



STOCKS & SENTIMENT

GROUP 55

WHY DO WE CARE?

- Higher, safer returns
- Better business strategy
- Better marketing



EFFICIENT MARKET HYPOTHESIS



A market where all past prices and data are incorporated into current stock prices

WEAK



A market where all publicly available information is almost immediately reflected in stock prices

SEMI-STRONG



A market where all information, both public and private, is reflected in stock prices.

STRONG

THE DATA

S & P 500



News headlines from
Kaggle





INVESTOR SENTIMENT ANALYSIS: OVERVIEW, APPROACHES, AND IMPLEMENTATION

SENTIMENT ANALYSIS



Overview

- Sentiment analysis is a technique to identify and extract the polarity (positive, negative, or neutral) of text data, such as news headlines.
- In this context, sentiment analysis examines how investor perception of news headlines might impact stock prices.
 - However, sometimes pure sentiment isn't enough - it must be contextualized.

Method	Shortcomings	Strengths
VADER	<ul style="list-style-type: none">• Can't contextualize headlines relative to US markets• Can't capture nuanced sentiments	<ul style="list-style-type: none">• Very fast to run• Very easy to implement
BERT	<ul style="list-style-type: none">• Can't contextualize headlines relative to US markets• Only covered positive/negative ratings	<ul style="list-style-type: none">• Fast to run• Using existing models, it is easy to implement• Able to process raw sentiment well
LLMs (e.g., Vicuna-7B)	<ul style="list-style-type: none">• Powerful hardware is needed• Takes very long to run• Output is messy, as it's a text model	<ul style="list-style-type: none">• Highly flexible• Understands context/language the best• Highly active area of research with great funding

LLM IMPLEMENTATIONS



With LLMs exploding in use cases, one must consider how they should be leveraged to bring benefit to a project

Approach 1 - **Basic**

LLMs (like Vicuna-7B) have online chat interfaces. Users can easily, with no technological expertise, ask questions to trial how LLMs handle various tasks.

Approach 2 - **Intermediate**

Open-source models like Vicuna-7B are free to use, provided you have the hardware. Great results can be achieved for anyone. We setup a locally hosted model with a RESTful API to query the model.

Approach 3 - **Professional**

The most powerful models are proprietary, with a cost to their API. They are easy to use, but can get quite expensive. However, the results are easily worth it in a professional setting.

OUTPUTS AND RATIONALES FROM EACH LLM MODEL

Headline: "Fox News Ripping Content Straight from the Depths of Reddit"

Vicuna-7B (8-bit mode)

This headline suggests that Fox News is plagiarizing content from Reddit, which could potentially have negative impacts on the US financial markets. The news could lead to a decline in trust in Fox News as a reliable source of information, which could affect the stock prices of companies that advertise on the network or have business relationships with them.

Negative

ChatGPT (GPT-3.5)

This headline pertains to a media outlet, Fox News, allegedly using content sourced from Reddit. While it may attract attention and generate discussion within the media landscape, the direct impact on the US financial markets is likely to be minimal. Media controversies or sourcing practices typically do not have a significant influence on large cap stock prices unless they involve severe legal or ethical violations that result in substantial legal consequences or damage to the reputation of companies involved. Additionally, since the sentiment analysis is focused on financial markets, the potential impact of this headline on stock prices or investor sentiment is likely to be neutral.

Neutral

Determining the Relevance of Sentiment

B

LSTM with only historical data only

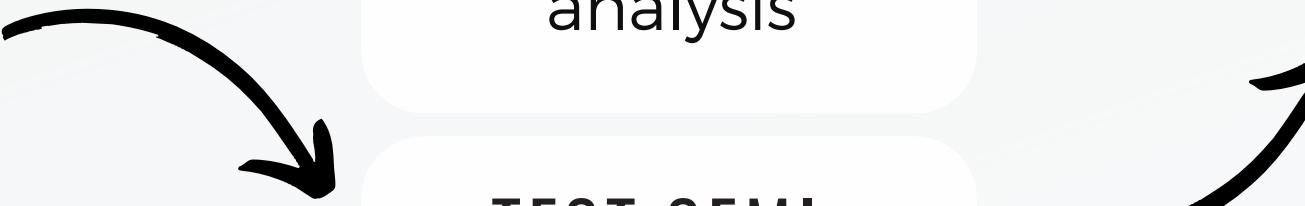
TEST WEAK EMH

LSTM with historical data and sentiment analysis

TEST SEMI-STRONG EMF

Does sentiment analysis improve predictive performance?

