Predicting **Recession Trends** with Machine Learning

Tony Hua, Faisal Karim, Maggie Mano,
Oscar Mercado - Dec 2022

Abstract

Our research applies machine learning techniques to predict S&P 500 prices surrounding recession events using economic indicator data such as commodities, consumer behavior, monetary supply, credit markets, real estate, labor, and other macroeconomic indexes.

About Us

Meet The **Team**

Tony Hua
M.S. Computer Science

Faisal Karim
M.S. Data Science

Maggie Mano
B.S. Computer Science

Oscar Mercado
B.S. Computer Science

Predicting Recession Trends with Machine Learning

Background

What is a Recession? S&P 500?

- Recessions are periods of temporary economic decline
- S&P 500 is a stock market index that tracks the prices of 500 largest companies
- S&P 500 is a great proxy for U.S. economic conditions.

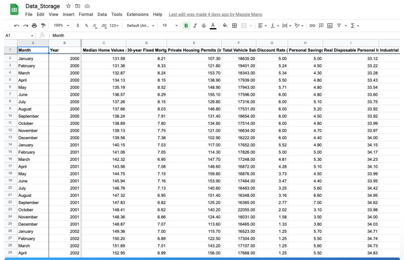
Why should researchers care?

Recessions impact:

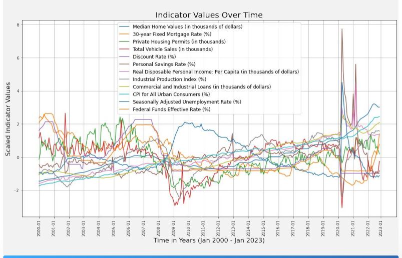
- Retirement accounts
- Business decisions
- Career changes
- Policy making
- Spending behavior
- and more...
-

Predicting Recession Trends with Machine Learning

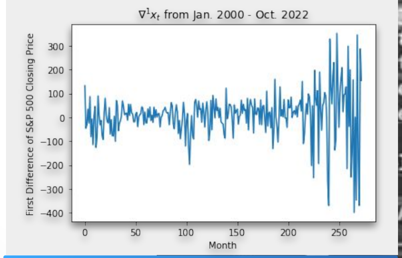
Methodology



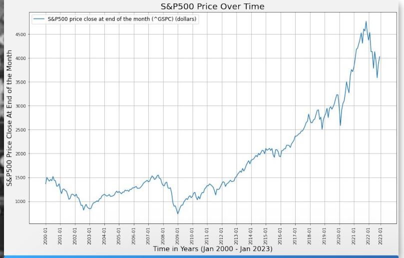
Sourcing & Cleaning Data



Exploratory Data Analysis



Model Fitting & Tuning



Interpret Results

Predicting Recession Trends with Machine Learning

Sourcing **Predictor Data**

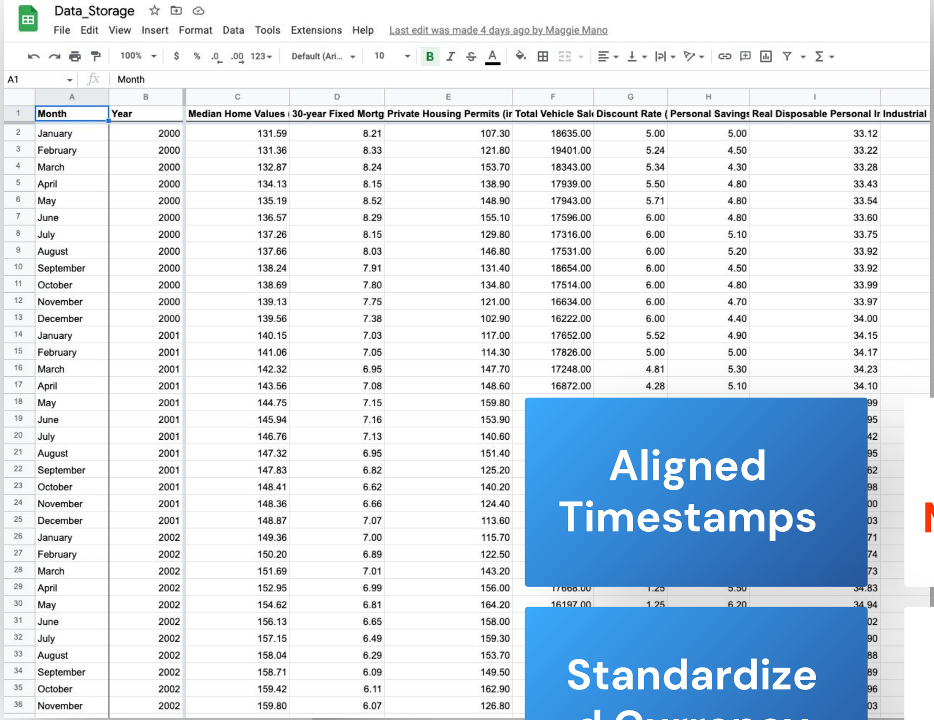
Prior to training machine learning models, we source a breadth of economic indicators for use as our predictor and training data. Our data encompass a wide range of economic information, from consumer spending habits to federal interest rates. With a wider net, the chances of discovering trends and correlations amongst the data will help inform the studies we can design.

