

Jannick Drechsler  
Simon Papekyan  
Jack Jürgens  
Konstantinos Papanagiotou

July 8, 2025

# STOCK PRICE PREDICTION

Using Financial Metrics coupled with  
Sentiment Analysis

**For the course**

Data Science In Practice  
taught by :

Professor Dr. Anne Lauscher, Pranav Agrawal

# INTRODUCTION



## THE GOAL

TO PREDICT FUTURE STOCK PRICE MOVEMENTS USING A COMBINATION OF

- TECHNICAL ANALYSIS
- SENTIMENT ANALYSIS BASED ON FINANCIAL NEWS

## APPROACH

WE USE AN EXPERIMENTAL METHOD COMBINING HISTORICAL PRICE DATA WITH REAL-WORLD NEWS SENTIMENT TO IMPROVE FORECASTING ACCURACY.

# FINANCIAL DATA SELECTION

**yfinance** was used to import the basic financial metrics, as in, the open/close prices & volume for each of **4 volatile stocks**, namely **NVDA, AAPL, MSFT, TSLA**, over a period of 14 years.

After that, having imported **matplotlib** and **pandas** made calculations of financial indicators possible. The indicators calculated for Fundamental Analysis were the ones listed below :

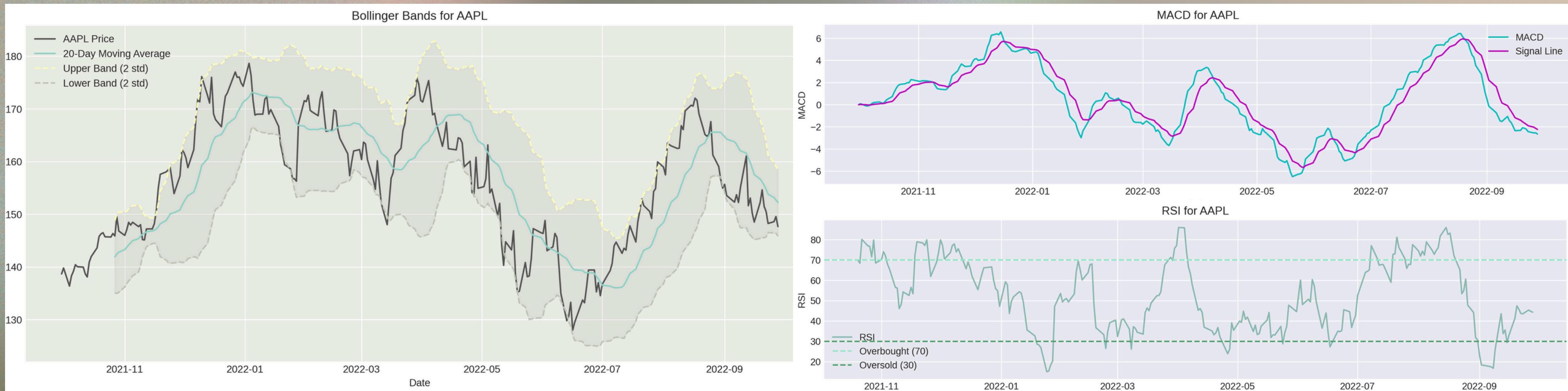
**MACD:** trend-following indicator that smooths out short-term fluctuations to highlight longer-term price trends.  
Moving Average Convergence/Divergence

**RSI:** momentum oscillator, identifies changes in momentum and price direction.  
Relative Strength Indicator

**BOLLINGER BANDS:** gauge market volatility and identify when securities are poised to rise or fall.

**DAILY RETURNS:** measure the percentage change in the stock's closing price from one day to the next. Volatility & performance insight.

