# Transformers and Multi-features Time2Vec for Financial Prediction

Bui Nguyen Kim Hai, Nguyen Duy Chien

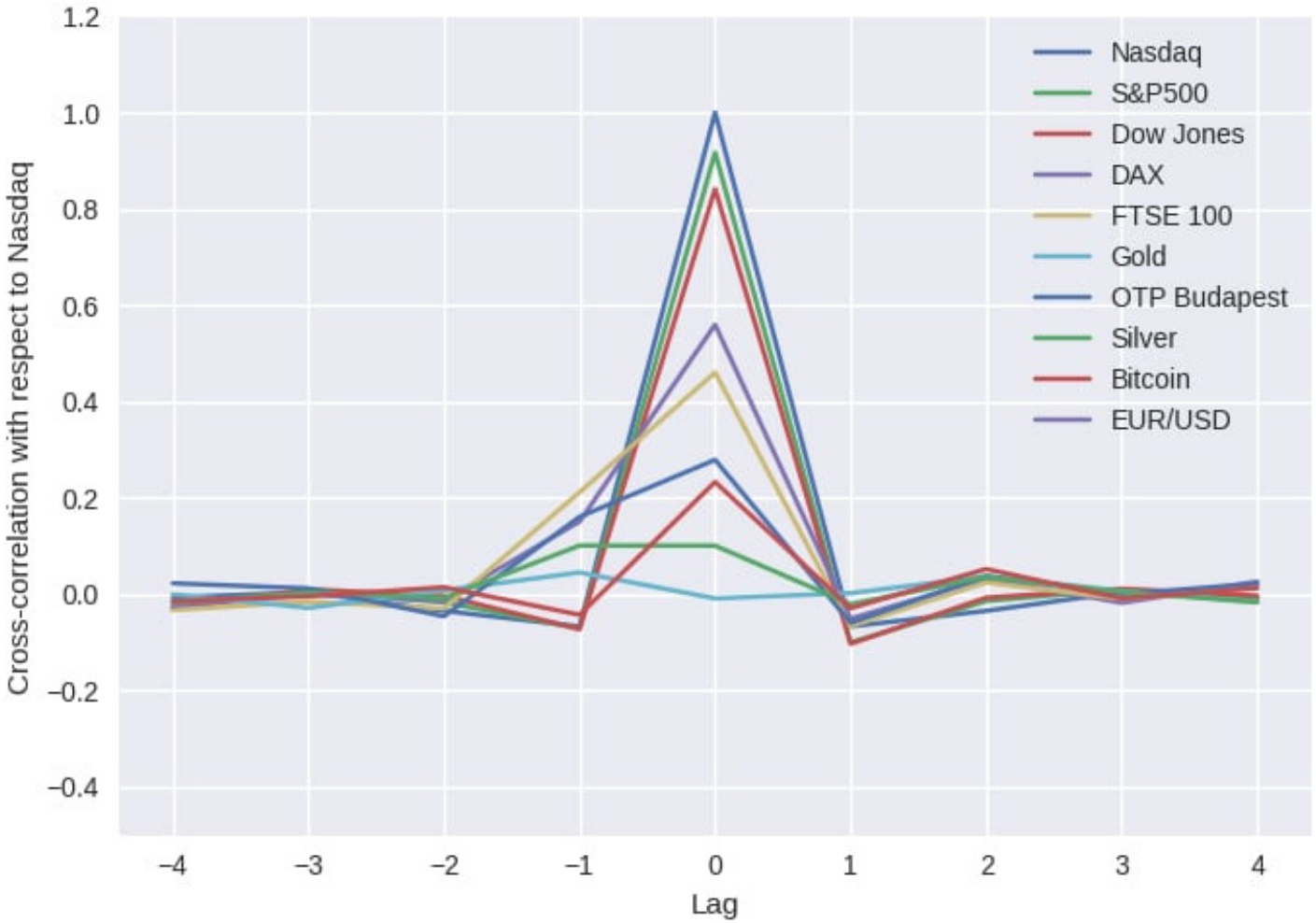
*TDK CONFERENCE – IT SCIENCE SECTION, 2024 SPRING*

