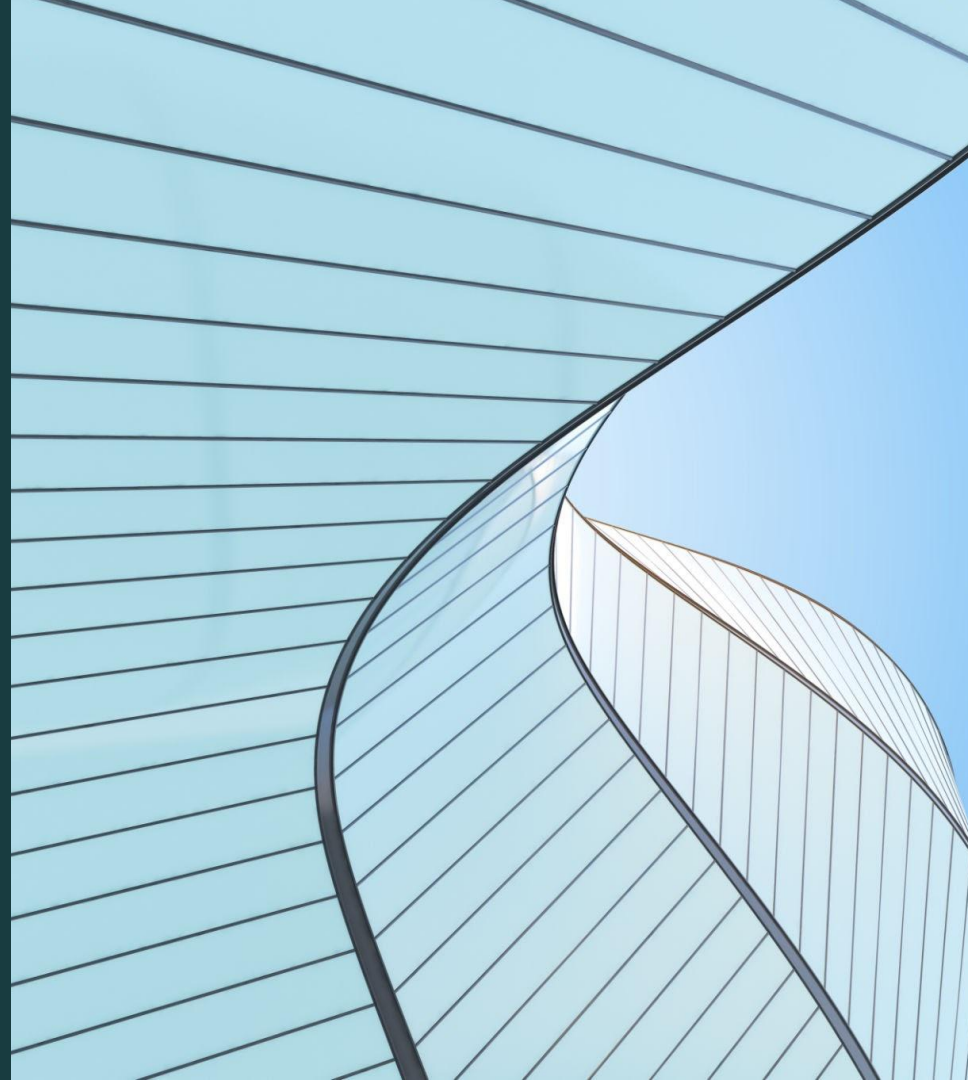


# Stats 418 Final Project Nikkei 225 Predictor

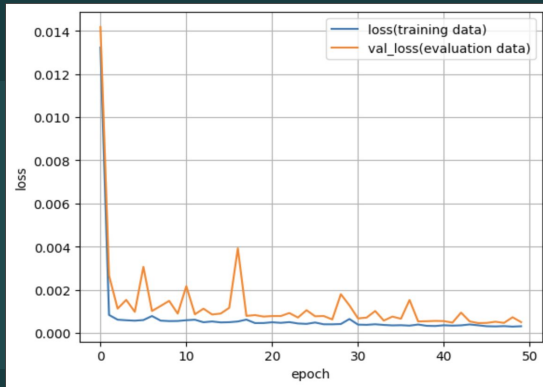
By Yifan Shen



# Model Training and Performance

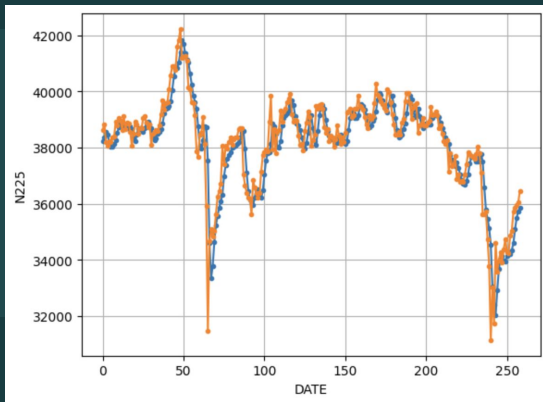
## Training Metrics (50 epochs, 10% validation split)

- Loss (MSE) and MAE tracked on both training & validation sets
- Training  $\rightarrow$  MSE  $\approx 0.000296$ , MAE  $\approx 0.0135$
- Validation  $\rightarrow$  MSE  $\approx 0.000497$ , MAE  $\approx 0.0168$
- Blue = Training MSE, Orange = Validation MSE
- Rapid drop in early epochs, then plateau



## Forecast VS Actual

- After training, predicted next-day closes against held-out test data
- Blue = LSTM predictions (inverse-scaled)
- Orange = actual Nikkei 225 closes
- Final  $R^2 = 0.804$ , Model explains ~80% of variability in unseen data  $\rightarrow$  strong predictive power



# Deployment & Demo

## Cloud Run API

- Endpoint:  
<https://nikkei-api-609056186247.us-central1.run.app/predict>
- Accepts JSON payload with last-10 closes → returns next-day forecast

```
michaelshen — zsh — 80x24
(base) michaelshen@Mac ~ % curl -X POST https://nikkei-api-609056186247.us-centr
all.run.app/predict \
-H "Content-Type: application/json" \
-d '{"history": [19619.35, 19650.12, 19675.00, 19700.44, 19725.11, 19750.33, 19775.50,
19800.22, 19825.99, 19850.13]}'
{"predicted_close": 20087.845703125}
(base) michaelshen@Mac ~ %
```

## Shiny App

- Hosted on shinyapps.io at  
<https://michaelshen25.shinyapps.io/nikkei-forecast/>
- Simple form for entering 10 past closes → displays prediction

Nikkei 225 LSTM Predictor

Enter the last 10 daily closes

Day 1 close:

Day 2 close:

Day 3 close:

Day 4 close:

Day 5 close:

Day 6 close:

Day 7 close:

Day 8 close:

Day 9 close:

Day 10 close:

Predicted Next Close:

# Conclusion & Next Steps

- **Strong predictive power:** Our LSTM captures ~80% of variance ( $R^2 \approx 0.80$ ) with low MAE on unseen data
  - **Serverless, scalable architecture:** Model runs as a Cloud Run API backed by a simple Shiny front-end
- 
- **Enrich feature set:** Add technical indicators (e.g. RSI, MACD) and external data (news sentiment, macro factors)
  - **Model enhancements:** Experiment with stacked/bi-directional LSTMs or attention layers
  - **Pipeline automation:** Set up CI/CD for model retraining, container builds, and data ingestion





Thank you