Machine Learning vs Initial Public Offerings

AUTHORS: AMANDEEP KAUR, NITIN THOMAS, PRABHDYAL SINGH, RABIH RAYMOND ANTOUN, MATTHEW PHAN

Your best quote that reflects your approach… “It’s one small step for man, one giant leap for mankind.”

- NEIL ARMSTRONG

Project Goals

To use various machine learning methods in order to predict the returns on initial public offerings after 100 days to see if they are profitable to trade or invest in or not.

Different deep learning methods will provide a the option for a better choice for performance.

Approach

- Information started with gathering data from iposcoop.com and this was manually done into a csv file which was then imported into a pandas dataframe.
- Data cleaning involved removing the ttm column, making seperate dataframes of 3 different time points, and encoding the 100 day return target column into non negative values for processing.
- The machine learning libraries used were Adaboost, XGBoost, random forest, and logistic regression.

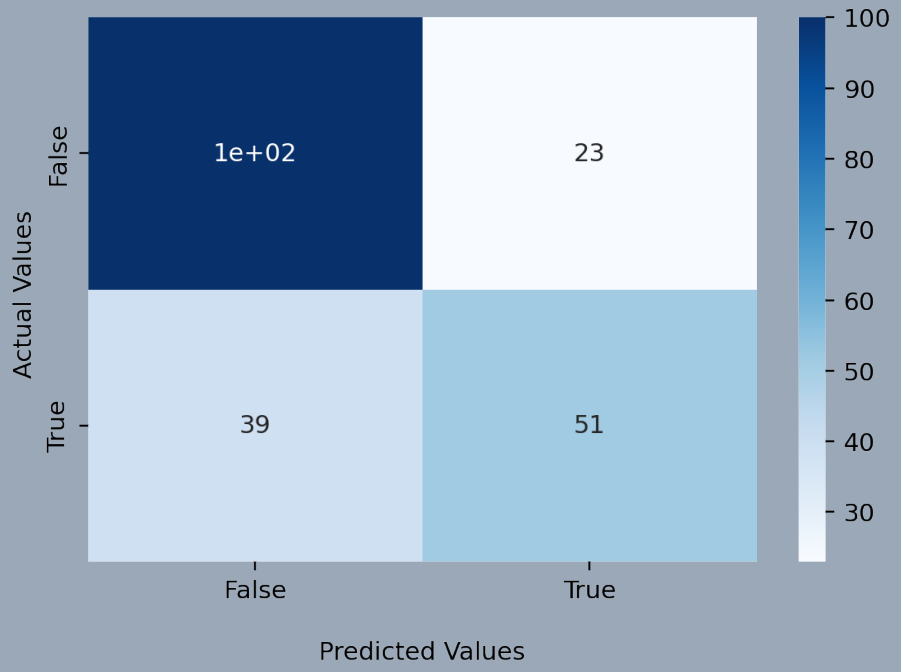
Difficulties

Dropping company symbol and company name, and use the encoded industry as a number, because there was problems reading the string values.

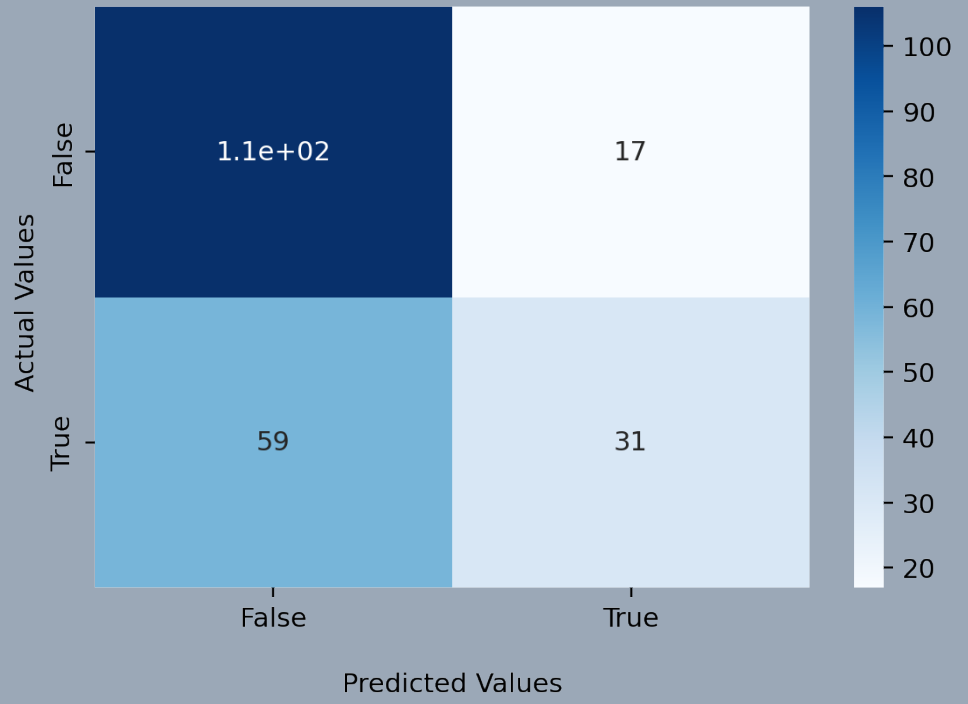
- Problems with data gathering, getting data older than 2018 and having it to work with the code was problematic, the old csv with dates all the way back to 2000 file wasn't computing properly with the notebook.

Outcome

AdaBoost Seaborn Confusion Matrix with labels



Logistic Regression Seaborn Confusion Matrix with labels



Conclusion

Next Steps

Add more features to see if accuracy is better.
(market cap, 1st day close)

Get more data from years predating 2018 and covid

More machine learning models for comparison
(LSTM, Support Vector)

Paid subscription info/data