



PROJECT TITAN-X

Team Ace





Table of Contents

- **ABOUT TEAMAce**
- **INTRODUCTION**
- **APPROACH**
- **EDA**
- **DATA PRE-PROCESSING**
- **ARCHITECTURE PIPELINE**
- **CHOSEN ALGORITHMS**
- **APPLICATION DEPLOYMENT**
- **LIVE DEMO**
- **CONCLUSION**



1

About TeamAce

The People Behind the Project

“



KENG HIN CHEONG

100711777

AI Specialist



AU QUANG LOC NGUYEN

100741710

AI Specialist



CHUTU LI

100765238

AI Specialist



*We always aim to conquer any difficulties and challenges.
“Just do it, make it happen, and ACE it!” is our team’s core spirit.*



2

Introduction



- ◊ Stock trading has always been a task done by trading professionals in finance, since it can be very difficult, risky, time consuming and emotionally affected.
- ◊ Needs professional knowledge and experience.
- ◊ Hard to make money at the end. 7/10 ordinary players lose money.
- ◊ Therefore, our project aims to address the problems mentioned.
- ◊ Focused on SP500 IT Sector of US Stock Market.



3

Approach

- ◊ The project has three main parts:
 - Long Term Investment
 - Short Term Investment
 - NLP Stock Recommendations
- ◊ Main tools and libraries used:



4

EDA





Stock Price Prediction

- ◇ Using an API called 'pandas_datareader.data' to get the stock data from Yahoo Finance

Out[99]:

	Date	High	Low	Open	Close	Volume	Adj Close
0	2016-01-04	105.370003	102.000000	102.610001	105.349998	67649400.0	98.446655
1	2016-01-05	105.849998	102.410004	105.750000	102.709999	55791000.0	95.979675
2	2016-01-06	102.370003	99.870003	100.559998	100.699997	68457400.0	94.101387
3	2016-01-07	100.129997	96.430000	98.680000	96.449997	81094400.0	90.129868
4	2016-01-08	99.110001	96.760002	98.550003	96.959999	70798000.0	90.606438
5	2016-01-11	99.059998	97.339996	98.970001	98.529999	49739400.0	92.073563
6	2016-01-12	100.690002	98.839996	100.550003	99.959999	49154200.0	93.409874
7	2016-01-13	101.190002	97.300003	100.320000	97.389999	62439600.0	91.008270
8	2016-01-14	100.480003	95.739998	97.959999	99.519997	63170100.0	92.998695



NLP

- ❖ Scrapped from Yahoo Finance

Out[50]:

Symbol	Name	Sector	Industry	Full Time Employees	CEO	Address	Description
AAPL	Apple Inc	Technology	Consumer Electronics	137,000	Timothy D. Cook	One Apple Park Way\nCupertino, CA 95014\nUnited States	Apple Inc designs, manufactures and markets computer hardware and software, iPhone, iPad, iPod touch, Apple Watch and Apple TV, and sells them through its own retail stores and online store.
ACN	Accenture Plc	Technology	Information Technology Services	492000	Julie T. Spellman Sweet	1 Grand Canal Square\nGrand Canal Harbour\nDublin 2, Ireland	Accenture provides consulting, technology, and outsourcing services.
ADBE	Adobe Systems Inc	Technology	Software - Application	21357	Shantanu Narayen	345 Park Avenue\nSan Jose, CA 95110\nUnited States	Adobe Inc. operates a diversified software company.
ADI	Analog Devices	Technology	Semiconductors	15,800	Vincent T. Roche	One Technology Way\nPO Box 9106\nNorwood, MA 02061-9106	Analog Devices, Inc. designs and manufactures analog integrated circuits.
ADP	Automatic Data Procs	Industrials	Business Services	58,000	Carlos A. Rodriguez	One ADP Boulevard\nRoseland, NJ 07068\nUnited States	Automatic Data Processing Inc. provides cloud-based HR and payroll services.