# IMPLEMENTATION OF TECHNICAL ANALYSIS



Using the output of our technical analysis module, demonstrated above, we are given valuable visual insights on how these factors are relevant to the future price of a stock.

the plotting of the financial indicators in python was made possible by using pandas and matplotlib features, along with dynamic file naming, in a 'for' loop

# DATA SELECTION FOR SENTIMENT ANALYSIS

FNSPID: Dataset including over 90 million news headlines

General market information

Global coverage of trends

	Date	Article_title
0	2023-12-16 23:00:00 UTC	Interesting A Put And Call Options For August ...
1	2023-12-12 00:00:00 UTC	Wolfe Research Initiates Coverage of Agilent T...
2	2023-12-12 00:00:00 UTC	Agilent Technologies Reaches Analyst Target Price
3	2023-12-07 00:00:00 UTC	Agilent (A) Enhances BioTek Cytation C10 With ...
4	2023-12-07 00:00:00 UTC	Pre-Market Most Active for Dec 7, 2023 : SQQQ,...
5	2023-12-05 00:00:00 UTC	A Quantitative Stock Analysis

# DATA PREPARATION FOR SENTIMENT ANALYSIS

CLEANING: Add rows for stock count

	NVDA	MSFT	AAPL	TSLA
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

Drop unwanted empty or faulty rows

Unify date format



Date
2023-12-16 23:00:00 UTC
2023-12-12 00:00:00 UTC
2023-12-12 00:00:00 UTC
2023-12-07 00:00:00 UTC
2023-12-07 00:00:00 UTC
2023-12-05 00:00:00 UTC

2007-11-29
2007-11-30
2007-12-01
2007-12-02
2007-12-03

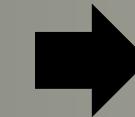
# SENTIMENT ANALYSIS

Model: Vader lexicon

Calculation: Sentiment score by headline and daily mean

Output: Daily overall market sentiment & Stock mention count

	Date	Article_title
0	2023-12-16 23:00:00 UTC	Interesting A Put And Call Options For August ...
1	2023-12-12 00:00:00 UTC	Wolfe Research Initiates Coverage of Agilent T...
2	2023-12-12 00:00:00 UTC	Agilent Technologies Reaches Analyst Target Price
3	2023-12-07 00:00:00 UTC	Agilent (A) Enhances BioTek Cytation C10 With ...
4	2023-12-07 00:00:00 UTC	Pre-Market Most Active for Dec 7, 2023 : SQQQ,...
5	2023-12-05 00:00:00 UTC	A Quantitative Stock Analysis



1	date, sentiment_score_mean_day
2	2007-11-29, 0.10356455990694068
3	2007-11-30, 0.08844748657881894
4	2007-12-01, -0.007009915611814344
5	2007-12-02, 0.012057918552036199
6	2007-12-03, 0.10727528134254688



Stock mentions found:  
NVDA: 23845 articles  
MSFT: 43828 articles  
AAPL: 76902 articles  
TSLA: 30101 articles

	NVDA	MSFT	AAPL	TSLA
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

# ALGORITHM SELECTION FOR STOCK PREDICTION

... THE RIGHT ALGORITHMS FOR FOLLOWING REQUIREMENTS:

- TIME-SERIES HANDLING WITH MEMORY
- MULTIMODAL DATA INTEGRATION
- SEQUENCE HANDLING
- DATA QUALITY AND VOLUME
- PREDICTION GOAL: PRICE VALUE (REGRESSION)



# ALGORITHM OVERVIEW

LINEAR REGRESSION

LINEAR-RELATIONSHIPS

WHY NOT USED

K-NN

NEAREST HISTORICAL DATA POINTS

FAILS ON NON-LINEAR MARKET BEHAVIOR

DECISION TREE  
RANDOM FOREST  
XGBOOST

NON-LINEAR WITH STRUCTURED DATA

NO SEQUENCE HANDLING

CNN

CAPTURES PATTERNS WITH CONVOLUTION

LACK SEQUENCE MEMORY

REINFORCEMENT LEARNING

REWARD-BASED LEARNING (TRADING)

