

多模態神機妙算 - AI帶你在股市賺

Using AI to Develop
Stock Trading Strategy

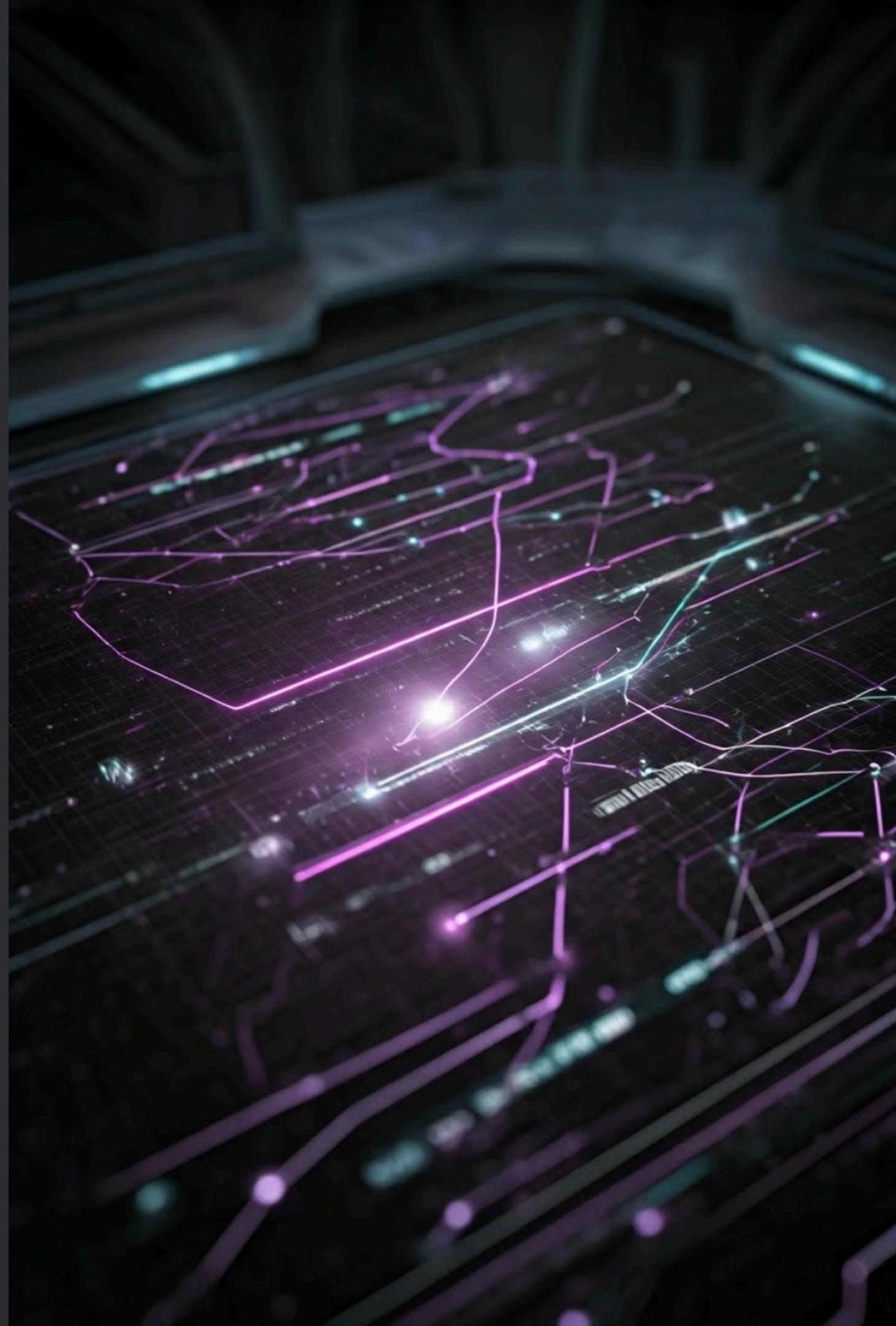
Members: Ryan Kuo, Lee Po-Ting,
Cheng-Jih Chang, Johnson Chang



