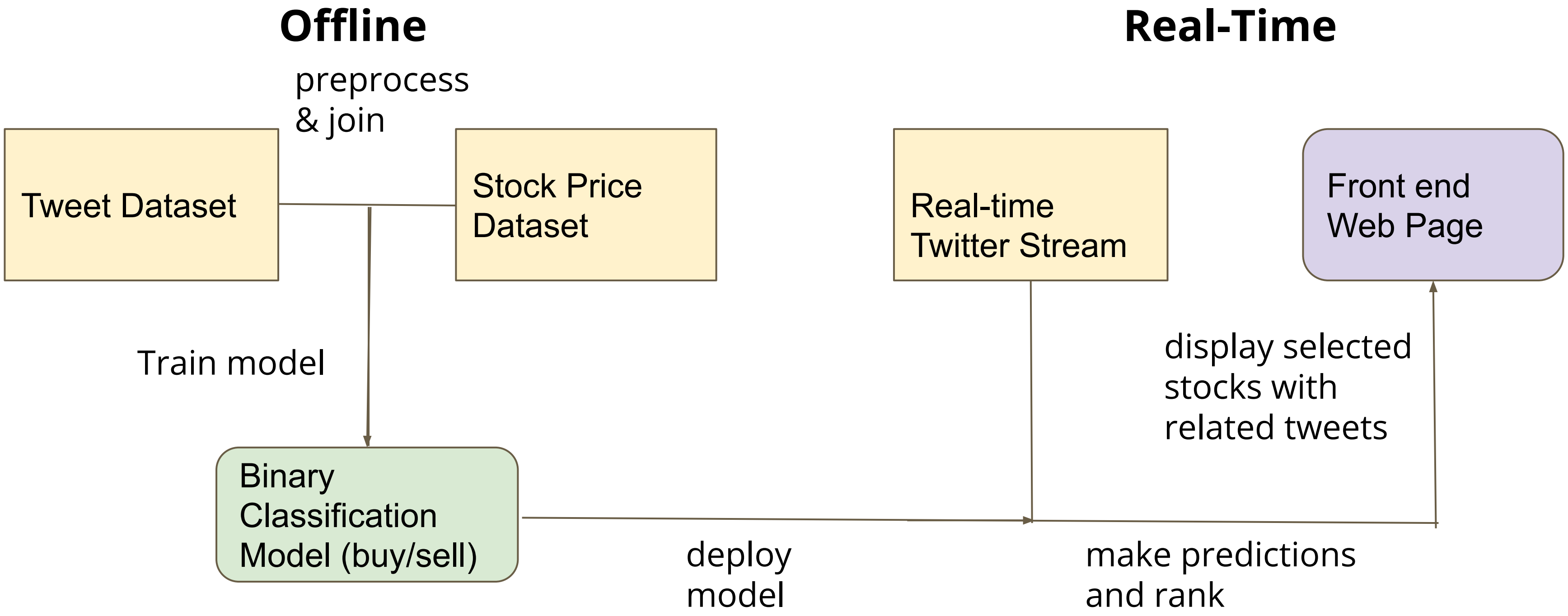
Forecast of Stock Price Using Twitter Sentiment

Group 12
Tianchun Huang, Haoxiong Su, Yi Chen

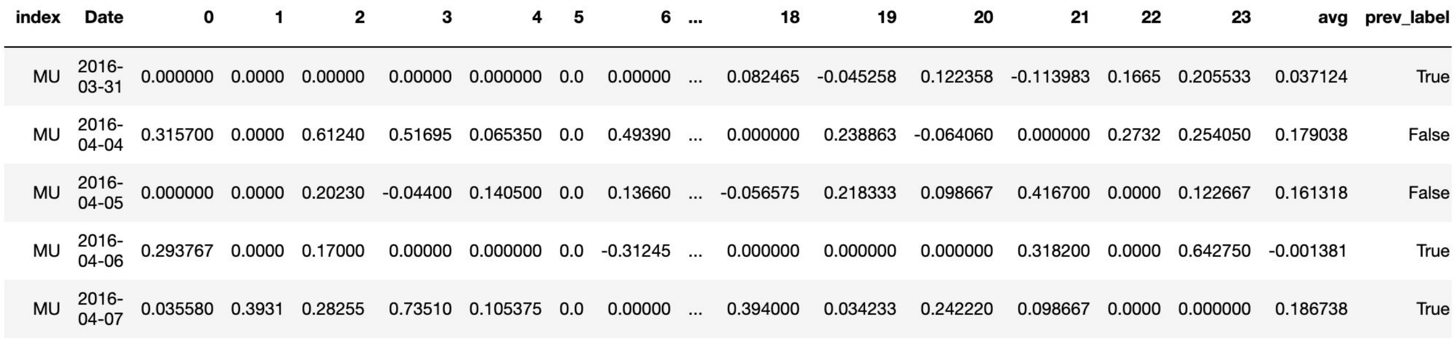
Recap of Data Pipeline



Data Preprocessing

- Used VADER sentiment analyzer to compute sentiment scores per tweet
- Computed sentiment scores per stock per hour weighted by #followers
- Merged tweet data and stock data

