

Holiday Gold Rush

Junaid Qamer

Christopher Robinson

Nathalie Kelley

What the heck are we doing? And why are we doing this to ourselves?

- Holiday sales have long been regarded as a predictor of Growth in GDP. We found holiday sales and GDP have a correlation of 98%.
- During times of GDP downturn, investors turn to buying gold as a safeguard.

Using 28 years of past holiday sales data, NYT news sentiment(NLP), GDP raw data, unemployment numbers, gas prices, consumer sentiments, inflation, retail index, and gold returns we use this data to predict 2020 holiday sales. If the holiday sales are high, we can assume there will be growth in GDP, therefore gold prices will drop and vice versa. Using this information we will build a trading algorithm for purchasing gold depending on our predictions.

Gathering Data

- NYT API to evaluate 28 years of holiday sales sentiment in the news.
- GDP data from bea.
- Retail sales numbers, gas prices for FRED- St. Louis
- Michigan University for consumer sentiment

PROCEDURE

Using 28 years of past holiday sales data, NYT news sentiment(NLP), GDP raw data, unemployment numbers, gas prices, consumer sentiments, inflation, retail index, and gold returns we use this data to predict 2020 holiday sales. If the holiday sales are high, we can assume there will be growth in GDP, therefore gold prices will drop and vice versa.

1. Using the NYT API, we gathered 28 years of holiday related articles from November 92- 2019 to December 92-2019. We then ran them through the Vader sentiment analyzer to analyze sentiment and built a data frame.
2. We joined our dataframes NYT sentiment and features together creating Main_df. We then ran a Hodrick-Prescott filter on columns Retail sales, CPI, and GDP growth to eliminate trend and keep the noise. A new Dataframe was created using this called MAIN_df. We then built 5 machine learning models, linear regression, Decision tree regressor, random forest regressor, gradient boost regressor, and LSTM. Our y is retail sales (what we are trying to predict) and the rest X were our features.
3. After analyzing the results of each model we need to evaluate the best working model using out of sample data. The original model used in- sample data but for our purpose of predicting holiday numbers for 2020 we needed to run the model using out of sample data. To do this we shifted our y (retail sales), dropped 2 columns from our features and ran it through our model.
4. To get this year's y, we plugged in our X in our best predictive model(shifted y). After doing this, we put the trend back into our number and obtain our 2020 predictions.