Publicly Available Economic Indicators (Predictor Variables)			**Response Variable**
Commercial and Industrial Loan Value	CPI for All Urban Consumers	Seasonally Adjusted Unemployment Rate	S&P 500 (^GSPC) Closing Price
30-year Fixed Mortgage Rate	Personal Savings Rate	Federal Funds Effective Rate	
Private Housing Permits	Real Disposable Personal Income Per Capita	Industrial Production Index	
Median Home Values	Total Vehicle Sales	U.S. Federal Discount Rate	

Data sourced from reputable providers: Census.gov, Freddiemac.com, FRED.com (Federal Reserve Economic Data), Yahoo Finance

Cleaning **Data**

Prior to modeling, our raw data were collected from many sources and therefore were on different time scales, frequencies, represented value in different scales, and more. Data preprocessing was needed to standardize our data for comparison.



Data_Storage

File Edit View Insert Format Data Tools Extensions Help Last edit was made 4 days ago by Maggie Mano

A1 Month

	A	B	C	D	E	F	G	H	I	
1	Month	Year	Median Home Values	30-year Fixed Mortg	Private Housing Permits (ir	Total Vehicle Sal	Discount Rate (	Personal Savings	Real Disposable Personal Ir	Industrial
2	January	2000	131.59	8.21	107.30	18635.00	5.00	5.00	33.12	
3	February	2000	131.36	8.33	121.80	19401.00	5.24	4.50	33.22	
4	March	2000	132.87	8.24	153.70	18343.00	5.34	4.30	33.28	
5	April	2000	134.13	8.15	138.90	17939.00	5.50	4.80	33.43	
6	May	2000	135.19	8.52	148.90	17943.00	5.71	4.80	33.54	
7	June	2000	136.57	8.29	155.10	17596.00	6.00	4.80	33.60	
8	July	2000	137.26	8.15	129.80	17316.00	6.00	5.10	33.75	
9	August	2000	137.66	8.03	146.80	17531.00	6.00	5.20	33.92	
10	September	2000	138.24	7.91	131.40	18654.00	6.00	4.50	33.92	
11	October	2000	138.69	7.80	134.80	17514.00	6.00	4.80	33.99	
12	November	2000	139.13	7.75	121.00	16634.00	6.00	4.70	33.97	
13	December	2000	139.56	7.38	102.90	16222.00	6.00	4.40	34.00	
14	January	2001	140.15	7.03	117.00	17652.00	5.52	4.90	34.15	
15	February	2001	141.06	7.05	114.30	17826.00	5.00	5.00	34.17	
16	March	2001	142.32	6.95	147.70	17248.00	4.81	5.30	34.23	
17	April	2001	143.56	7.08	148.60	16872.00	4.28	5.10	34.10	
18	May	2001	144.75	7.15	159.80				99	
19	June	2001	145.94	7.16	153.90				95	
20	July	2001	146.76	7.13	140.60				42	
21	August	2001	147.32	6.95	151.40				95	
22	September	2001	147.83	6.82	125.20				62	
23	October	2001	148.41	6.62	140.20				98	
24	November	2001	148.36	6.66	124.40				00	
25	December	2001	148.87	7.07	113.60				03	
26	January	2002	149.36	7.00	115.70				71	
27	February	2002	150.20	6.89	122.50				74	
28	March	2002	151.69	7.01	143.20				73	
29	April	2002	152.95	6.99	156.00	17668.00	1.25	5.50	34.83	
30	May	2002	154.62	6.81	164.20	16197.00	1.25	6.20	34.94	
31	June	2002	156.13	6.65	158.00				02	
32	July	2002	157.15	6.49	159.30				90	
33	August	2002	158.04	6.29	153.70				88	
34	September	2002	158.71	6.09	149.50				89	
35	October	2002	159.42	6.11	162.90				96	
36	November	2002	159.80	6.07	126.80				03	

Cleaned data storage

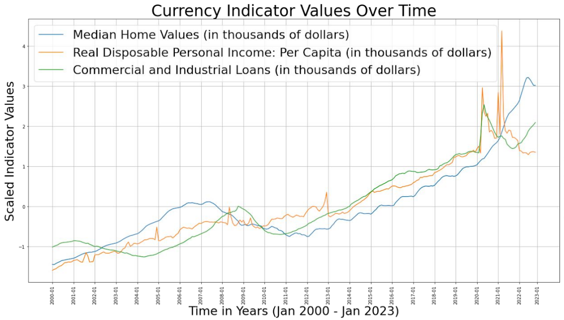
A spreadsheet of our cleaned data storage, with 12 economic indicators realigned on a common time scale.