Budapest, Hungary
May 29, 2024

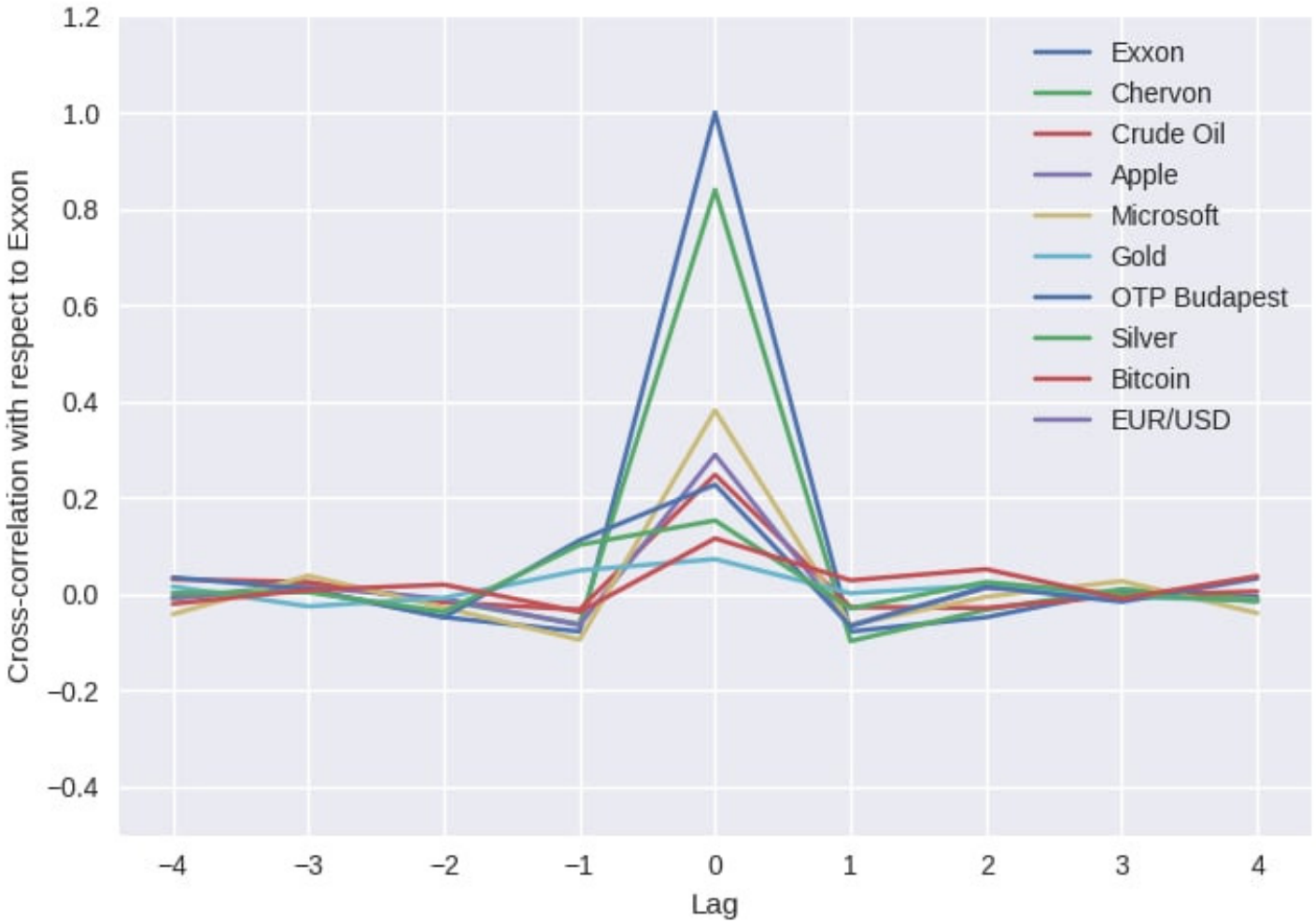
# Outline

# Outline

# Motivation: Cross-correlation to NASDAQ



# Motivation: Cross-correlation to Exxon Mobil



# Motivation

## By other works

- Researchers try to combine Time2Vec with CNN, RNN, LSTM, and Attention mechanism
- For instances:
  - Aeroengine Risk Assessment
  - Predicting Production in Shale and Sandstone Gas Reservoirs
  - Stock Price Forecasting

## In finance area

- Studies primarily rely on one dataset

## By observing trends

- Stock's trend is a Markov process
- Historical data offers limited foresight
- Stocks having similar trend is more promising

# Related work

## ARIMA

Making one-step-ahead predictions

## RNN

Handling temporal problems in sequential data and time-series analysis.

