

# Capstone project: Sentiment analysis of EV maker's stocks

Sunisa Poonpipatgul, DSIF-10





# Table of contents

**01**

**Background**

**02**

**Data  
acquisition &  
cleaning**

**03**

**EDA &  
Sentiment  
Analysis**



**04**


**Modeling**

**05**

**Conclusions**

**06**

**What's next?**



# Background

The automotive industry is rapidly embracing electric vehicles (EV) as the forefront of the future. With each passing year, EV markets are seeing exponential growth with a growing number of manufacturers substantially expanding their electric lineup.

The growth in EV sector is driven by climate change and the reduction of carbon footprints generated from transportation.

As an emerging sector with growth potential of 22.6% per year between 2020-2027, Investing in this EV sector could yield substantial returns in the future.







# Problem statement

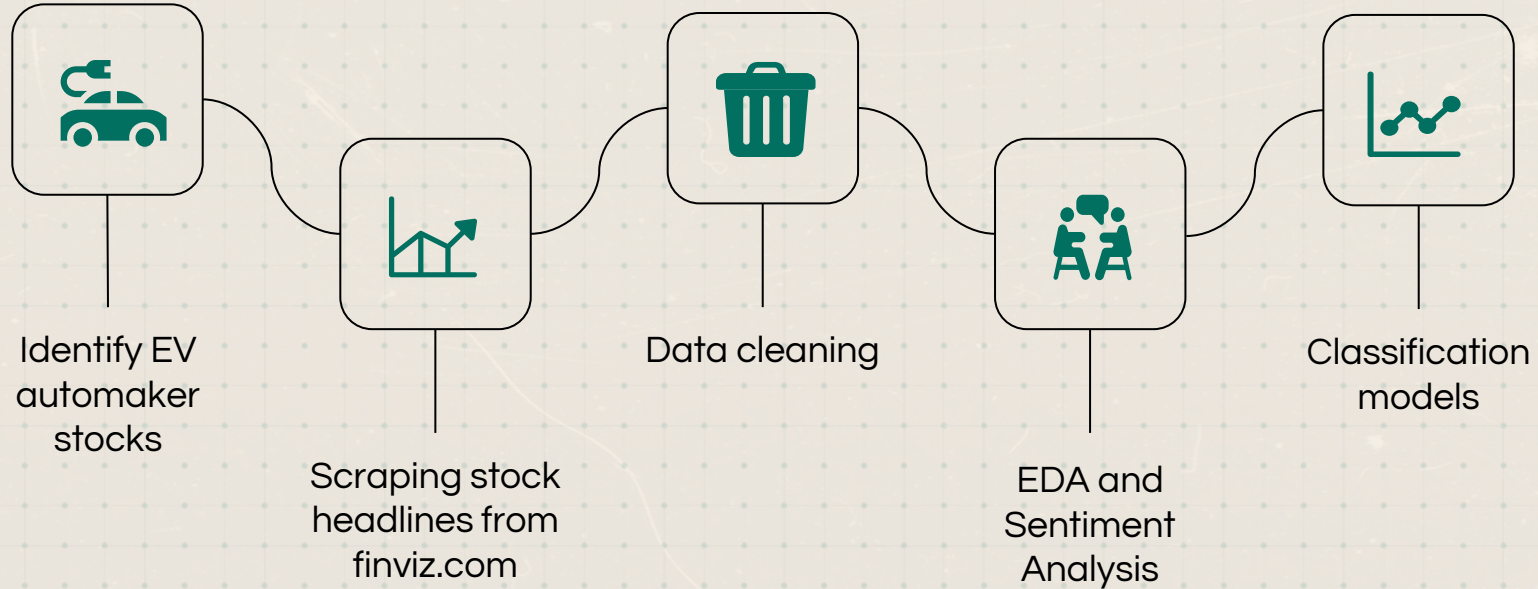
In this era, financial news happens in seconds. We can leverage this data to gain an edge in investment with insight from the news headlines.

Goal as independent investors who are interested in investing in EV manufacturers:

- To understand the emotion behind the headlines and predict whether the market feels good or bad about stocks in this segment.
- Possible to educated guess on how specific stocks will perform and trade accordingly.