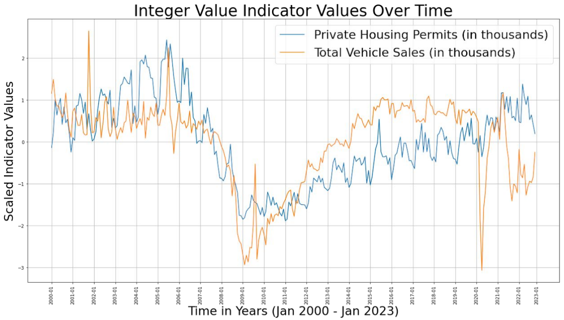
Aligned Timestamps

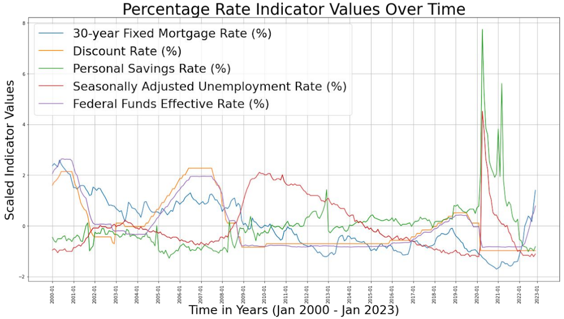
Imputed Missing Values

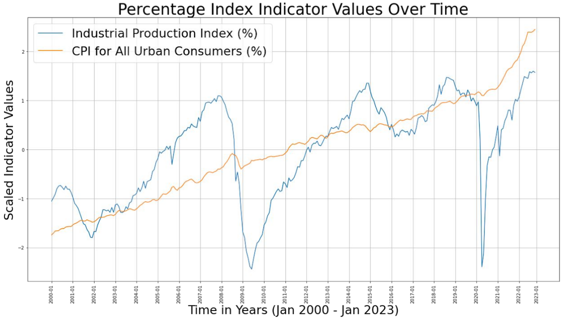
Standardized Currency

Grouping Indicators







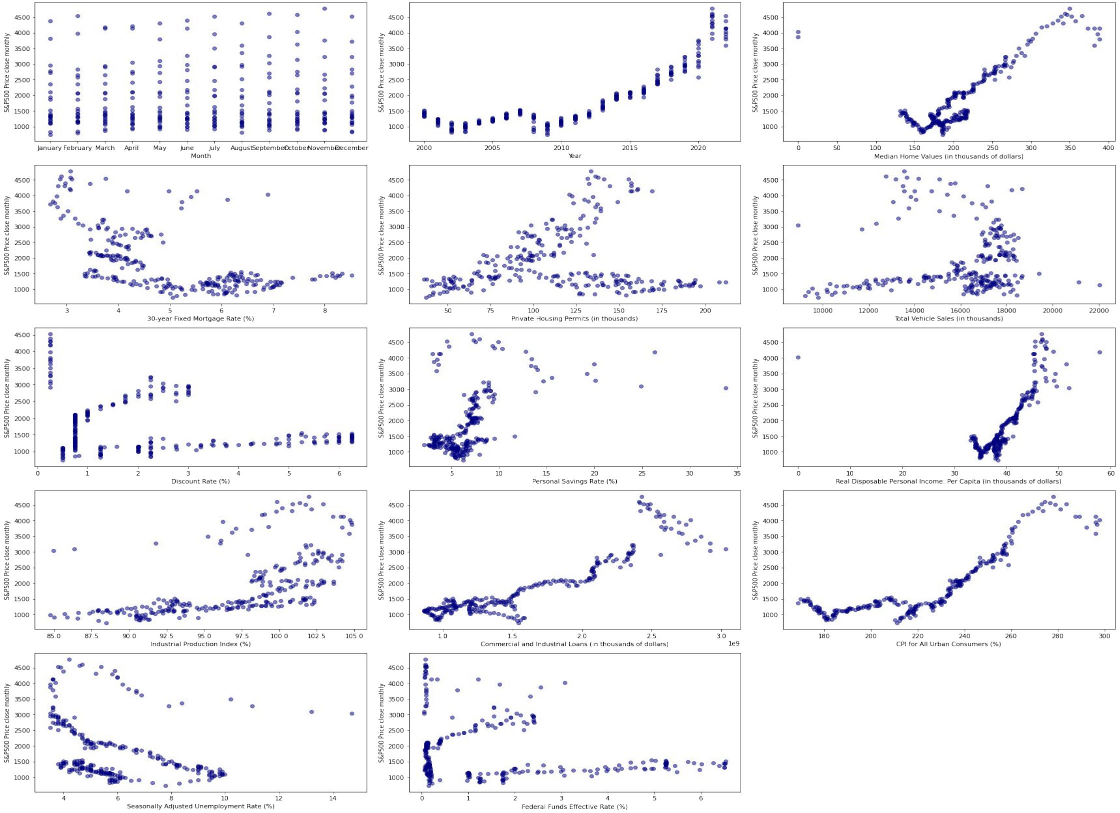


Early Data **Analysis**

From our EDA, compared the indicators with each other and observed interesting correlations:

- Home values, personal disposable income, and real estate loan values were very tightly correlated
- Housing permits issued and vehicle sales are correlated
- Unemployment rate and federal discount rates are negatively correlated
- Industrial production is highly cyclical, and recession periods strongly affect production

Relationship between S&P 500 Price and predictors



Early Data **Analysis**

We also plotted each economic predictor against our response variable. Here, we see a strong relationship between inflation (CPI) and stock performance, with the relationship best fitted by a polynomial regression model.

Additionally, most predictors appear to have a linear relationship with S&P 500 price performance, therefore including those indicators will be beneficial to predicting our S&P 500 price.