Introduction



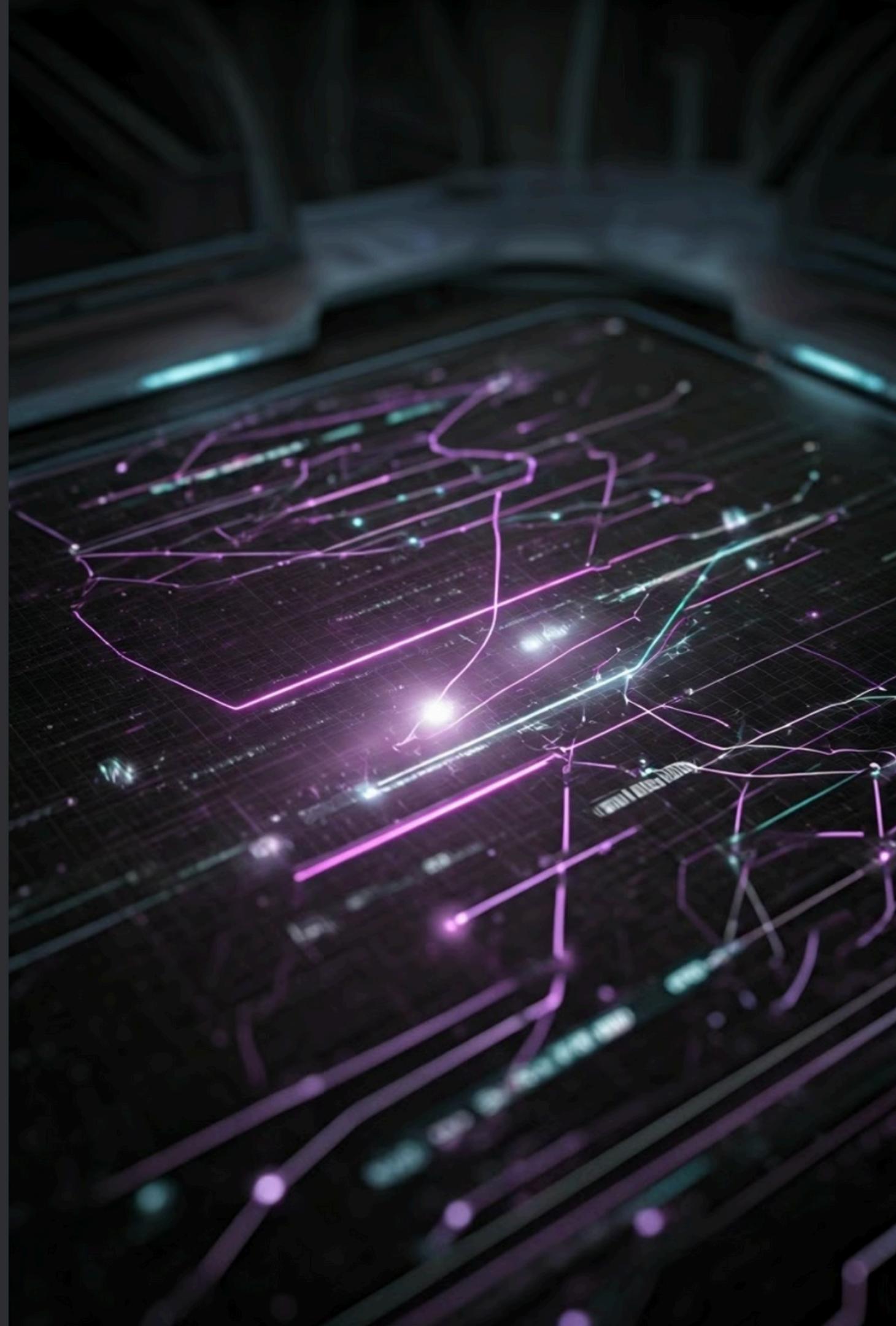
- Fundamental Analysis



Introduction



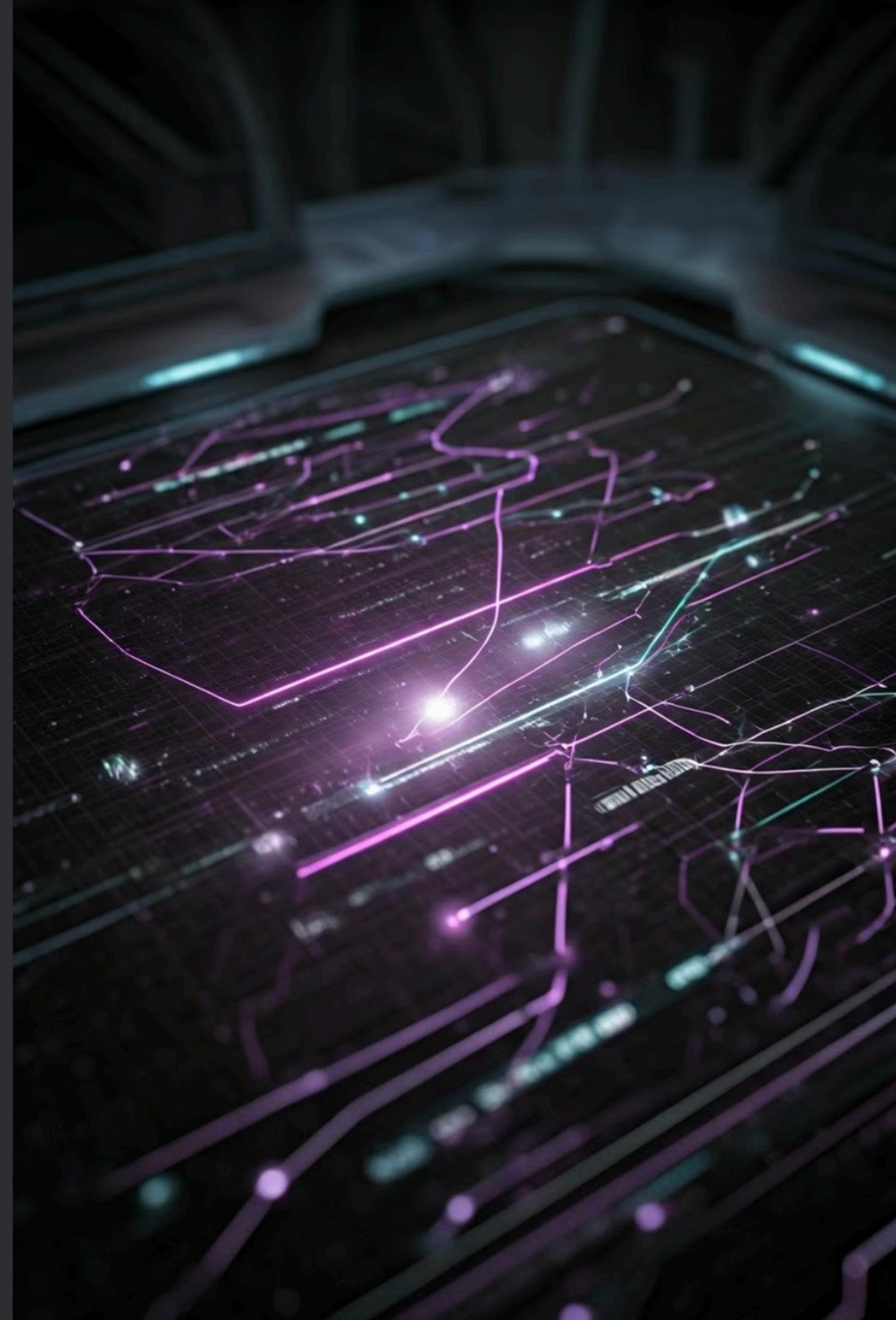
- Fundamental Analysis
- Market Analysis



Introduction



- Fundamental Analysis
- Market Analysis
- Technical Analysis



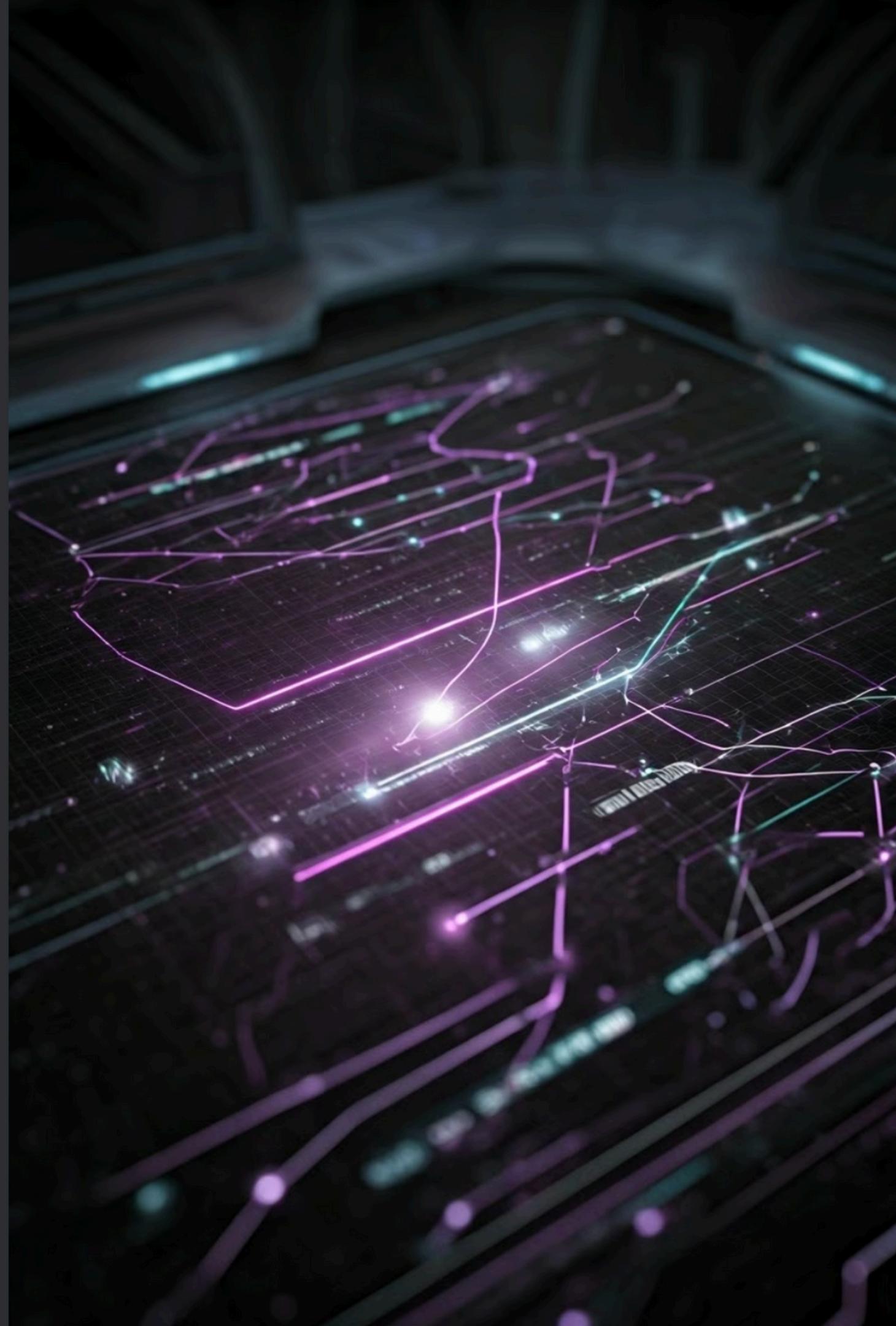
Introduction



- Fundamental Analysis
- Market Analysis
- Technical Analysis

Single data source limit

- Synergy benefit



Approach



Multimodal

- News
- Candle graph
- GNN

Output

- Back testing



Project Goal

- What's in scope?

Top 60 Stocks from 2014

Stock trend for portfolio creation



Project Goal

- **What's in scope?**

Top 60 Stocks from 2014

Stock trend for portfolio creation

- **Hypothesis**

Inefficient market



Project Goal

- **What's in scope?**

Top 60 Stocks from 2014

Stock trend for portfolio creation

- **Hypothesis**

Inefficient market

- **Profit from market dynamic**

Alpha from information

New perspective for projection



News Content Text Processing



News Content Text Processing

Collecting and processing relevant news texts

Extract keywords and train

Pass the trained parameters to multimodal for training



Collecting and Processing Relevant News Texts

1. Collect 58 stock news texts from 2014 to 2024
2. Forecast every 12 days
3. Use news from the past 60 days to predict stock price trends in the next 10 days

Collecting and Processing Relevant News Texts

1. Collect 58 stock news texts from 2014 to 2024
2. Forecast every 12 days
3. Use news from the past 60 days to predict stock price trends in the next 10 days



60 days

training: 2014/1/1-2021/12/30
test: 2022/1/1-2024/6/30

10 days later



Stock prices rise or fall

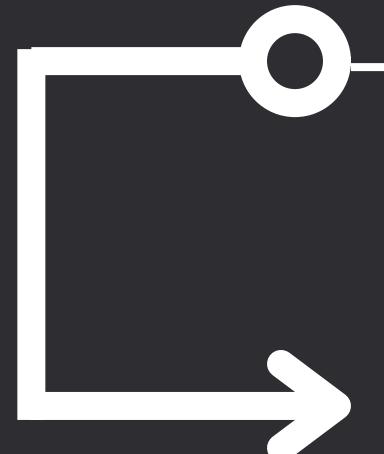
Extract keywords and train: BERT-Based News Classification Model

input data Use @ to separate each news item within 60 days

code	time	content	label
2884	2022/1/4	全家、玉山銀等合營電子支付 公平會放行 公平會不禁止全家便利商店、玉山銀行、拍付國際公司等3家事業合資新設事業全盈支付公司，專營電子支付機構業務。雖然全家便利商店、玉山銀行、拍付國際公司分別經營連鎖式便利商店、銀行、第三方支付等業務，並分屬不同領域，但因各參與結合事業皆有提供行動支付服務，且玉山銀行、拍付國際公司的子公司，與新設事業全盈支付公司都已獲得金管會許可，得兼營或專營電子支付機構業務。公平會也因此從水平及多角化結合的角度評估，認為三家合資所獲取的使用者資料及數據的類型、特性，以及市場狀態，競爭對手並非無法透過其他來源蒐集或以其他資料取代。 @ 110/12/15關係人徐輝添申讓 80 張占原持股 1,110 張之 7.2 %。 nan @ 110/12/15關係人徐輝添申讓 80 張占原持股 1,190 張之 6.7 %。 nan @ 110/12/13經理人謝冠仁申讓 650 張占原持股 1,721 張之 37.8 %。 nan @ 110/12/13經理人謝冠仁申讓 500 張占原持股 1,071 張之 46.7 %。 nan @玉山金財務公告未經會計師核閱數，110/11 合併稅前盈餘 206,526 萬元，累計 110/01 至 110/11	1
2884	2022/1/5	全家、玉山銀等合營電子支付 公平會放行	1
2884	2022/1/6	全家、玉山銀等合營電子支付 公平會放行	1
2884	2022/1/7	金管會核准玉山銀行 胡志明市設分行	0
2884	2022/1/10	金管會核准玉山銀行 胡志明市設分行	0
2884	2022/1/11	金管會公布同意玉山銀行向越南申請設立胡志明市代表人	0
2884	2022/1/12	金管會公布同意玉山銀行向越南申請設立胡志明市代表人	0
2884	2022/1/13	金管會公布同意玉山銀行向越南申請設立胡志明市代表人	1

input
embedding

avg_tensor =



tensor([[101, ..., 102]]) @

+

@

@

@

tensor([[101, ..., 102]])