# Methodology





# Here are six important EV stocks



**Tesla, Inc.**

TSLA



**NIO Inc.**

NIO



**Li Auto Inc.**

LI



**Xpeng Inc.**

XPEV



**Rivian Automotive**

RIVN



**Lucid Group Inc.**

LCID





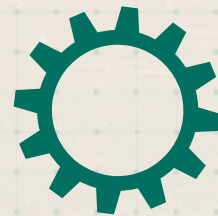


# Data acquisition and data dictionary

- Scraping from finviz.com on 23<sup>rd</sup> August 2023

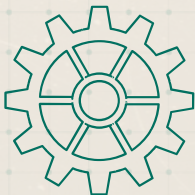
| # | Data     | Number of data | Data type | Description  |
|---|----------|----------------|-----------|--|
| 0 | stock    | 600            | object    | Stock tickers  |
| 1 | headline | 600            | object    | News headline that display on the finviz.com per stock, maximum 100 news headlines were scraped per time |
| 2 | source   | 600            | object    | Sources of news  |
| 3 | date     | 600            | datetime  | News published date  |
| 4 | Time     | 600            | datetime  | News published time  |



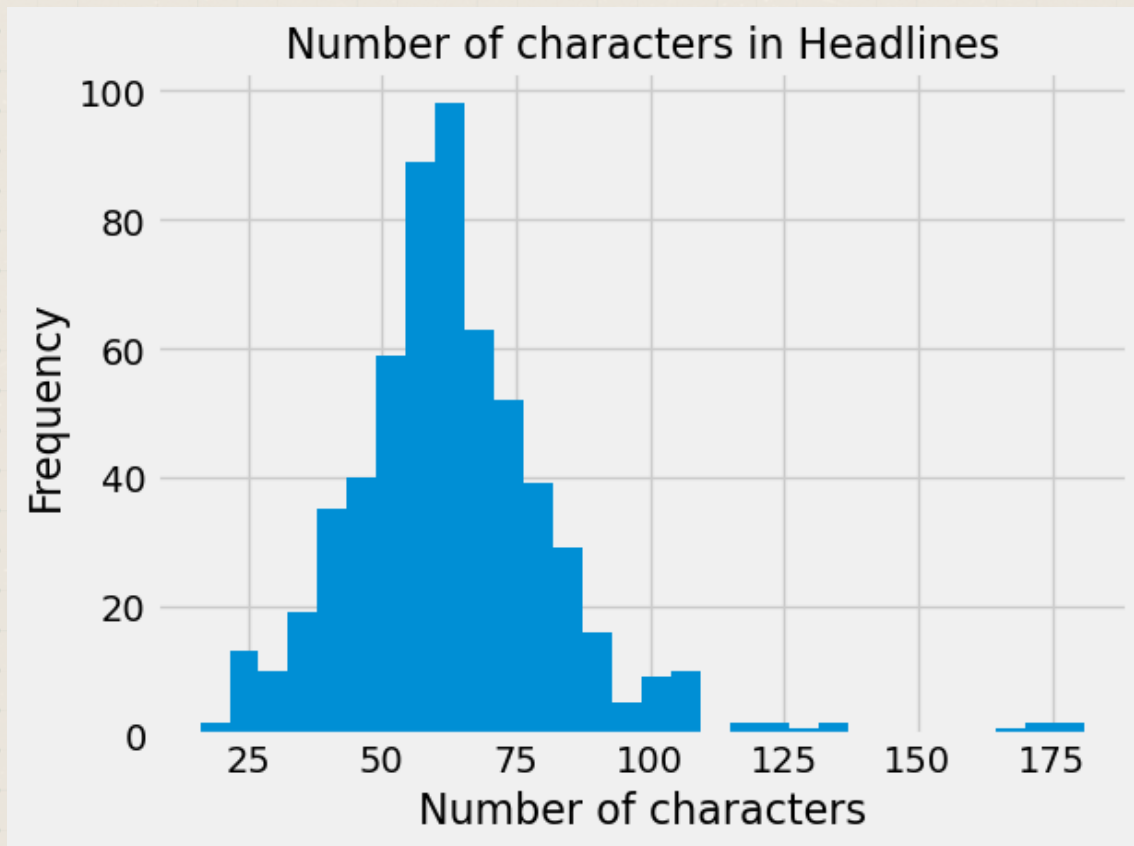


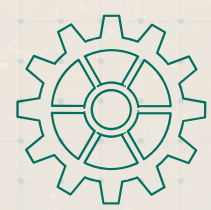
# EDA & Sentiment Analysis





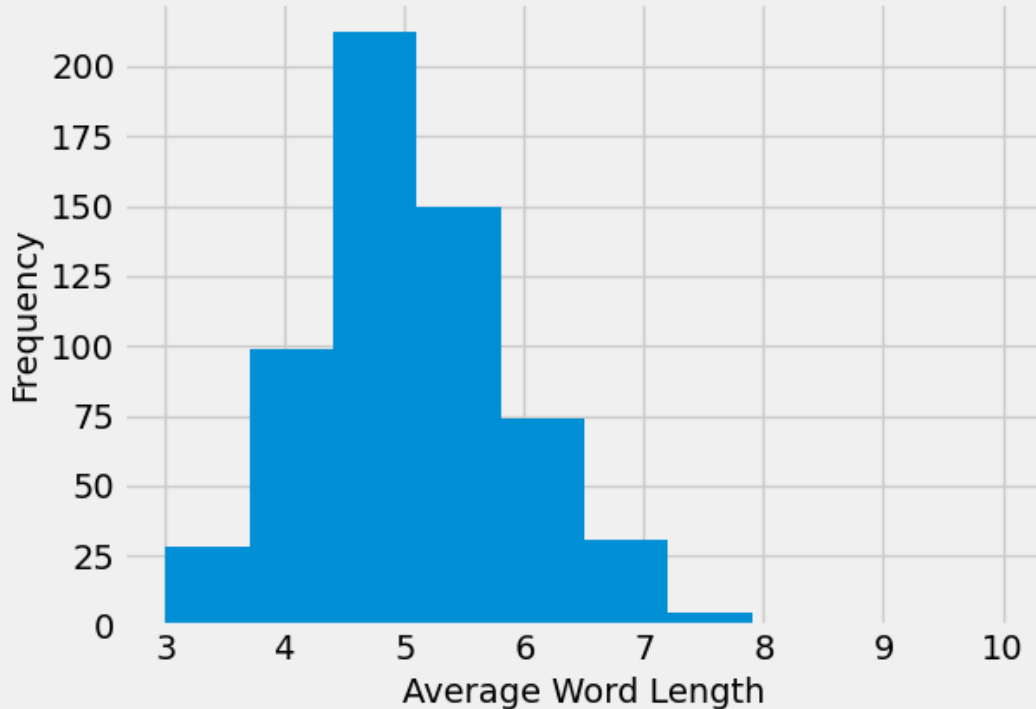
# Number of characters in Headlines

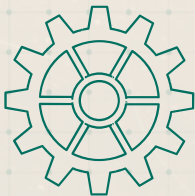




