

Optimized Stock Price Prediction from News

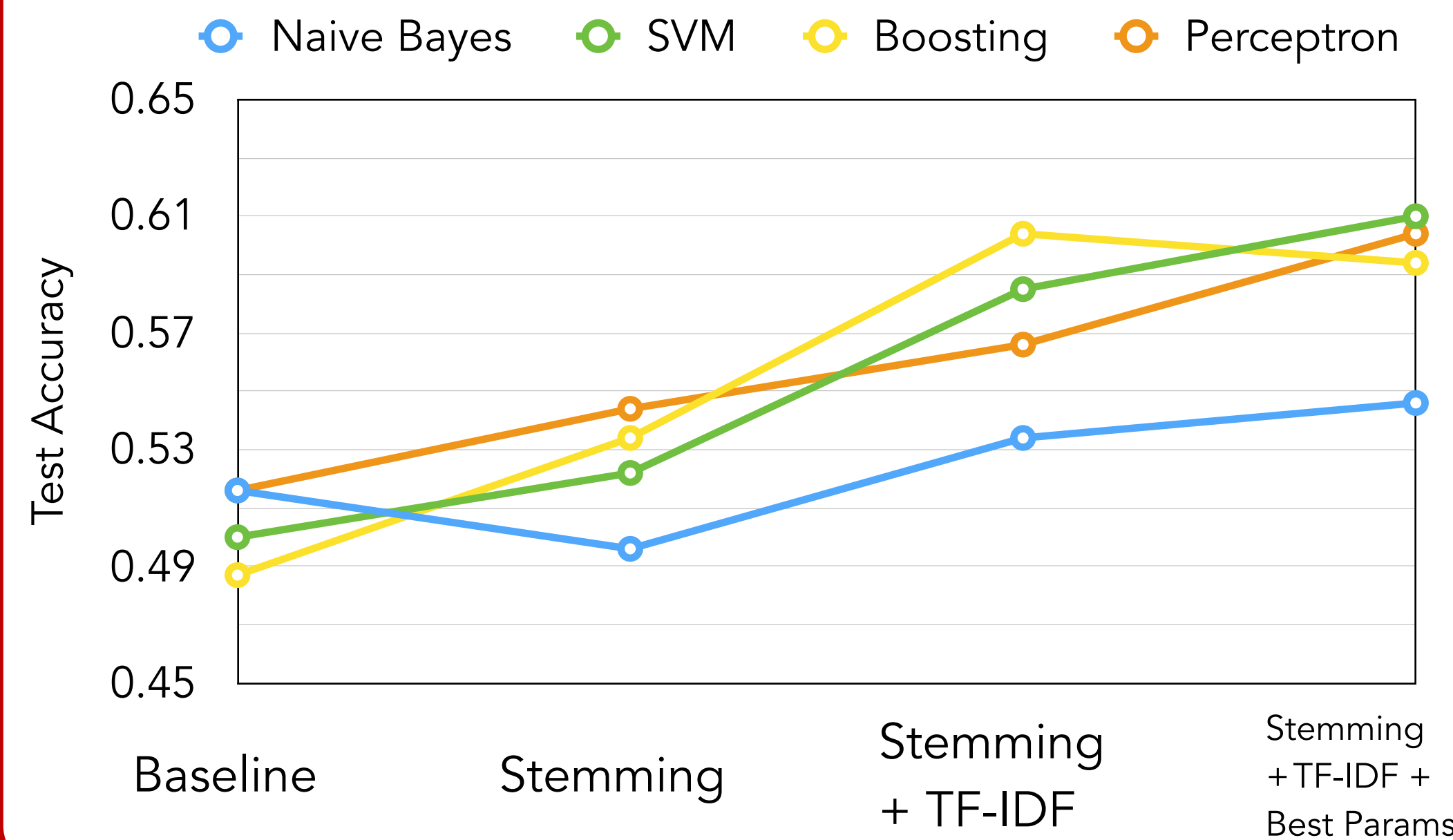


Bryan Cheong, Christopher Yuan, Stephen Ou

Abstract

We use several supervised learning algorithms in conjunction with a 'bag-of-words' representation of **reputable news articles to predict stock price movement**. With model selection and parameter tuning, we were able to achieve 60% accuracy on our test set.