## LSTM

Using gates, LSTM enables network to learn long-term dependencies and prevent the vanishing gradient problem.

## Transformer

The SOTA architecture that works well in many area such as NLP, and time-series

## Time2Vec

Use to embed the time-series data to vector

# Outline

# Data collection

## Where to collect?

Yahoo Finance

## What will be collected?

Date, Open, High, Low, Close, Volume columns

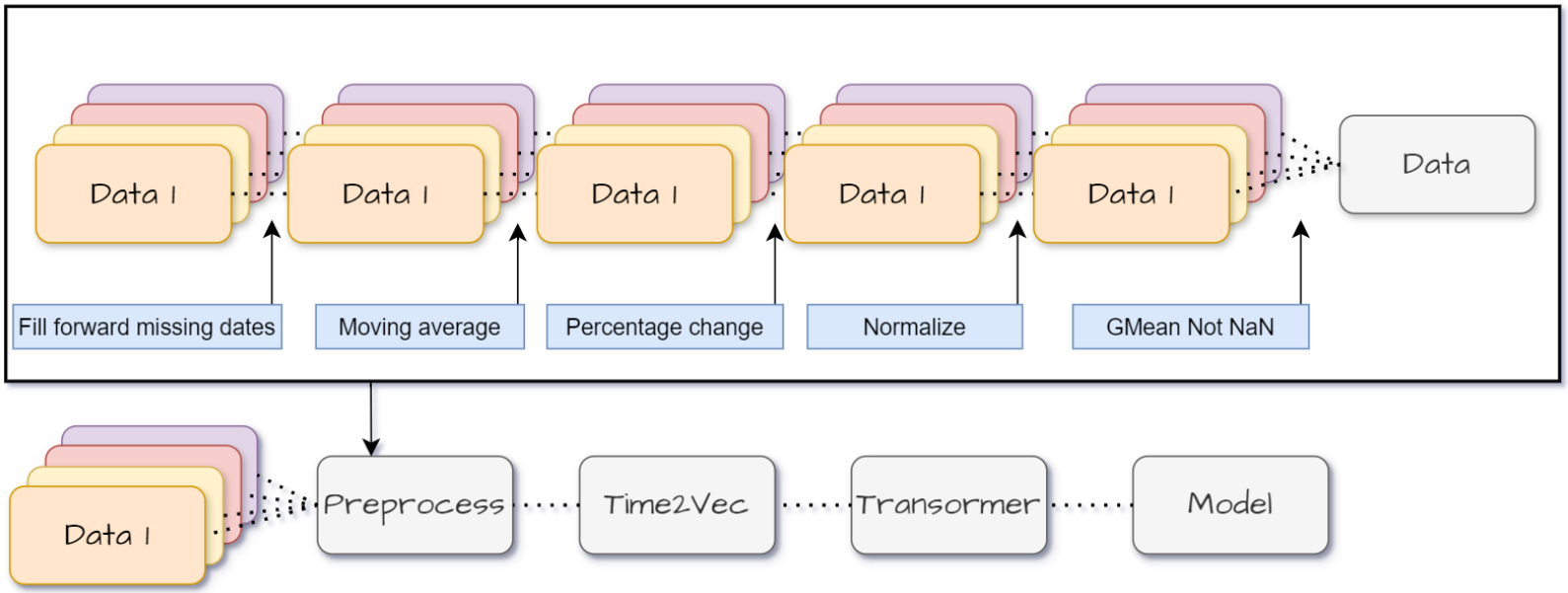
## How many datasets should we collect?