NOT BUILT FOR LONG-TERM DEPENDENCIES

FOCUSED ON TRADING STRATEGY,  
NOT PRICE PREDICTION ACCURACY

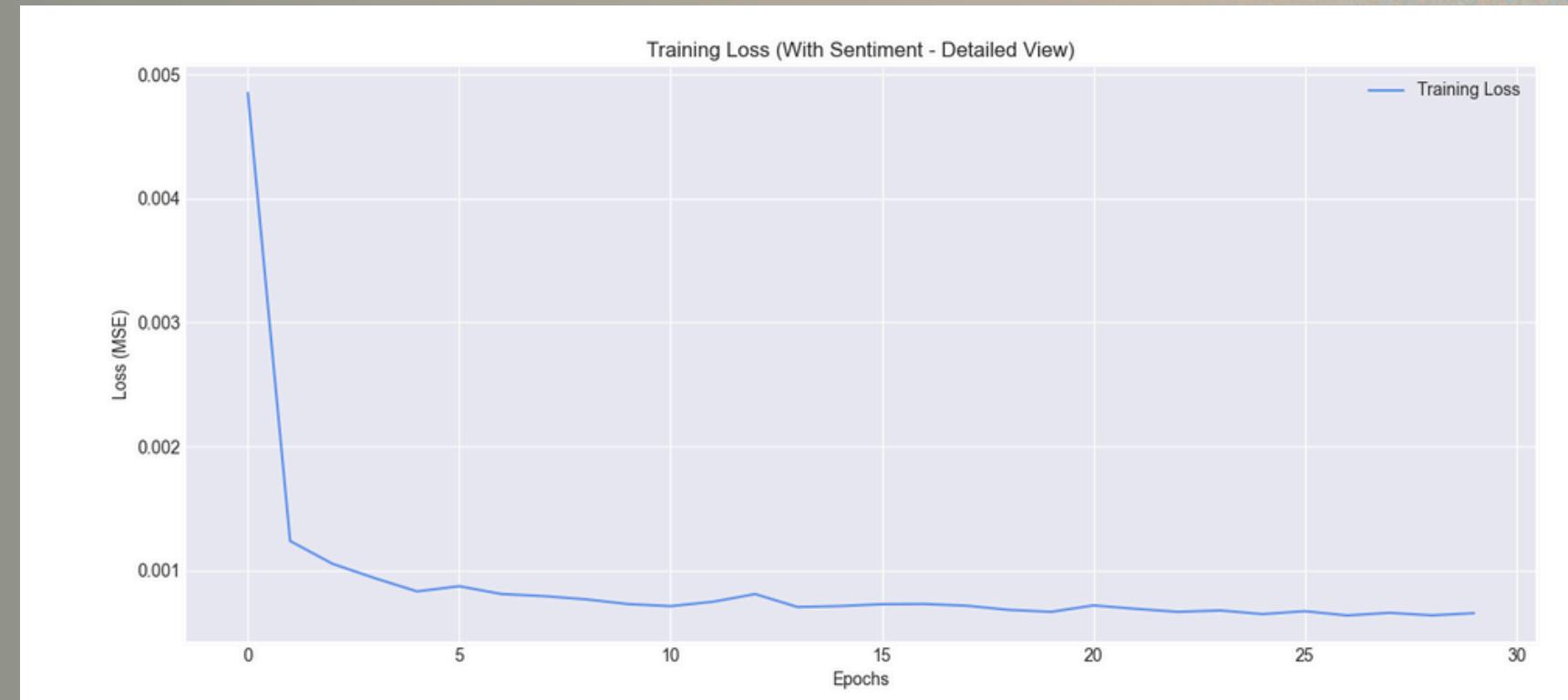
# ALGORITHM SELECTION: LSTM



- HANDLES TIME-BASED DATA  
LSTMS ARE DESIGNED TO SPOT PATTERNS OVER TIME.
- COMBINES MULTIPLE DATA TYPES  
WORKS WITH TECHNICAL INDICATORS, AND SENTIMENT DATA
- LEARNS FROM HISTORICAL NON-LINEAR TRENDS
- REGRESSION  
PRICE VALUE - NOT CLASSIFICATION (UP/DOWN)