With this EDA, we believe a linear and polynomial regression model will best represent our data.

Predicting Recessions with Machine Learning

Evaluating Performance

Mean Absolute Percentage Error

MAPE is a loss function that is calculated by finding the absolute difference between the actual and predicted values, divided by the actual value. The ratios are added for all values and the mean is taken. Overall a robust way of forecasting accuracy score in regression models.

$$\mathrm{MAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$

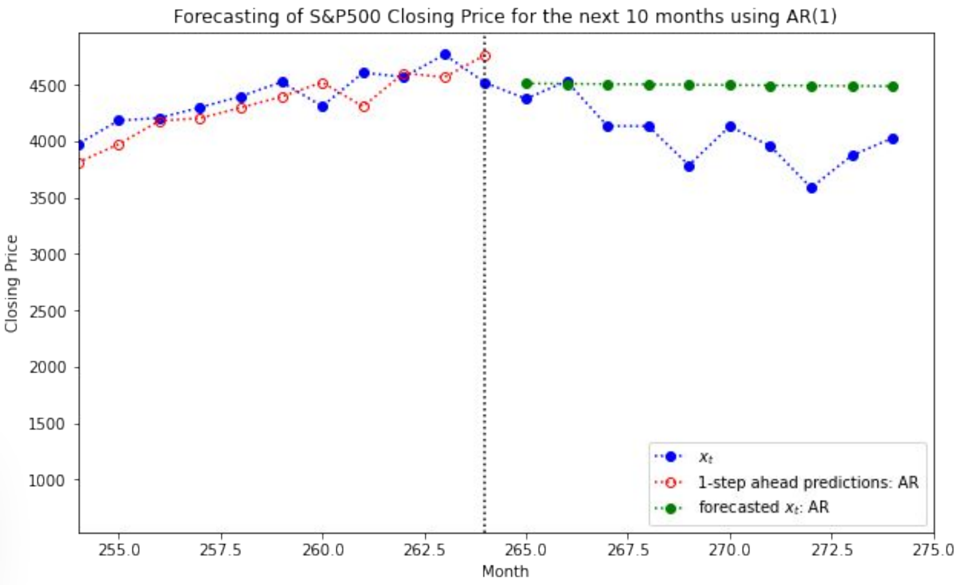
Predicting Recessions with Machine Learning

Time Series Model

First difference, mean squared errors, worse than baseline, simple AR(1) model,

no GARCH / ARCH things,

MAPE: 88.46%
MSE: 263452.63

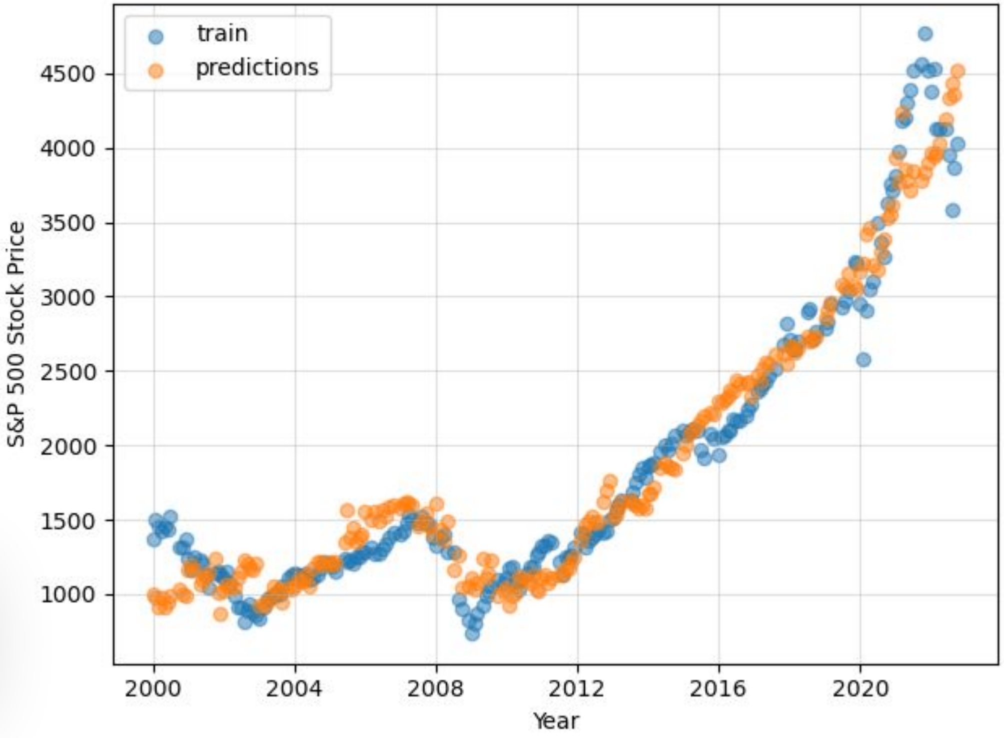


Predicting Recessions with Machine Learning

Linear Regression

Quantitative response variable -> Regression problem. Fit a simple linear regression model to predict stock price using all predictor variables.

