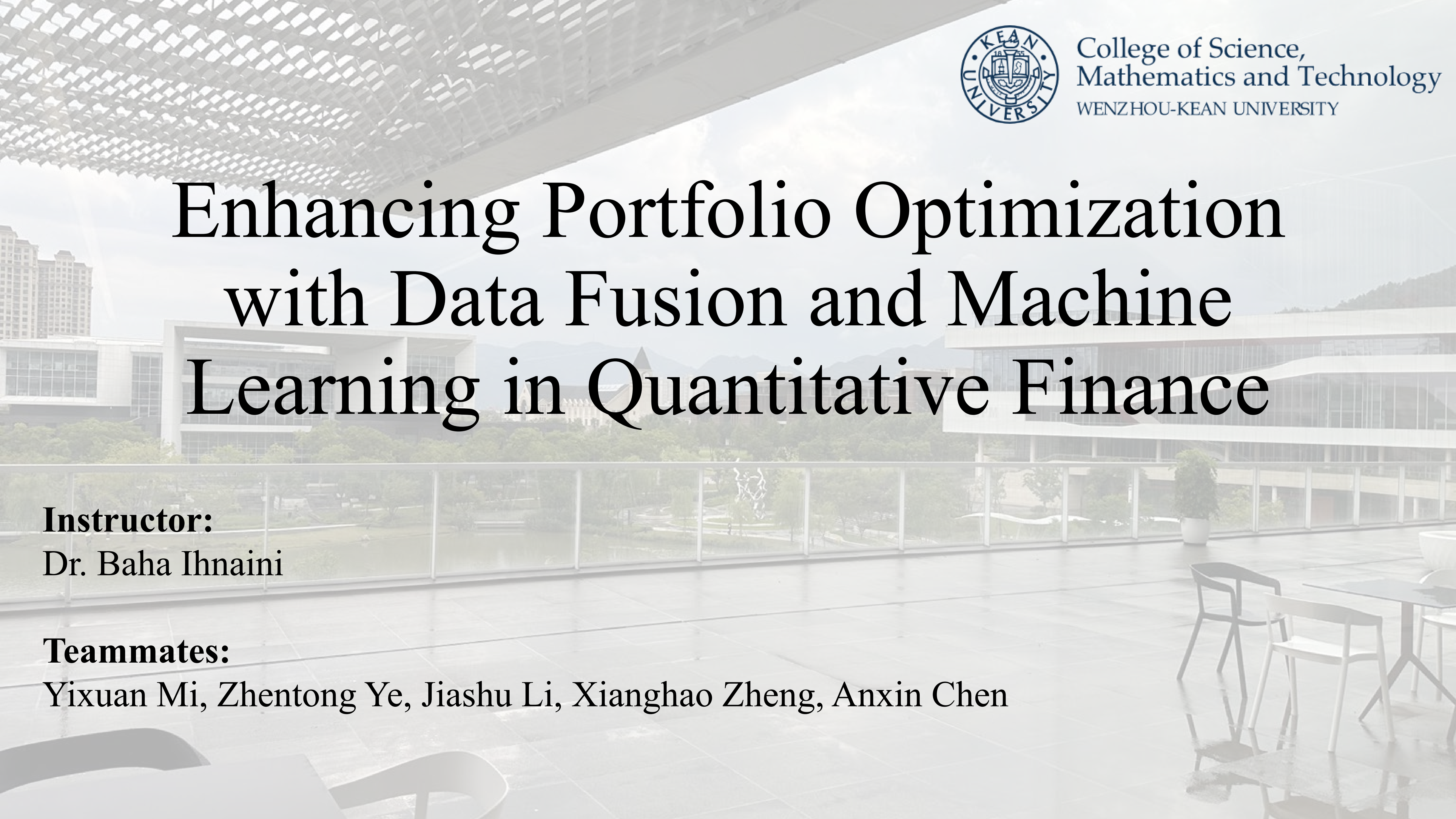


College of Science,
Mathematics and Technology
WENZHOU-KEAN UNIVERSITY

Enhancing Portfolio Optimization with Data Fusion and Machine Learning in Quantitative Finance