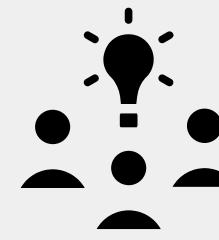
COMPARE



LSTM OVERVIEW



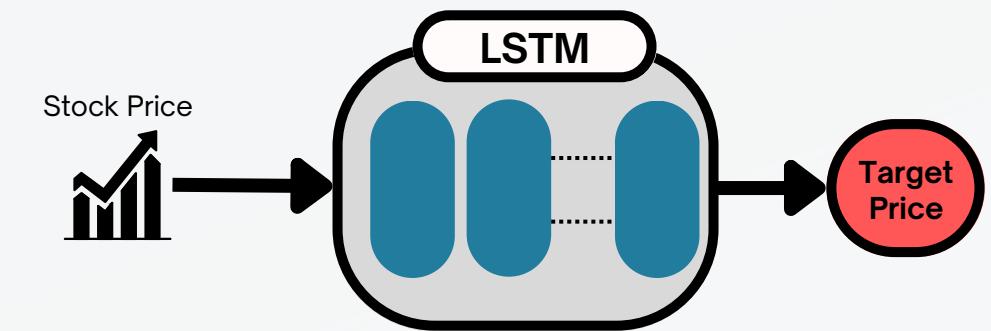
LSTM stands for long short-term memory, a type of neural network that excels with sequential data



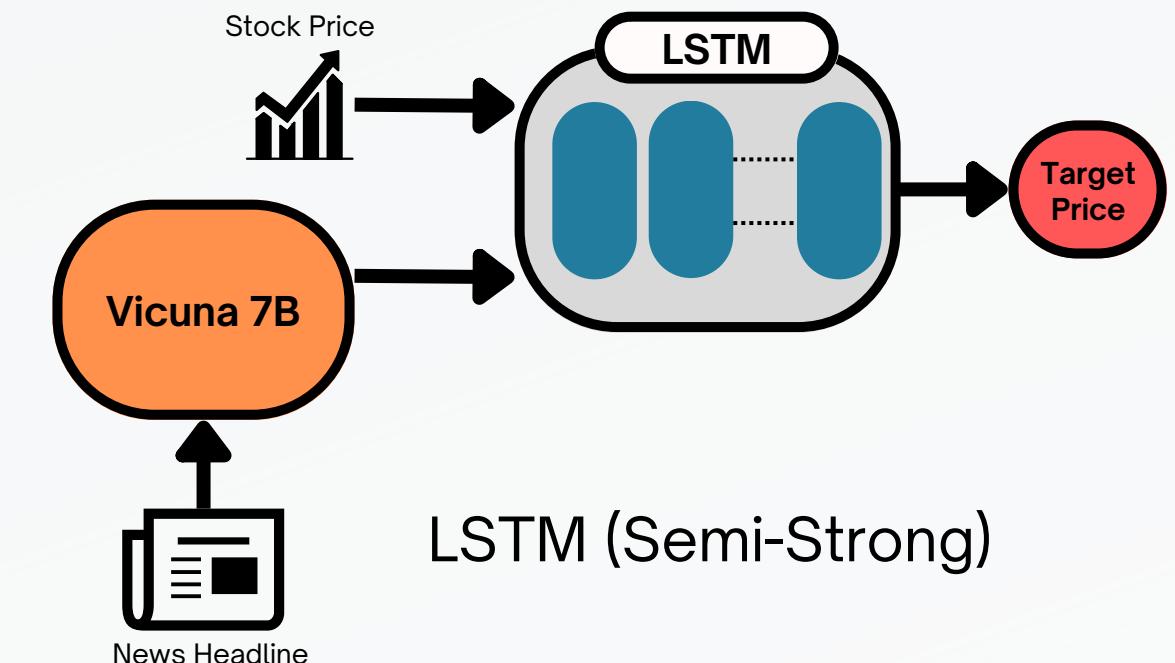
Sentiment analysis assigns a polarity or emotion score to headline news, such as positive, negative, or neutral



We tested the weak and semi-strong form of the efficient market hypothesis (EMH)



