487	67.97%
NY	77	29	106	72.64%
MA	64	19	83	77.11%
WA	24	18	42	57.14%
CO	14	5	19	73.68%
OR	6	1	7	85.71%
MD	5	2	7	71.43%
MN	3	2	5	60.00%
RI	2	1	3	66.67%
TN	2	1	3	66.67%

EDA industry

We can see that the industries with the most startups belong to software, web, mobile, enterprise. This gives us insight about the technology industry being the most important when it comes to startups. Let's see the success rate based on industry.
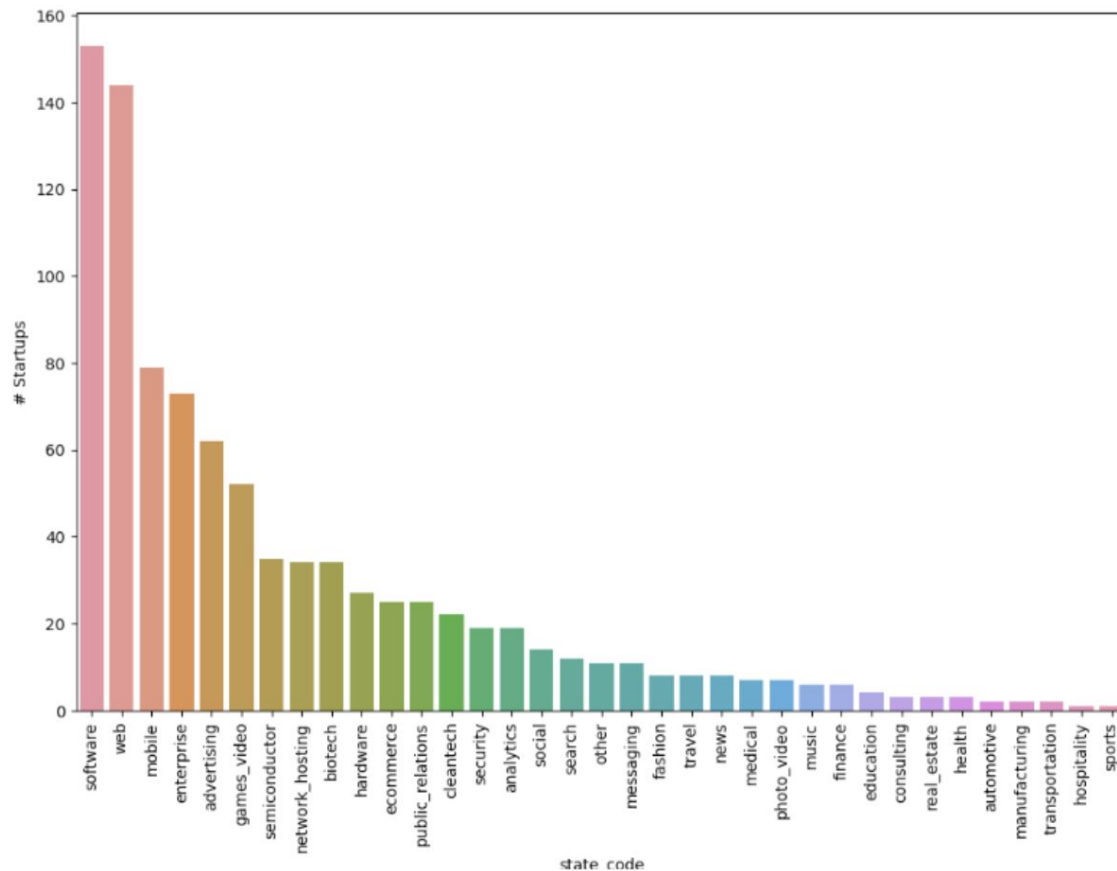
EDA industry

I add the 'success_rate' statistic which represents the percent of Acquired startups, we can see positive numbers for most of the tech industry, with advertising having the highest 'success_rate'. software, web and mobile have the highest count of startups with their success_rate being really similar. This tells us that tech companies have a high success rate and also a high count, making it a competitive market but with opportunity.

category_code	Acquired	Not Acquired	total_count	success_rate
software	101	52	153	66.01%
web	93	51	144	64.58%
mobile	52	27	79	65.82%
enterprise	56	17	73	76.71%
advertising	45	17	62	72.58%
games_video	31	21	52	59.62%
semiconductor	24	11	35	68.57%
network_hosting	24	10	34	70.59%
biotech	22	12	34	64.71%
hardware	11	16	27	40.74%

Feature importance

It is clearly visible which are the top features with more predicting weight, the top 3 are 'relationships', 'funding_total_usd' and 'age_first_funding_year'. These features alone represent around 42% of the predictive weight of the modeling.

While analyzing deeper we found out that having **more than 4 relationships** is essential to being successful in a startup. We also found out that **having more than 6 million usd** in total funding is a key component. The last key finding was the receiving **funding in an early stage (1 to 3 years)** is also essential for success.