5

Data Pre-Processing



Stock Price Prediction-Long Term

- ◇ Extract the year, month, day from Date
- ◇ Take log of Close

Out[11]:

	Date	High	Low	Open	Close	Volume	Adj Close	Year	Month	Day	logClose
0	2016-01-04	105.370003	102.000000	102.610001	105.349998	67649400.0	98.446655	2016	1	4	4.666736
1	2016-01-05	105.849998	102.410004	105.750000	102.709999	55791000.0	95.979675	2016	1	5	4.641599
2	2016-01-06	102.370003	99.870003	100.559998	100.699997	68457400.0	94.101387	2016	1	6	4.622027
3	2016-01-07	100.129997	96.430000	98.680000	96.449997	81094400.0	90.129868	2016	1	7	4.579339
4	2016-01-08	99.110001	96.760002	98.550003	96.959999	70798000.0	90.606438	2016	1	8	4.584559



Stock Price Prediction-Short Term

- ◆ Create a column date and remove unnecessary columns. We will work with Adj Close

```
data['days'] = range(len(data))

data = data.drop(columns = ['Date', 'Close', 'High', 'Low', 'Open', 'Volume'])
```

```
data.head()
```

	Adj Close	days
0	171.917709	0
1	171.771851	1
2	170.177322	2
3	172.034363	3
4	174.990051	4



NLP

- ◇ Select and combine need columns
- ◇ Cleaning the fields

Out[59]:

	Id	Symbol	Name	Text
0	1	AAPL	Apple Inc	technology consumer electronics apple inc desi...
1	2	ACN	Accenture Plc	technology information technology service acce...
2	3	ADBE	Adobe Systems Inc	technology software application adobe inc oper...
3	4	ADI	Analog Devices	technology semiconductors analog devices inc d...
4	5	ADP	Automatic Data Procs	industrials business service automatic data pr...



6

Architecture Pipeline



Stock Price Prediction- Long Term

Get Stock Data
Using API

Feature
Engineering

Model
Training

Model
Evaluation
& Selection,
Final Stock
Price
Prediction

Model
Evaluation
& Selection,
Final Stock
Price
Prediction





Stock Price Prediction- Short Term

Get Stock Data
Using API

Feature
Engineering

Fuzzification

Rule
generation

Forecasting





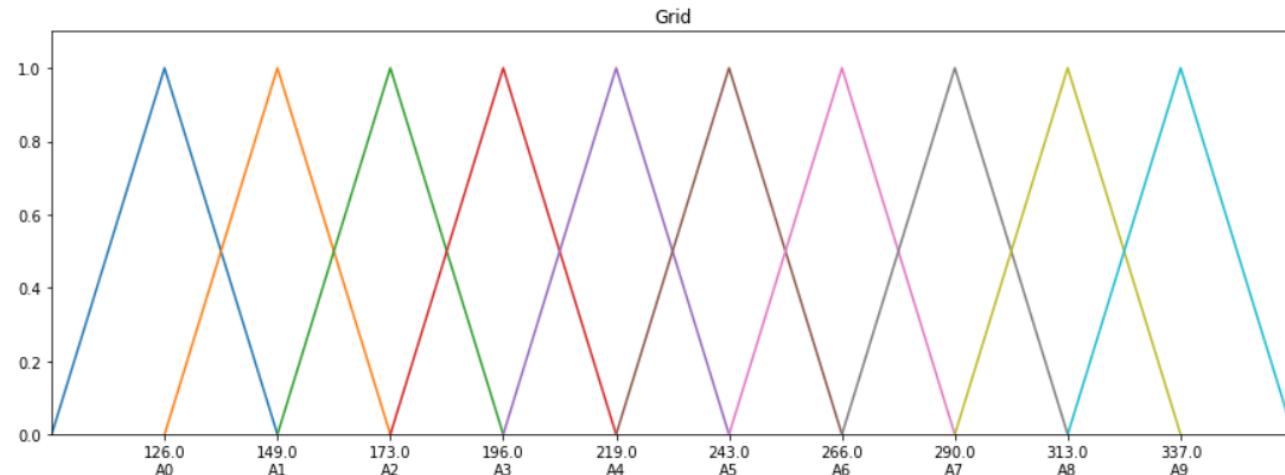
Visualization of the Fuzzy sets of the Linguistic Variable Adj Close

```
from pyFTS.partitioners import Grid

fs = Grid.GridPartitioner(data=data,npart=10)

fig, ax = plt.subplots(nrows=1, ncols=1, figsize=[15,5])

fs.plot(ax)
```