Instructor:
Dr. Baha Ihnaini

Teammates:
Yixuan Mi, Zhentong Ye, Jiashu Li, Xianghao Zheng, Anxin Chen

温州肯恩大学
WENZHOU-KEAN UNIVERSITY

Instructor:
Dr. Baha Ihnaini

Designed by:
Yixuan Mi
Zhentong Ye
Jiashu Li
Xianghao Zheng
Anxin Chen

CST
Wenzhou-Kean University
COLLEGE OF SCIENCE AND TECHNOLOGY

Enhancing Portfolio Optimization with Data Fusion and Machine Learning in Quantitative Finance

Introduction:

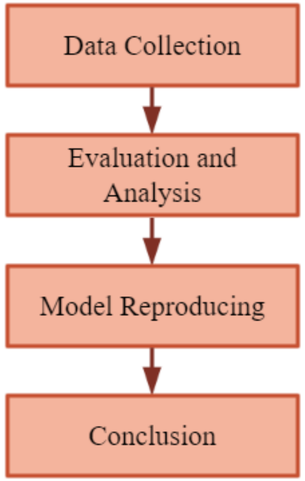
This project aims to revolutionize portfolio optimization in quantitative finance by integrating advanced machine learning techniques and data fusion methodologies. Our team strives to improve the problem that traditional methods will face: limitations in handling market complexities and integrating diverse data effectively.

Data Collection:

CSMAR (China Stock Market & Accounting Research) is a comprehensive database jointly established by Chinese regulatory authorities and the Shanghai Stock Exchange, providing data on Chinese A-share market companies for research and analysis.

Based on the CSMAR platform, we use CSMAR's downloadable CSV files to collect historical financial asset data and market sentiment data such as financial statements, investor sentiment indices, and sentiment consistency.

Research Flowchart:


Data Collection
Evaluation and Analysis
Model Reproducing
Conclusion

Platforms Support:

PyCharm
JetBrains IDE

CSMAR

Qlib

SSH Connection

MobaXterm

Model Reproducing:

Once we collected all the data, we used "Qlib" to choose the high-quality model (like LSTM) to fit our data; then, we used remote development tools (MobaXterm, Pycharm, etc.) to reproduce and improve the model.

- Signal-based evaluation: IC, ICIR, Rank IC, Rank ICIR
  - $\mathrm{corr}(\mathbf{x}, \mathbf{y}) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$
  - $\mathrm{IC}^{(t)} = \mathrm{corr}(\hat{\mathbf{y}}^{(t)}, \mathbf{ret}^{(t)})$
  - $\mathrm{ICIR} = \frac{\mathrm{mean}(\mathbf{IC})}{\mathrm{std}(\mathbf{IC})}$
  - $\mathrm{Rank\ IC}^{(t)} = \mathrm{corr}(\mathrm{rank}(\hat{\mathbf{y}}^{(t)}), \mathrm{rank}(\mathbf{ret}^{(t)}))$
  - $\mathrm{Rank\ ICIR} = \frac{\mathrm{mean}(\mathbf{Rank\ IC})}{\mathrm{std}(\mathbf{RankIC})}$

Results and Conclusion:

Model Name	Dataset	IC	ICIR	Rank IC	Rank ICIR
LSTM	Alpha360	0.0478±0.01	0.3620±0.05	0.0585±0.00	0.4578±0.04
ADD	Alpha360	0.0419±0.00	0.3066±0.04	0.0550±0.00	0.4205±0.03
ADARNN	Alpha360	0.0468±0.01	0.3706±0.08	0.0544±0.01	0.4416±0.07

Through trial and error, our team has reproduced three models, and each model was successfully run 50 times. It proves that our plan is working, paving the way for further advancements in our field.

Reference:

Song, C. (2023). Portfolio Optimization Based on Machine Learning. Advances in Economics, Management and Political Sciences. https://doi.org/10.54254/2754-1169/25/20230500.

