NUS Databusters Final Presentation

Team: **Data Cleaners In DeMand**

Members

Celine Tan Yen Xiu
Damaen Tan Teck Hwe
Dawn Koh Wen Xi
Martha Henrietta Soetedjo

Which Dataset Did We Use?

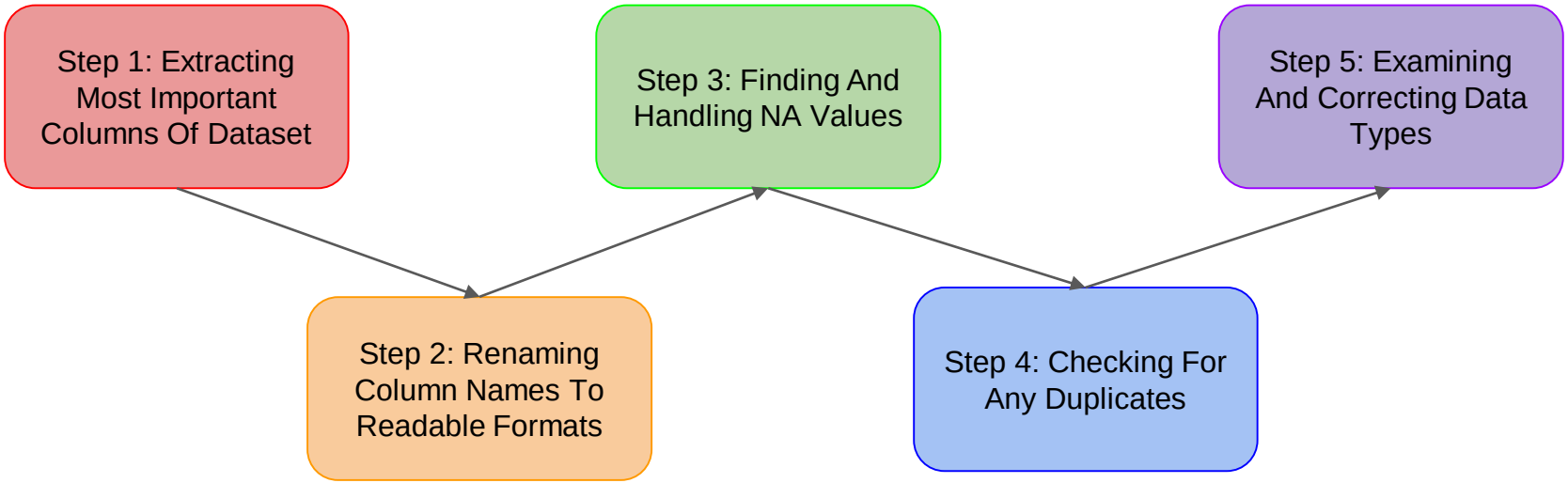
Quarterly Data

- ★ Contains **more relevant columns** used for analysis (246 columns)
- ★ **Reduces noise** (smooths out random variations and provides a clearer trend)
- ★ **Reducing overfitting & Model complexity**

Monthly Data

- ★ **Much less relevant columns** available (127 columns)
- ★ **More noise** due to higher probability of outliers
- ★ More **prone to overfitting and complex models**

Data Cleaning Process



Initial Dataset:

- 246 columns
- Unreadable column names
- NA values for certain columns
- Incorrect data type for sasdate

Cleaned Dataset:

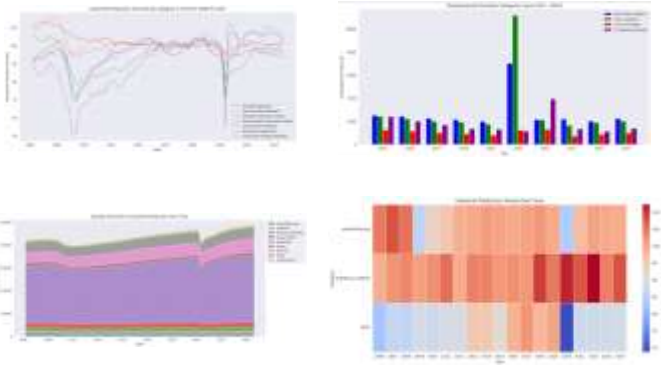
- 32 columns
- Readable column names
- No more NA values
- Correct data types

Data Visualizations

Performed 10 Different Visualizations:

1) Stock Prices Over Time
2) Imports And Exports Of Goods & Services
3) Real Disposable Personal Income
4) Industrial Production Amounts
5) Number Of Employees
6) Civilian Unemployment Rates
7) Personal Consumption Expenditures
8) Consumer Price Index
9) Industrial Production
10) Exchange Rates

★ Data Cleaning
★ Graph Plotting
★ Graph Types



Interesting Visualizations!

