

What is the relationship between news and stock price trends

MACS 30250 Project

Tianxin Zheng

April 9

References

- Bollen, J., Mao, H. & Zeng, X. **Twitter mood predicts the stock market.** *Journal of Computational Science* 2, 1–8 (2011)
- Robert P. Schumaker and Hsinchun Chen. 2009. **Textual analysis of stock market prediction using breaking financial news: The AZFin text system.** *ACM Trans. Inf. Syst.* 27, 2, Article 12 (March 2009), 19 pages. DOI: <https://doi.org/10.1145/1462198.1462204>
- Levenberg, A., Pulman, S., Moilanen, K, Simpson, E. & Roberts1, S. **Predicting Economic Indicators from Web Text Using Sentiment Composition,**. *International Journal of Computer and Communication Engineering* vol. 3, no. 2, pp. 109-115, (2014).
- Lee, H., Surdeanu, M., MacCartney, B., & Jurafsky, D. (2014). **On the importance of text analysis for stock price prediction.** In *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014* (pp. 1170-1175). European Language Resources Association (ELRA).

Basic Idea

- Use NLP and machine learning techniques to conduct sentiment analysis on financial news and use it to predict the future trends of the stock index



Data

News From...

- Yahoo Finance
- Wall Street Journal
- Financial Time

Stock Price Index:

- DJIA (Dow Jones Industrial Average) from Yahoo Finance

2018/01/01 – 2019/01/01 Training data

2019/01/01 – 2019/03/01 Test data

Textual Analysis Techniques

- Bag-of -Words
- TF-IDF
- Word2Vec
- N-grams for context analysis

The background of the slide features a series of thin, curved lines in light gray and white, creating a sense of motion and depth. These lines are more prominent on the left side and fade out towards the right.

Textual Analysis Techniques

- Bag-of -Words
- TF-IDF
- Word2Vec
- N-grams for context analysis

The background of the slide features several thin, curved lines in a light gray color, some solid and some dashed, creating a sense of motion or a stylized globe. On the left side, there is a red speech bubble with a tail pointing towards the bottom left.

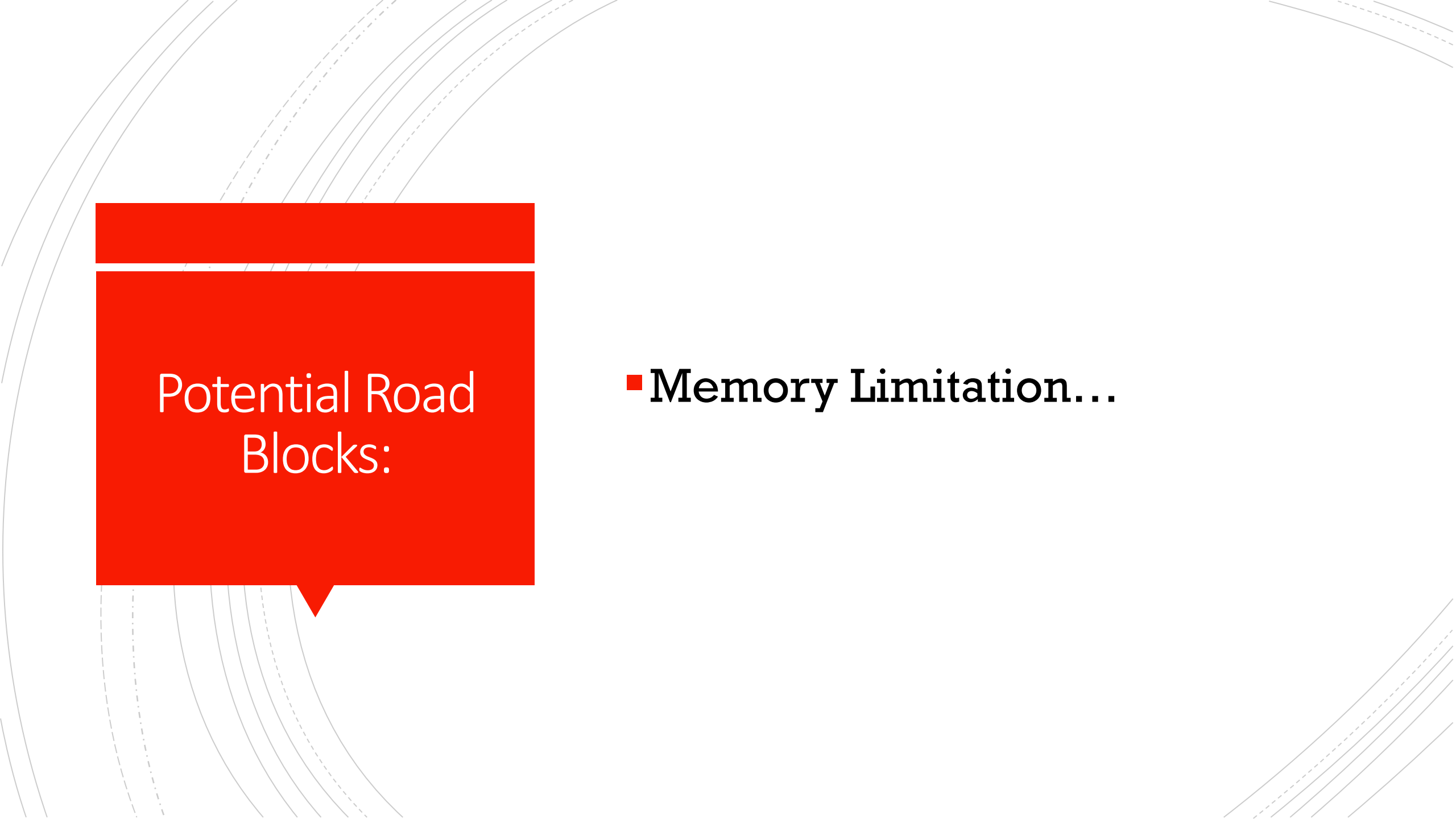
Machine Learning Models

- **Logistic Regression**
- **Naïve Bayes**
- **SVM**
- **Potentially neural nets...**

The background of the slide features several thin, curved lines in a light gray color, some solid and some dashed, creating a sense of motion or a stylized globe. A large red speech bubble is positioned on the left side of the slide.

New Contribution

- **Comparing different text analysis techniques**
- **Combining various text streams**

The background of the slide features several thin, curved lines in a light gray color, some solid and some dashed, creating a sense of motion or a stylized road. A large red speech bubble is positioned on the left side of the slide.

Potential Road
Blocks:

- **Memory Limitation...**

A red speech bubble with a white outline, containing the text "Potential Result".

Potential Result

■ **Can't say**