Feature Weights
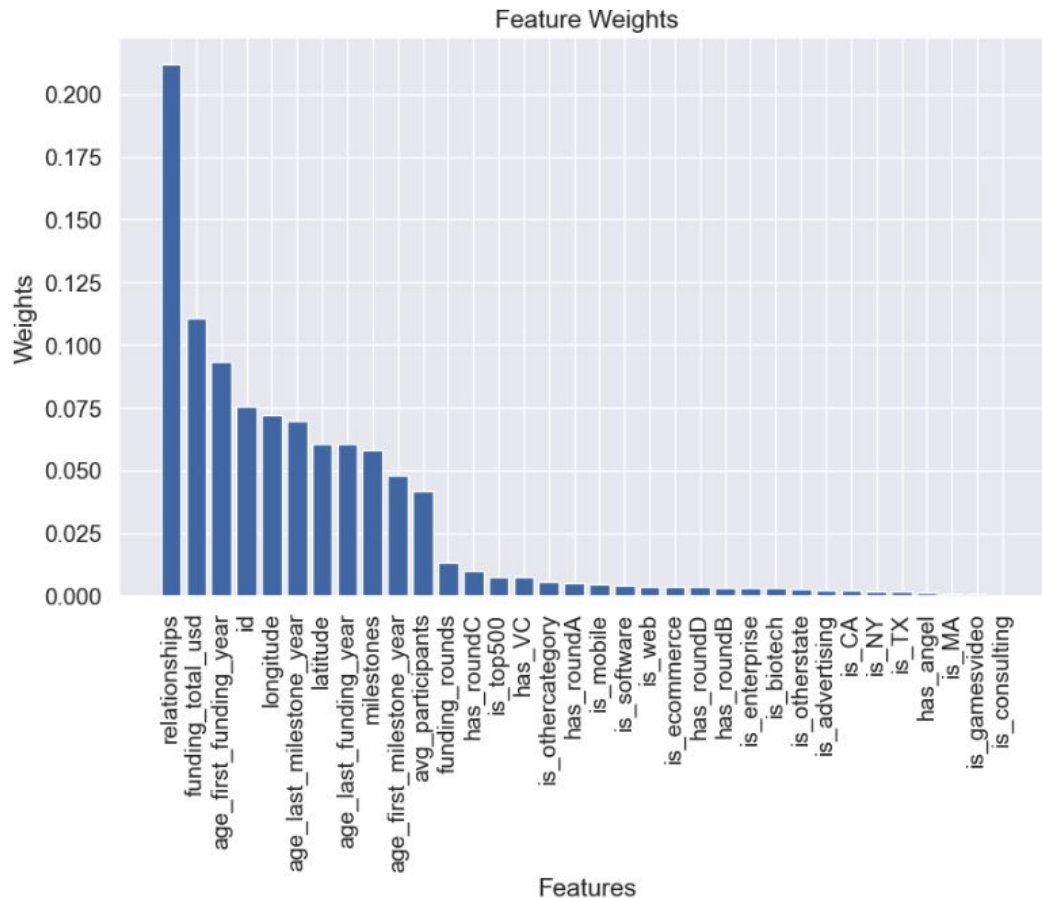
Modeling

The performance of the models will be evaluated using 'classification_report', 'confusion_matrix' and 'accuracy_score'. Given that the classification_report gives us several metrics to analyze for the purpose of this project we will focus on the f1-score given that this metric is a more general one. This metric will be the one to help us choose our best performing model.

Of the 5 models applied our best performing model is Gradient Boosting with an f1_score of .849, which is significantly higher than the other models. This is the recommended model for deployment and for using if we want to predict if a startup will be acquired or not be acquired.

Model Ranking in Test by f1_score
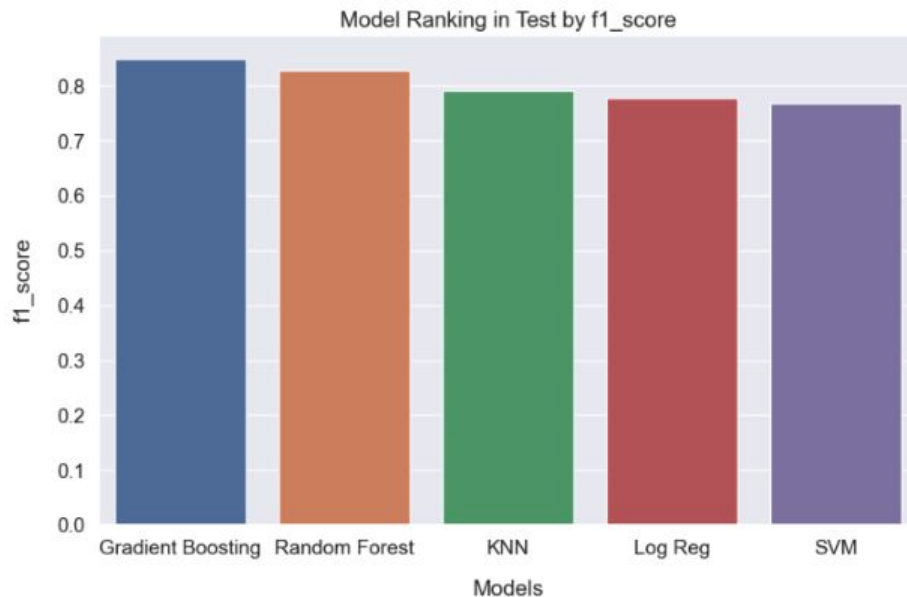
Model	f1_score
Gradient Boosting	0.849206
Random Forest	0.827309
KNN	0.791667
Log Reg	0.778243
SVM	0.767932

Conclusion

After doing all the analysis we get a model with strong predictive power, our Gradient Boosting model can help an investor or someone analyzing the startup environment to be more prepared and a step further from the competition. This model can be used to accurately decide if a startup will be acquired or not.