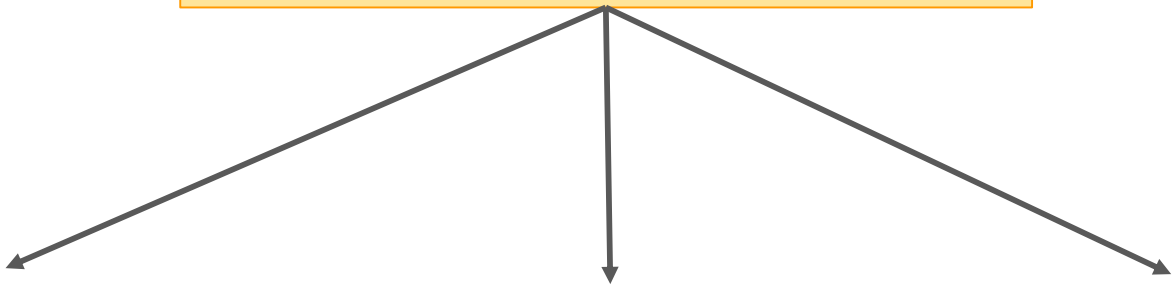
Time-Series Analysis

- Line Plots (Trend)
- Area Charts (Trend)
- Bar Graphs (Year)
- Heat Maps (Year)

Analyze and suggest key findings from the visualizations afterwards

Machine Learning Model Implementation

Autoregressive Distributed Lag (ARDL) Model



Extends the Autoregressive (AR) model to predict economic downturns by capturing dynamic interactions with quarter on quarter real GDP growth (gdp_qq)

1) **Ensure stationarity**
2) **Obtain optimal lag for Y variable (gdp_qq)**
3) **Obtain optimal lag for X variables**

```
total_consumption: 1 lags, AIC: 260.32957968749366
real_exports: 1 lags, AIC: 256.9335593550977
real_imports: 2 lags, AIC: 253.68782266739345
production_index: 2 lags, AIC: 253.00861607196686
capacity_util: 3 lags, AIC: 253.96565055898262
unemployed_rate: 4 lags, AIC: 206.39845622404647
manu_trade_sales: 3 lags, AIC: 202.381736600795
```

Akaike Information Criterion (AIC) used to optimize model selection by choosing the best lag structure for improved forecasting accuracy

Machine Learning Model Evaluation

- Durbin-Watson statistic (**2.415**): no significant autocorrelation
- Omnibus (**p = 0.000**) and Jarque-Bera (**p = 3.48e-21**) tests: significant deviations from normality
- Skew test (**-1.057**): extreme negative residuals
- Kurtosis (**8.320**): a heavy-tailed distribution

Machine Learning Model Evaluation

- Removed outliers using Standardized Residuals, Cook’s distance, and Leverage
- Durbin-Watson statistic (**2.124**) confirms no autocorrelation
- Omnibus (**p = 0.593**) and Jarque-Bera (**p = 0.823**) suggests residuals are normal
- Skew test (**-0.295**): slight left skew
- Kurtosis (**2.934**): mildly light tails

Predicting Forecast GDP Growth with ARDL

Using the optimal lag of independent and dependent variables obtained

↓

Train ARDL Model

↓

Predict forecast for 2025 quarter on quarter GDP growth

	gdp_qq_pred	economic_contraction
2025Q1	0.455470	no
2025Q2	0.609999	no
2025Q3	0.662249	no
2025Q4	0.550359	no