91.63% MAPE
42927.78 MSE

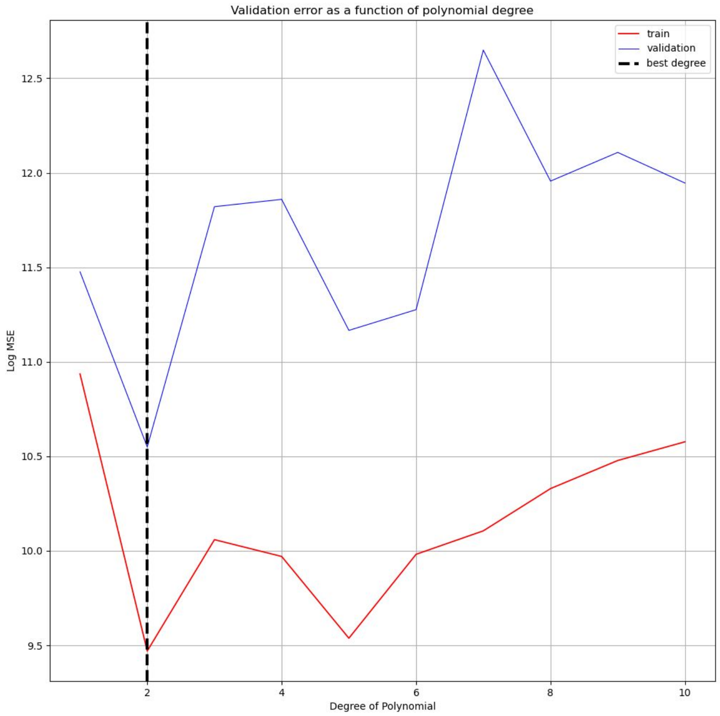


Predicting Recessions with Machine Learning

Polynomial Regression

Expand on SLR, relationship did not appear to be strictly linear. Initially fit polynomial regression, using single validation set we found that the best degree for polynomial regression was 5. Next used CV with 5 folds to confirm best degree; results indicate the best degree was 2.

93.90% MAPE
16399.29 MSE

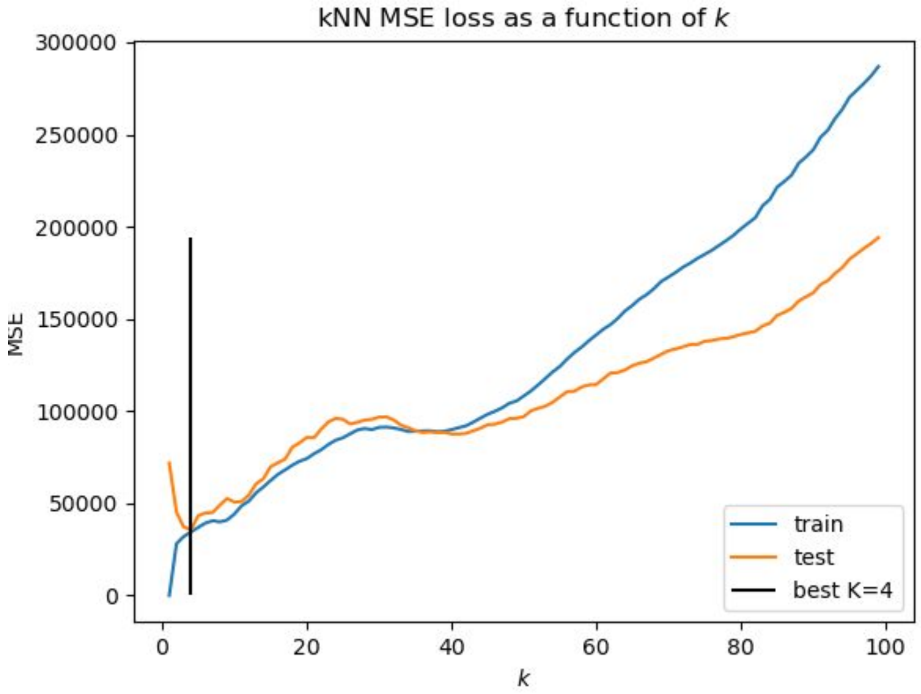


Predicting Recessions with Machine Learning

KNN Regression

Intuition: Nearest time intervals are best for predicting stock price. Fit a KNN model, tuning for our parameter of K neighbors. MSE lowest out of simple regression models. Paradoxically, MSE value better than SLR.