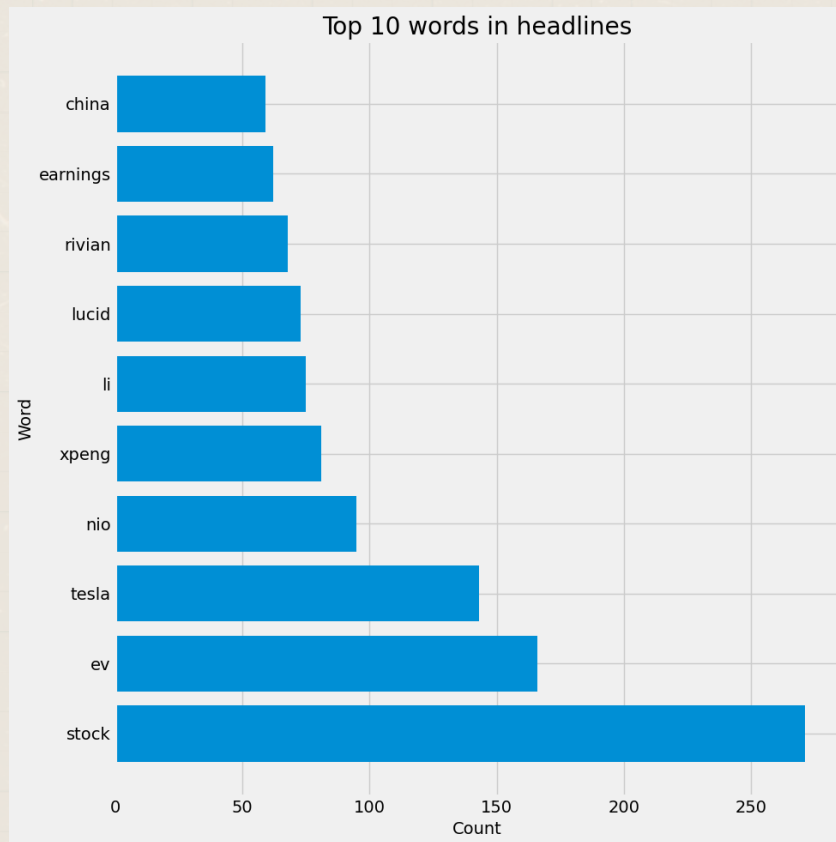
# Average word length in headlines

Distribution of Average Word Lengths in Headlines





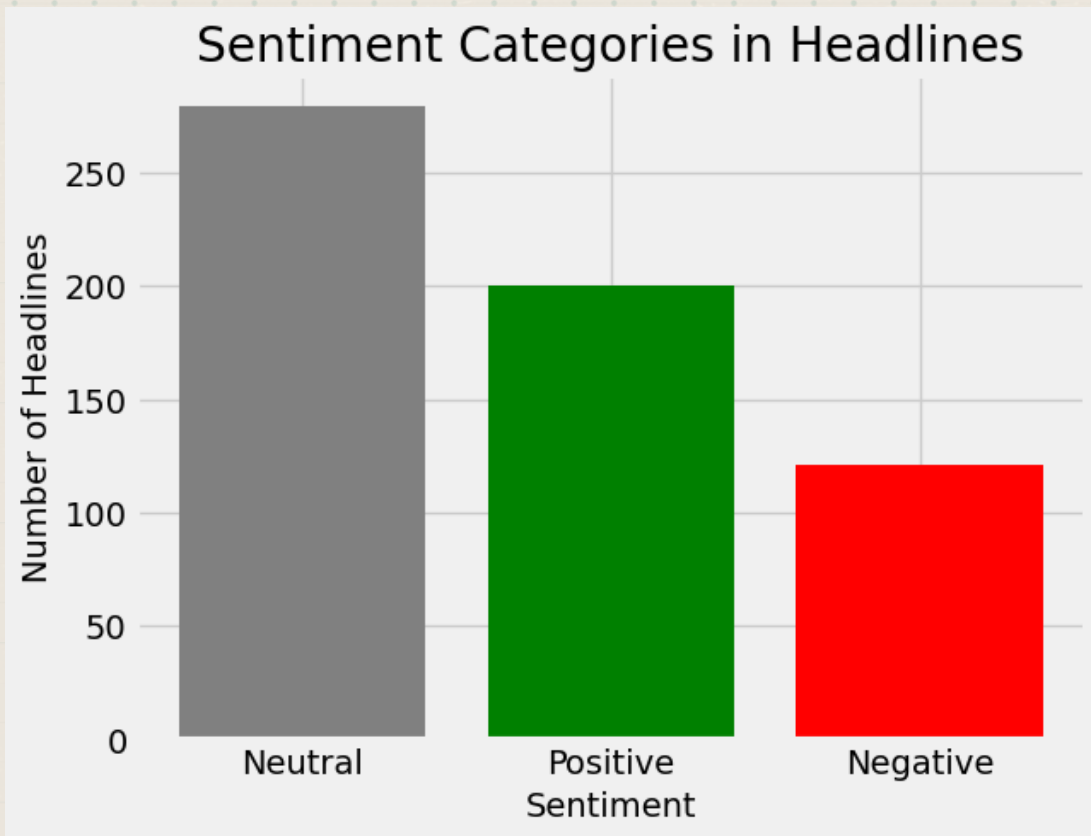
# Top words

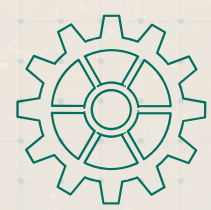




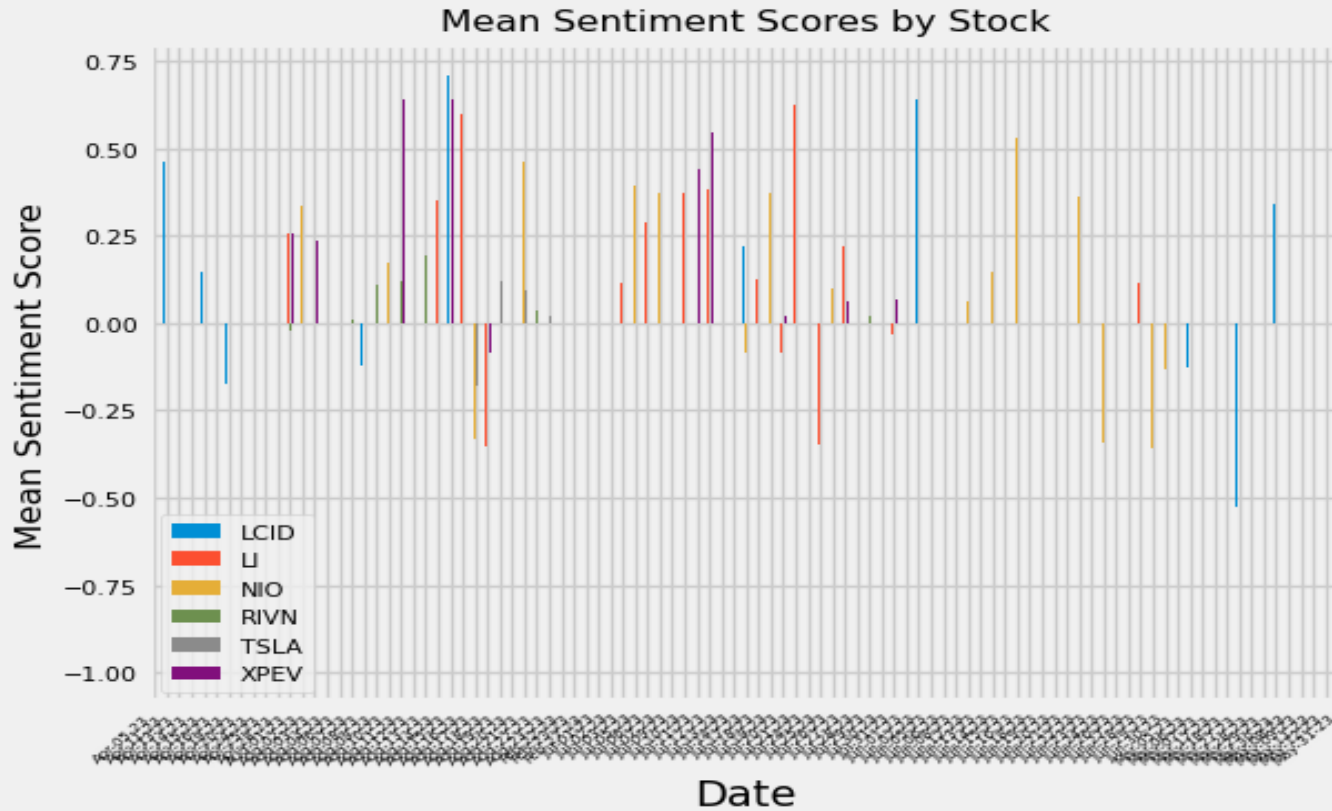


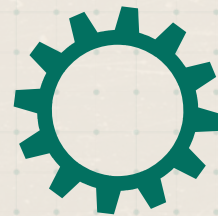
# sentiment categories





# Stock sentiment





# Modeling





# Modeling

## Models

- Naive Bayes (NB) model
- Random Forest (RF) model

## Feature extraction techniques

- CountVectorizer (Cvec)
- TF-IDF Vectorizer (Tvec)

# Model Performances

| Model       | NB-Cvec | NB-Tvec | RF-Cvec | RF-Tvec |
|-------------|---------|---------|---------|---------|
| Best score  | 0.693   | 0.664   | 0.690   | 0.690   |
| Train score | 0.993   | 0.910   | 1.000   | 1.000   |
| Test score  | 0.639   | 0.556   | 0.656   | 0.667   |
| Specificity | 0.607   | 0.129   | 0.266   | 0.276   |

Baseline = 0.466667



# Model selection decisions

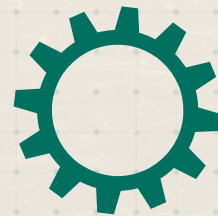
After constructing and analyzing various models, all of the constructed models showed an improvement over the baseline model.