Machine Learning Model Optimization

Dataset Preparation

Converting real_gdp to gdp_qq to show quarter-on-quarter growth

Inclusion of financial market variables due to leading nature

Introduce train-test-split

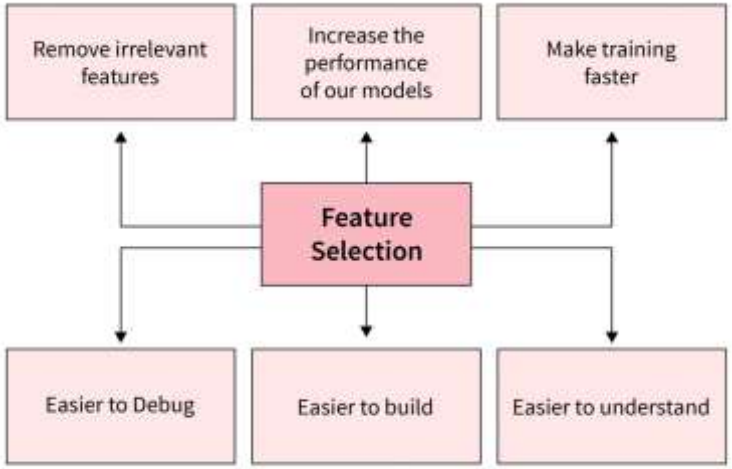
(2000 to 2017 - train)
(2018 to 2024 - test)

Removing features highly correlated to gdp_qq

Feature Selection

Feature selection using importances, defined as the contribution of variable to RF model

Top 9 variables were selected for refitting with train data



Variable	Importance
baa_corporate_bond_yield	0.329887
new_priv_housing_unit_perms	0.117855
consumer_sentiment	0.115558
unemployed_rate	0.102421
inventory_sales_ratio	0.060350
average_weekly_hours	0.055167
treasury_maturity	0.043194
policy_uncertainty	0.042466
sp500_price	0.040925
production_index	0.034321
cpi	0.022335
real_exports	0.021278
real_imports	0.014242

Random Forest Model Evaluation

Evaluation Metrics Used

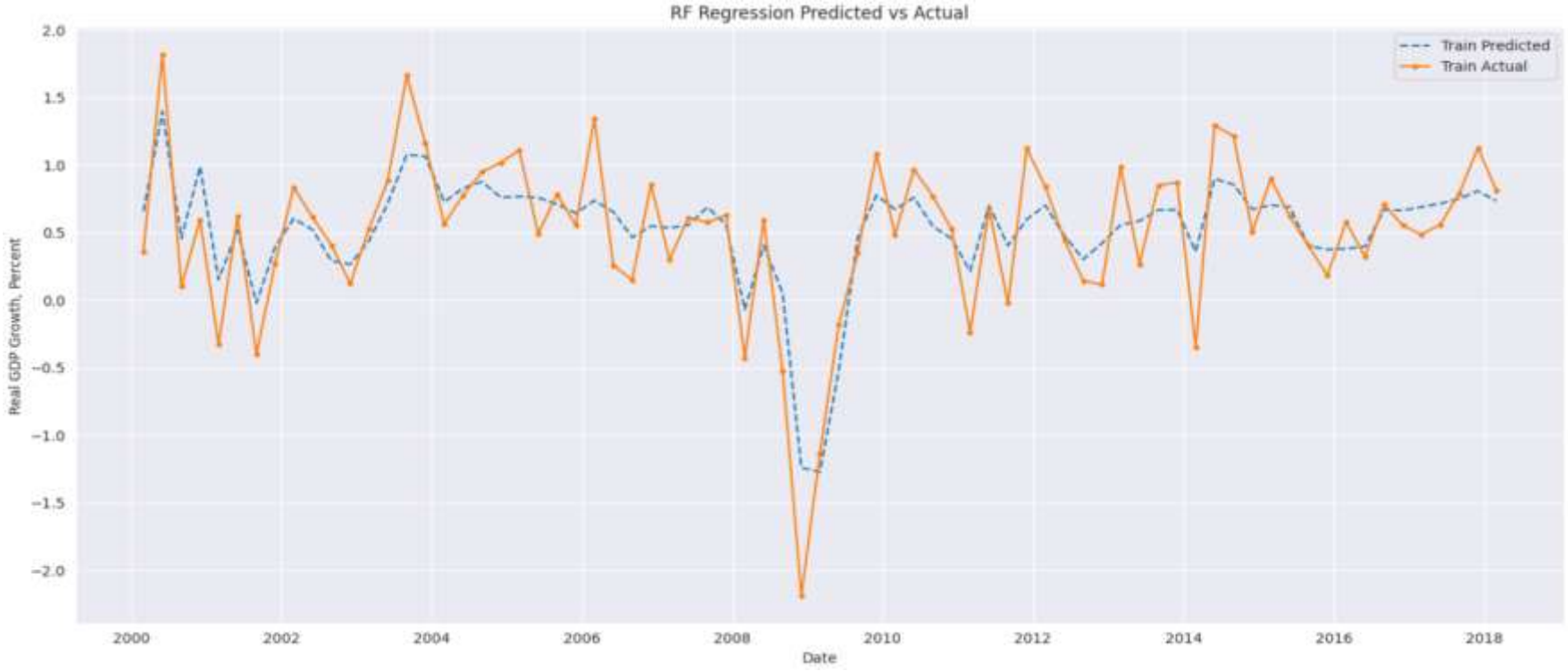
- Mean Absolute Percentage Error (MAPE)
- R-Squared (R^2)
- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)