+

news number

Positional
Encodding

Extract keywords and train: BERT-Based News Classification Model

Using Chinese BERT pre-training with full-word masking

Key Features of the Chinese BERT Model:

1. Bidirectional Encoder

Extract keywords and train: BERT-Based News Classification Model

Using Chinese BERT pre-training with full-word masking

Key Features of the Chinese BERT Model:

1.Bidirectional Encoder

2.Pre-training and Fine-tuning

Extract keywords and train: BERT-Based News Classification Model

Using Chinese BERT pre-training with full-word masking

Key Features of the Chinese BERT Model:

- 1.Bidirectional Encoder
- 2.Pre-training and Fine-tuning
- 3.Masked Language Model (MLM)

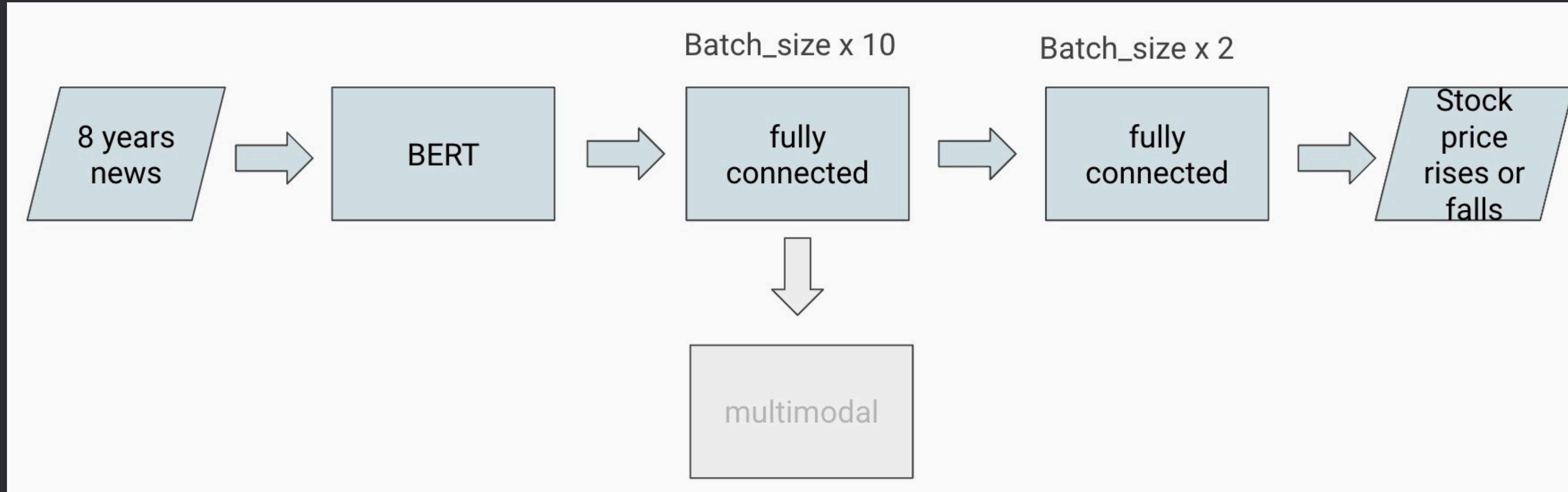
Extract keywords and train: BERT-Based News Classification Model

Using Chinese BERT pre-training with full-word masking

Key Features of the Chinese BERT Model:

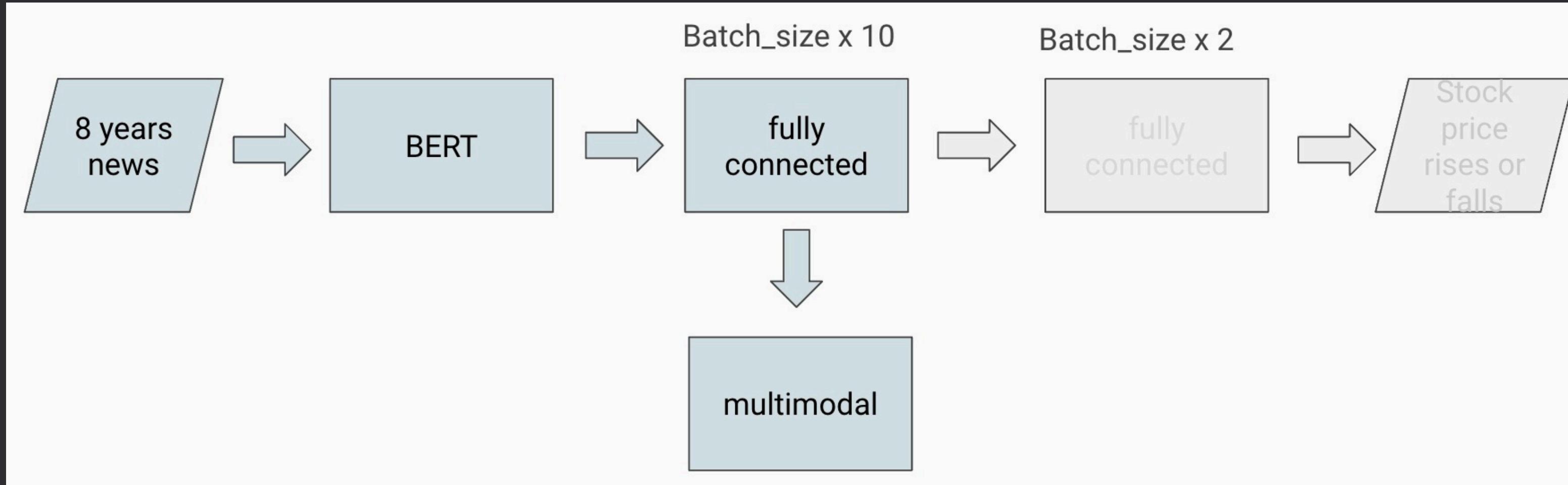
- 1.Bidirectional Encoder
- 2.Pre-training and Fine-tuning
- 3.Masked Language Model (MLM)
- 4.Next Sentence Prediction (NSP):

Pass the trained parameters to multimodal for training



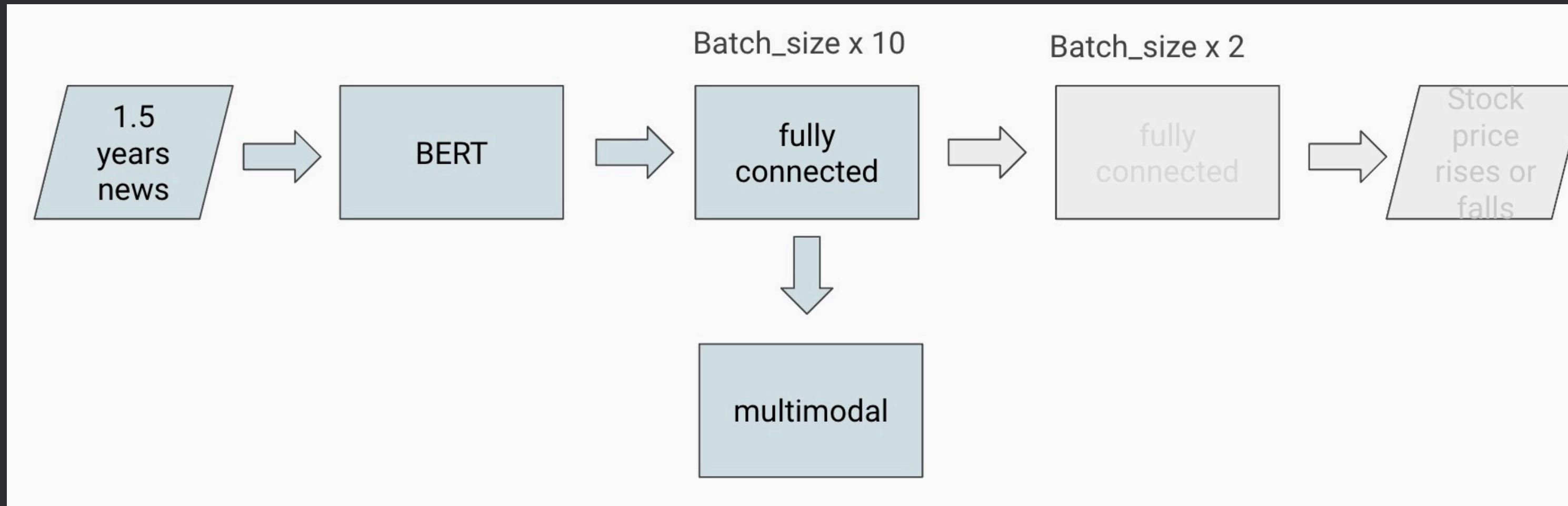
Step 1: Train the news classification model