Fuzzification & Temporal Patterns

```
# Using the method maximum, just the maximum membership fuzzy set is chosen.  
# However in other FTS methods all fuzzyfied values are considered.  
fuzzyfied = fs.fuzzyfy(data, method='maximum', mode='sets')  
fuzzyfied  
['A3',  
 'A3',  
 'A4',  
 'A4',  
 'A4',  
 'A4',  
 'A4',  
 'A4']
```

```
from pyFTS.common import FLR  
  
patterns = FLR.generate_non_recurrent_flrs(fuzzyfied)  
print([str(k) for k in patterns])  
  
['A2 -> A2', 'A2 -> A1', 'A1 -> A2', 'A1 -> A1', 'A2 -> A3', 'A3 -> A2', 'A3 -> A3',  
 'A5', 'A5 -> A5', 'A5 -> A6', 'A6 -> A6', 'A6 -> A7', 'A7 -> A7', 'A7 -> A8', 'A8 ->  
 8 -> A7', 'A7 -> A6', 'A6 -> A5', 'A5 -> A4']
```



Rule generation

```
from pyFTS.models import chen

#Chen Fuzzy Time Series method
model = chen.ConventionalFTS(partitioner=fs)
model.fit(data)
print(model)
```

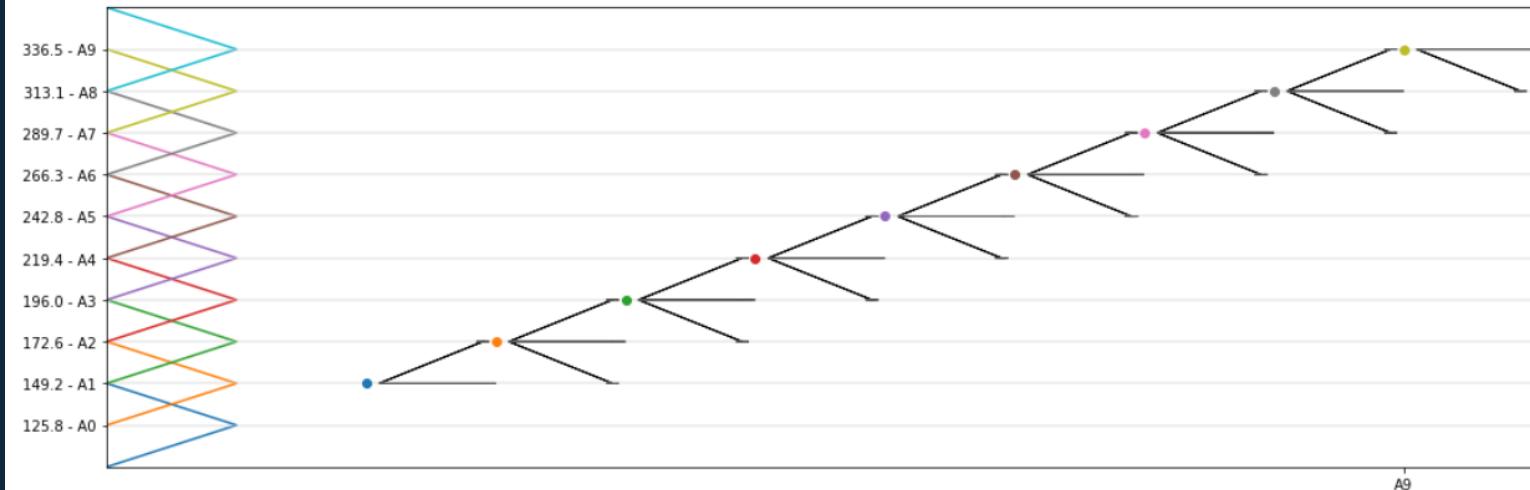
Conventional FTS:

A1 -> A1,A2
A2 -> A1,A2,A3
A3 -> A2,A3,A4
A4 -> A3,A4,A5
A5 -> A4,A5,A6
A6 -> A5,A6,A7
A7 -> A6,A7,A8
A8 -> A7,A8,A9
A9 -> A8,A9