Results Of Random Forest Model

Evaluation Metric	Train data	Test Data
MAPE	80.154%	145.004%
R^2	0.744	0.004
RMSE	0.300	2.224
MAE	0.242	1.122

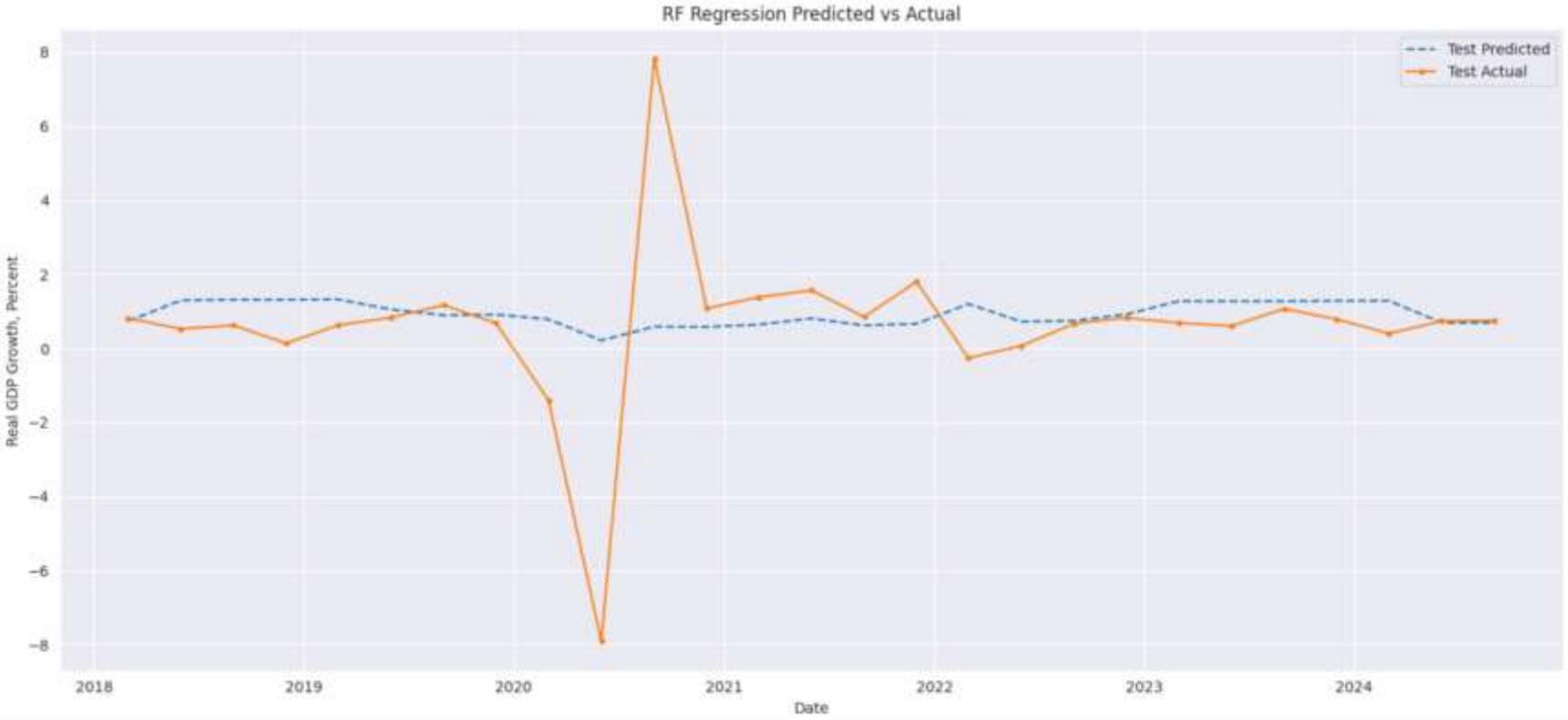
Random Forest Model Evaluation

Training Data

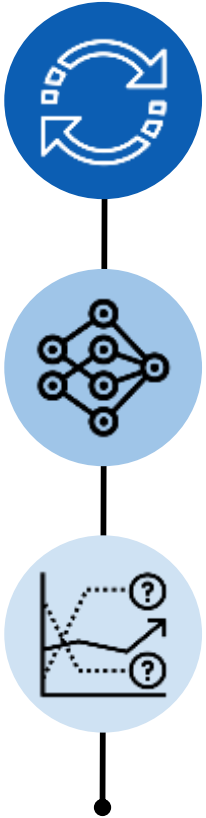


Random Forest Model Evaluation

Test Data



Predicting Forecast GDP Growth with ARIMA



Replace anomalous COVID-19 values (2020 Q1 - 2021 Q4) using train data up to 2019 Q4

Fit ARIMA model on amended dataset, choosing best p, d, q values

Forecast future GDP growth for next 5 quarters using RF model for 2025 forecasts

Predicting Forecast GDP Growth with ARIMA

Quarter	Forecast of GDP growth
2024-12-01	0.719780
2025-03-01	0.719780
2025-06-01	0.562613
2025-09-01	0.751898
2025-12-01	0.648779