Pass the trained parameters to multimodal for training



Step 2: Train the multimodal

Pass the trained parameters to multimodal for training



Step 3: Test data
input

Graphical Processing for Stock Analysis

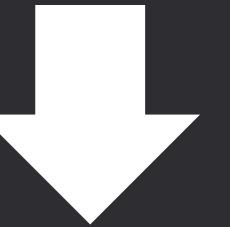




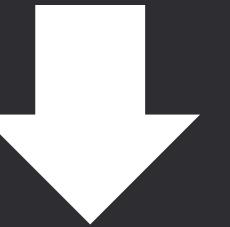
Fetching and Processing Data

yahoo!finance

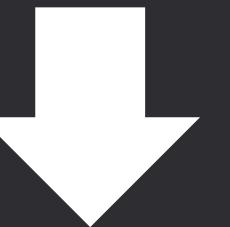
yahoo!finance



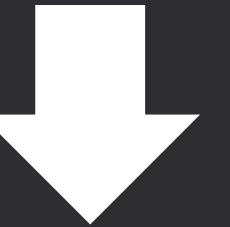
Stock OHLC, Volume



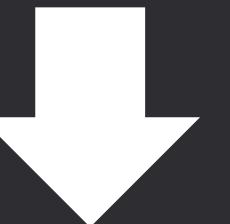
Stock OHLC, Volume



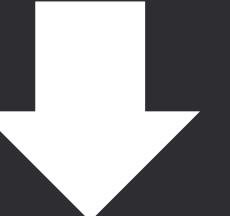
2013.7.1 - 2024.7.20



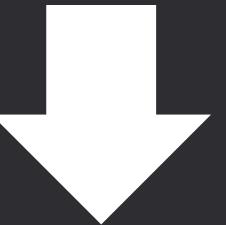
Stock OHLC, Volume



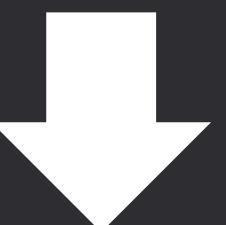
2013.7.1 - 2024.7.20



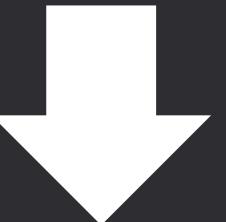
Indicator (20 MA)



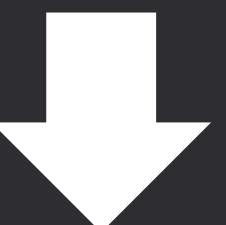
Stock OHLC, Volume



2013.7.1 - 2024.7.20



Indicator (20 MA)



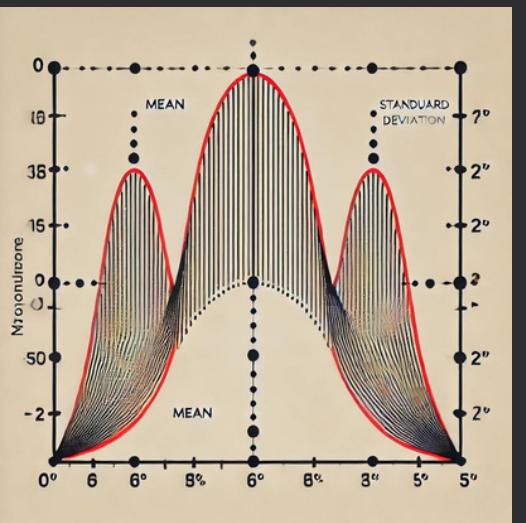
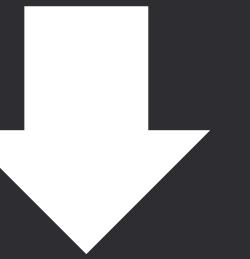
Labeling



Image Generation

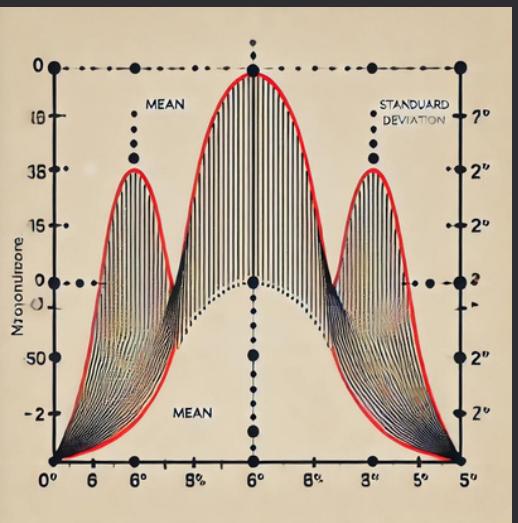
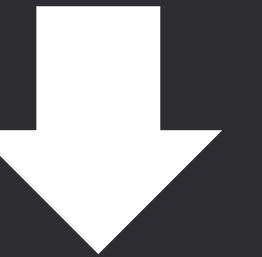
Retrieve 60 days data