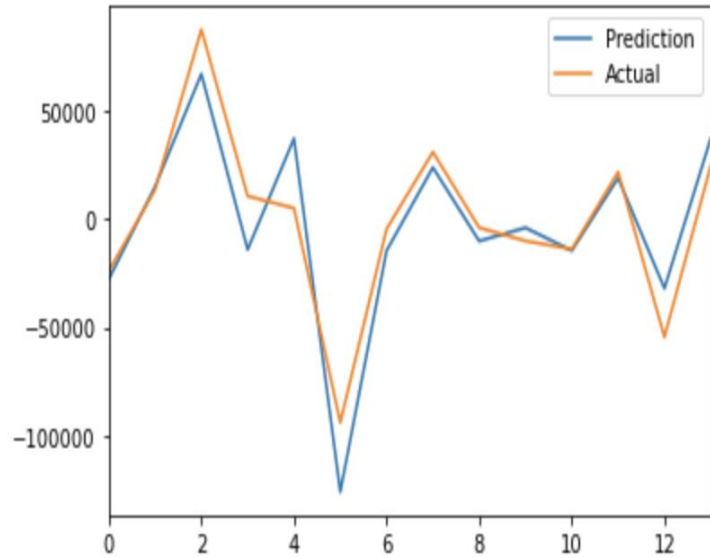
MACHINE LEARNING MODELS

LINEAR REGRESSION

DECISION TREE

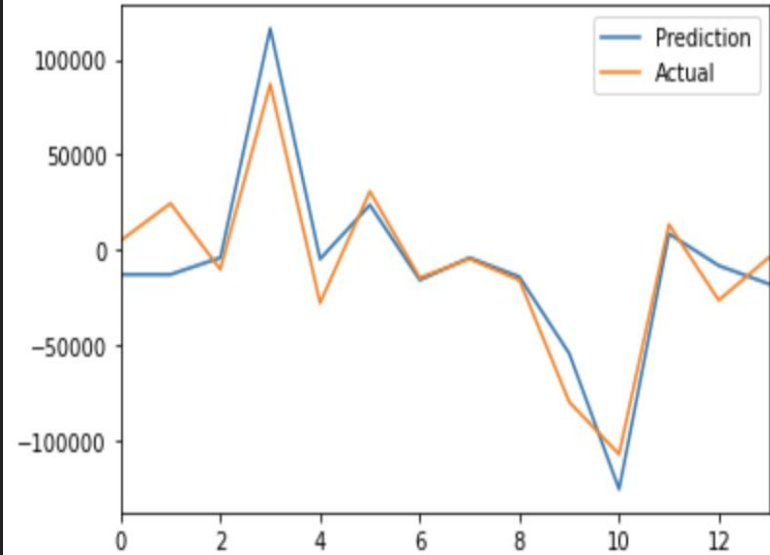
ORIGINAL MODEL

<AxesSubplot:>



PREDICTIVE MODEL

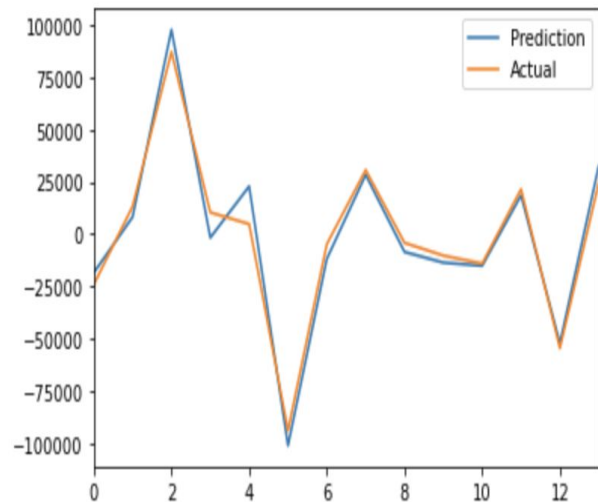
<AxesSubplot:>



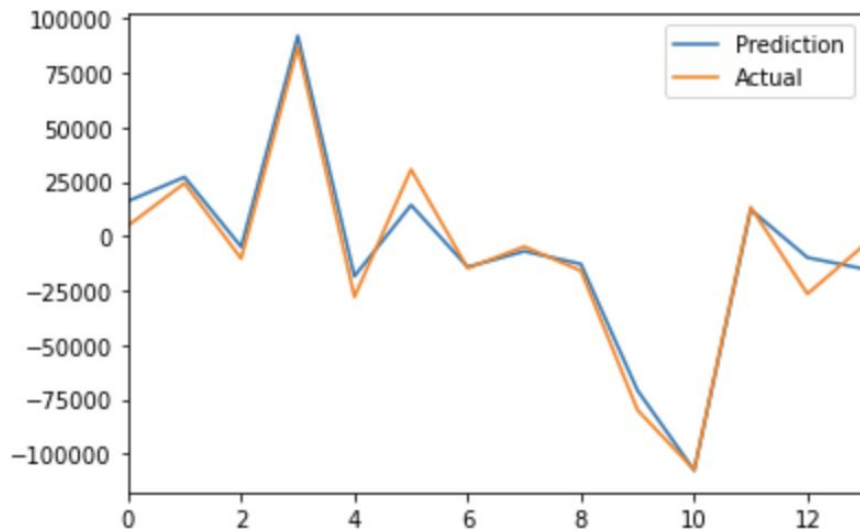
RANDOM FOREST

ORIGINAL MODEL

<AxesSubplot:>



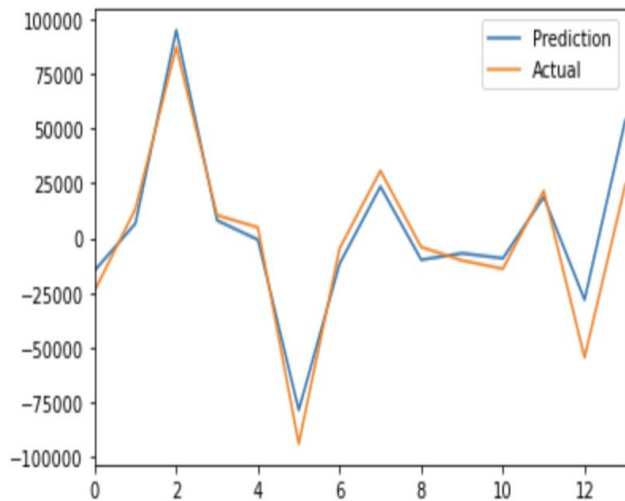
<AxesSubplot:>



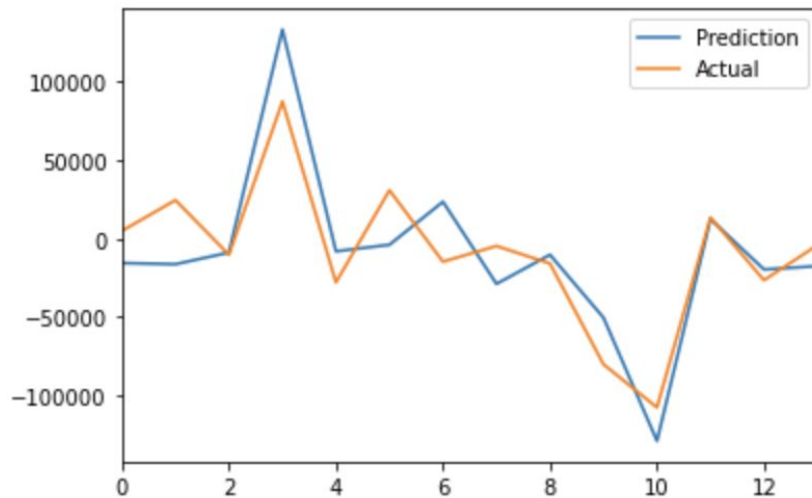
GRADIENT BOOST

ORIGINAL MODEL

<AxesSubplot:>



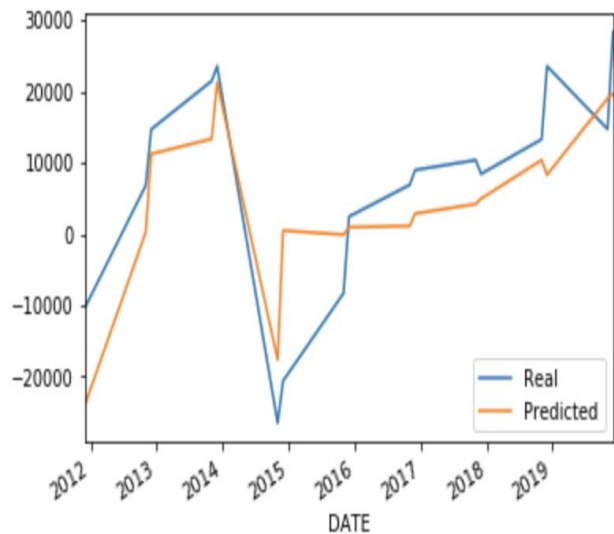
<AxesSubplot:>



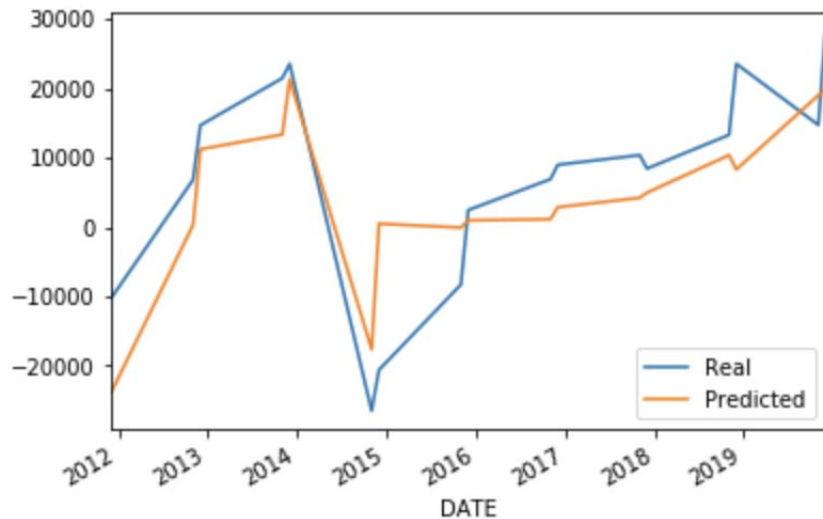
LSTM

ORIGINAL

<matplotlib.axes._subplots.AxesSubplot at 0x7fbd4c1c3b10>



<matplotlib.axes._subplots.AxesSubplot at 0x7fbd504bb810>



CORRELATION CHART

	NYT_sentiment	gas_prices	consumer_sentiments	unemployment_rate	CPI	retail_index	Consumer_debt_service	GDP_growth	sales