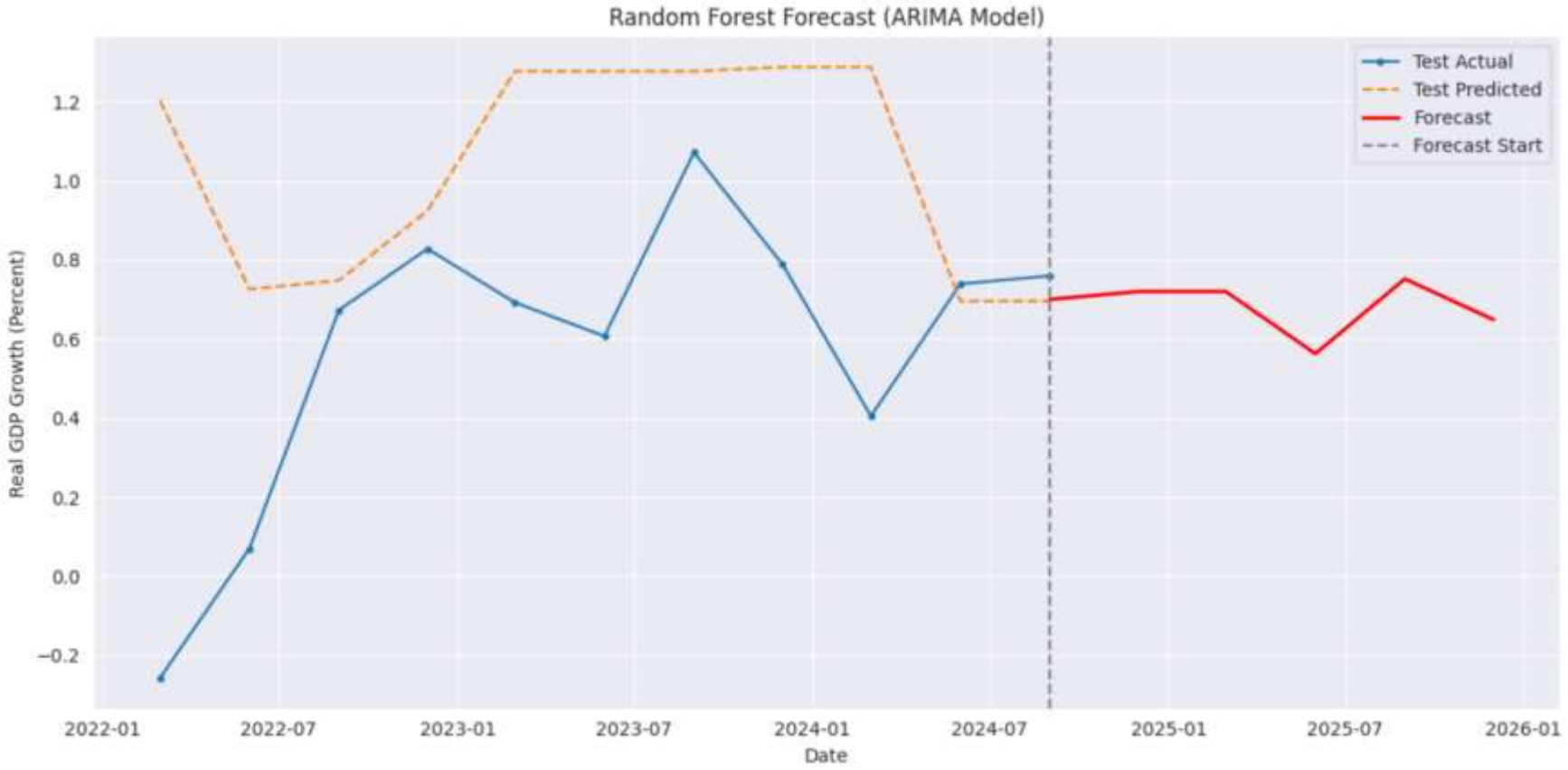
Preliminary Findings

- General trend of positive GDP growth over the next few quarters
  - While positive, the growth is relatively small
- In line with positive growths in bond yield, private housing permits, and treasury maturity

Conclusion

The forecast shows a strengthening economy rather than downturn in the next few quarters

RF Forecast Predictions



Possible Improvements To Models

1. ARDL Model

- Addressing Non-Linearity:
  - ARDL assumes a linear relationship between variables, which may not capture complex macroeconomic effects.