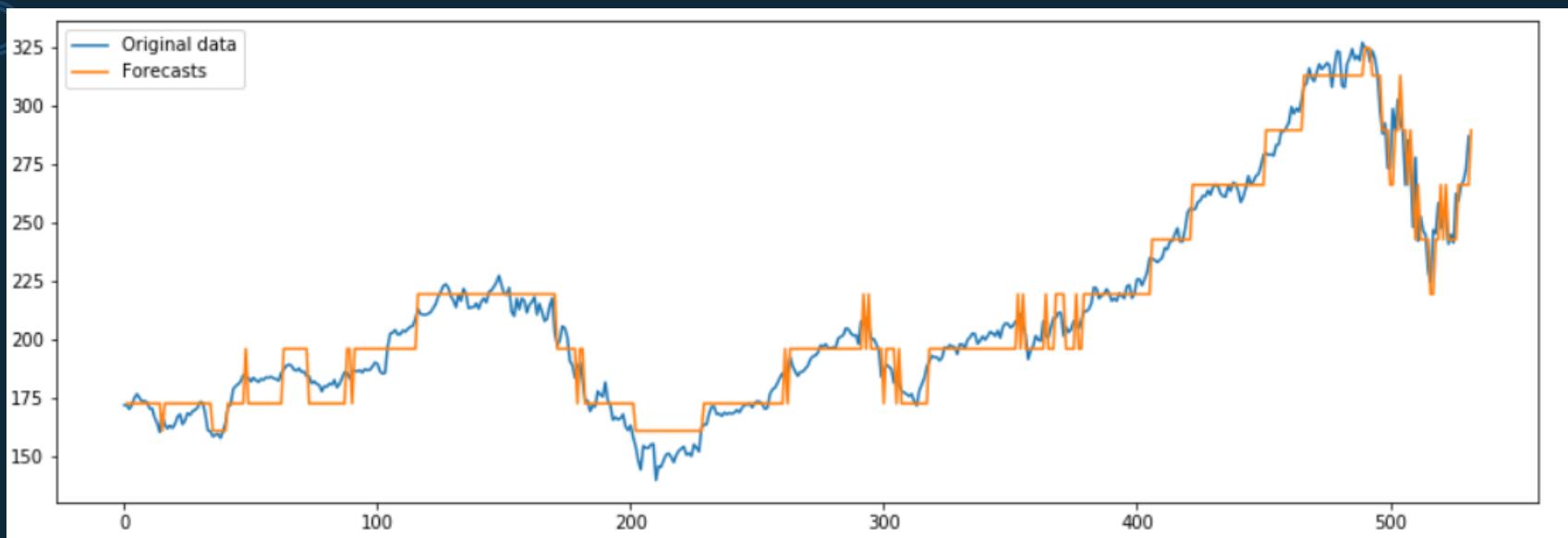


Rules plot

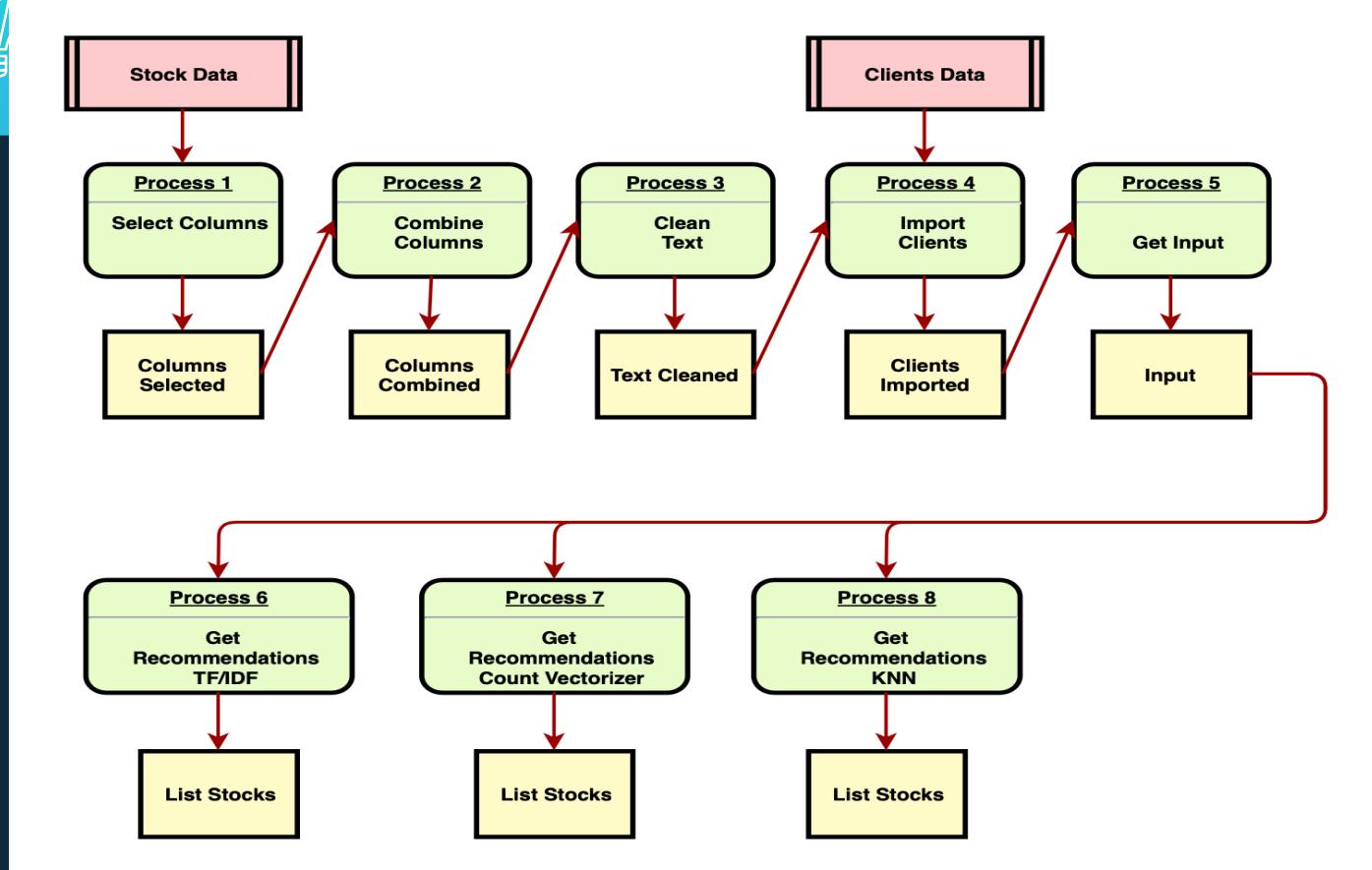
```
from pyFTS.common import Util  
Util.plot_rules(model, size=[15,5] , rules_by_axis=10)
```



