U.S. Gas Price Prediction

IE 434: Deep Learning

By: Ruoyu Sun, Narathip Rodwarna and Liam Harding

Agenda

- Motivation
- Data Characteristics & Statistics
- Model Comparison
- Challenges
- Conclusion

Why Predict Gas Price?

- Gas prices are uncertain with a general increasing trend.
- Gas price predictions are crucial for consumer budgeting and travel planning to allocate resources effectively.
- Industries reliant on transportation and logistics benefit from accurate predictions to optimize operations.
- Financial markets are influenced by gas prices, prompting investors to make informed decisions.
- Governments use gas price forecasts for economic and environmental policy planning, while anticipating global events and geopolitical factors that impact energy markets.

Data Characteristics & Statistics

Data Source

The U.S. Energy Information Administration has documented the Gasoline Price and Exportation data, encompassing 1142 records spanning from November 9, 2001, to September 22, 2023.

Date for US Imports and Exports	Weekly U.S. Exports of Crude Oil and Petroleum Products (Thousand Barrels per Day)	Weekly U.S. Exports of Crude Oil (Thousand Barrels per Day)	Weekly U.S. Exports of Total Petroleum Products (Thousand Barrels per Day)	Date for Retail Gas Price	Weekly U.S. All Grades All Formulations Retail Gasoline Prices (Dollars per Gallon)	Season
9-Nov-01	10772	9639	1133	12-Nov-01	1.224	autumn
16-Nov-01	10243	8879	1364	19-Nov-01	1.208	autumn
23-Nov-01	9576	8187	1389	26-Nov-01	1.168	autumn
30-Nov-01	11170	9856	1314	3-Dec-01	1.149	autumn
7-Dec-01	9885	8966	919	10-Dec-01	1.136	winter
14-Dec-01	10189	8772	1417	17-Dec-01	1.101	winter
21-Dec-01	9826	8657	1169	24-Dec-01	1.113	winter
28-Dec-01	9360	8132	1228	31-Dec-01	1.137	winter
4-Jan-02	10100	8628	1472	7-Jan-02	1.152	winter
11-Jan-02	9650	8483	1167	14-Jan-02	1.152	winter
18-Jan-02	10136	9153	983	21-Jan-02	1.146	winter

Data Characteristics & Statistics

Data Preprocessing

Three features are selected from the set of available features which are

1) **Date for US Imports and Exports**

   This is processed by filter out only year.

2) **Weekly U.S. Exports of Crude Oil and Petroleum Products (Thousand Barrels per Day)**

3) **Season**

   String information is converted into integer. We use “0, 1, 2, 3” to represent “spring, summer, autumn, winter” respectively.

	Year	Exports	Season	Prices (Dollars per Gallon)
0	2001	10772	2	1.224
1	2001	10243	2	1.208
2	2001	9576	2	1.168
3	2001	11170	2	1.149
4	2001	9885	3	1.136
...	...	...	...	...
1137	2023	-1684	1	3.931
1138	2023	-2593	2	3.925
1139	2023	431	2	3.941
1140	2023	-2290	2	4.001
1141	2023	-1706	2	3.963

1142 rows × 4 columns

Data Characteristics & Statistics

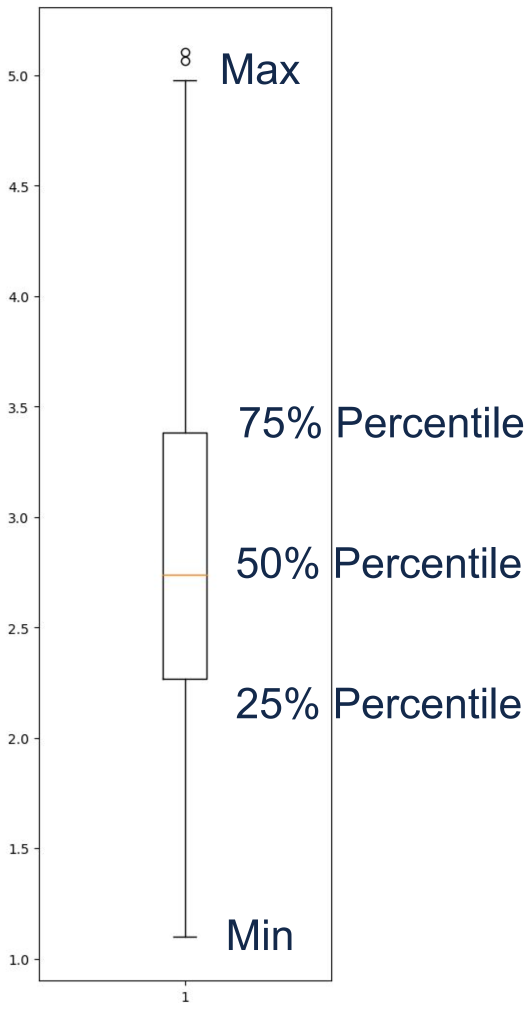
Descriptive Statistics

Gasoline Price (USD per Gallon)