  - Consider Threshold ARDL (TARDL) or Nonlinear ARDL (NARDL)

2. Random Forest Model

- Optimizing accuracy
  - Using ensemble methods such as bagging and boosting (XGBoost or AdaBoost)
- More validation processes (e.g., through rolling window)

Question 1

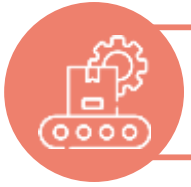
Possible key predictors of economic downturns include:

Rising Unemployment Rates

Decline in Consumer Confidence

Stock Market Declines

Inversion of Yield Curve



Decline in production activity

Question 1

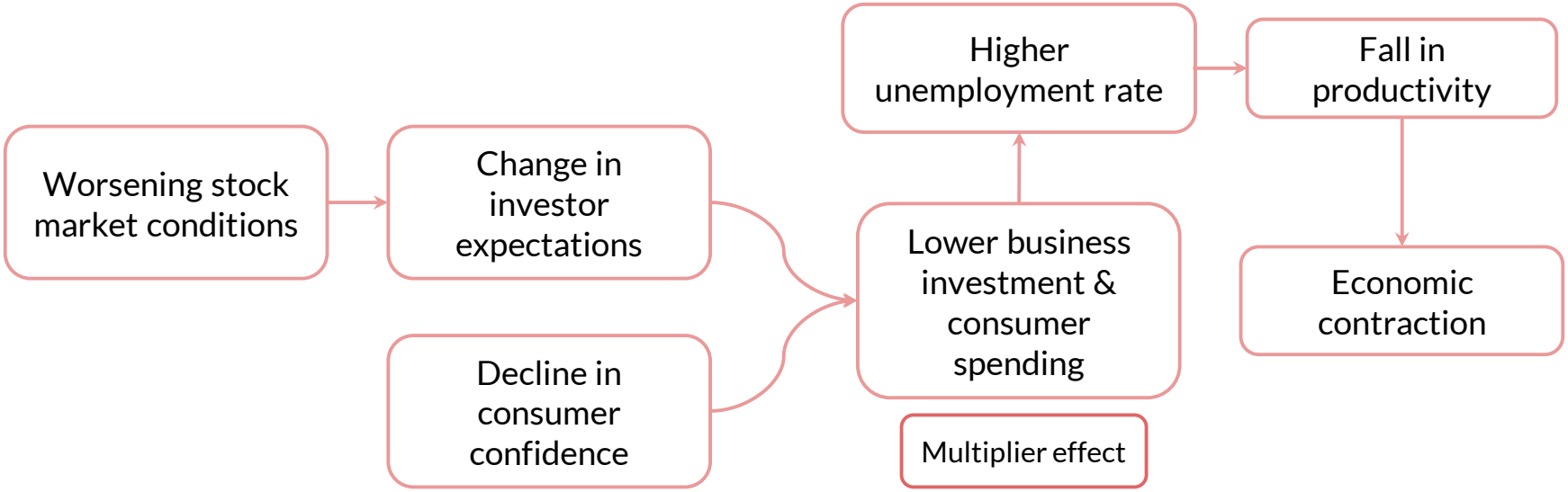
Forecast horizon and metrics to measure the change in each indicator