Forecasting



NLP





7

Chosen Algorithms



Stock Price Prediction-Long Term

-----Linear Regression(on Training Set)-----

Evaluation Metrics
MAE : 0.0476
MAPE : 0.9754
RMSLE : 0.0098

-----Linear Regression(on Test Set)-----

Evaluation Metrics
MAE : 0.2156
MAPE : 4.1032
RMSLE : 0.0383

-----LR_lasso Regression(on Training Set)-----

Evaluation Metrics
MAE : 0.0475
MAPE : 0.9747
RMSLE : 0.0098

-----LR_lasso Regression(on Test Set)-----

Evaluation Metrics
MAE : 0.2143
MAPE : 4.0778
RMSLE : 0.0381

-----LGBM Regression(on Training Set)-----

Evaluation Metrics
MAE : 0.0108
MAPE : 0.2207
RMSLE : 0.0025

-----LGBM Regression(on Test Set)-----

Evaluation Metrics
MAE : 0.1242
MAPE : 2.3143
RMSLE : 0.0238

-----RandomForest Regression(on Training Set)-----

Evaluation Metrics
MAE : 0.0068
MAPE : 0.1392
RMSLE : 0.0033

-----RandomForest Regression(on Test Set)-----

Evaluation Metrics
MAE : 0.1213
MAPE : 2.2620
RMSLE : 0.0233

-----XGB Regression(on Training Set)-----

Evaluation Metrics
MAE : 0.0143
MAPE : 0.2931
RMSLE : 0.0032

----- XGB Regression(on Test Set)-----

Evaluation Metrics
MAE : 0.1255