90.43% MAPE
35727.26 MSE

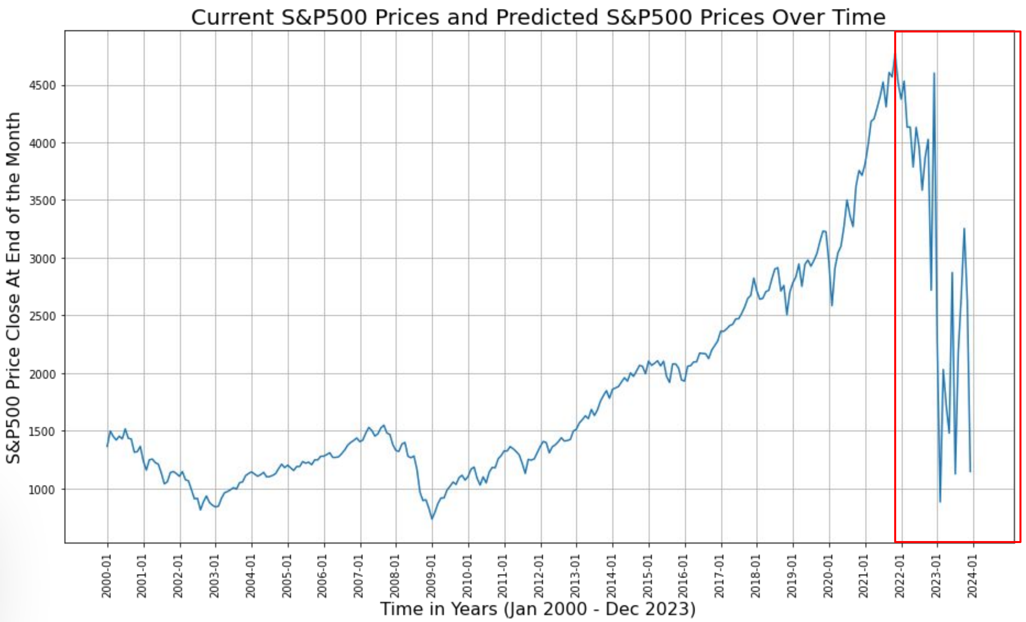


Predicting Recessions with Machine Learning

Random Forest

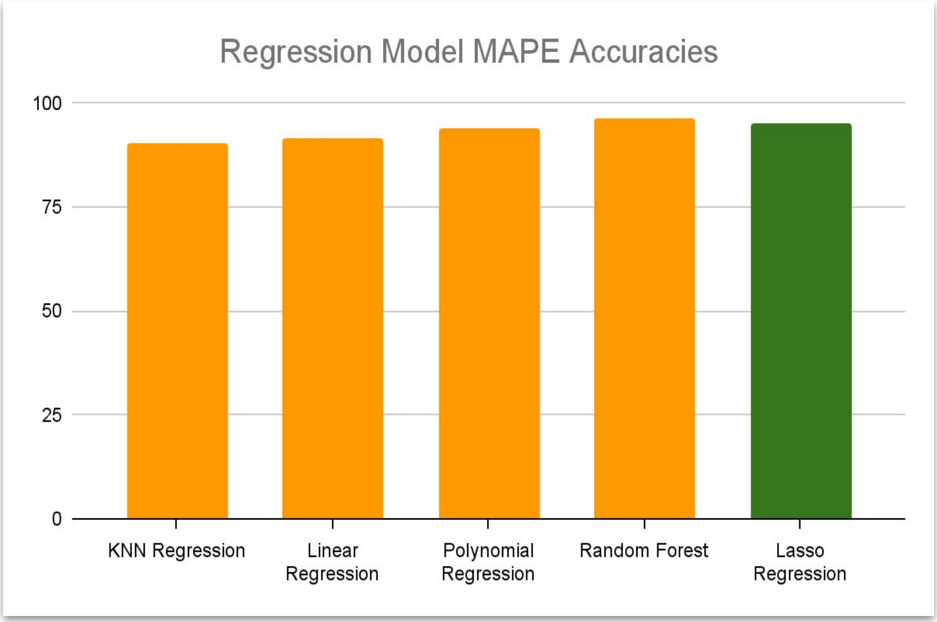
The Random forest model resulted in the lowest MSE score due to it being an ensemble regression method. We did not use it as a main model because Random Forest is unable to extrapolate values outside the training set. This means Random Forest won't be able to formulate data in time series form because it cannot identify trends!