Indicator	Forecast horizon	Possible metrics
Rising Unemployment rate	Very short term	Unemployment rate, unemployment claims
Stock market decline	Short term	SP500, VIX
Decline in consumer confidence	Short term	Consumer sentiment, Policy uncertainty
Decline in production activity	Short-medium term	Industrial production, sales to inventory ratio
Inversion of yield curve	Medium-long term	corporate bond yield, treasury bond yield

Sentiments, unemployment and Productivity

The stock market is a leading indicator, relating closely to investor expectations on future returns. Similarly, consumer confidence affects spending decisions

These affect production and hence unemployment rates

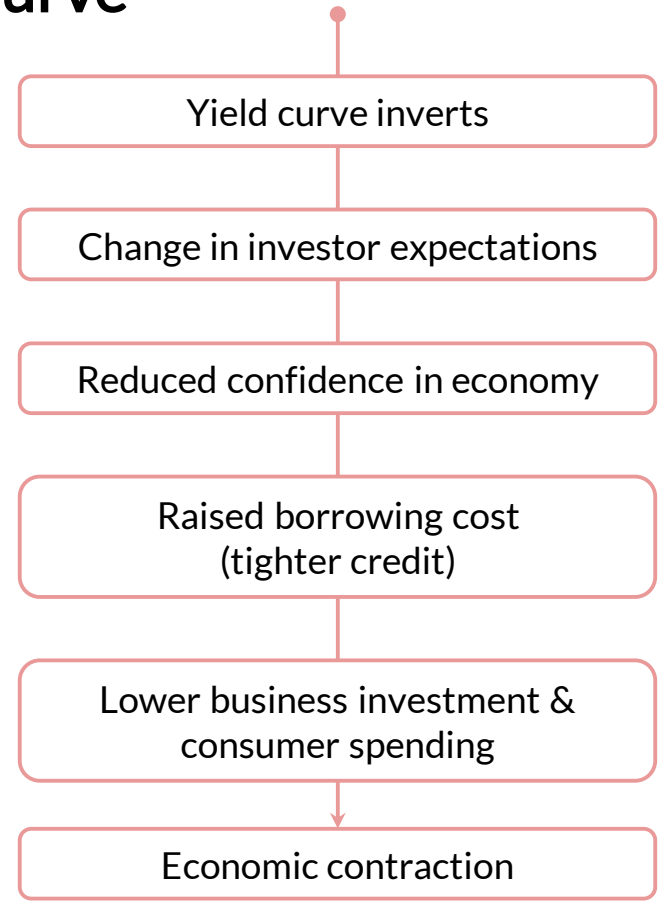


Inversion of Yield Curve

Inversion of the yield curve occurs when **short-term interest rates exceed long-term interest rates**

Possible Causes Of Inverted Yield Curves:

- Inflation Expectations From Consumers
- Interest Rate Hikes
- Supply And Demand Imbalances
- Geopolitical Tensions
- Financial Crisis



Question 2

How would you expect policymakers in Singapore to act based on your forecasts and what are the associated risks they might face?

Singapore is a small, open economy, and is heavily reliant on international trade, finance, and investments

Forecast Results In 2025:

	Forecast
2024-12-01	0.719780
2025-03-01	0.719780
2025-06-01	0.562613
2025-09-01	0.751898
2025-12-01	0.648779

- GDP growth is projected to hover around **0.563%** from Q4 2024 to around **0.752%** by the end of 2025
- 2 significant dips during June and December 2025 — suggesting the economy might **face headwinds**
- Fluctuations for the forecast is **less comparable** to the actual test from 2022 to 2024

Policy Responses

Monetary Policy

- If inflation is **low and steady**: MAS can keep the S$NEER policy band stable or loosen it slightly to support exports
- If inflation **continues to increase** with GDP growth: MAS can tighten the S$NEER policy band to prevent overheating

Fiscal Policy

- Increase government spending on areas such as healthcare, infrastructure tand digital transformation o boost productivity
- Hold back fiscal stimulus to avoid contributing to rising inflation

Structural Policy

- Workforce upskilling, digital transformation, and R&D incentives to sustain longer-term productivity growth

Risks of Policies

Monetary Policy

- Over-tightening S$NEER policy band could harm competitiveness
- Over-loosening S$NEER policy band could cause imported inflation
- Trade-off between balancing GDP growth and inflation

Fiscal Policy

- Crowd out private investments
- Fiscal policies might have a smaller impact on GDP growth as compared to countries with closed economies
- Rising inflation might cause MAS to tighten the policy band, which counteracts fiscal policies

Structural Policy

- May take time for results to be seen, may be costly and might cause short term disruptions

Question 3

Considering the proposed tariffs and other plans likely to be enacted in 2025 by the newly elected U.S. President Trump, what judgmental adjustments might you make to your forecasts based on this qualitative information not yet reflected in the data?