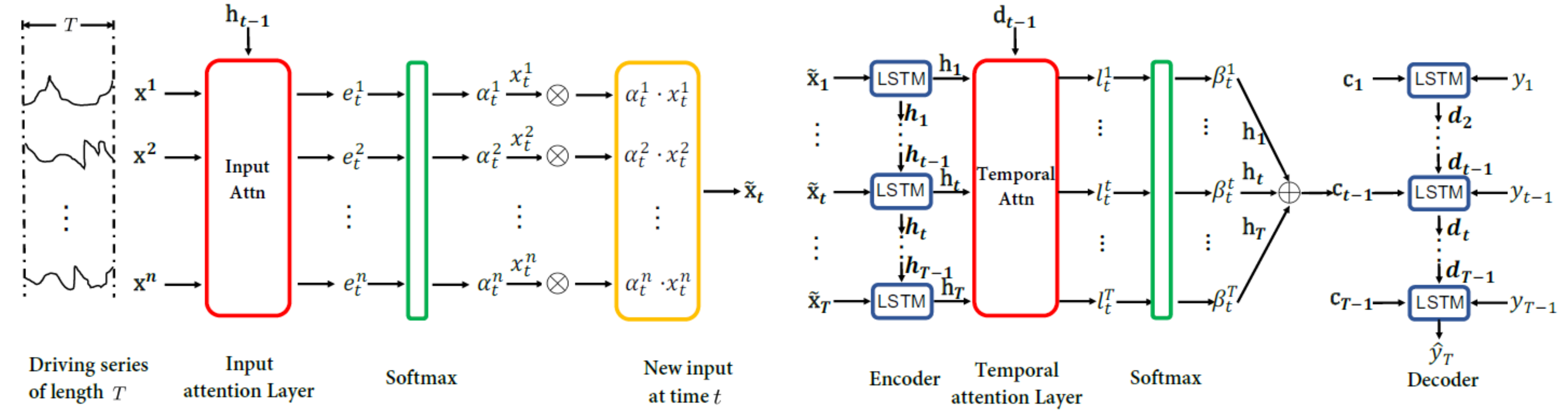
Wang, Y. (2023). Review: Application of Machine Learning to Investment Portfolios. BCP Business & Management. https://doi.org/10.54691/bcpbm.v38i.4351.

Our GitHub Repository:

https://github.com/EthanYixuanMi/Machine-Learning-in-Quantitative-Finance

Model Reproducing: ALSTM

College of Science, Mathematics and Technology
WENZHOU-KEAN UNIVERSITY



(a) Input Attention Mechanism

(b) Temporal Attention Mechanism

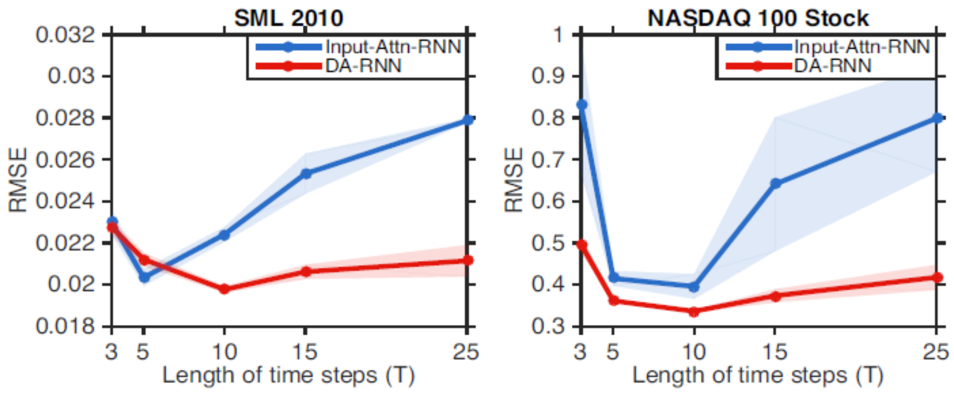


Figure 4: **RMSE** vs. length of time steps T over SML 2010 (left) and NASDAQ 100 Stock (right).