# LSTM ARCHITECTURE

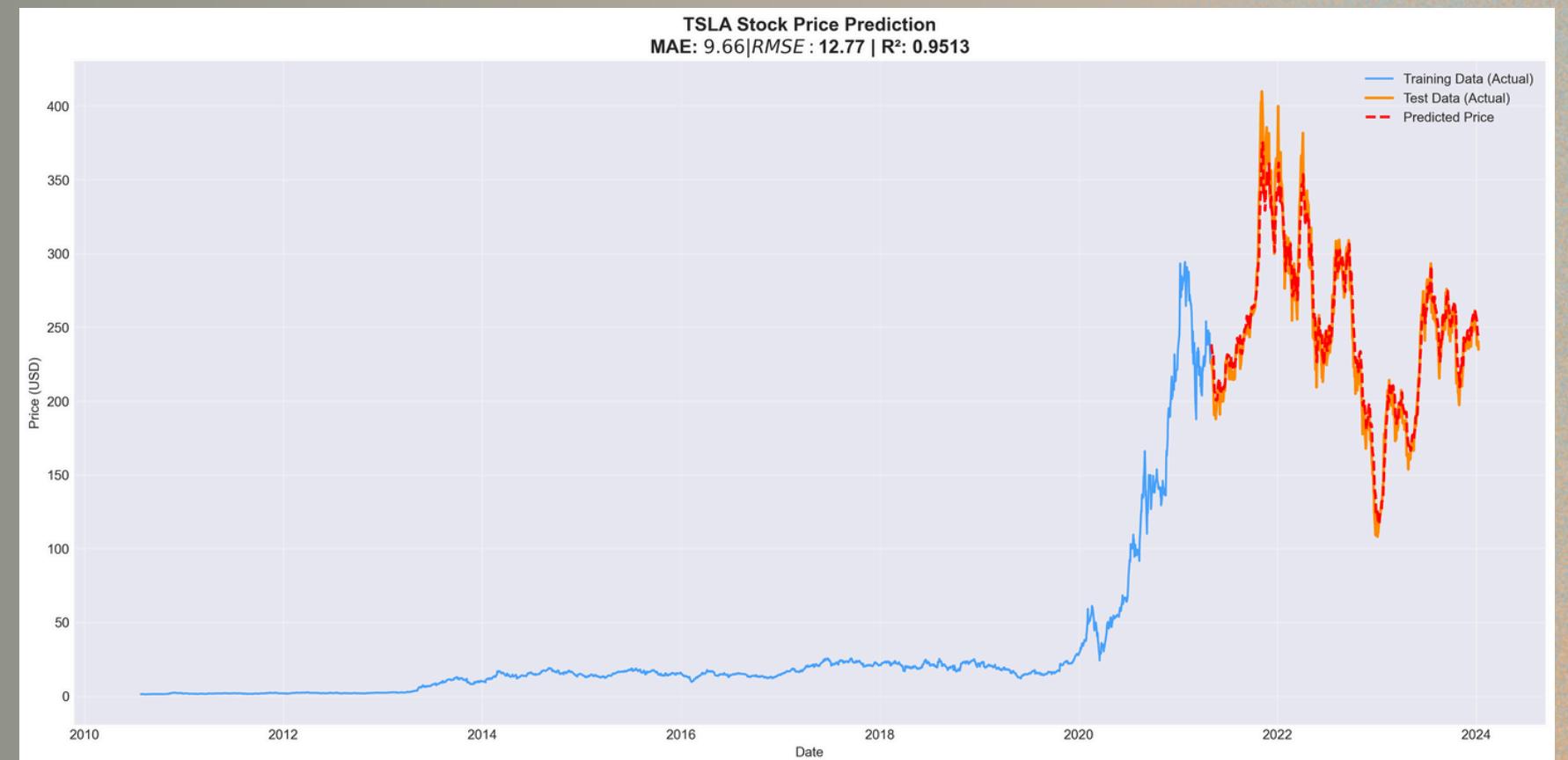
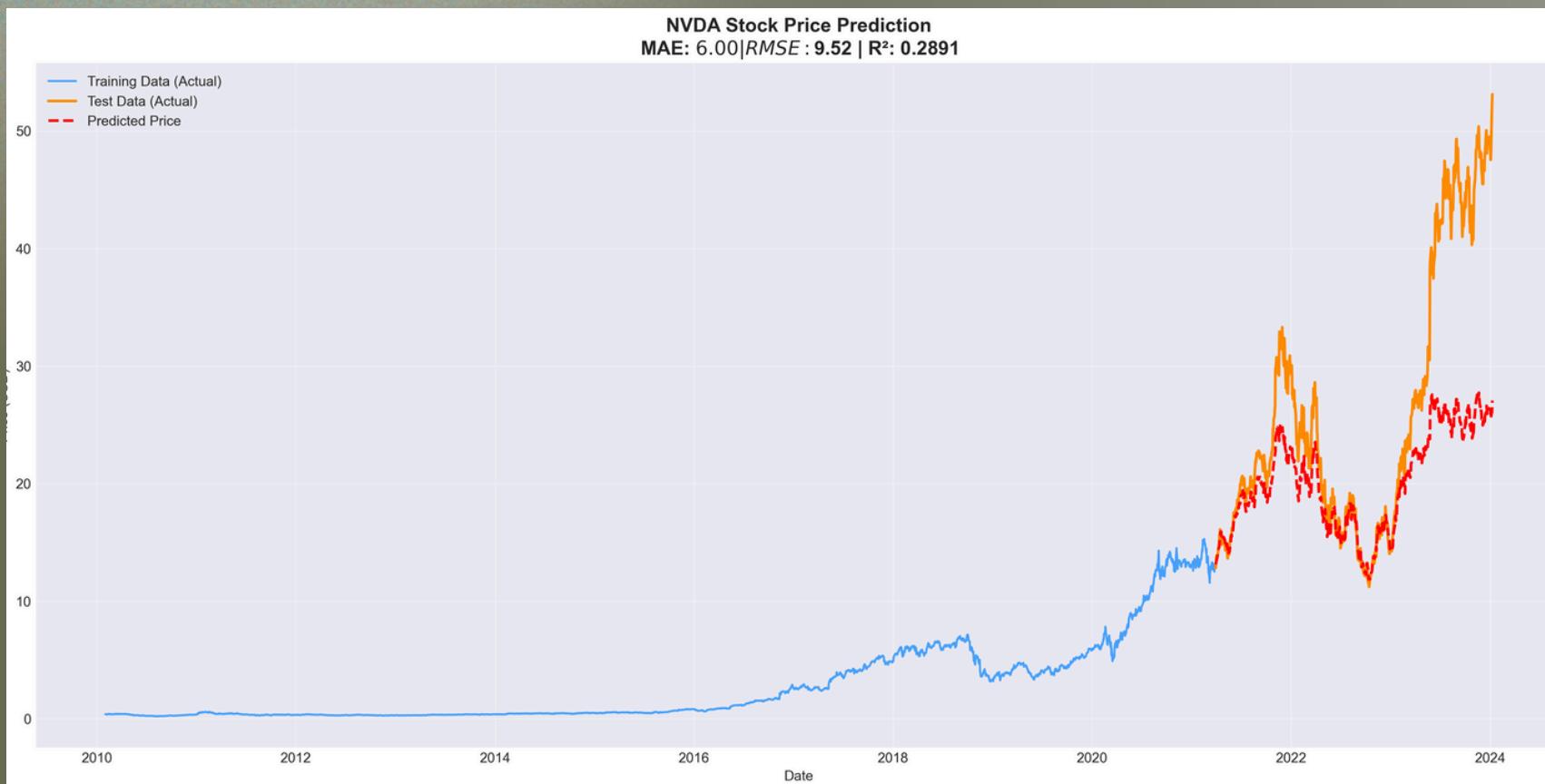
- multi stock Regression
- 60 days of data for next day prediction
- 12 Features
  - closing price, market sentiment, financial metrics
- Trained the model over 30 epochs
- 10.894 total sequences



# EVALUATION

- coefficient of determination  $\varnothing = 0.7147$
- $\varnothing$  absolute error = 7.09\$
- best performance → TSLA
- worst performance → NVDA

Stock	MAE (\$)	RMSE (\$)	R <sup>2</sup> Score
TSLA	9.34	12.42	0.9513
NVDA	5.31	8.70	0.2891
AAPL	3.56	4.54	0.8629
MSFT	10.17	13.81	0.7557



Jannick Drechsler  
Simon Papekyan  
Jack Jürgens  
Konstantinos Papanagiotou

(these guys)

THANK YOU FOR YOUR  
ATTENTION !