Two, three, four ..., as long as they are highly correlated to each other

## Collected datasets

- Group1: NASDAQ, S&P500, DJI, DAX
- Group2: Exxon Mobil, Chervon

# Preprocessing data: The pipeline

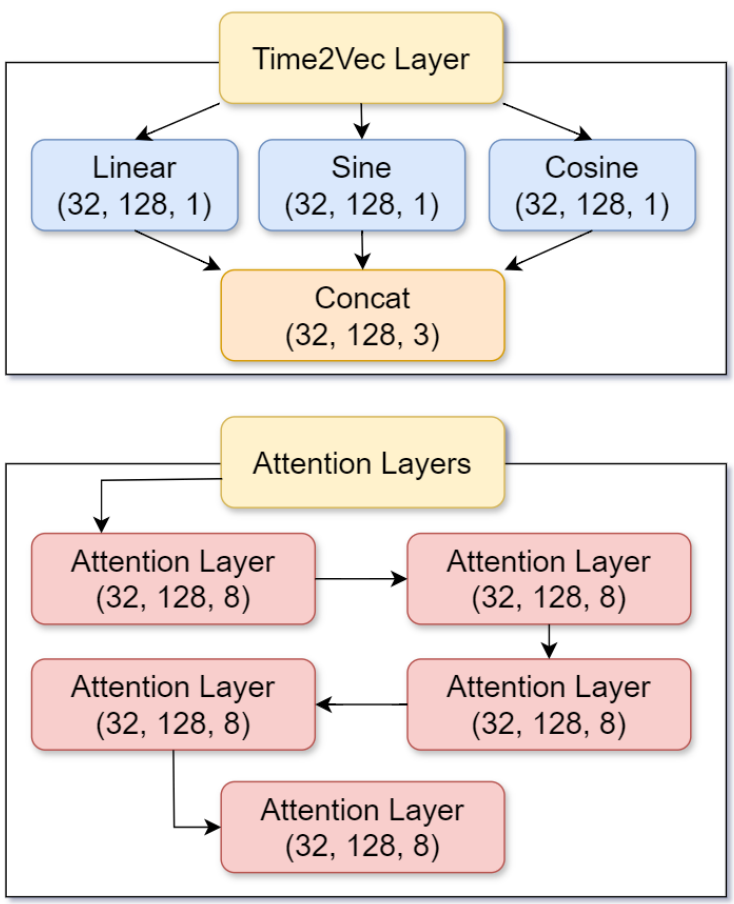


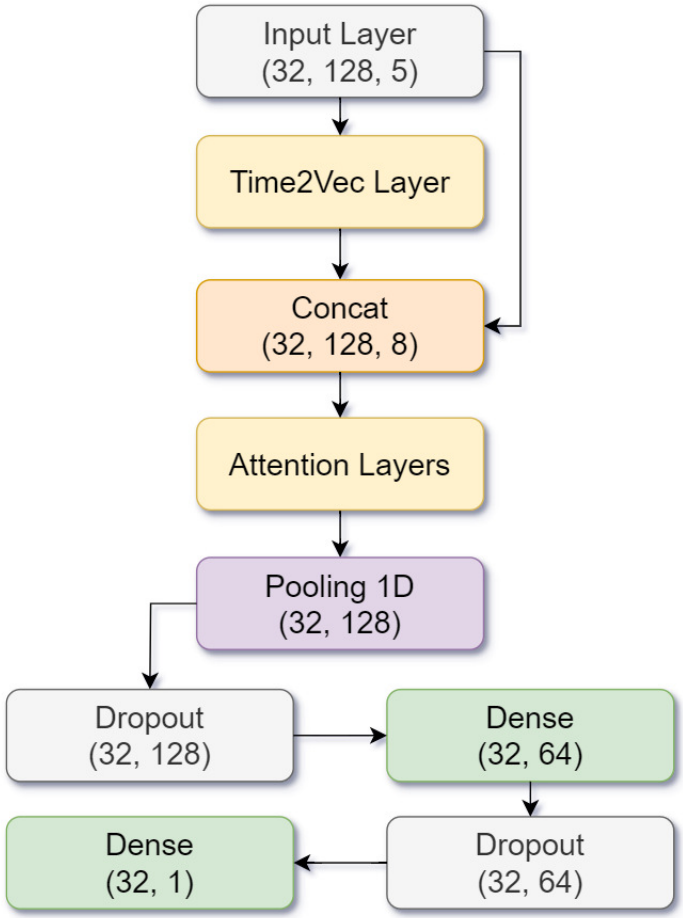

The preprocessing data pipeline.