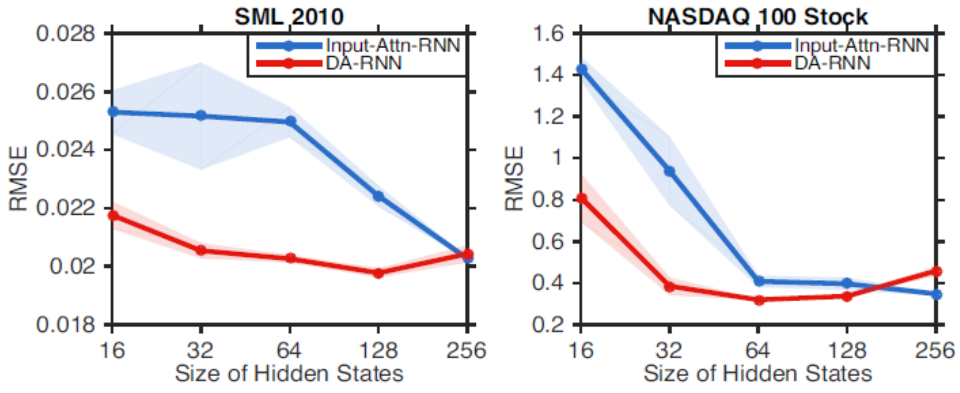
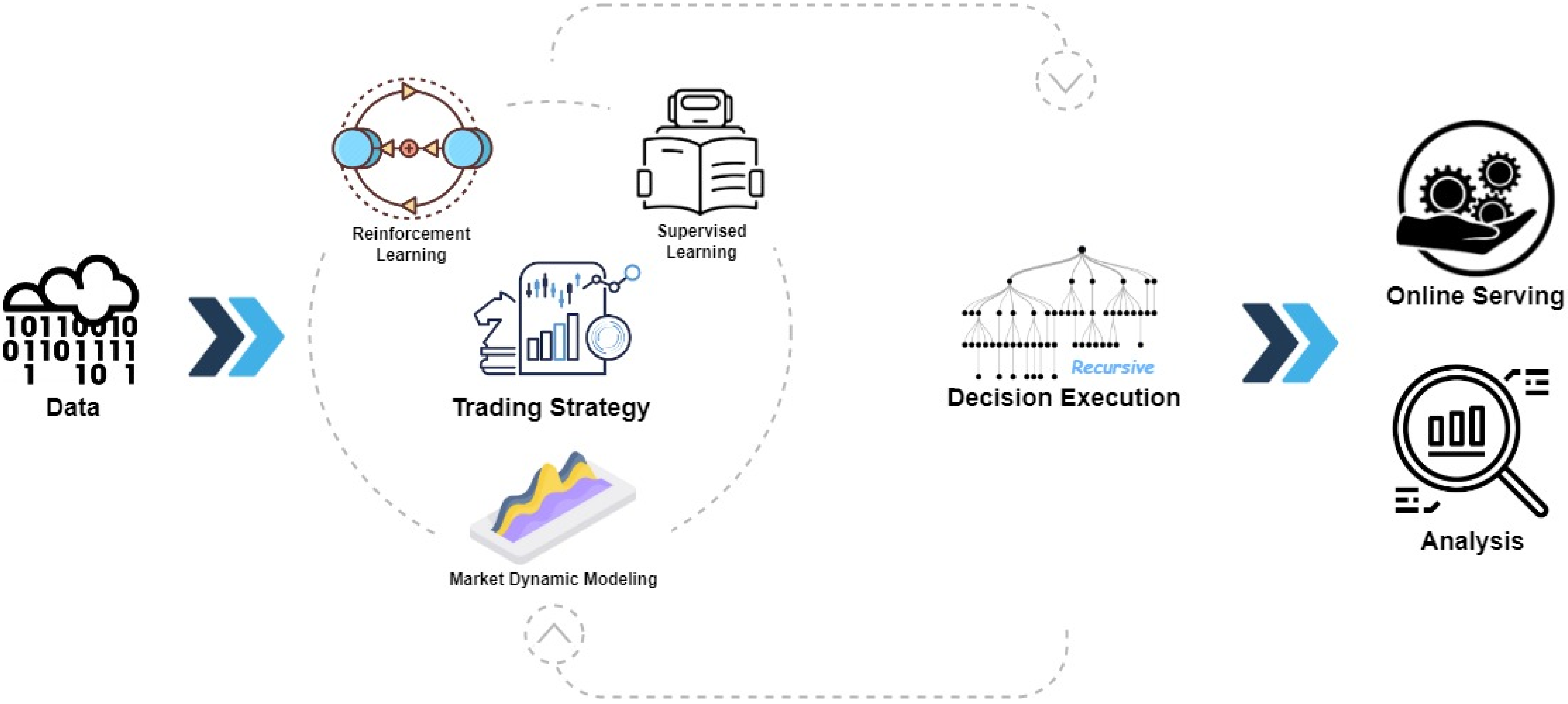