Retrieve 60 days data

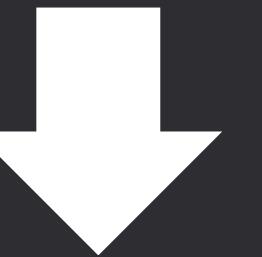


Normalization

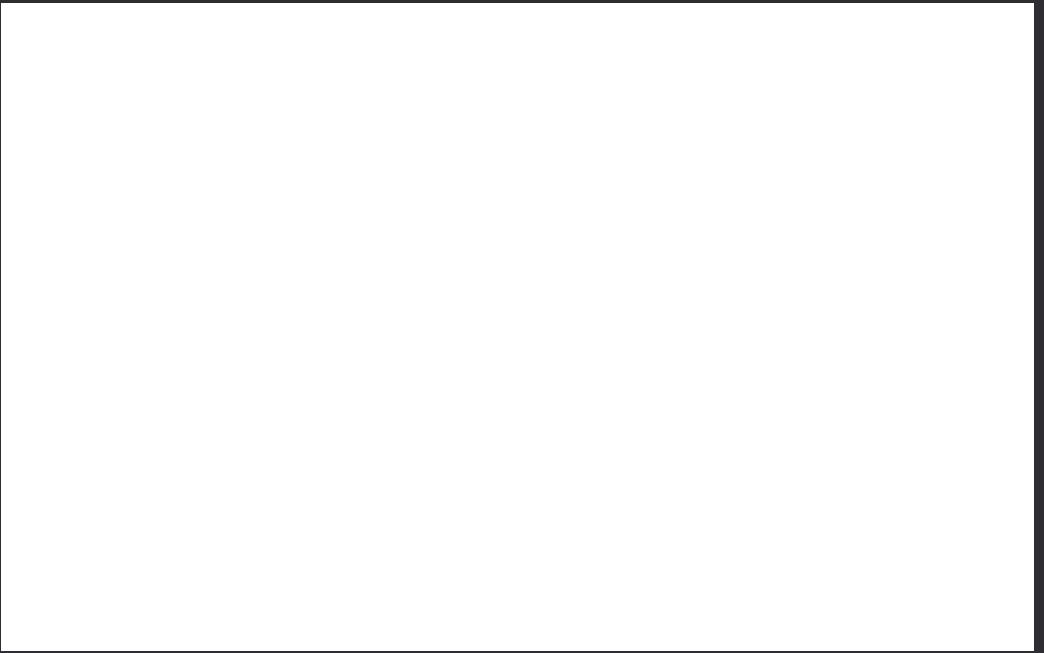
Retrieve 60 days data



Normalization



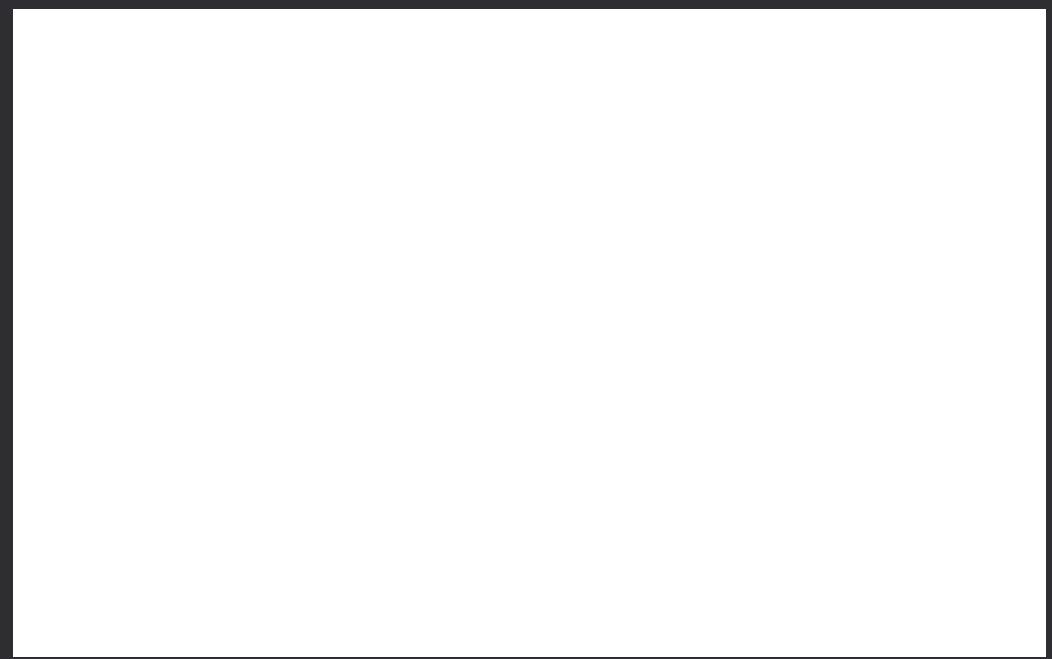
Span → 12 days



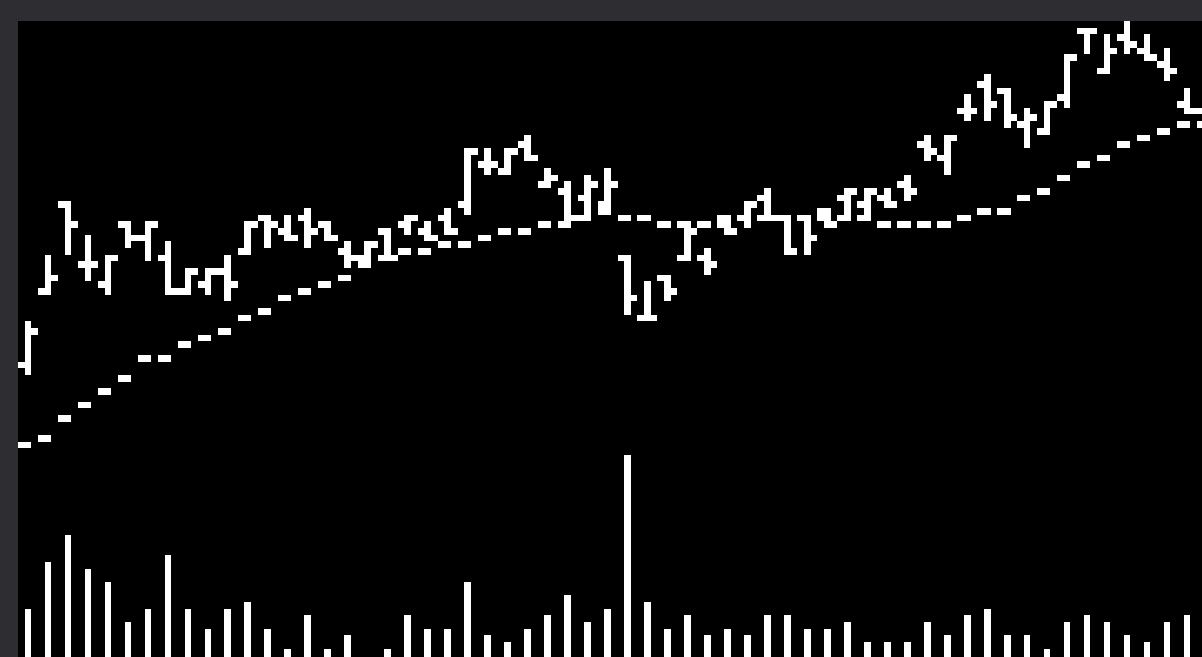
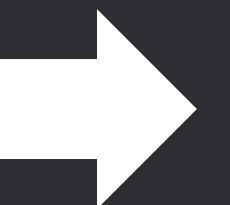
96



180



96



96



180

Model Training



CNN

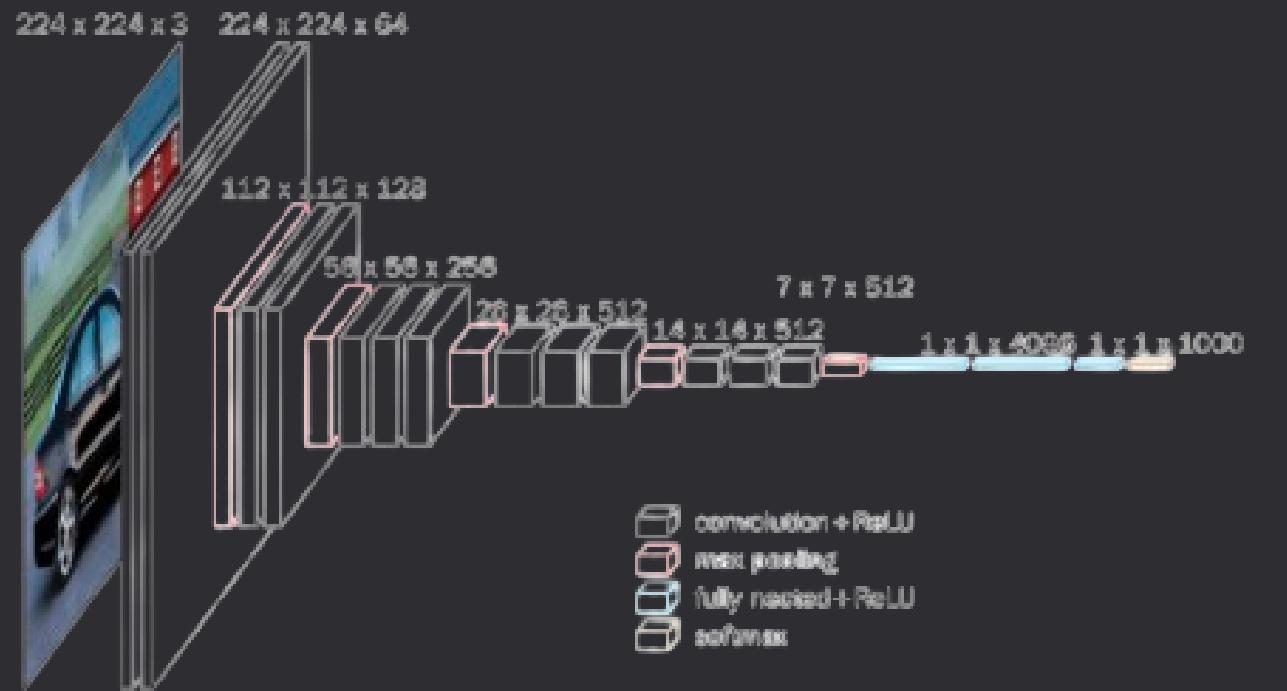


CNN



Other Models

Other Models

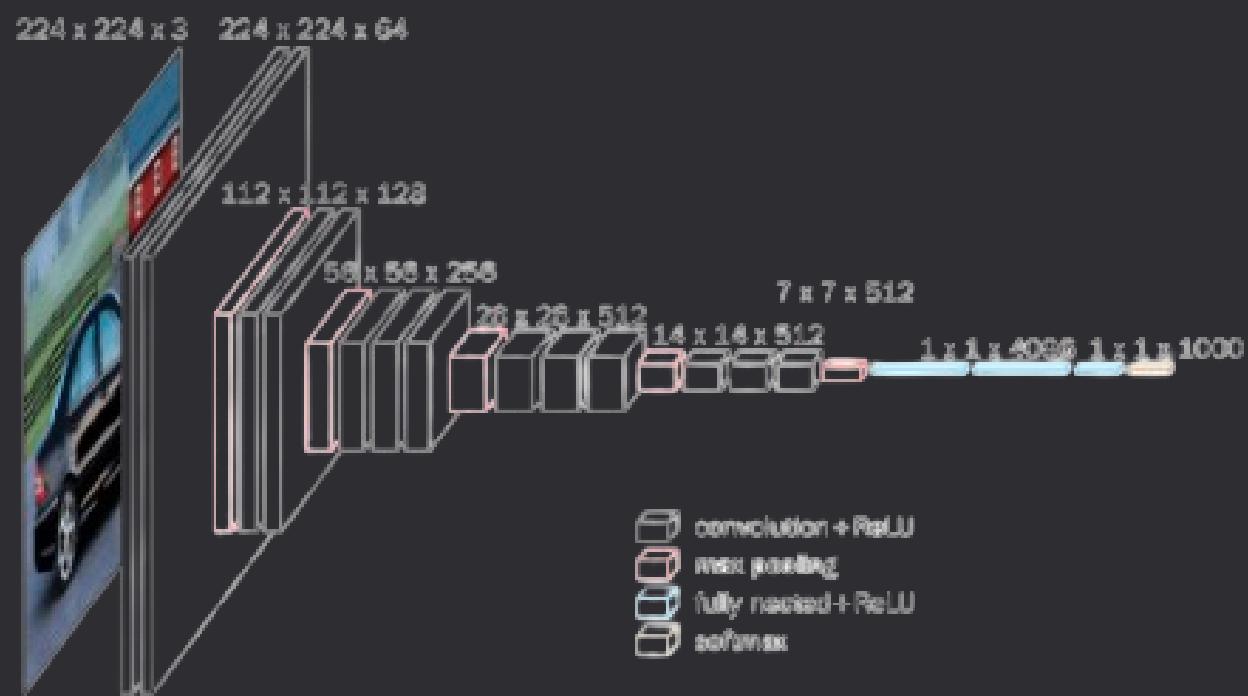


VGGNet

Reference:

<https://neurohive.io/en/popular-networks/vgg16/>

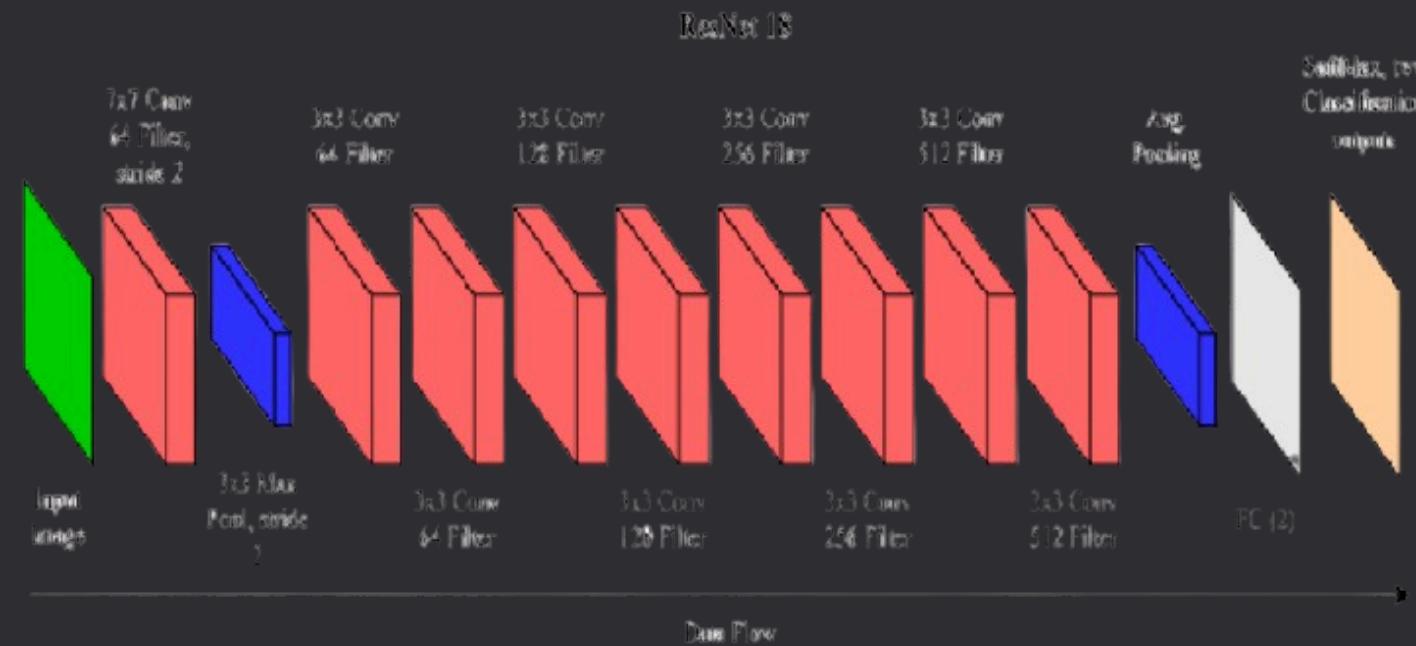
Other Models



VGGNet

Reference:

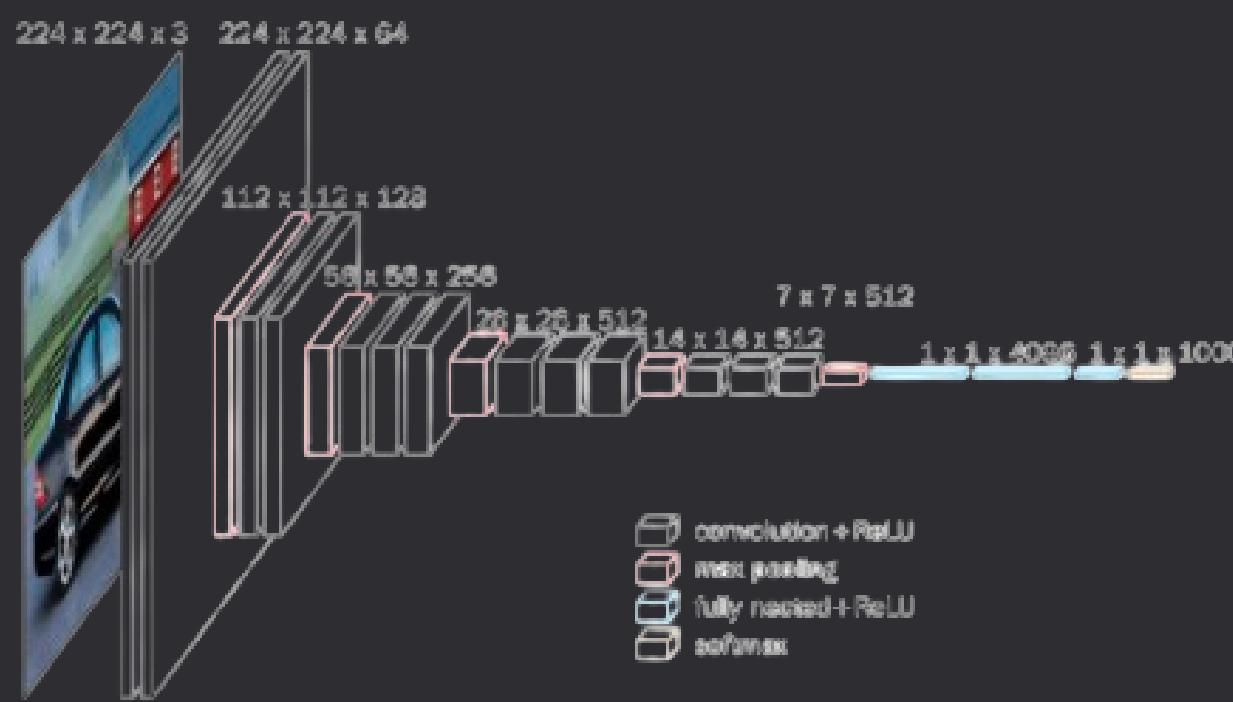
<https://neurohive.io/en/popular-networks/vgg16/>



ResNet

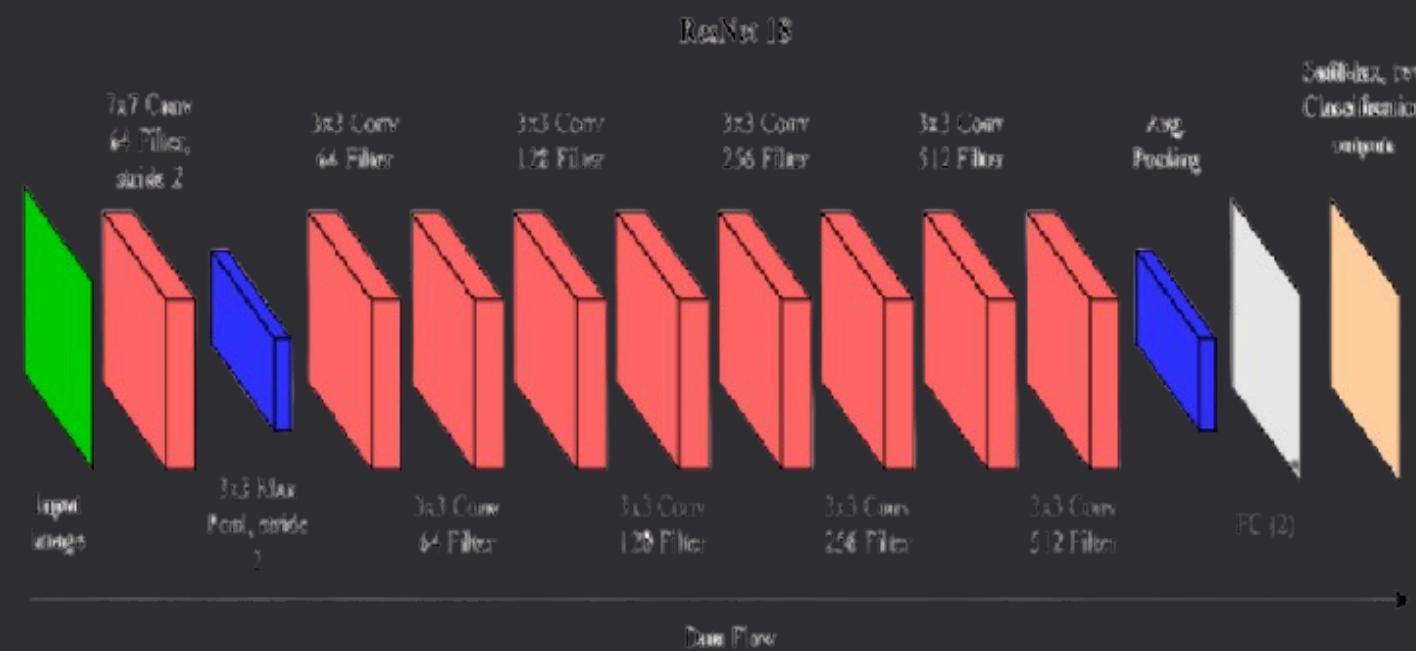
Reference:
https://www.researchgate.net/figure/ResNet-18-architecture-20-The-numbers-added-to-the-end-of-ResNet-represent-the_fig2_349241995

Other Models



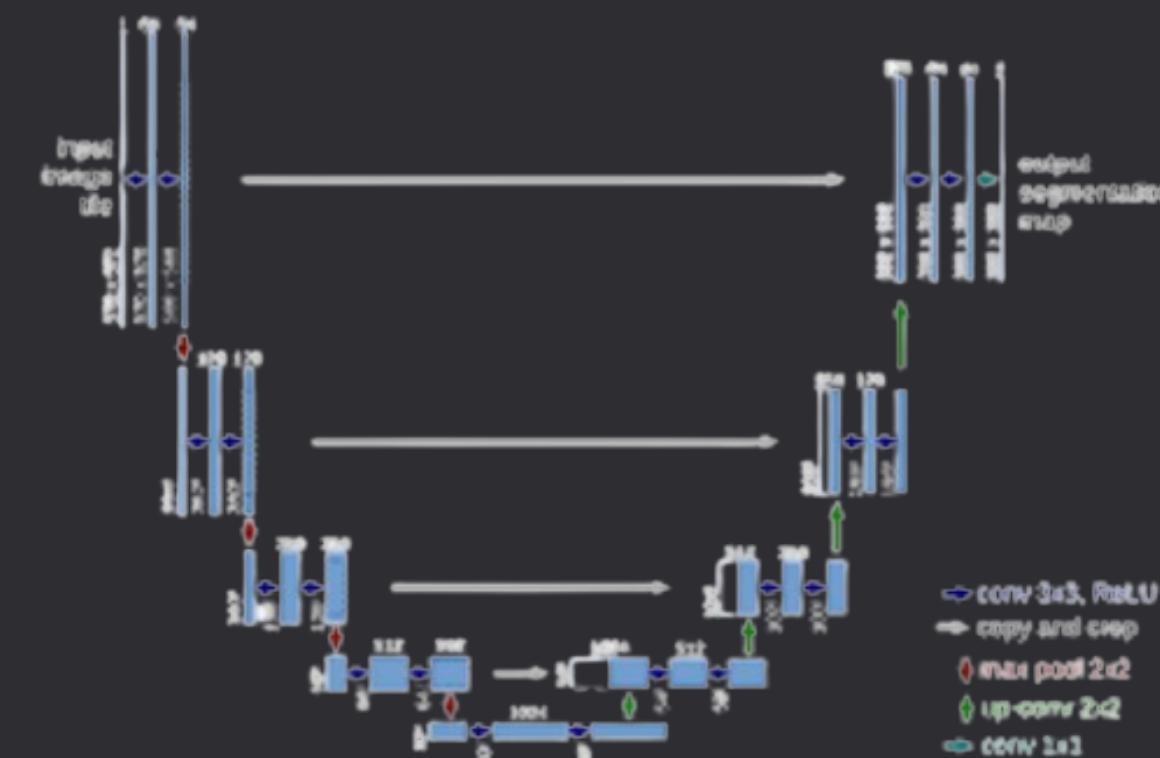
VGGNet

Reference:
<https://neurohive.io/en/popular-networks/vgg16/>



ResNet

Referecne:
https://www.researchgate.net/figure/ResNet-18-architecture-20-The-numbers-added-to-the-end-of-ResNet-represent-the_fig2_349241995