MAPE : 2.3477
RMSLE : 0.0234

-----GradientBoosting Regression(on Training Set)-----

Evaluation Metrics
MAE : 0.0102
MAPE : 0.2083
RMSLE : 0.0025

----- GradientBoosting Regression(on Test Set)-----

Evaluation Metrics
MAE : 0.1295
MAPE : 2.4144
RMSLE : 0.0246



Stock Price Prediction-Long Term

Enter the current date (dd/mm/yyyy)(for example: 01/02/2025):03/04/2020

Enter the current price of the stock: 241

Using model of LR, the future stock price is: [249.4154833]

Using model of LR_Lasso, the future stock price is: [248.54759035]

Using model of XBG, the future stock price is: [181.56859]

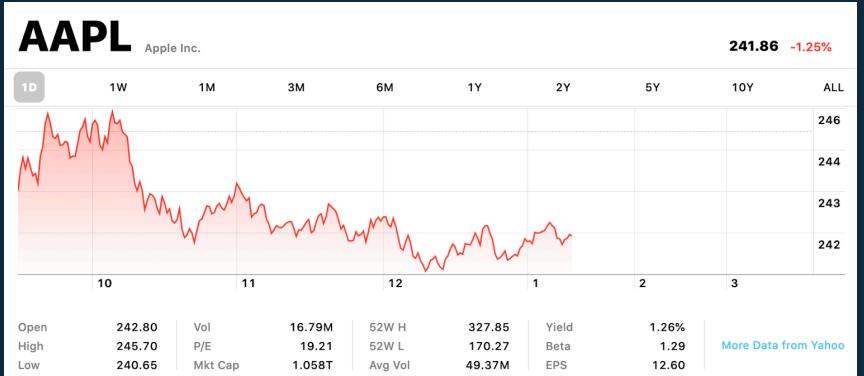
Using model of RFR, the future stock price is: [193.08299866]

Using model of LGBM, the future stock price is: [186.26974494]

Using model of GB, the future stock price is: [185.31150929]

The best prediction above is 96.51 % accurate compared to the current actual price of the stock!

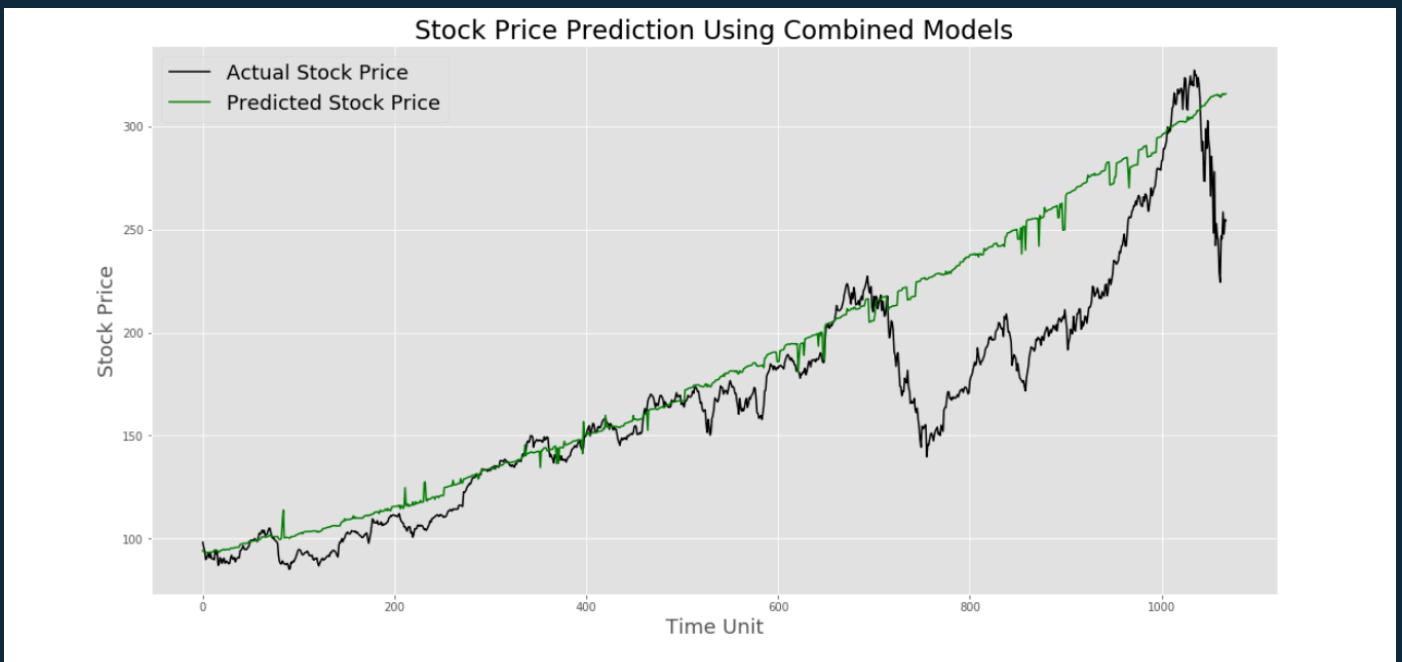
- ◆ Linear Regression Chosen
- ◆ Fast and easy
- ◆ The prediction error to the real-time stock price is less than 10%!
- ◆ Amazing result!