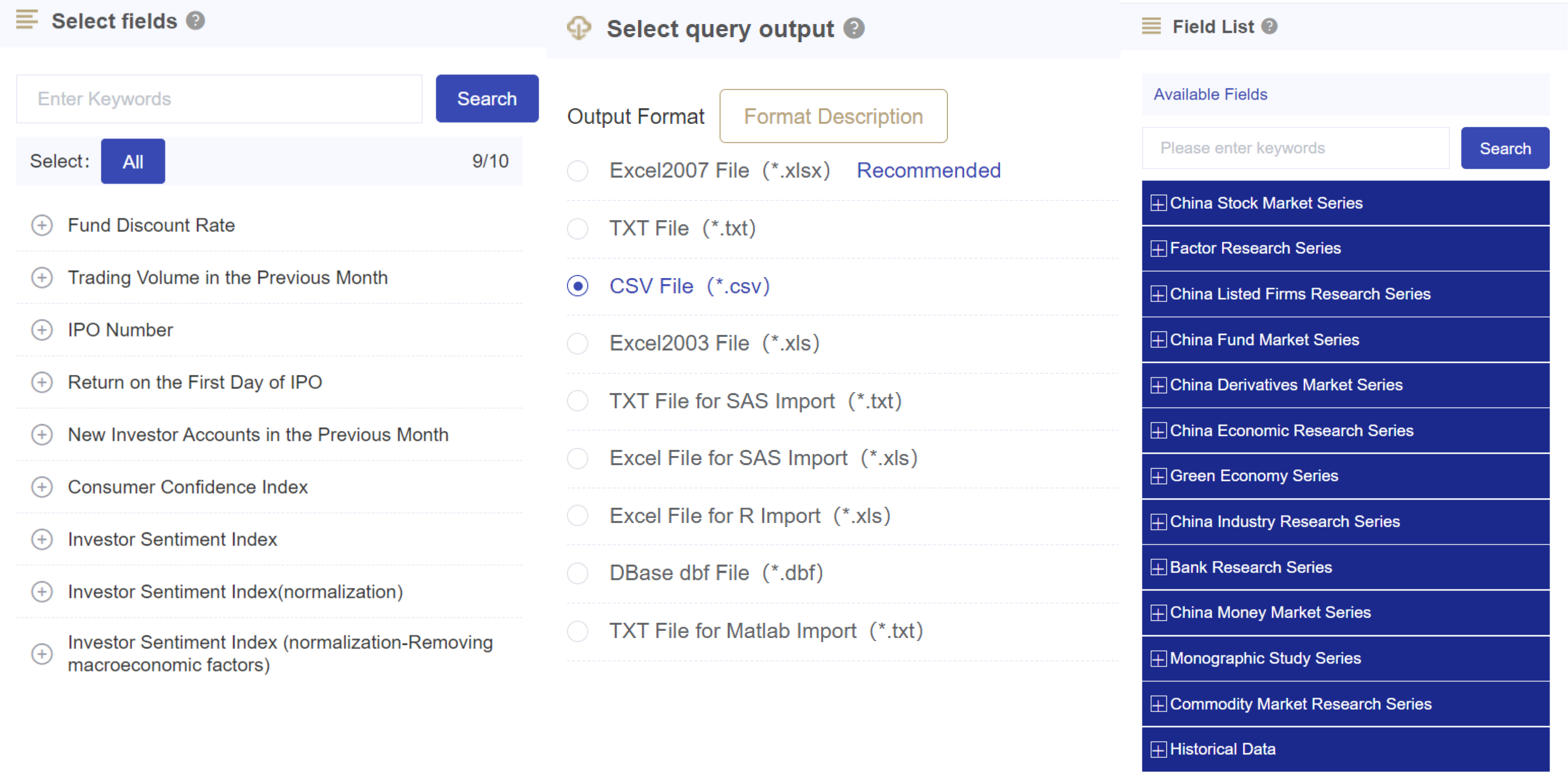