MAPE: 96.33%.
MSE: 7545.561



Predicting Recessions with Machine Learning

Comparing Model **Performances**

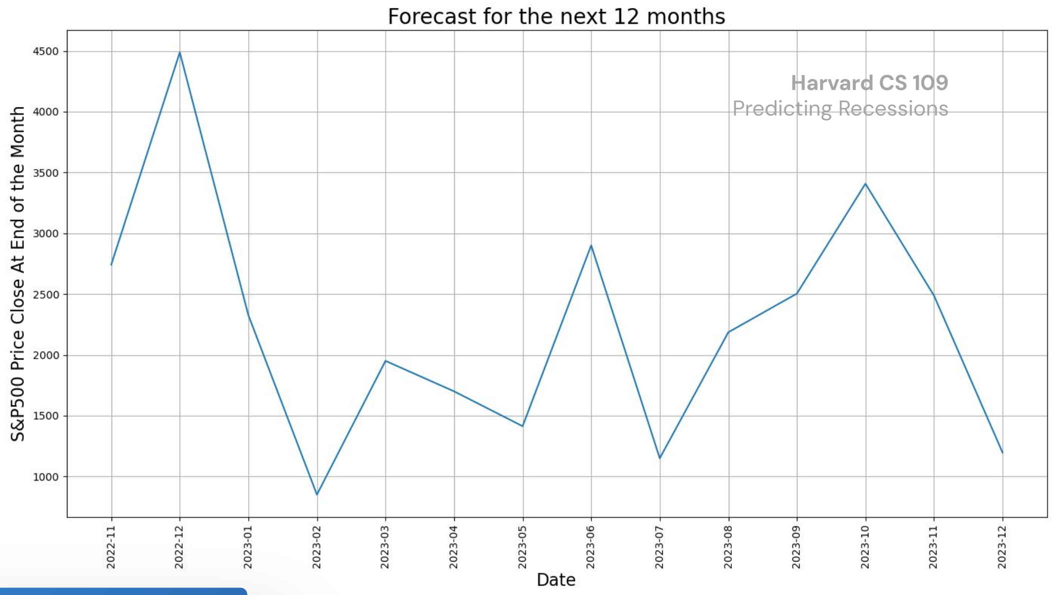


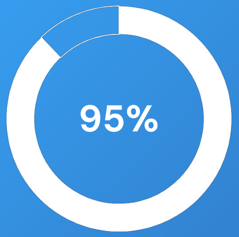
A Comparison of our models

The Lasso and polynomial regression models resulted in the lowest MSE value out of the regression techniques (not including random forest). Although not included in the chart to the right, they have the lowest MAPE accuracies as well.

Predicting Recessions with Machine Learning

Final LASSO Regression Model





Test Accuracy

Sampled with unseen data, this model accurately predicted S&P 500 prices with 95% accuracy.

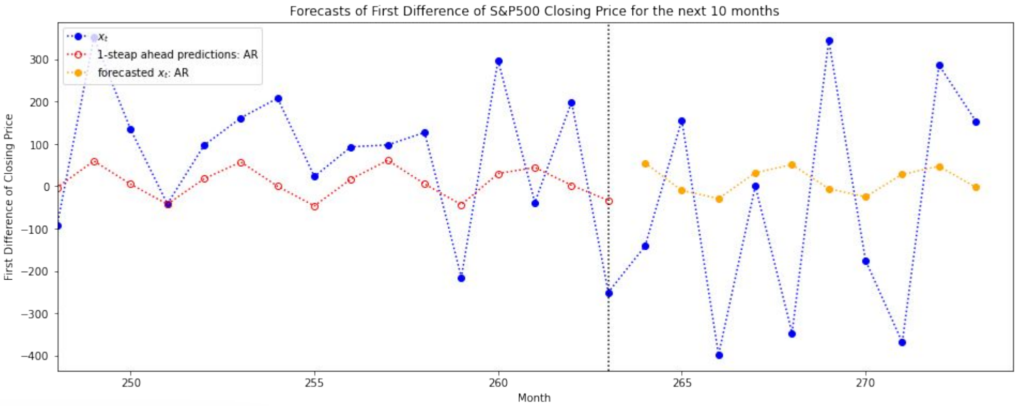
MSE: 13034.38

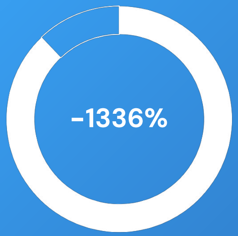
MAPE: 95.23

Given best polynomial regression model, we turned to using LASSO methods to optimize feature selection. Only hyperparameter to fit was alpha level; using 20 fold LASSO CV, best alpha level is 0.1. LASSO method increased overall model performance and robustly handles multicollinearity in our data.

Predicting Recessions with Machine Learning

Final Time Series Model





Test Accuracy

Sampled with unseen data, this model predicted first difference in S&P 500 prices with -1336% accuracy - accuracy is not a suitable measurement here, and the output is also impossible.

Using an ARMA(2, 2)-GARCH(1,1) model to forecast the last 10 data points of the first difference of Closing Price, the MSE was calculated to be 75900.20, a significant improvement from the baseline time series AR(1) model with an MSE value of 263452.63 Test Accuracy values do not make much sense, which is why MSE value is a better standard of comparison here.