Stock Price Prediction-Long Term

- ◊ The combined model with Linear Regression & Random Forest Regression works the best





NLP

- ◊ TF/IDF Chosen
- ◊ All similar results, but TF/IDF easiest to tweak and the most accurate

Recommendation using KNN

	User Id	Id	Symbol	Name	Score
0	25.0	22	DXC	Dxc Technology Company	1.257184
1	25.0	25	FISV	Fiserv Inc	1.264603
2	25.0	24	FIS	Fidelity National Information Services Inc	1.270372
3	25.0	30	GPN	Global Payments Inc	1.277175
4	25.0	5	ADP	Automatic Data Procs	1.279105
5	25.0	53	PAYX	Paychex Inc	1.282093
6	25.0	20	CTSH	Cognizant Technology Solutions Corporation	1.291415
7	25.0	48	MU	Micron Technology	1.304110
8	25.0	38	JKHY	Jack Henry & Associates Inc	1.324938
9	25.0	15	BR	Broadridge Financial Solutions Inc	1.332240

Recommendations using TF/IDF

User Id	Id	Symbol	Name	Score
0	25.0	33	IBM	International Business Machines 0.212730
1	25.0	22	DXC	Dxc Technology Company 0.209744
2	25.0	25	FISV	Fiserv Inc 0.200390
3	25.0	24	FIS	Fidelity National Information Services Inc 0.193077
4	25.0	30	GPN	Global Payments Inc 0.184412
5	25.0	5	ADP	Automatic Data Procs 0.181945
6	25.0	53	PAYX	Paychex Inc 0.178119
7	25.0	20	CTSH	Cognizant Technology Solutions Corporation 0.166124
8	25.0	48	MU	Micron Technology 0.149649
9	25.0	38	JKHY	Jack Henry & Associates Inc 0.122269

Recommendations using CountVectorizer

User Id	Id	Symbol	Name	Score
0	25.0	33	IBM	International Business Machines 0.462993
1	25.0	53	PAYX	Paychex Inc 0.460409
2	25.0	22	DXC	Dxc Technology Company 0.458339
3	25.0	25	FISV	Fiserv Inc 0.448208
4	25.0	30	GPN	Global Payments Inc 0.403641
5	25.0	20	CTSH	Cognizant Technology Solutions Corporation 0.383516
6	25.0	5	ADP	Automatic Data Procs 0.377964
7	25.0	24	FIS	Fidelity National Information Services Inc 0.324127
8	25.0	44	MA	Mastercard Inc 0.324091
9	25.0	2	ACN	Accenture Plc 0.307339