Figure 5: **RMSE** vs. size of hidden states of encoder/decoder over SML 2010 (left) and NASDAQ 100 Stock (right).

Platform Using: Qlib



Data Collecting: CSMAR



Select fields

Enter Keywords | Search

Select: All 9/10

- ⊕ Fund Discount Rate
- ⊕ Trading Volume in the Previous Month
- ⊕ IPO Number
- ⊕ Return on the First Day of IPO
- ⊕ New Investor Accounts in the Previous Month
- ⊕ Consumer Confidence Index
- ⊕ Investor Sentiment Index
- ⊕ Investor Sentiment Index(normalization)
- ⊕ Investor Sentiment Index (normalization-Removing macroeconomic factors)

Select query output

Output Format | Format Description

- ○ Excel2007 File（*.xlsx） Recommended
- ○ TXT File（*.txt）
- ● CSV File（*.csv）
- ○ Excel2003 File（*.xls）
- ○ TXT File for SAS Import（*.txt）
- ○ Excel File for SAS Import（*.xls）
- ○ Excel File for R Import（*.xls）
- ○ DBase dbf File（*.dbf）
- ○ TXT File for Matlab Import（*.txt）

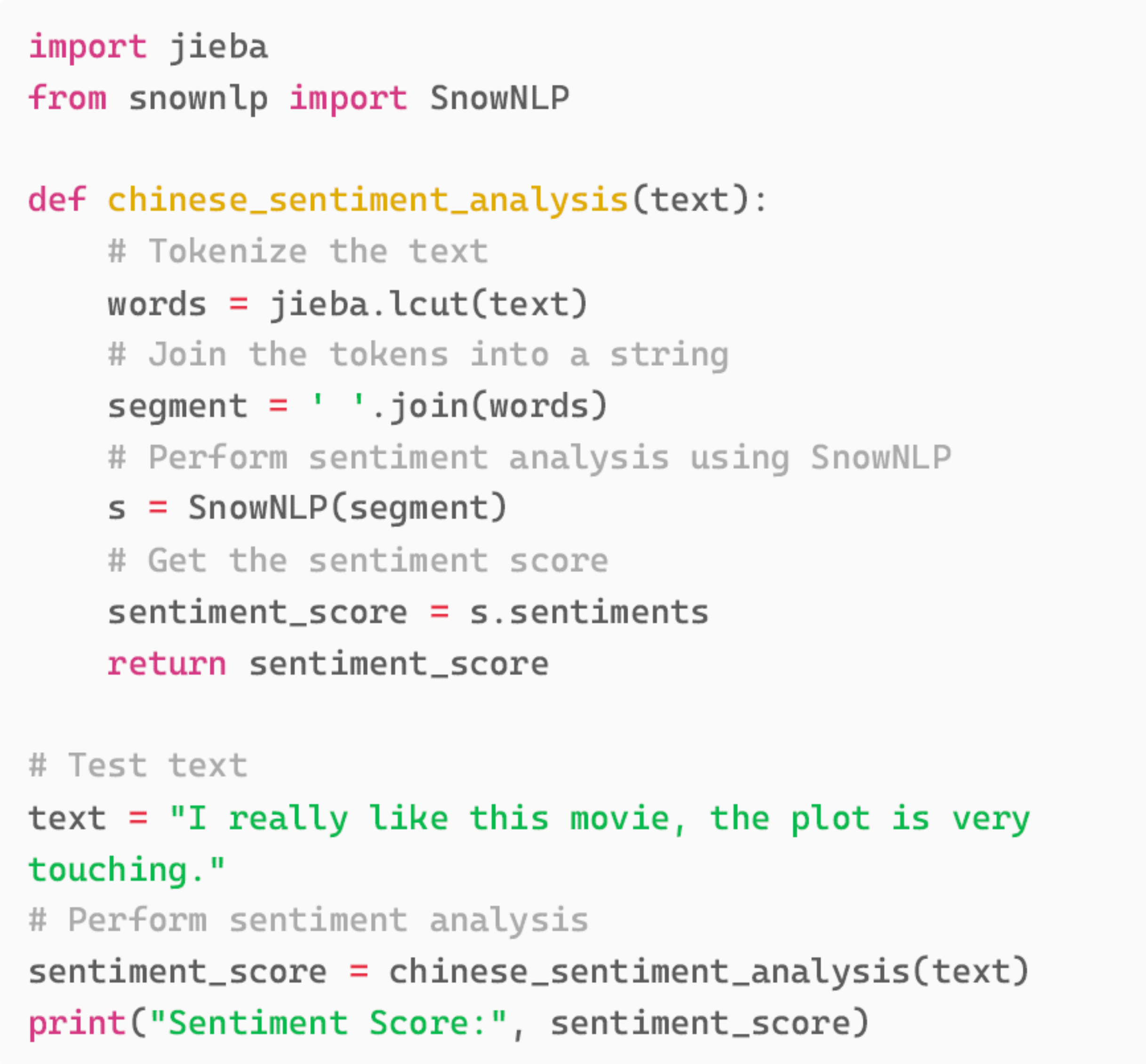
Field List

Available Fields

Please enter keywords | Search

- ⊞ China Stock Market Series
- ⊞ Factor Research Series
- ⊞ China Listed Firms Research Series
- ⊞ China Fund Market Series
- ⊞ China Derivatives Market Series
- ⊞ China Economic Research Series
- ⊞ Green Economy Series
- ⊞ China Industry Research Series
- ⊞ Bank Research Series
- ⊞ China Money Market Series
- ⊞ Monographic Study Series
- ⊞ Commodity Market Research Series
- ⊞ Historical Data