- A Number of Record: 1,142
- Mean: 2.767
- Median: 2.741
- Standard Deviation(SD): 0.755
- Range
  - Min 1.101
  - Max 5.107
- Percentile

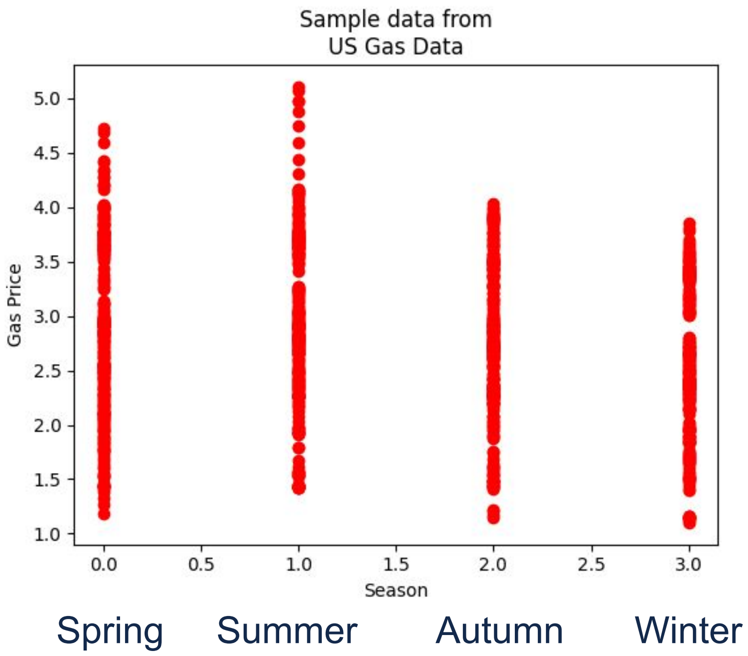
  The 25% 50% 75% percentile of dataset
  - 0.25 2.267
  - 0.50 2.741
  - 0.75 3.385

Gas Price (USD/Gallon)



Data Characteristics & Statistics

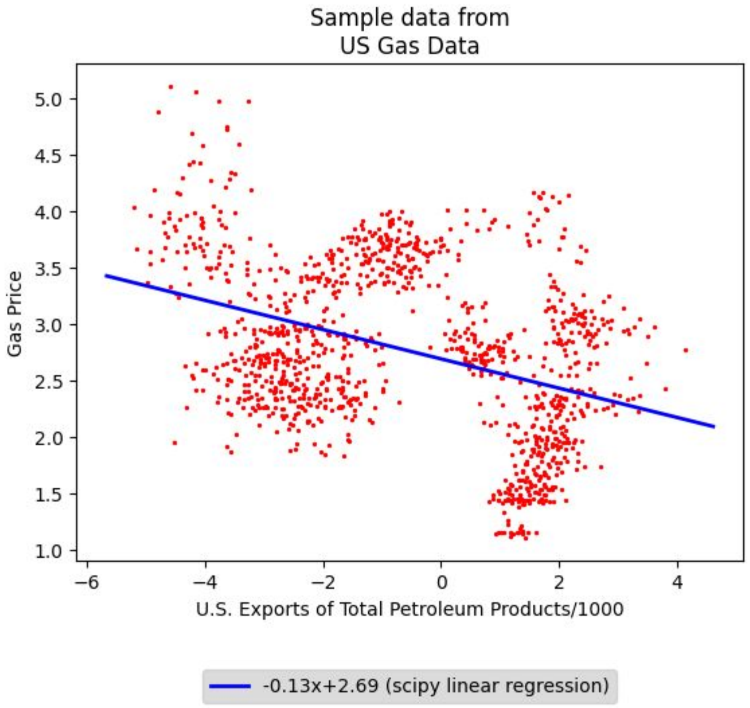
Descriptive Statistics



- The maximum price occurs in Summer and Summer also has the widest range of price
- Autumn and Winter have lower gasoline prices compared to Summer and Spring.

