Predicting Employment Trends and Outlook

Team: Recession Regression
Members: April Nellis, Boyang Wu, Preetham Mohan, Shirlyn Wang, Tejaswi Tripathi, Zhengjun Liang

Overview

- **Guiding Research Question**: predict future change in total private sector employment
- **Motivation**: Employment is a major measure of economic health, corresponding to the strength of private sector companies and the overall well-being of Americans.
- **Key Stakeholders**: government policymakers and market investors
- **KPIs**:
  - Volatility in future trends through R2 score
  - Outlook classification accuracy
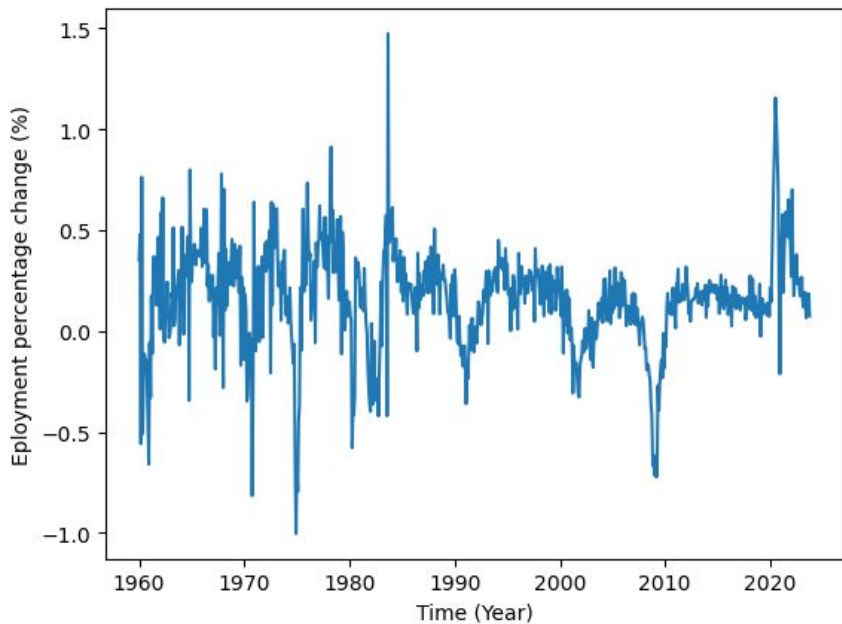
Data Collection

- Primary Resources:
    - Federal Reserve Economic Data Database
    - Yahoo Finance

- Types of Data:
    - Gross Domestic Product (GDP)
    - Consumer Price Index (CPI)
    - Loans (consumer, commercial, overall borrowing)
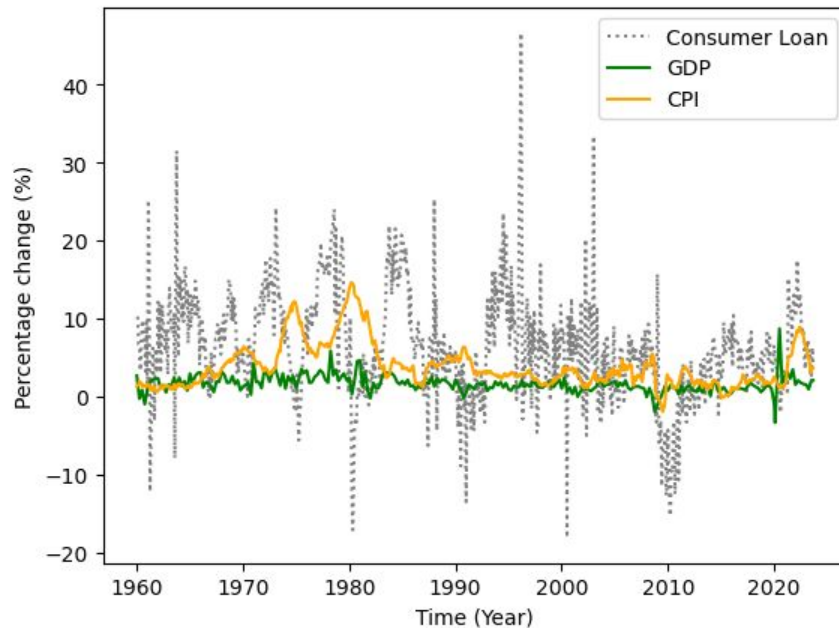   - Federal funds rate and federal deficit
    - S&P 500 index

