# EV Sentiment Analysis

Josh Everts, Xuan Jiang

# Question and Motivation

**Research Question: Are there sentiment and/or topic differences between News coverage of Electric Vehicles (EVs) and EV owner/user reviews.**

**Hypothesis: Yes, we think there are differences in sentiment between these groups.**

Additionally:

- Can NLP help us question and/or confirm our hypothesis or bring to light new possibilities we haven't considered?
- Could this analysis help to indirectly observe what the major avenues and barriers to widespread EV adoption may be?

# Our News and Consumer Data

- News Data
  - Bloomberg, CNBC, Reuters, WSJ, Fortune
  - https://www.kaggle.com/jeet2016/us-financial-news-articles [1]
  - Collected January-May 2018
  - Filtered for EV-related bi-grams in article text
    - After filtering: ~1700 articles
- Consumer/User Data
  - Scraped from InsideEVs.com Tesla Model 3 Forum: 'https://www.insideevsforum.com/community/index.php?forums/model-3.32/' [2]
  - ~1900 comments from ~280 threads.

# Approach and Methods

Collect and Filter Data → SemAxis [3] Model → Gensim Topic Model

- News Dataset: Filter with bi-gram term dictionary
- Consumer Reviews: Scraped from insideEVs.com Tesla Model 3 Owners Forum

- Pre-trained GloVecs
- Corpus-trained word2vec
- Decide on positive/negative sentiment vectors using test embeddings
- Find average token-wise sentiment for each corpus

- Tokenize on News and Review Corpora with K=10
- SemAxis Analysis of Topic Tokens

# Results -- Sentiment Analysis

GloVe embeddings vs word2Vec corpora embeddings flip the sentiment order -- this suggests that our analysis is highly dependent on embedding and semantic vector choice.

Significance
Note: SemAxis produces scores between [-1,1]

| Model | News Score | Comments Score | Significance (t-test) |
|---|---|---|---|
| gloVe | .0496 | .0167 | P < .0001 |
| word2Vec | -.035 (news-com) | -.012 (com-com) | P < .0001 |

Conclusion: we found that the two corpora have statistically different sentiments in both an unsupervised and semi-supervised case. But the practical significance is very low as seen by the near-0 sentiments over each corpus.

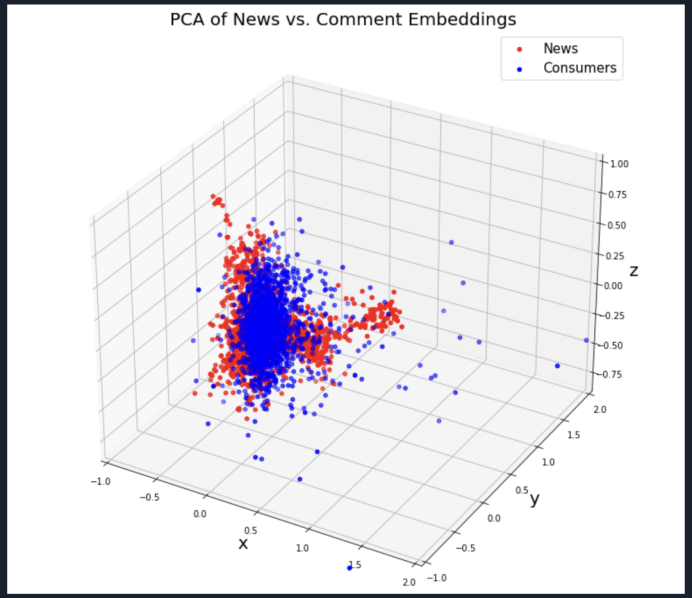
# Results -- Topic Model and Sentiments

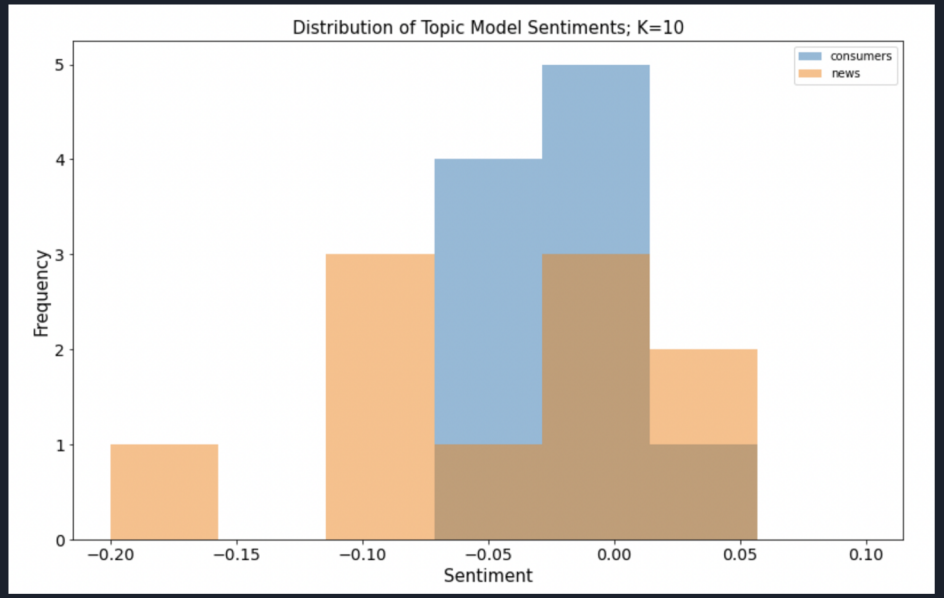
News

```
topic 0:        china cobalt billion mining chinese supply batteries million lithium world demand global battery
topic 1:        's after stocks prices or shares oil stock investors fund earnings down european
topic 2:        oil energy battery power we technology solar years bp million cars global shell
topic 3:        cars daimler german he vw diesel selfdriving euros volkswagen technology billion germany their
```

Consumers

```
topic 0:          model ’ 3 that s on tesla for with we more as but
topic 1:          not but have at would was trip on n't you for power as
topic 2:          that tesla 's for model but 3 will n't as they on this
topic 3:          was for on charge charging but range this battery we 3 mi model
```



# Conclusions

## Summary

- There is a statistically significant difference in sentiment between Tesla Model 3 Owners and business-news headlines, though it is very small in magnitude. Does it generalize?
- Topic Models reveal big differences in discussions; consumers discuss a variety of issues that are hard to capture in a single sentiment.

## Next Steps

- We each performed our own test-labelling of 100 random forum comments with a positive, negative, or neutral sentiment. We obtained moderate agreement with a Cohen's Kappa = 0.521.
- This suggests that a supervised sentiment-classification method using BERT is a good next step.
- Generally need more data, and will add non-Tesla forum comments.
- Remove repeat text in forum comments using sequence alignment.

# Questions?

# Data and References

## Sources

### Data From:

[1]: Jeet.J. “US Financial News Articles.” *Kaggle*, 5 Sept. 2018, https://www.kaggle.com/jeet2016/us-financial-news-articles.

[2]: *Inside EVS Forum*, https://www.insideevsforum.com/community/index.php?forums%2Fmodel-3.32%2F.

### References:

[3]: An, Jisun, et al. “Semaxis: A Lightweight Framework to Characterize Domain-Specific Word Semantics beyond Sentiment.” *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, https://doi.org/10.18653/v1/p18-1228.