Data Characteristics & Statistics

Descriptive Statistics



Looking at the the data set there is a general negative trend when comparing gas price to U.S. Exports of Total Petroleum but for the U.S. Export of Total Petroleum products the variance is relatively high with a of 2.1 or higher. In the graph, there are outliers which increase the overall loss and accuracy of the neural network.

Model Comparison

There are 3 models studied in this project

1. Polynomial Regression (Baseline)
2. Convolutional Neural Network (CNN)
3. Recurrent Neural Network (RNN)

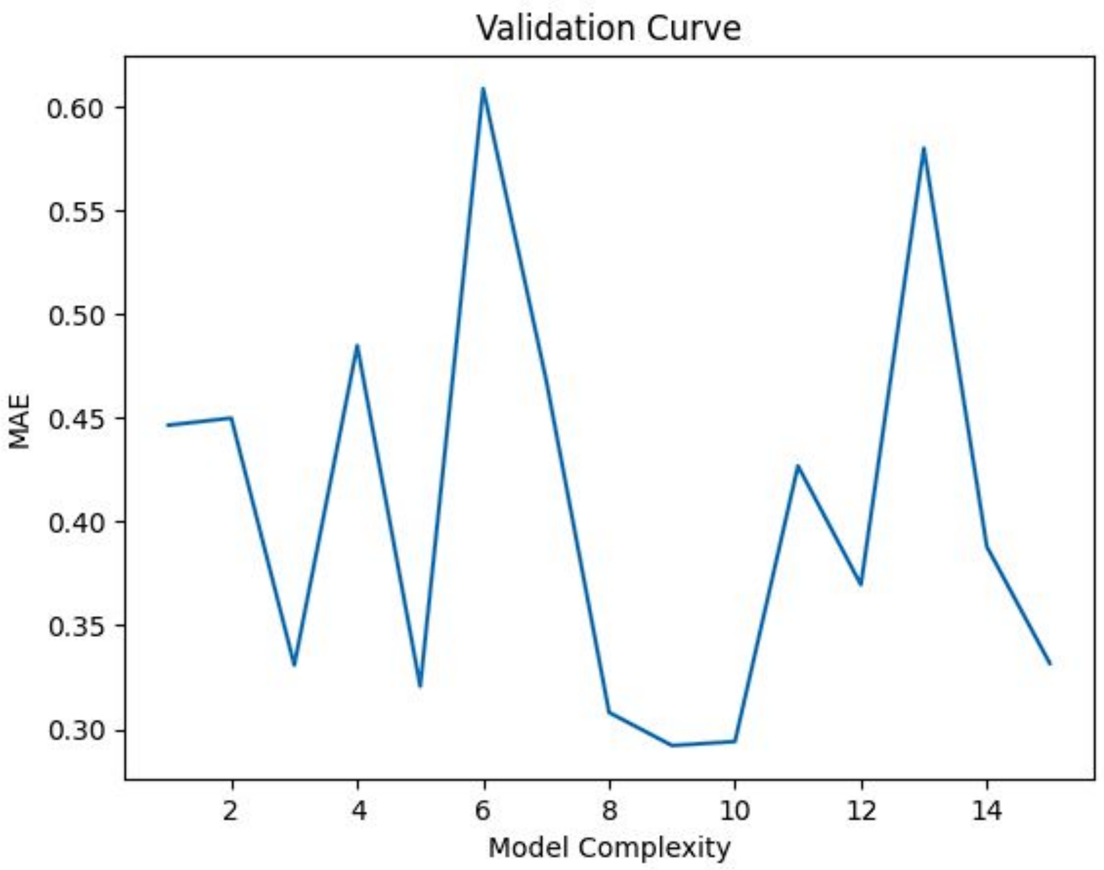
Baseline Model

Polynomial Regression

- By using Polynomial Regression with different degrees of polynomial, 9th degree provides the lowest Mean Average Error(MAE).
- Mean Average Error(MAE) and Mean Squared Error(MSE) on test dataset are 0.34 USD and 0.19 USD respectively.