Predicting Recession Trends with Machine Learning

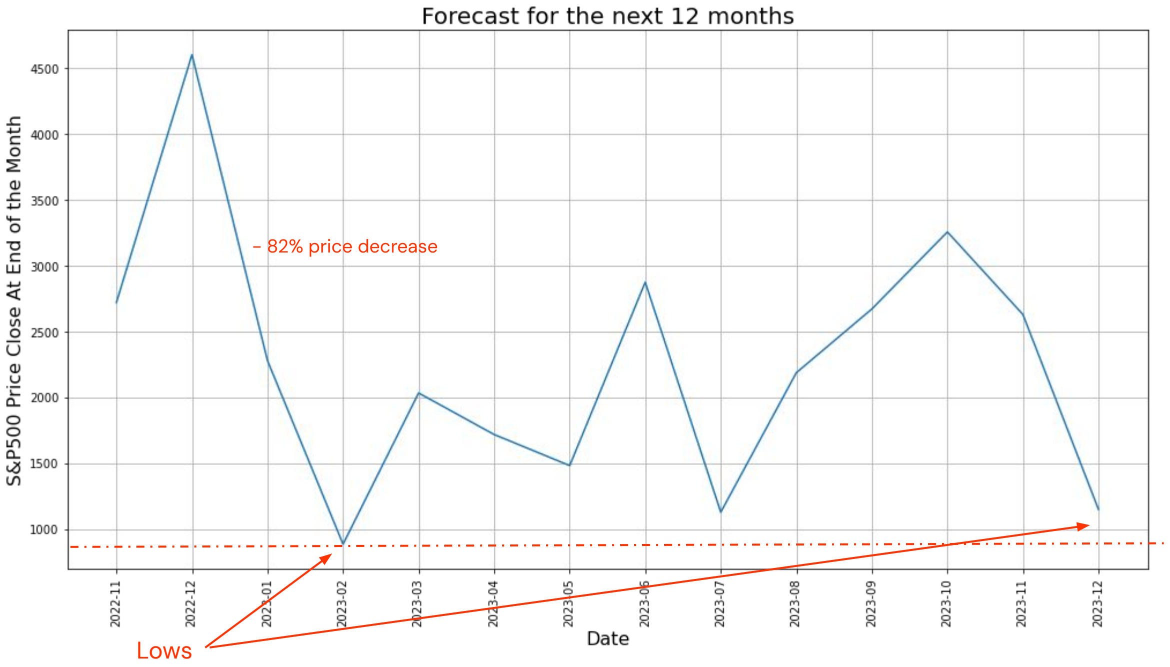
Price Forecast

1 year low - Feb & Dec 2023

~$800/share

Small recovery - Oct 2023

~$3300/share

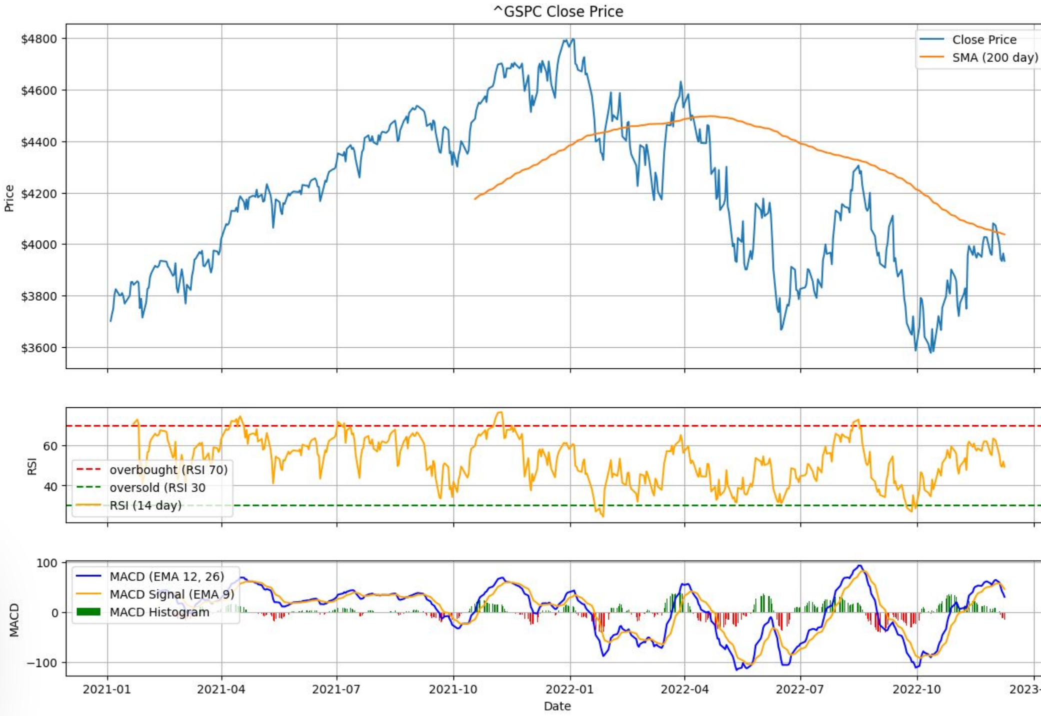


Predicting Recessions with Machine Learning

SMA, RSI, & MACD

As a reference check, we modeled S&P 500 prices with common stock trading technical analysis indicators such as the 200 day Simple Moving Average, the 14 day Relative Strength Index, and a Moving Average Convergence and Divergence indicator. We can S&P 500 price is underneath the SMA curve and being rejected by the SMA line, the RSI is near over brought territory, and MACD is red / negative. The indicators implies the price action will continue to decline, therefore supporting the prediction our machine learning models forecasted.

Confirmation: price will continue to decline



Predicting Recessions with Machine Learning

Takeaways

Economy will continue to decline for next 12 months.

Investment strategies

From our earlier EDA, we observed that recessions highly impact industrial productivity, so we can expect radical strategy shifts from industry

Spending Habits

From our price forecast, we see that the next 12 months will have periods of gain and periods of losses. Consumer spending behavior will also slow down.

Predicting Recessions with Machine Learning

Future Work

Tune model

Time series data is ever evolving and we are limited by the scope of our CS109A knowledge. Other researchers have shown success with neural networks and deep learning models

Explore Tangential Trends

Other useful trends such as home prices, individual stock prices, might be predictable with a similarly tuned model.

Share Results

Creating a library, api endpoints, or even a hosted inference web application to continually predict trends with new time data and share our findings with the world.

Harvard CS 109
Predicting Recessions

Dec 5
2022

Thank You

Presented by
Tony Hua, Faisal Karim, Maggie Mano, Oscar Mercado