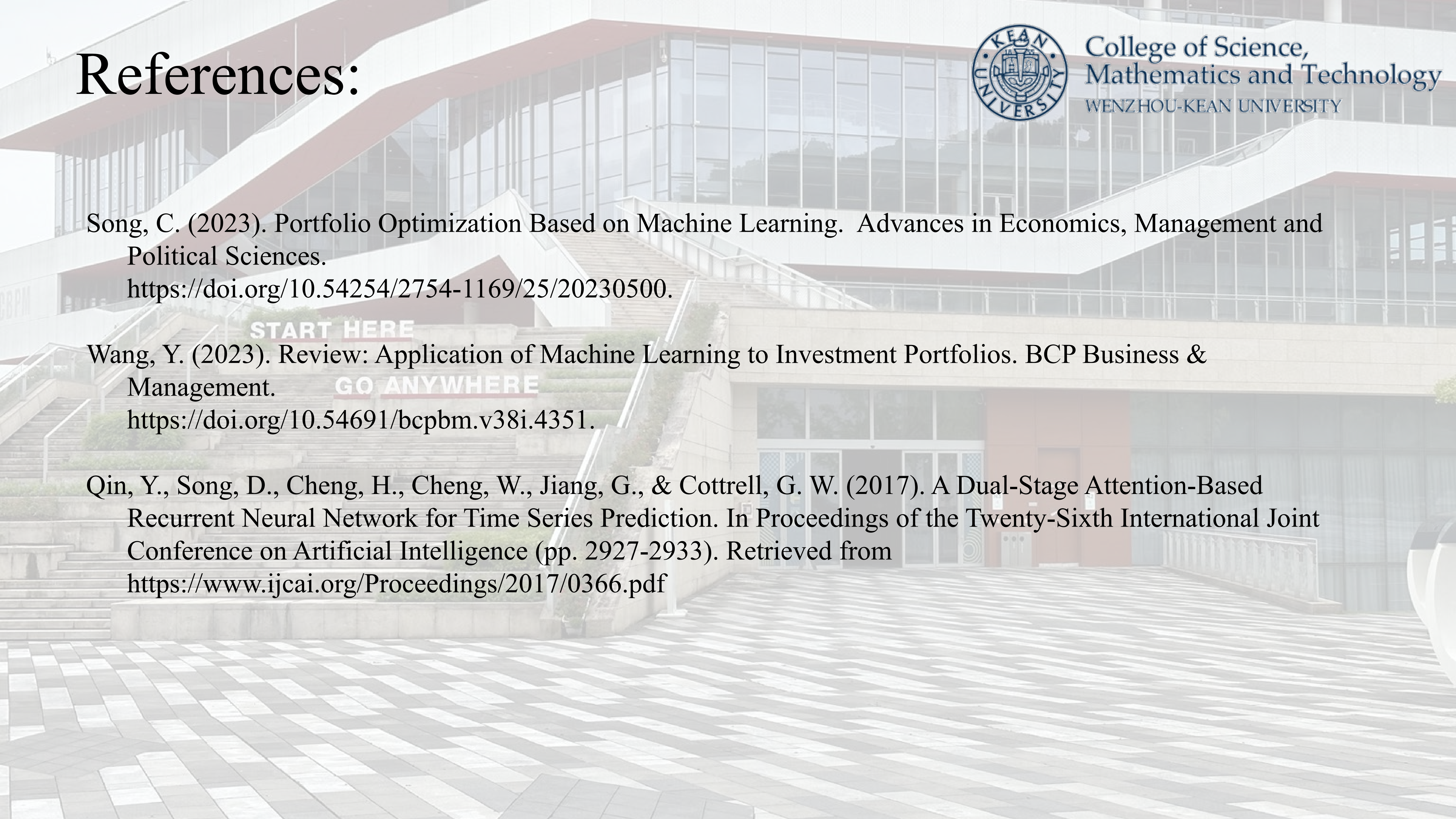
Future: Sentiment Analysis



```python
import jieba
from snownlp import SnowNLP

def chinese_sentiment_analysis(text):
    # Tokenize the text
    words = jieba.lcut(text)
    # Join the tokens into a string
    segment = ' '.join(words)
    # Perform sentiment analysis using SnowNLP
    s = SnowNLP(segment)
    # Get the sentiment score
    sentiment_score = s.sentiments
    return sentiment_score

# Test text
text = "I really like this movie, the plot is very touching."
# Perform sentiment analysis
sentiment_score = chinese_sentiment_analysis(text)
print("Sentiment Score:", sentiment_score)
```



References:

Song, C. (2023). Portfolio Optimization Based on Machine Learning. Advances in Economics, Management and Political Sciences. https://doi.org/10.54254/2754-1169/25/20230500.

Wang, Y. (2023). Review: Application of Machine Learning to Investment Portfolios. BCP Business & Management. https://doi.org/10.54691/bcpbm.v38i.4351.

Qin, Y., Song, D., Cheng, H., Cheng, W., Jiang, G., & Cottrell, G. W. (2017). A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (pp. 2927-2933). Retrieved from https://www.ijcai.org/Proceedings/2017/0366.pdf