## Techniques

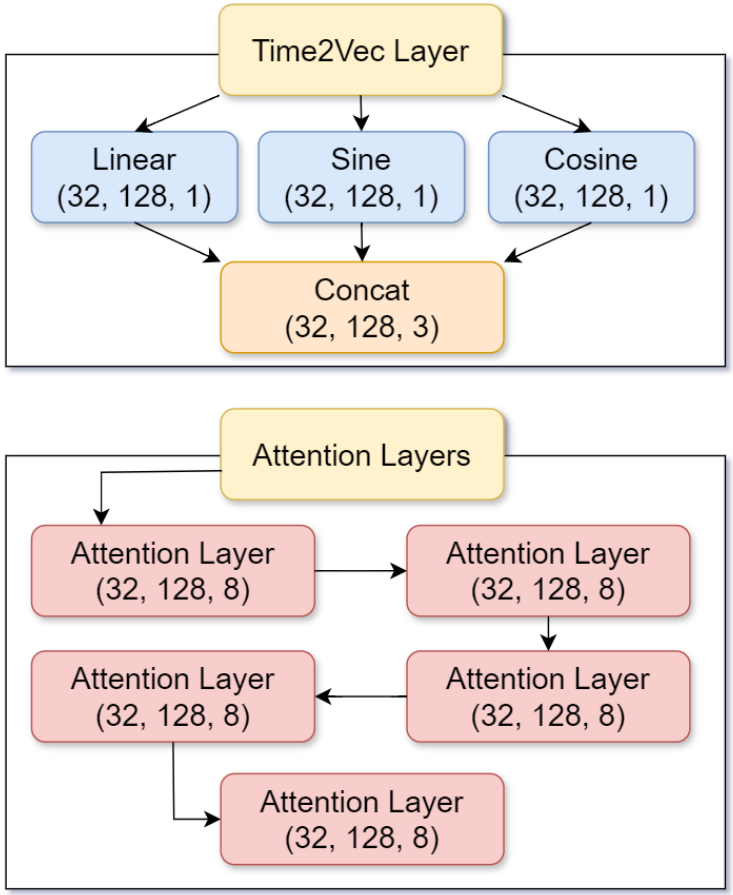
- Fill-forward: Filling missing data in dataset
- Moving Average: Smoothing dataset by averaging data
- Percentage Change: Compute the difference in the data
- Min-Max Normalization: Normalizing dataset
- Geometry Mean Not NaN (GMNN): Combining multiple datasets