LSTM (weak)



LSTM (Semi-Strong)

LSTM COMPONENTS

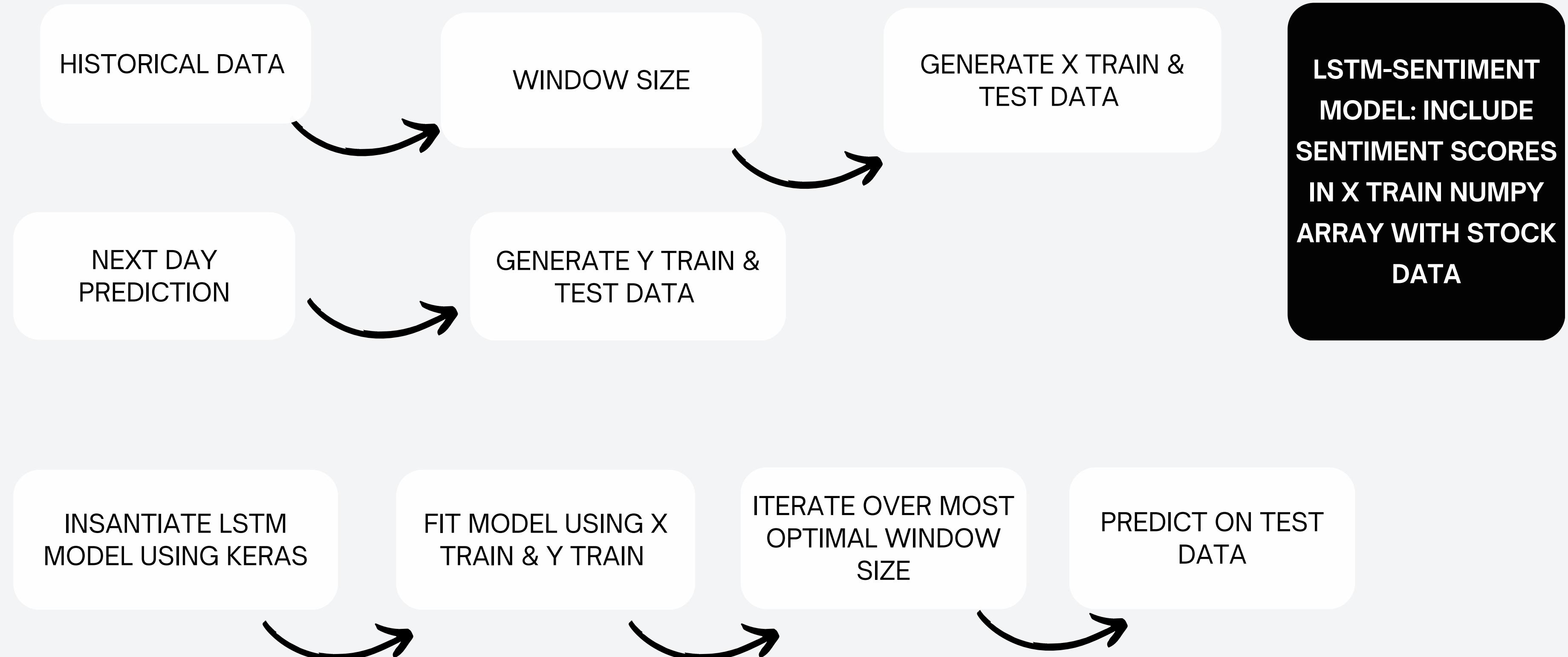
WINDOW SIZE

LSTM CELLS

DENSE LAYERS

ACTIVATION

LSTM IMPLEMENTATION



SENTIMENT IMPROVES PERFORMANCE

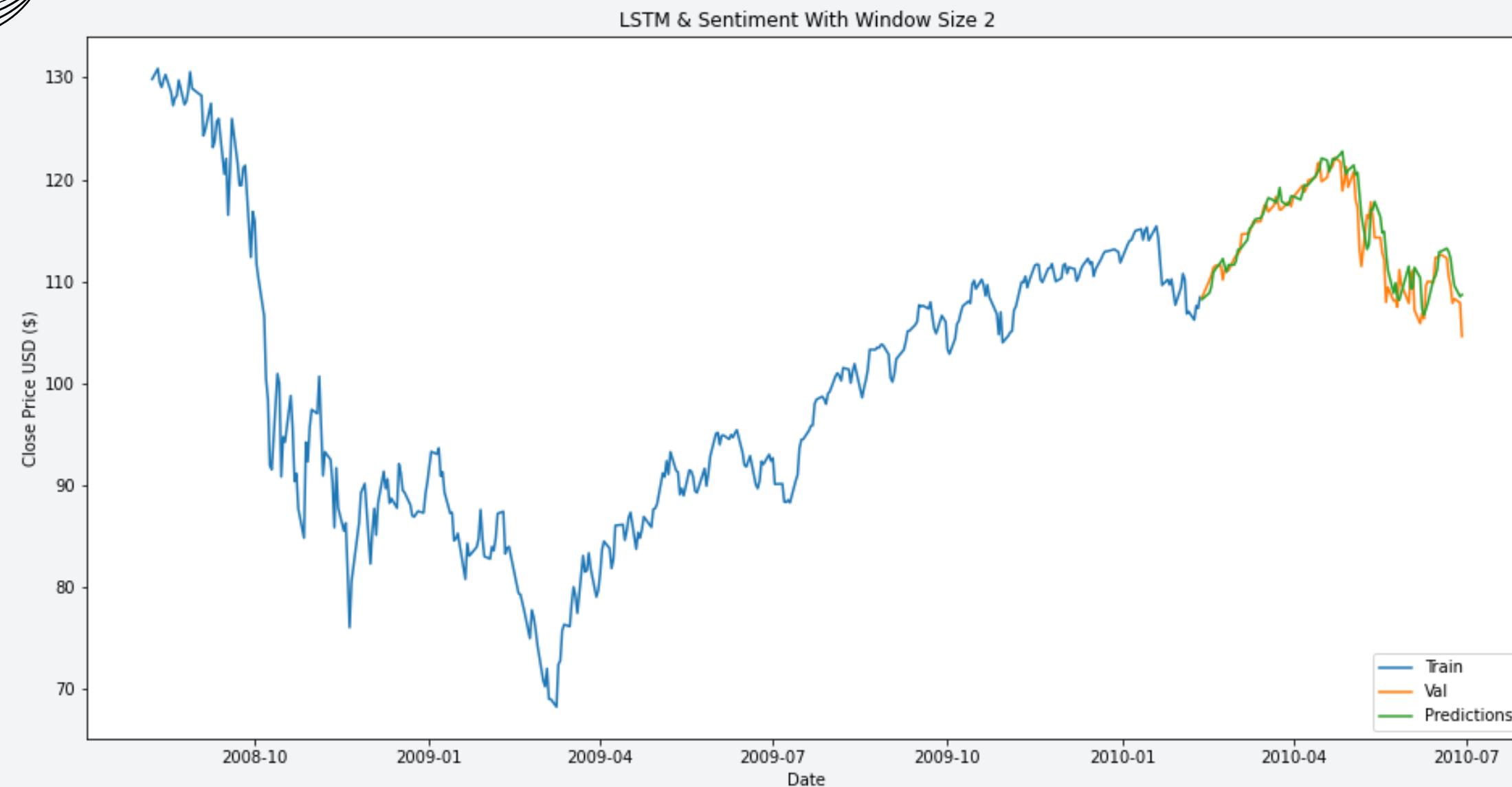
No Sentiment

With Sentiment

Window Size	RMSE
2	2.37
4	2.40
6	3.36
8	3.10
10	2.40
12	2.66

Window Size	RMSE
2	1.86
4	2.64
6	3.28
8	2.60
10	2.83
12	2.17

BEST MODEL



- A window size of 2 is optimal, as larger window sizes lead to higher RMSE and more prediction errors.
- The LSTM sentiment model shows a gap between validation and prediction data around April 2010, due to a string of negative sentiment scores that affect the model's sensitivity

FUTURE DIRECTIONS

- Q Experiment with different sentiment analysis sources and methods
- Q Explore other analytical techniques or models, such as attention-based or transformer-based LSTM or deep reinforcement learning
- Q Provide evidence that sentiment analysis can improve stock price prediction and that public information can influence stock prices



THANK YOU

OUR TEAM

Ben
Kacikanis

Brett
Gileau

Jonah
Maroszek

Shaya
Shakib