UNet

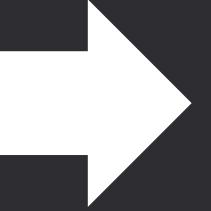
Referecne:
<https://arxiv.org/pdf/1505.04597.pdf>

Macroeconomic Indicators

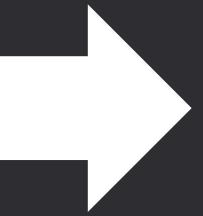




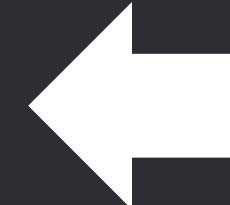
yahoo!finance



- USD/TWD
- VIX
- Crude Oil
- Gold
- CPI
- Unemployment Rate
- Interest Rate
- M2 Supply



60 days
Normalization



- USD/TWD
- VIX
- Crude Oil
- Gold
- CPI
- Unemployment Rate
- Interest Rate
- M2 Supply



Pass the Trained Vectors to
GNN for Training



Graph Neural Networks (GNN)

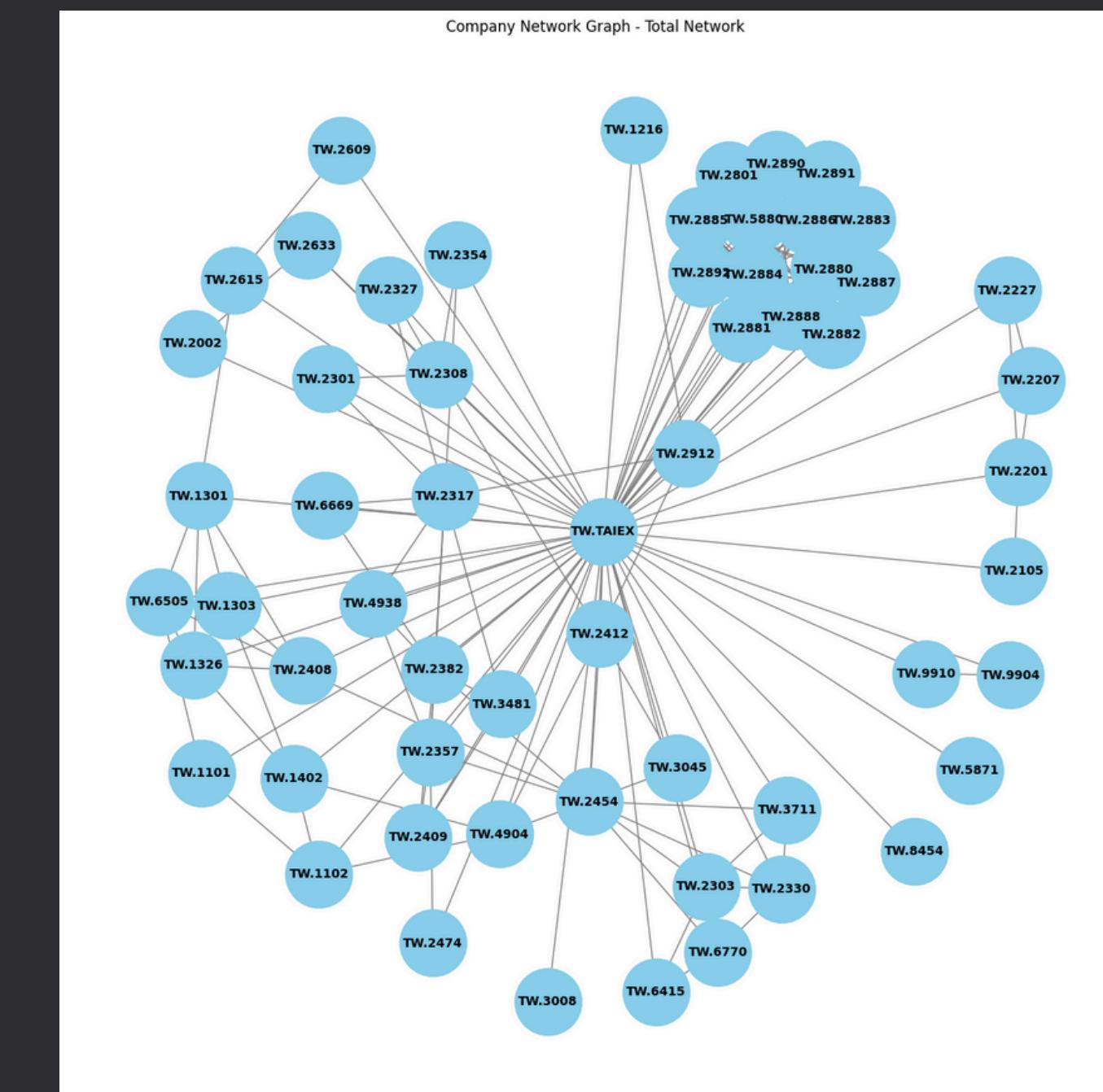
Why GNN?