# Model architecture: Proposed model



# Model architecture: Role of layers



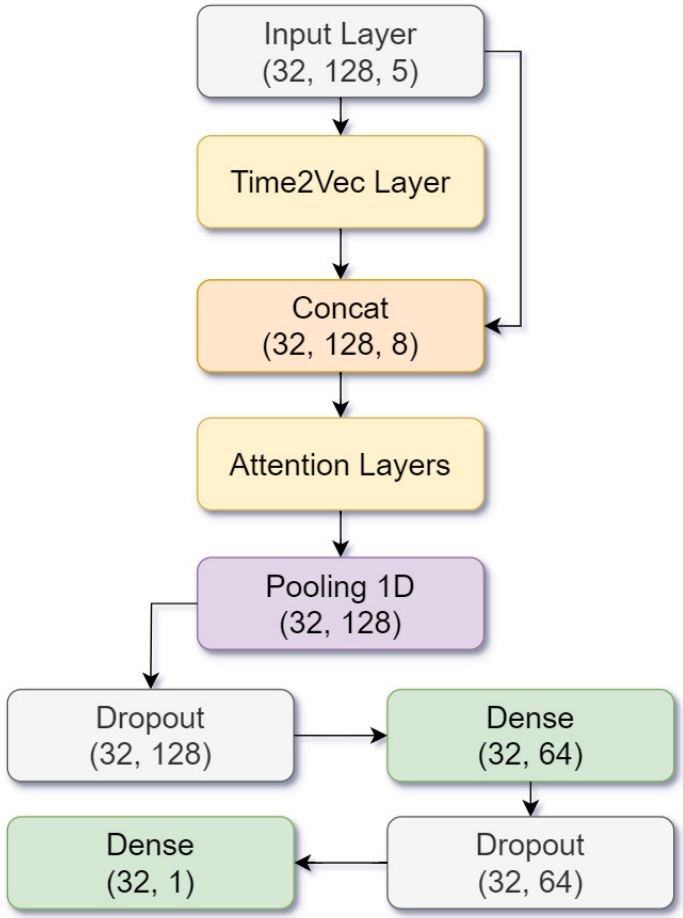

## Roles

- Time2Vec
  - Linear: Capturing linear trends
  - Sine, Cosine: Encoding positions and capturing periodic behaviors
  - Concat: Concatenating above three layers
- Attention Layers
  - To study the trend from different aspects, positions

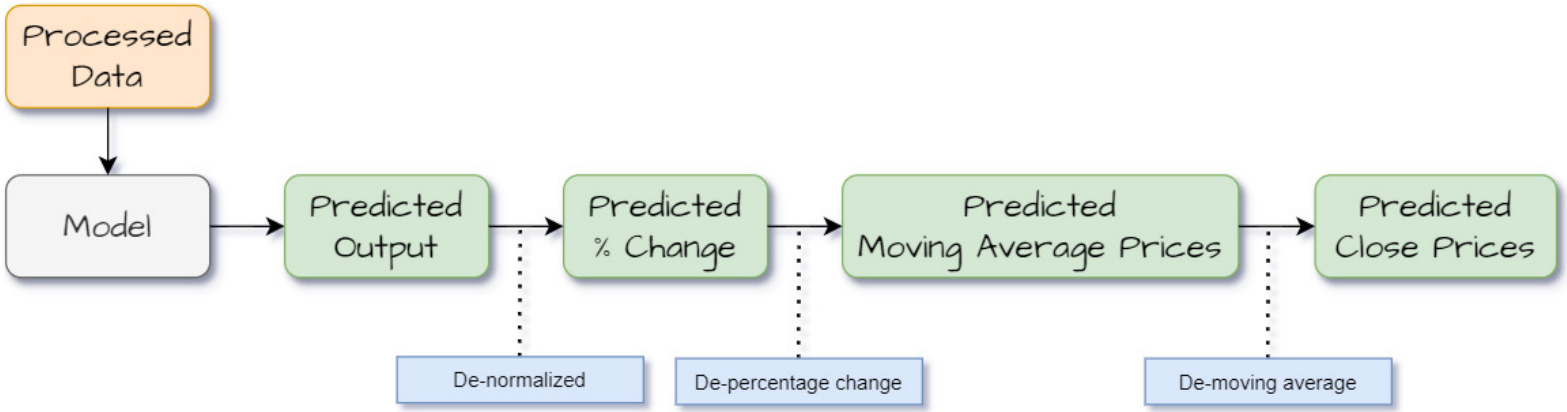
# Model architecture: Role of layers

## Roles

- Time2Vec: Catch continuous attribute of time
- Concat: Apply Residual Connection
- Attention: Deep understanding trend movements
- Pooling: Reducing dimension
- Dropout: Prevent over-fitting
- Dense: Apply activation functions (ReLU)



# Decoding engineering




The decoding pipeline.

## Techniques

- De-normalized
- De-percentage change
- De-moving average

## Why don't we use De-GMNN step?

- Output is **normalized** (Invariant)
- Target is **one** dataset, output only reflects that one