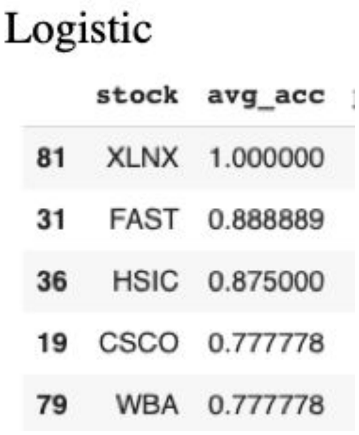
Preprocessed data



index	Date	0	1	2	3	4	5	6	...	18	19	20	21	22	23	avg	prev_label
MU	2016-03-31	0.000000	0.0000	0.00000	0.00000	0.000000	0.0	0.00000	...	0.082465	-0.045258	0.122358	-0.113983	0.1665	0.205533	0.037124	True
MU	2016-04-04	0.315700	0.0000	0.61240	0.51695	0.065350	0.0	0.49390	...	0.000000	0.238863	-0.064060	0.000000	0.2732	0.254050	0.179038	False
MU	2016-04-05	0.000000	0.0000	0.20230	-0.04400	0.140500	0.0	0.13660	...	-0.056575	0.218333	0.098667	0.416700	0.0000	0.122667	0.161318	False
MU	2016-04-06	0.293767	0.0000	0.17000	0.00000	0.000000	0.0	-0.31245	...	0.000000	0.000000	0.000000	0.318200	0.0000	0.642750	-0.001381	True
MU	2016-04-07	0.035580	0.3931	0.28255	0.73510	0.105375	0.0	0.00000	...	0.394000	0.034233	0.242220	0.098667	0.0000	0.000000	0.186738	True

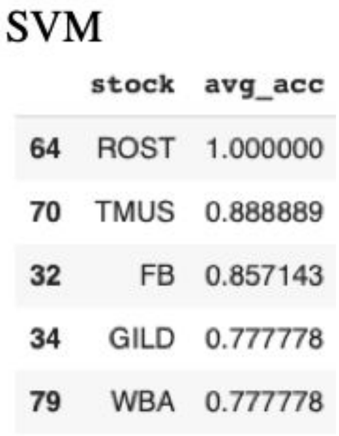
Current Results

- Trained binary classification using **Logistic Regression**, **Support Vector Machine**, **K-Nearest Neighbours**, and **Decision Trees**
- Got good results on some of the stocks while the others not
- Future improvements: time-series model and more careful feature engineering, other sentiment analyzer such as TextBlob



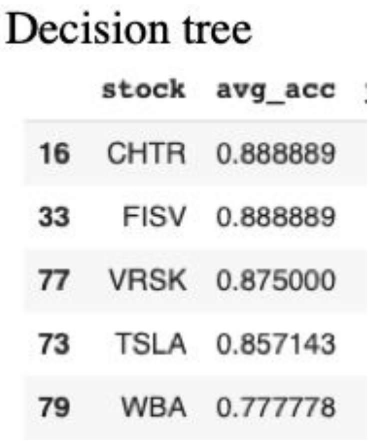
Logistic

	stock	avg_acc
81	XLNX	1.000000
31	FAST	0.888889
36	HSIC	0.875000
19	CSCO	0.777778
79	WBA	0.777778



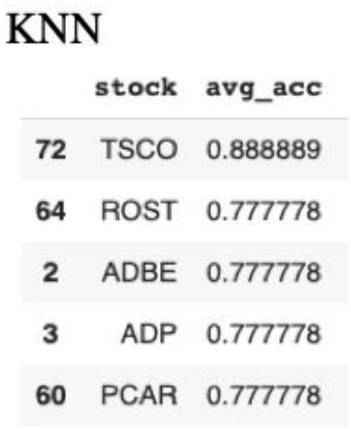
SVM

	stock	avg_acc
64	ROST	1.000000
70	TMUS	0.888889
32	FB	0.857143
34	GILD	0.777778
79	WBA	0.777778



Decision tree

	stock	avg_acc
16	CHTR	0.888889
33	FISV	0.888889
77	VRSK	0.875000
73	TSLA	0.857143
79	WBA	0.777778



KNN

	stock	avg_acc
72	TSCO	0.888889
64	ROST	0.777778
2	ADBE	0.777778
3	ADP	0.777778
60	PCAR	0.777778

Evaluation Methods

- Evaluate the binary classification model using both accuracy and AUC score.

- First evaluate using test set, then evaluate on real-time stream data and daily stock price.

- When displaying results on front end web page, display some of the tweets to convince the user that our predictions are reasonable

Plan for Next Steps

- Improve the current offline result by model improvement and feature engineering. (Haoxiong)

- Build Twitter streaming pipeline for real-time prediction. (Yi)

- Write front end web page for result visualization and user interaction. (Tianchun)

- Report and video.

References

1. Mittal, Anshul. “Stock Prediction Using Twitter Sentiment Analysis.” (2011).
2. Serban, Iulian et al. “Prediction of changes in the stock market using twitter and sentiment analysis.” (2014).
3. Bollen, Johan et al. “Twitter mood predicts the stock market.” *ArXiv* abs/1010.3003 (2011): n. pag.