Data Visualization
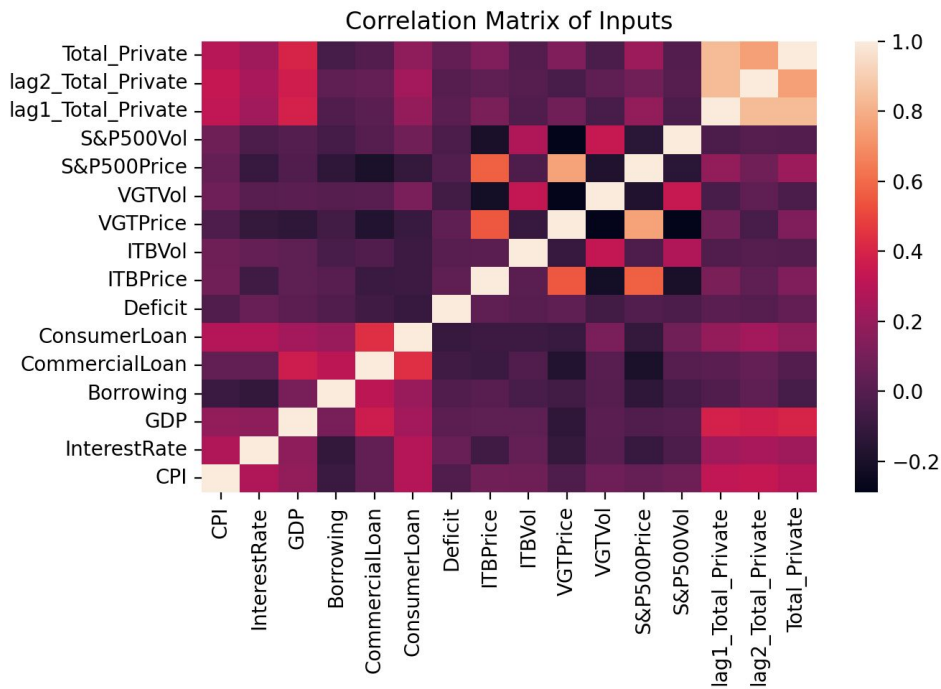

Output

Some Input Features
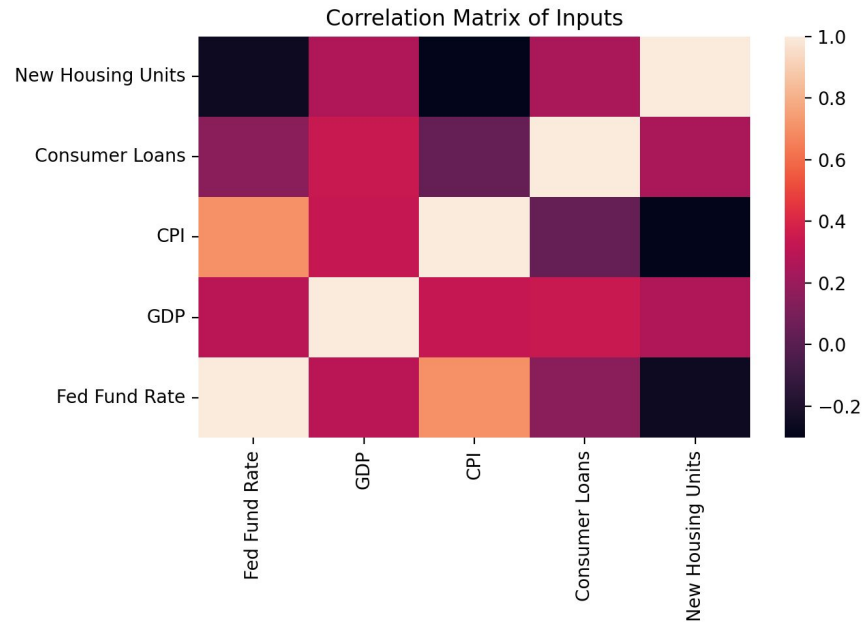
Exploratory Data Analysis

All features from 2006 onward

Correlation Matrix of Inputs

All features from 1960 onward

Correlation Matrix of Inputs

Regression vs. Classification?