# Outline

# Conclusion

## Conclusion

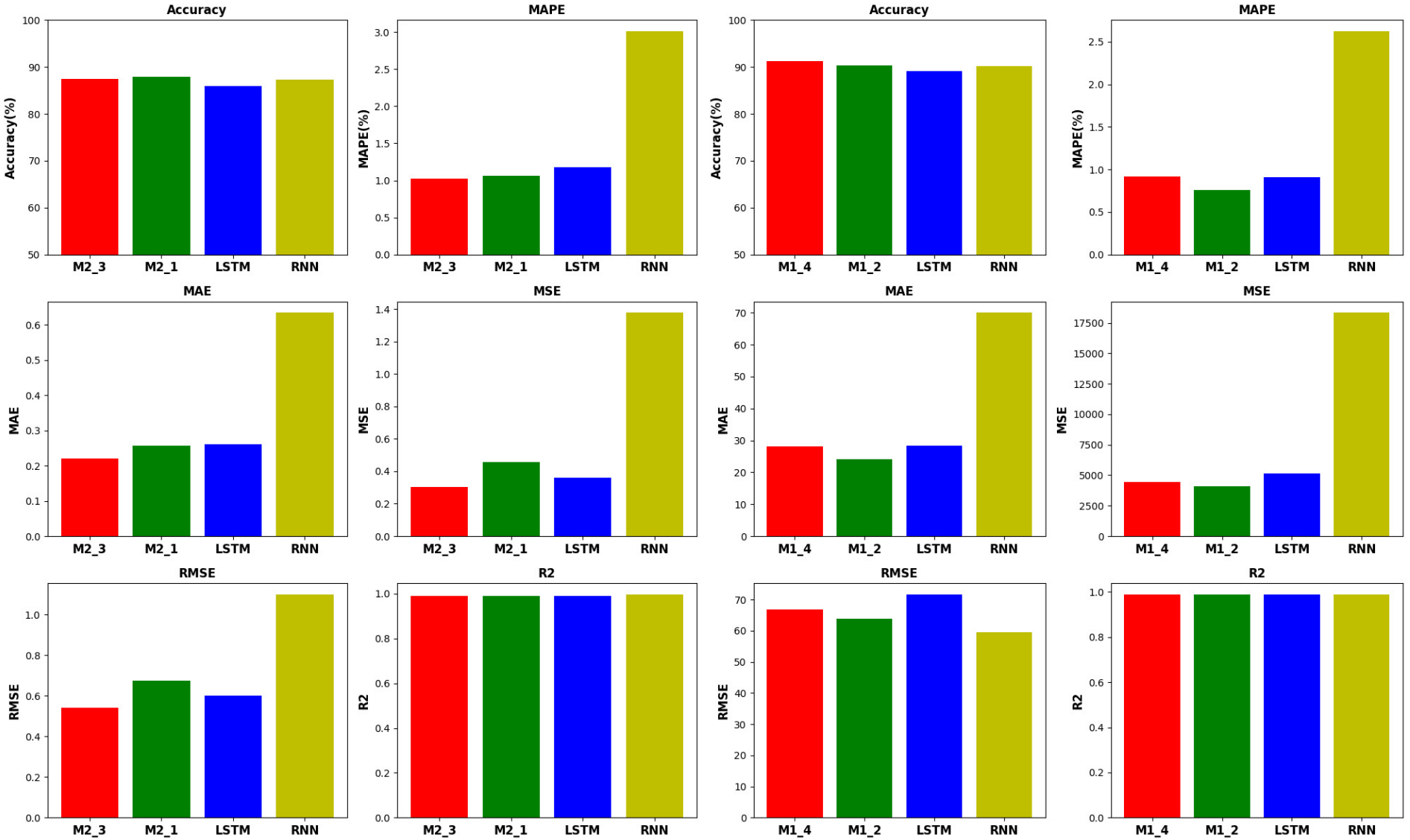
By leveraging multiple criteria to evaluate the proposed model such as

- MAE, MAPE, RMSE, MSE, R2-score (price prediction task)
- Accuracy (trend forecasting task)

We can proudly say that, the multi-feature model

- **Outperforms** the single-feature one in most cases and they are **extremely close** to each other in other scenarios.
- Usually yields **better** result than the SOTA in almost every contexts.