Donald Trump's proposed tariffs

Tariffs On Other Countries

- 25% tariff on all imports from Mexico and Canada, reduced 10% tariff on exports
- Additional 10% tariff on Chinese Imports
- Potential 25% tariff on imports in the EU

Industry-Specific Tariffs

- 25% tariff on all steel and aluminium imports
- 25% tariff on automobiles, semiconductors and pharmaceuticals

Direct Impacts

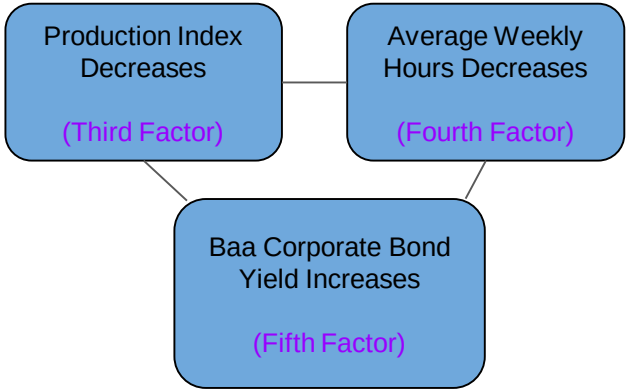
Direct Impact Of Tariffs On Economy: Imports & Exports

Addition or increase in tariffs on imports in USA	→ Countries less willing to export to USA as cost of imported goods rises	→ USA will import less from other countries as fewer foreign goods are sold	→ Overall imports decline (First Factor)
Reduction in tariffs on certain exports in USA	→ Goods from USA will be cheaper for foreign buyers	→ Increases demand for US exports as they are more affordable	→ Overall exports Increase (Second Factor)

Other Impacts Of Tariffs On Economy

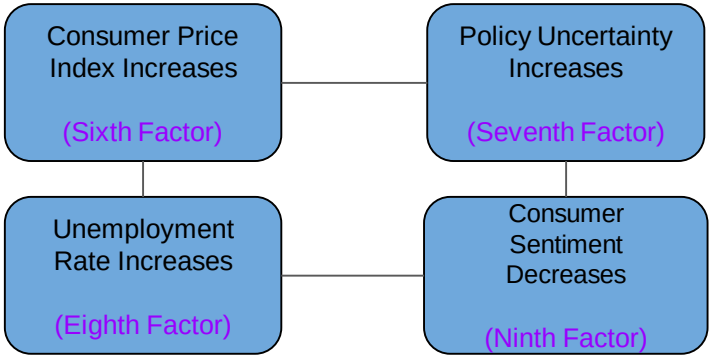
Higher production costs faced by industries

- Industries less willing to produce goods
- Industries will cut costs
- Reduced employee work hours due to job cuts
- Company earnings will decrease
- Corporate bonds will be riskier and this pushes yields higher



Higher costs passed on to consumers in the form of higher prices

- Reduction in purchasing-power
- Lowering consumer sentiment
- Higher prices for goods and services
- Demand for goods and services decline
- Might lead to potential layoffs
- Unemployment rate increases
- Contribute to economic uncertainty



How Tariffs Will Affect Our Model

Variable	Importance	After Tariffs Values
baa_corporate_bond_yield	0.329887	→ Increases
new_priv_housing_unit_perms	0.117855	→ Decreases
consumer_sentiment	0.115558	→ Decreases
unemployed_rate	0.102421	→ Increases
inventory_sales_ratio	0.060350	
average_weekly_hours	0.055167	→ Decreases
treasury_maturity	0.043194	
policy_uncertainty	0.042466	→ Increases
sp500_price	0.040925	
production_index	0.034321	→ Decreases
cpi	0.022335	→ Increases
real_exports	0.021278	→ Increases
real_imports	0.014242	→ Decreases

We can impute values of these features from December 2024 onwards

- ★ Set a higher value than September 2024 if feature increases after tariffs
- ★ Set a lower value than September 2024 if feature decreases after tariffs
- ★ Set the same value as September 2024 if there is no noticeable change in feature value after tariffs


Thank
you!