	Regression Models	Classification Models
Strengths	• Predict both direction and magnitude of employment change • More descriptive • Easy to visualize predicted results	• Simpler problem to solve • More general overview of employment outlook
Weaknesses	• Predictions are very sensitive to noisy data • Less straightforward metric of model performance	• Susceptible to unbalanced data • Cannot capture magnitude of changes

Model Overview

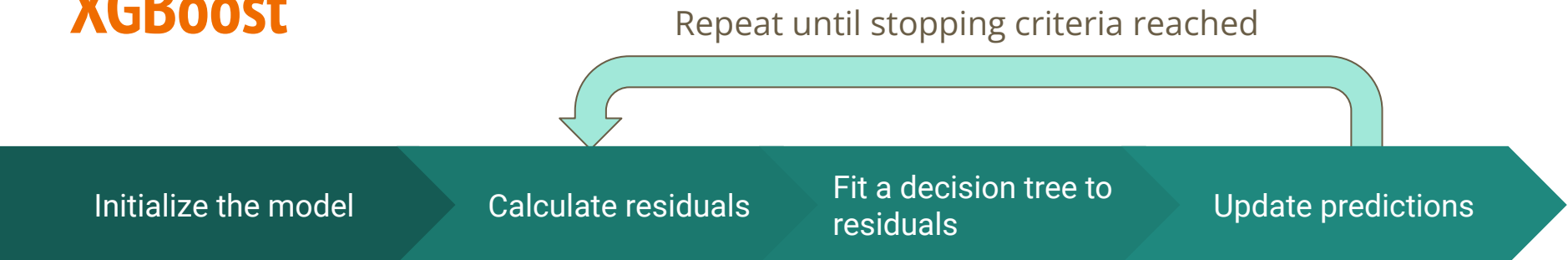| Classification	Regression
Naive Forecasting (Next month's outlook = This month's outlook)	Linear Regression
XGBoost	XGBoost
Recurrent Neural Network	Recurrent Neural Network

Recurrent Neural Network (RNN)

- Type of neural network adapted to work for time series or sequential data
- Input: data since 1960-01-01
  - Lookback period: 3 years
- Output:
  - Employment percentage change (Regression)
  - Employment increase or decrease (Classification)
- Key parameters of RNN
  - Number of layers: 6
  - Hidden layer size: 64
- Results:
  - Regression R2 score: **0.606**
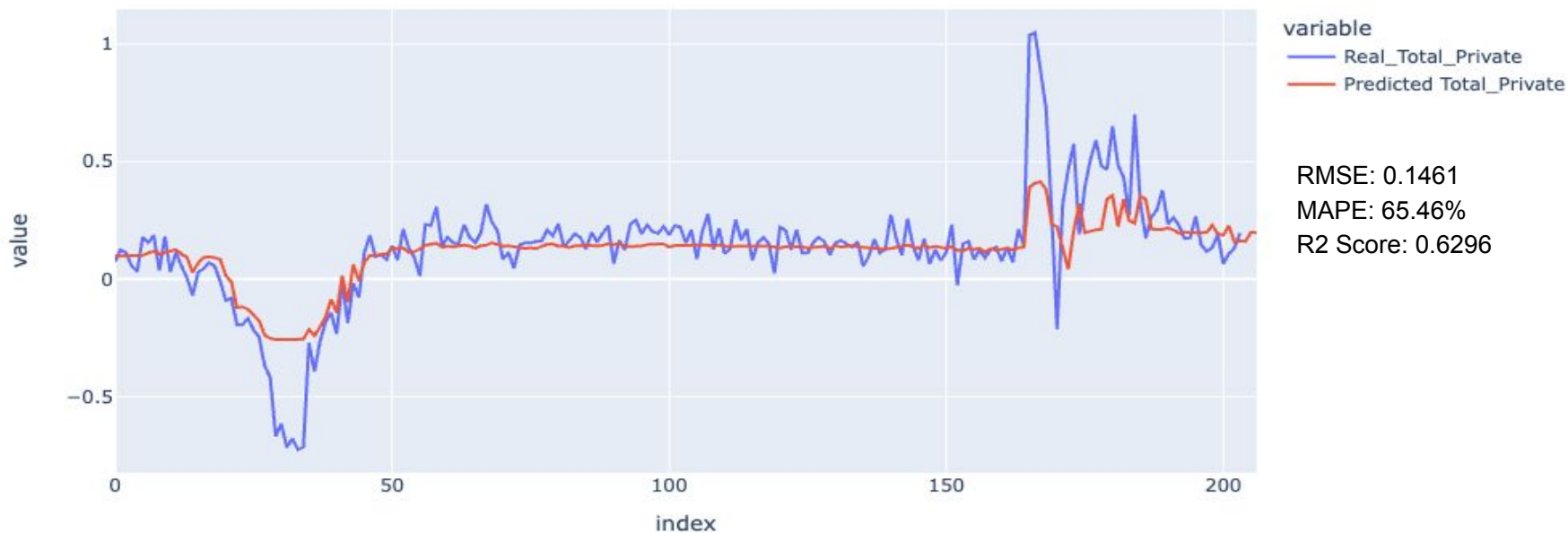 - Classification accuracy: **98.2%**