# Results




Comparing 6 metrics with respect to Exxon (Left), NASDAQ (Right)

# Outline

# Summary

## Summary

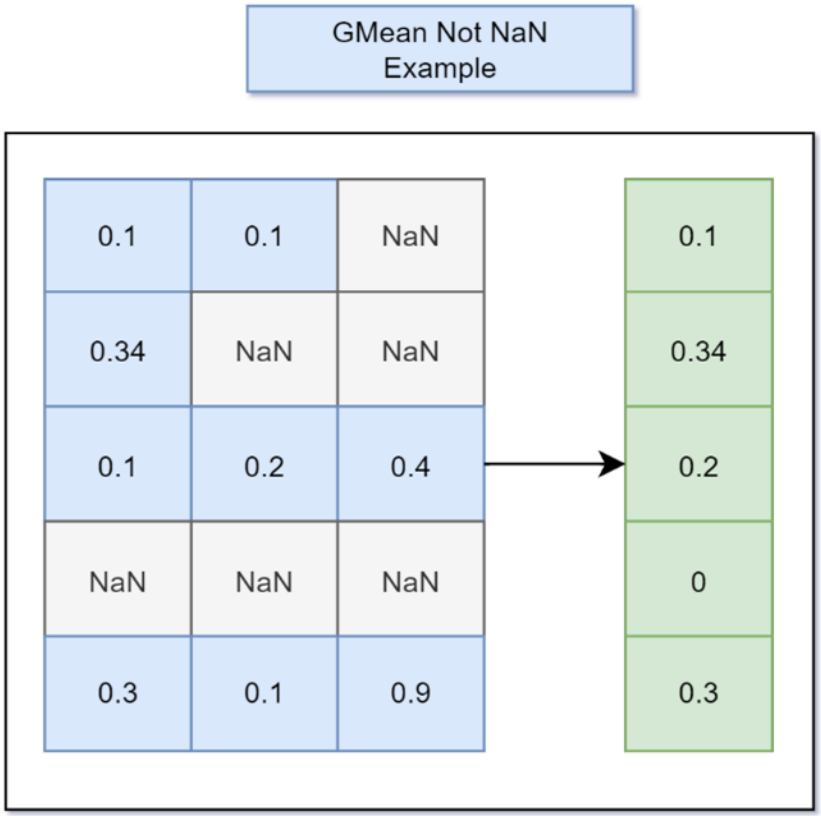
- We explore deep learning for challenging stock price prediction
- Paving the way for new feature studies and applications in various deep learning models
- Demonstrates correlation-based features and innovative neural networks improve stock price prediction

## Further Research

- Fine-tuning the architecture
- Continuing improving processing methods
- Comparing to other SOTA neural networks like KAN
- Applying the architecture to other areas

Thank you for your attention!

# But... What is GMNN?




## GMNN Attributes

- Union: Handling length difference when combining datasets
- Invariant: Keeping the data stays normalized
- Representation: The output reflects the whole datasets

A simple sample of applying GMNN transformation

# But... What is Time2Vec[1]?

## Time2Vec

An approach providing a model – agnostic vector representation for time

## Time2Vec Function

$$\mathbf{t2v}(\tau)[i] = \begin{cases} w_i\tau + \varphi_i & \text{if } i = 0 \\ F\left(w_i\tau + \varphi_i\right), & \text{if } 1 \leq i \leq k \end{cases}$$

$w, \varphi$: learnable parameters $\tau$: time

$F$: periodic activation function (eg. $\sin, \cos$)

## Time2Vec Attributes

- Capturing both periodic and non periodic patterns
- Being invariant to time re-scaling
- Being simple enough so it can be combined with many models

[1] Seyed Mehran Kazemi et al. "Time2Vec: Learning a Vector Representation of Time" (2019)

# Splitting data

## Can we just put the processed data into model straight forward?

No, we can not feed the whole data straight forward into the model, because that action will cause **over-fitting** problem which no one want it to be happened when training model

## Do we need to shuffle the data before splitting?

No, we don't want to shuffle the data because it has **continuous** attribute of time then shuffling will make the data lose this special and important property which cause a big problem for the model to learn the pattern

## How will we feed input data to our model?

We will split the data into three categories:

- Train: The first 80% data of the input
- Validation: The next 10% data
- Test: The last 10% data

# Applications

## Can it be used in production?

Of course, but to be more accurate and precise, the architecture must be fine-tuned to fit the expectation

## Is it easy to set up and use in production?

Yes it is, developer just need to find the appropriate datasets and train the model

## Can it be used in other areas?

Yes it can be used in other areas, we just need to find the dimension sizes and appropriate hyper-parameters to fit the area expectation

## Can it predict the tomorrow stock price?

Yes it can, but we need to apply some more techniques to decode back to real values