NYT_sentiment	1.000000	-0.359748	0.141274	0.041893	-0.183963	-0.019460	-0.114457	0.345912	-0.197558
gas_prices	-0.359748	1.000000	-0.491512	0.325284	0.682532	-0.553542	-0.302524	-0.367889	0.599383
consumer_sentiments	0.141274	-0.491512	1.000000	-0.772300	-0.201631	0.563745	0.267027	0.732948	-0.003984
unemployment_rate	0.041893	0.325284	-0.772300	1.000000	-0.021729	-0.540780	-0.483871	-0.381928	-0.243156
CPI	-0.183963	0.682532	-0.201631	-0.021729	1.000000	-0.151752	-0.059089	-0.324210	0.965211
retail_index	-0.019460	-0.553542	0.563745	-0.540780	-0.151752	1.000000	0.417569	0.143131	-0.027074
Consumer_debt_service	-0.114457	-0.302524	0.267027	-0.483871	-0.059089	0.417569	1.000000	-0.060483	0.076767
GDP_growth	0.345912	-0.367889	0.732948	-0.381928	-0.324210	0.143131	-0.060483	1.000000	-0.179655
sales	-0.197558	0.599383	-0.003984	-0.243156	0.965211	-0.027074	0.076767	-0.179655	1.000000
GDP	-0.181612	0.590662	-0.054932	-0.204064	0.969799	-0.033127	0.128035	-0.242846	0.988492
close	-0.350953	0.884859	-0.397444	0.331468	0.682911	-0.409803	-0.497224	-0.371098	0.576120

CONCLUSION

- THERE IS A 98% CORRELATION BETWEEN HOLIDAY SALES AND GDP.
- THERE IS ALSO A NEGATIVE CORRELATION BETWEEN RETAIL SALES NUMBERS AND GOLD PRICES AS WELL A NEGATIVE CORRELATION BETWEEN GOLD PRICES AND GDP.
- WE ARE PREDICTING AN ESTIMATED 486 BILLION DOLLARS IN NOVEMBER 2020 RETAIL SALES AND 497 BILLION DOLLARS IN DECEMBER 2020 RETAIL SALES. ESTIMATING A TOTAL OF BILLION DOLLARS IN HOLIDAY SALES FOR 2020.
- November 2019 retail sales were 461 billion dollars.
- December 2019 retail sales were 460 billion dollars.