Naïve Bayes model with CountVectorizer (NB-Cvec) seem to be the best choice here in classifying the sentiment on the stock news headlines.

This model showed an improvement over the baseline model indicating that our model significantly outperforms a random guess or a simple heuristic.







# Conclusions and next steps



# Conclusions and recommendations

From sentiment analysis, most of the news has neutral sentiment. And more positive than negative.


Unsurprisingly, most top words are EV brands, as these news are curated related to the specific stocks.

The most mentioned brand is Tesla which is consider a pioneer and leader in this segment.

Followed by Nio, Xpeng, Li, lucid, and Rivian.

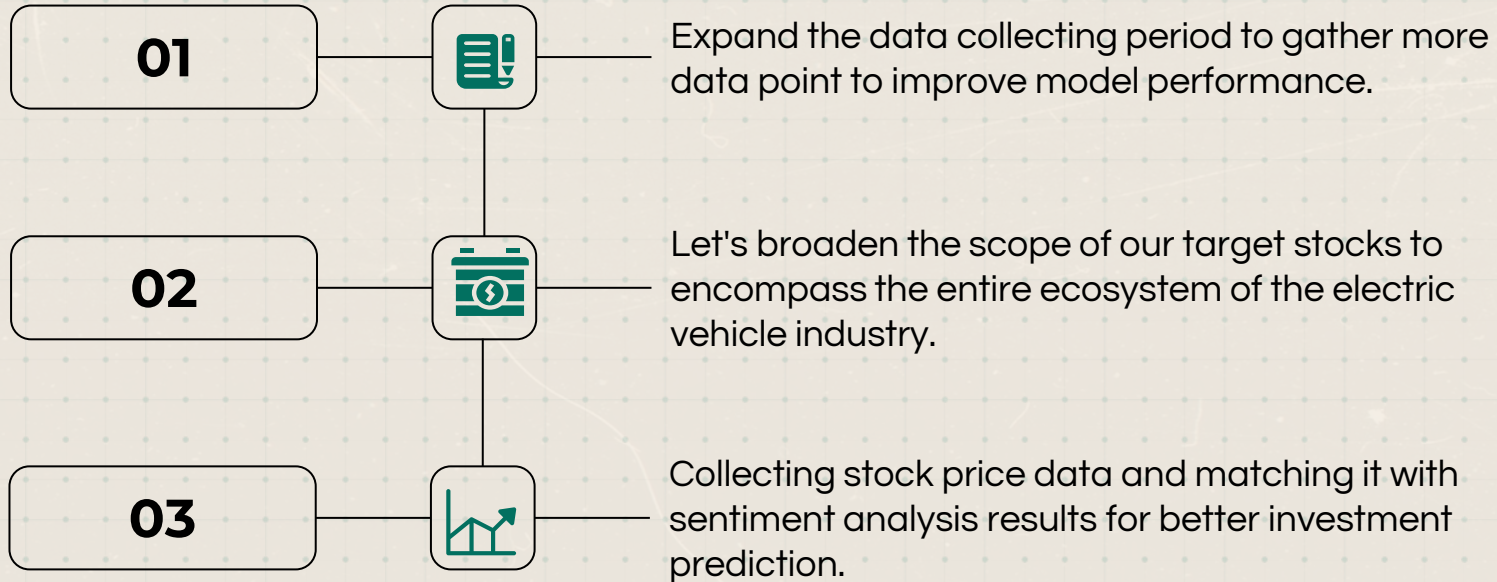
In top words, China is the only country mentioned in the news headlines. This might be due to China's dominance as World's EV leader.

Observation over a period of time, the distribution of news headlines for EV manufacturers' stock is toward the positive sentiment. Indicating a positive outlook towards the stocks in this segment.





# What's next





# Thanks!

Do you have any questions?

**CREDITS:** This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

Please keep this slide for attribution