Test Accuracy



Most Relevant Terms

Positive

Term	Weight
suggest	0.54
rumor	0.51
high	0.51
report	0.48
rise	0.45
express	0.44
staff	0.43
soar	0.42
diversified	0.40
reorganization	0.40

Negative

Term	Weight
wary	-0.51
discourage	-0.48
misinform	-0.48
reprehensible	-0.46
abusive	-0.45
dispute	-0.45
denounce	-0.45
cheapen	-0.43
rough	-0.41
takeover	-0.40

(Used unstemmed words for readability)

Classification Pipeline

1 The Wall Street Journal

RECESSION FEARED

The stock market took a major hit today as the Spring CS229 class presented their final projects. Supervised learning methods sh- and caused several companies including Apple, BP, and Honda to go bankrupt. President Obama issued a plea to students,

2 Preprocessing

the stock market took a major hit today as the spring cs229 class present their final project supervise learning method and cause several company include apple bp and honda to go bankrupt president obama issues a plea to student

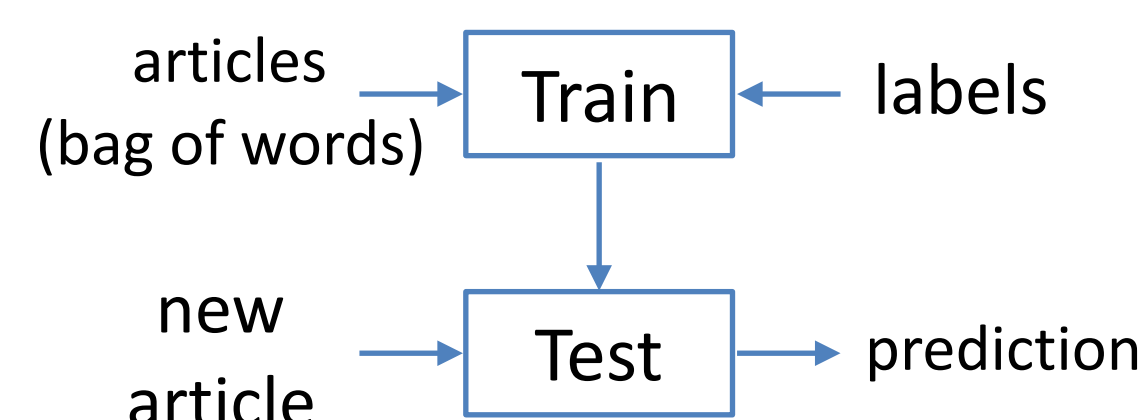
4 Labels (prices)

Date	BA	GM	GE	BP	FB
5/11	↑	↑	↑	↓	↑
5/12	↓	↑	↑	↓	↑
5/13	↓	↑	↑	↓	↑
5/14	↓	↓	↓	↓	↓

3 Bag of Words

Term	Article No.					
	1	2	3	4	5	6
company	1	0	0	4	3	7
rumor	0	1	0	0	2	3
stock	2	1	0	5	3	2
issues	1	2	0	2	8	6

5 Classification



Optimizations

- Stemming**
{discouraging, discouraged, discourage} → discourag
{diversified, diversify, diversifying} → diversifi
- TF-IDF term weighting**
- Decreases weights of common words
$$tfidf(t, d, D) = tf(t, d) \cdot idf(t, D)$$
$$tf(t, d) = 1 + \log(tf_{t,d})$$
$$idf(t, D) = \log\left(\frac{N}{df_t}\right)$$
- Parameter tuning**
- Naive Bayes: Laplace smoothing
- SVM: kernel functions, type of regularization
- Boosting: base estimator, number of estimators
- Perceptron: loss functions, type of penalty

Analysis & Future Work

- We can predict changes in companies' stock value using words from newspaper articles
- Refinements using **model selection tools** significantly improve the predictive power of the baseline bag-of-words model
- Future applications could focus on **simulating the performance** of making stock purchase and sale decisions using our optimized model

References:

- Gidofalvi, Gyoza. "Using news articles to predict stock price movements."
- Ma, Qicheng. "Stock Price Prediction Using News Articles."
- Timmons and Ryan. "Predicting the stock market with news articles."
- Chen, Jerry et. al. "Predicting Stock Prices from News Articles." (2015).
- Aase, Kim-Georg. "Text mining of news articles for stock price predictions."
- Ting and Zheng. "A study of adaboost with naive bayesian classifiers: Weakness and improvement."