- Illustrate the Relationship among all Companies



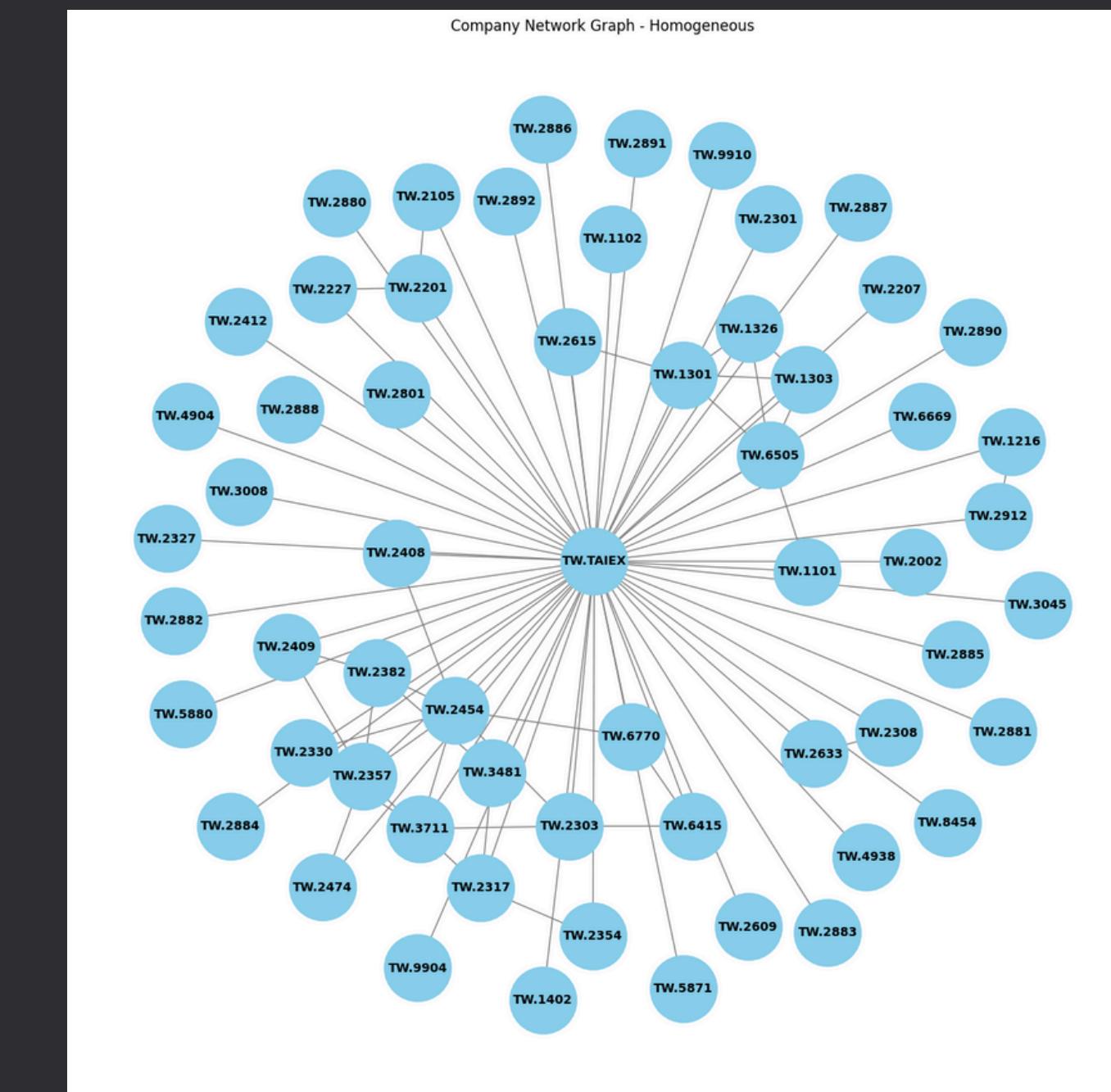
Relationship Types

- # • Total Network



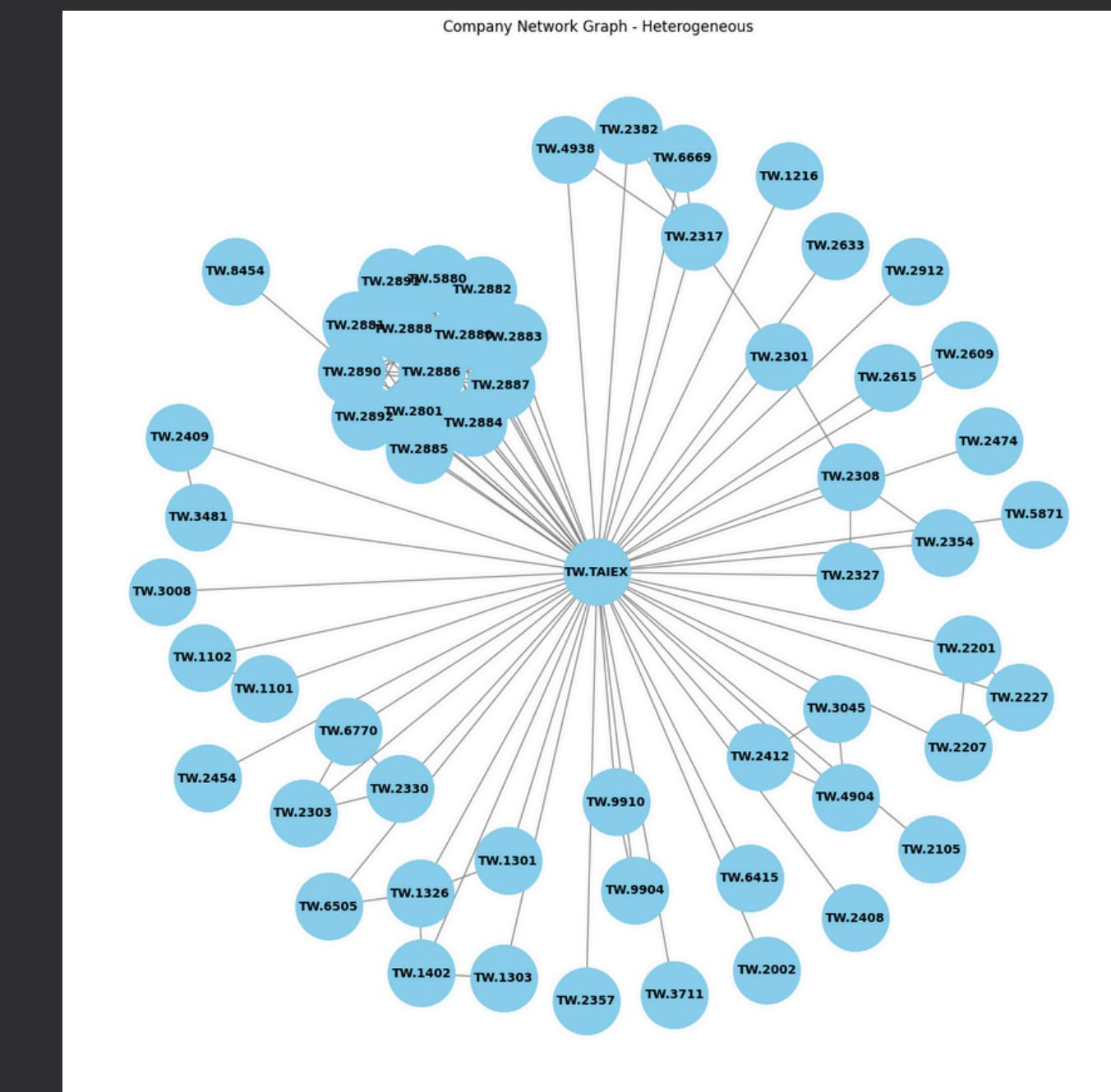
Relationship Types

- Homogeneous



Relationship Types

- Heterogeneous



Data Source



Process

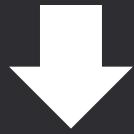
Input Data

1. image: 10 vectors
 2. news: 10 vectors
 3. macroeconomic data: 8 vectors
- => total: 28 vectors
+ 3 types of relationship



Process

Input Data



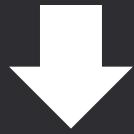
Train GAT

1. image: 10 vectors
 2. news: 10 vectors
 3. macroeconomic data: 8 vectors
- => total: 28 vectors
+ 3 types of relationship

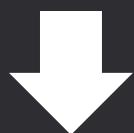


Process

Input Data



Train GAT



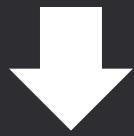
Test

1. image: 10 vectors
 2. news: 10 vectors
 3. macroeconomic data: 8 vectors
- => total: 28 vectors
- + 3 types of relationship

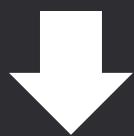


Process

Input Data



Train GAT



Test



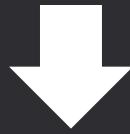
Squeeze to 1 Vector

- 1. image: 10 vectors
 - 2. news: 10 vectors
 - 3. macroeconomic data: 8 vectors
- => total: 28 vectors
- + 3 types of relationship

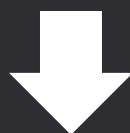


Process

Input Data



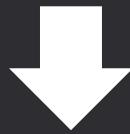
Train GAT



Test



Squeeze to 1 Vector



Rank Companies by their Vectors

1. image: 10 vectors
 2. news: 10 vectors
 3. macroeconomic data: 8 vectors
- => total: 28 vectors
- + 3 types of relationship



Testing Result

- **12 Models**

Image: 4 (CNN, VGGNet, ResNet, UNet)

News: 1 (BERT)

Network: 3 (Total Network, Homo, Hetero)

=> Total: $4 \times 1 \times 3 = 12$



Testing Result

- Testing Accuracy: 49.5% ~ 52%
- The most Accurate:
ResNet + BERT + GAT(Total Network) : 52%

```
# testing results (accuracy)

# ResNet_Total_output_vectors_testing: 52.04%
# ResNet_Homo_output_vectors_testing: 51.70%
# ResNet_Hetero_output_vectors_testing: 51.70%

# VGGNet_Total_output_vectors_testing: 49.53%
# VGGNet_Homo_output_vectors_testing: 50.51%
# VGGNet_Hetero_output_vectors_testing: 49.74%

# UNet_Total_output_vectors_testing: 50.17%
# UNet_Homo_output_vectors_testing: 49.49%
# UNet_Hetero_output_vectors_testing: 49.79%

# CNN_Total_output_vectors_testing: 50.72%
# CNN_Homo_output_vectors_testing: 49.83%
# CNN_Hetero_output_vectors_testing: 50.00%
```



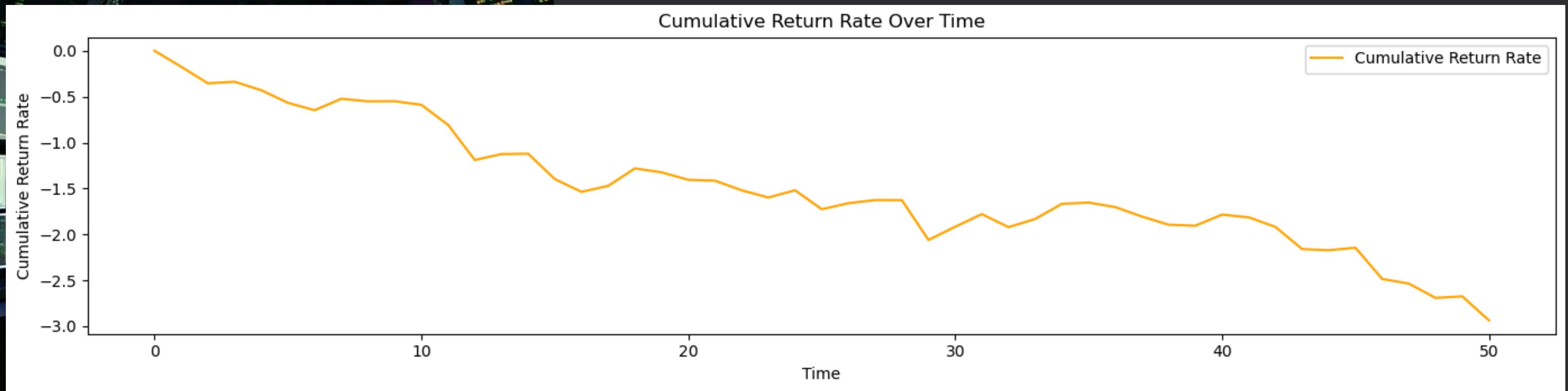


Trading Strategy

- Long 5 Companies with the Highest Vectors
- Short 5 Companies with the Lowest Vectors

Backtesting

- Negative Return Rate for all Models





4 Reasons for Failure

- Too Many Dimensions, but too Little Data



4 Reasons for Failure

- Too Many Dimensions, but too Little Data
- Not Total Market Analysis for Factor Investing



4 Reasons for Failure

- Too Many Dimensions, but too Little Data
- Not Total Market Analysis for Factor Investing
- Building Model by Closing Price





4 Reasons for Failure

- Too Many Dimensions, but too Little Data
- Not Total Market Analysis for Factor Investing
- Building Model by Closing Price
- Only Consider Return Rate, not Stock Price

Future Works

- Conduct Total Market Analysis



Future Works

- Conduct Total Market Analysis
- Collect More Data
 - 1. shorten time intervals
 - 2. enlarge sources of market information



Future Works

- Conduct Total Market Analysis
- Collect More Data
 1. shorten time intervals
 2. enlarge sources of market information
- Invest in Hardware of Higher Level





Thank for listening!



Johnson Chang

Team Member



Ryan Kuo

Team Member



Lee Po-Ting

Team Member



Cheng-Jih Chang

Team Member







DATA
EXTRACT
TRANSFORM
LOAD
PIPELINE

STOCK
MARKET
MOVEMENTS
COULD BE
USED TO
PREDICT

10 DAY PREDICTION

DATA
EXTRACT
TRANSFORM
LOAD
PIPELINE



