Predicting the next recession

Benjamin Green, Nate Medina, Julian Herazo

Goals for the project

- To identify which economic indicators best predict recessions
- To compare model performances and determining whether the models can reasonably forecast recession risk.
- Provide an estimate of the current recession probability.

Why does predicting the next recession matter?

- Recessions impact jobs, investments, and government budgets.
- Having an early detection helps businesses make a plan ahead of time.
- Policymakers can respond sooner to stabilize the economy.
- Consumers benefit from better financial preparedness.

Databases Used:

- Global Economic monitor (GEM)
- Federal Reserve Economic Data (FRED)

Features:

Many features contribute to a recession, but the features we found that gave the best early warning signal for recessions were:

- GDP (Economic growth)
- CPI (Inflation)
- Imports/Exports (Trade activity)
- Interest Rates
  - 10 Year Treasury Yield
  - 3 Month Treasury Yield
- Yield Curve Spread (10Y – 3M)
- And these features we collected came from China, Japan, Europe, UK, South Korea, US

Tools

- Excel to gather the datasets initially
- PostgreSQL
- VS Code to then sort through the datasets
- Github to collaborate as a group

Libraries

- Pandas
- Mathplotlib
- Scikit-learn
- Pathlib

Data Cleaning and Preprocessing

- Different countries reported GDP quarterly while others reported monthly
- Missing values appeared at different time periods depending on when countries started reporting their data

What we did?

- Converted quarterly GDP to monthly by forward-filling values so all features aligned on a monthly timeline
- Merged all countries into one unified monthly index (1990–2025)
- Forward-filled and backward-filled missing values to keep the timeline consistent
- Dropped features with too many missing values, but kept valid NaNs before resampling

Feature Engineering

- GDP growth rates (current, 3-month lag, 6-month lag)
- Yield spread (10-year minus 3-month interest rate)
- 12-month average & minimum yield spread
- International indicators (Japan GDP growth, EU/UK short rates)
- Rolling window features (12-month trends)

Models Used

- **Logistic Regression** – some input translate to the probability of recession given in that month
- **Random Forest** – builds complex trees
- **K-Nearest Neighbors** – compares to past condition (“no learning”)

How Each Model Works

Logistic Regression

- Learns a weighted rule (yield curve + GDP = recession risk)

Random Forest

- Builds many decision trees
- Good at capturing nonlinear patterns

KNN

- No training
- Predicts based on nearest historical months

Model Results

- AUC: **0.98** → excellent ability to separate recession vs normal
- Recession months are rare (only 2 in test set)
- F1 Score lower because rare events are harder to detect
- Model still learned key recession patterns

Model Result visualizations

```
Train: 297 obs, Test: 119 obs
Train recessions: 26, Test recessions: 2

AUC-ROC: 0.983
F1 Score: 0.154

              precision    recall  f1-score   support

         0.0       1.00      0.81      0.90       117
         1.0       0.08      1.00      0.15         2

    accuracy                           0.82       119
   macro avg       0.54      0.91      0.53       119
weighted avg       0.98      0.82      0.88       119
```

```
AUC-ROC: 0.940
F1 Score: 0.000

              precision    recall  f1-score   support

         0.0       0.98      1.00      0.99       117
         1.0       0.00      0.00      0.00         2

    accuracy                           0.98       119
   macro avg       0.49      0.50      0.50       119
weighted avg       0.97      0.98      0.97       119
```

1. Logistic Regression (Top left)

2. Random Forest (Bottom left)

3.K-Nearest Neighbor (Bottom right)

precision = when predicting a recession, how often is there actually a recession?

recall = Of all the actual recessions, how many did it detect?

F1 = balance average of precision and recall (prediction quality)