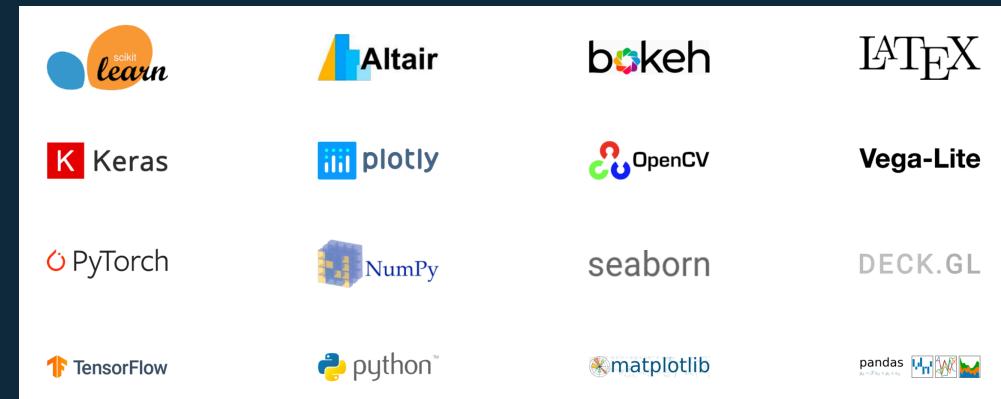
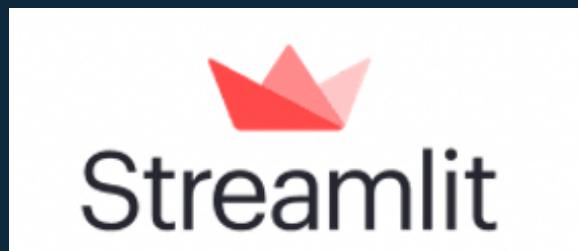
8

Application Deployment



Using Streamlit

- ◊ An open source app framework
- ◊ Specifically designed for ML engineers
- ◊ Working with Python
- ◊ Compatible with major libraries & frameworks of ML





Cloud Deployment-AWS



Running the App



```
cena2098 — ubuntu@ip-172-31-38-10: ~ /khc/AIDI/Winter2020/Capstone/final...  
(tensorflow2_p36) ubuntu@ip-172-31-38-10:~/khc/AIDI/Winter2020/Capstone/finalPr...  
ject$ streamlit run app.py  
  
You can now view your Streamlit app in your browser.  
  
Network URL: http://172.31.38.10:8501  
External URL: http://52.201.68.14:8501  
  
Warning - Certain functionality requires requests_html,  
which is not installed. Install using:  
pip install requests_html  
  
After installation, you may have to restart your Python session.  
/usr/local/lib/python3.6/dist-packages/pandas/_datareader/comput/_init__.py:7: FutureWarning: pandas.util.testing is deprecated. Use the functions in the public API at pandas.testing instead.  
from pandas.util.testing import assert_frame_equal  
/usr/lib/python3/dist-packages/h5py/_init_.py:36: FutureWarning: Conversion of the second argument of isubdtype from 'float' to 'np.floating' is deprecated. In future, it will be treated as 'np.float64 == np.dtype(float).type'.  
from _conv import register_converters as _register_converters  
Using TensorFlow backend.
```



The screenshot shows a web browser window titled "Not Secure — 52.201.68.14". The main content area is titled "AI Assistant for Stock Investing". It displays a message: "Please select what you want on the left! 😊". Below this, it says "Today's date: 17/04/2020". A section titled "Predict Long Term Investment of Stock" asks, "Which US stock are you interested in for long term? Please enter its stock symbol, for example, AAPL, IBM, AMZN, etc." A text input field contains "2020/04/17". Another text input field asks for "Enter the current price of the stock:" with a value of "0.00". At the bottom is a button labeled "Process Long Term".



9

Live Demo

Click this link to access the app-<http://52.201.68.14:8501>



10

Conclusion

- 
- ◊ With AI, predicting stock price is possible.
 - ◊ With AI, stock trading will be much easier.
 - ◊ The level of understanding of the dataset is the key to whether an AI project can be successful.
 - ◊ For future improvements, wider range of datasets in addition to stock prices are needed.
 - ◊ Also, possible combination with Neural Network models could lead to better predicted results.
- 



Thanks!

Any questions?