XGBoost

Repeat until stopping criteria reached

| Initialize the model | Calculate residuals | Fit a decision tree to residuals | Update predictions |
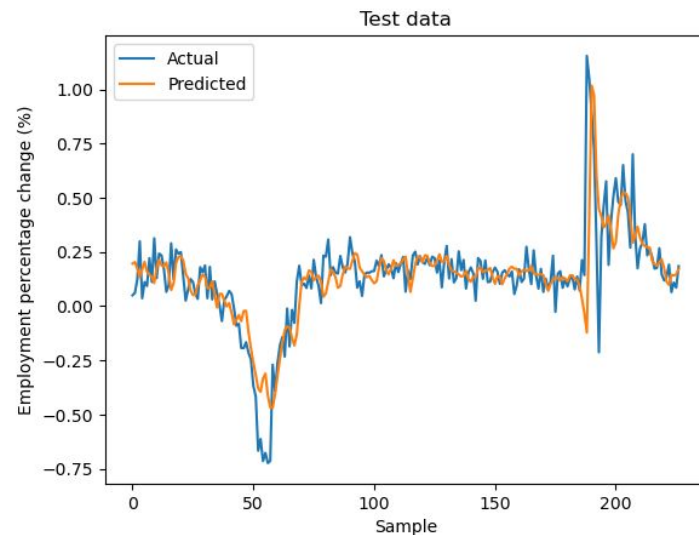
- A powerful ensemble learning technique
- Time Series Cross Validation for grid searching optimal hyperparameters
- Mean Absolute Percentage Error as benchmark to choose the best model
- Features selection via plot_importance and feature_importances_ to find 15 most important features to train the model
- Classification Accuracy: 0.966
- R2 Score: 0.6296

XGBOOST Overall Performance

XGBOOST vs Total_Private FULL

variable
— Real_Total_Private
— Predicted Total_Private

RMSE: 0.1461
MAPE: 65.46%
R2 Score: 0.6296

Model Comparison

Classification Models	Accuracy
Naive Forecasting (Baseline)	89.5%
XGBoost	96.6%
Recurrent Neural Network (RNN)	98.2%

Regression Models	R2 Score
Linear Regression (Baseline)	0.2071
Recurrent Neural Network (RNN)	0.6060
XGBoost	0.6296
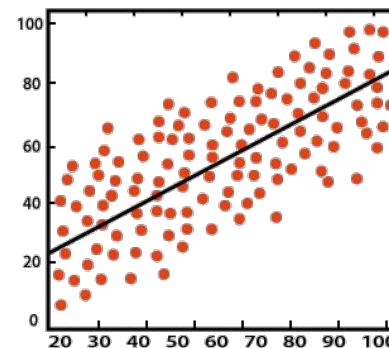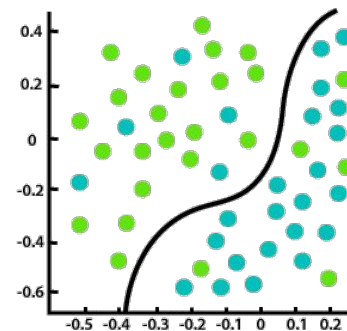
Possible Future Research

- Robustness to shocks:

  - We have removed some major shocks from the economy, such as the COVID-19 pandemic.

  - Idea: train on such "irregular" data and successfully capture the volatility arising from such unexpected events.

- Employer and Investor Sentiment

  - Natural language processing

Acknowledgements

- Erdos Institute
- Our mentor: Dr. Dyas Utomo

Thank you!