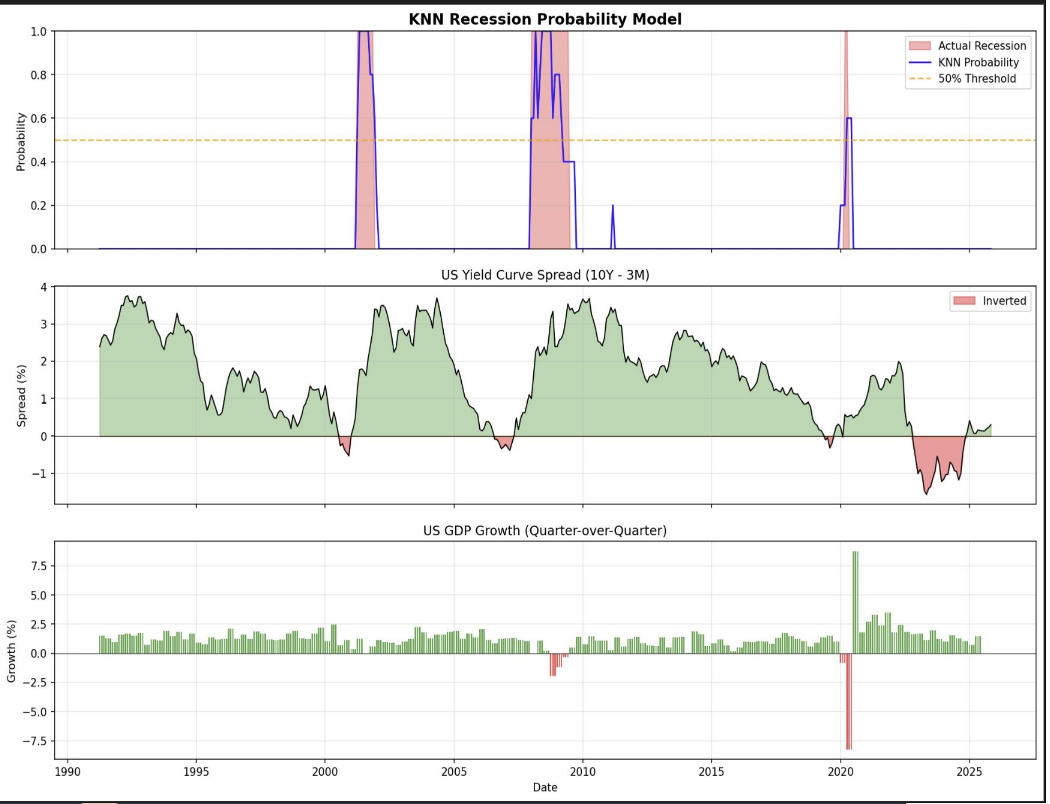
```
AUC-ROC: 0.987
F1 Score: 0.444

              precision    recall  f1-score   support

         0.0       1.00      0.96      0.98       117
         1.0       0.29      1.00      0.44         2

    accuracy                           0.96       119
   macro avg       0.64      0.98      0.71       119
weighted avg       0.99      0.96      0.97       119
```

Model plot



What does this mean?

-Top panel shows the probability of a recession happening and goes from a scale of 0-1. The blue line is what indicates the recession's probability at the time.

-Middle panel is a good warning indicator and more specifically shows the yield curve which is A picture of the same kind of loan but for different periods of time. We can see that there's an inverted representation before each recession occurs.

-Last panel is GDP growth and we can clearly see where each of the last 2 recessions occurred.

Feature Importance

```
Top 10 important features:
  us_gdp_growth: 0.1232
  us_gdp_growth_lag3: 0.1167
  us_yield_spread_lag12: 0.1105
  us_gdp_growth_lag6: 0.0900
  jp_gdp_growth: 0.0789
  uk_3m: 0.0538
  us_3m: 0.0502
  us_yield_spread_12m_min: 0.0448
  eu_3m: 0.0370
  us_trade_balance: 0.0363
```

Key:

- us_gdp_growth = the strongest predictor, gdp correlates with probability of a recession
- us_gdp_growth_lag3 = falling gdp consecutive quarters, recession more likely
- us_yield_spread_lag12 = inverted yield curve can be an early predictor of a recession

Forecasting

- Latest recession probability: **6%–8%**
- This signals to us that there is a mild risk but no strong warning indicated when we ran our dataset through.
- Matches real-world conditions in 2025. In other words we took data from the past couple of years all the way up until recently to make a good prediction.

Model Strengths and Limitations

Model	Strengths	Limitations
Logistic Regression	Detects warnings signs, catches recessions	Many false alarms
Random Forest	Identifies important features	Memorizes, can’t generalize
KNN	Can identify recession similarities	Doesn’t learn anything

Conclusion

- Economic data CAN predict recessions
- Our models detected real warning signs
- Yield spreads and GDP growth were strongest predictors
- Future work: more data, more countries, newer models