Note : Data is divided into 3 parts

1.Training 70% 2.Validation 15% 3.Test 15 %



Convolutional Neural Network (CNN)

Hyperparameter Tuning

The model is tuned with these set of hyperparameters

Number of Layer:	3 and 4
Optimizer:	Adam, Stochastic Gradient Descent(SGD) and Root Mean Squared Propagation(RMSprop)
Learning Rate:	0.0001-0.1

Number of Neuron in each Layer : 4,8,16,32

Convolutional Neural Network (CNN)

The Best Case Parameters

```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Linear-1                  [-1, 8]              32
            Linear-2                  [-1, 8]              72
            Linear-3                 [-1, 32]             288
            Linear-4                  [-1, 1]              33
================================================================
Total params: 425
Trainable params: 425
Non-trainable params: 0
```

Parameter

Number of Layer:	4
Optimizer:	Adam
Learning Rate:	0.0006

Result

The prediction on test dataset has MAE of 0.6397 USD.

Recurrent Neural Network (RNN)

Time Series Forecasting with LSTM

x(date)		y(price)	
0	2001-11-09	0	1.224
1	2001-11-16	1	1.208
2	2001-11-23	2	1.168
3	2001-11-30	3	1.149
4	2001-12-07	4	1.136
	...		...
1137	2023-08-25	1137	3.931
1138	2023-09-01	1138	3.925
1139	2023-09-08	1139	3.941
1140	2023-09-15	1140	4.001
1141	2023-09-22	1141	3.963

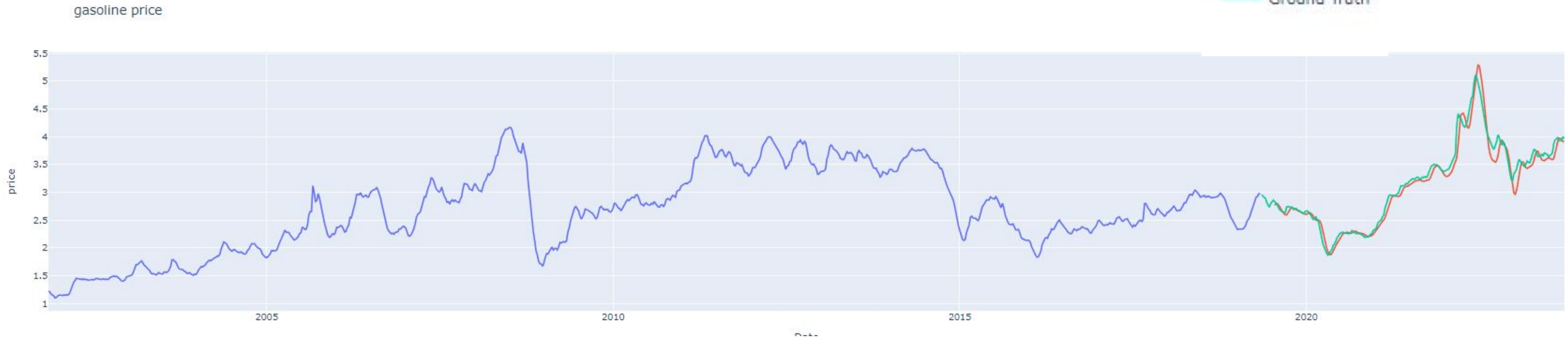
Key Idea: Use previous gasoline price data to predict future price which is also known as Time Series Forecasting

Hyperparameter: Number of previous price used

Dataset: 80% Training data, 20% Test data

Recurrent Neural Network (RNN)

Result



- 10 Lookback data points can generate good predictions with MAE 0.1 USD.
- Although RNN can predict the price data well, the model can be complex and may need longer time for training given different types of data. The input data are time-series data so RNN provides good predictions for this set of sequential data.

Model Comparison

Model	Architecture	Features	Error [USD]
Polynomial Regression	Degree 9 Polynomial	Year, Weekly U.S. Exports, Season	0.34
CNN	4-Layer Neural Network	Year, Weekly U.S. Exports, Season	0.64
RNN	LSTM-based model with 10 lookback data	Gas Price	0.10

- RNN can outperform baseline by using only 1 feature while CNN does not perform well as shown by error in the table.
- Gasoline price is time series dataset. It is particularly suitable to RNN which can predict the future data points well even without other features.

Challenges and Lessons Learned

Challenges

- Technical Challenges
  - Notable technical challenges arise from the limitation of a dataset with only three features, leading to difficulties in achieving accurate predictions.
- Risk of Overfitting
  - The constrained dataset raises concerns about overfitting, where a model may perform well in training but struggle with unseen data during testing.
- Trade-off Management
  - There is a crucial need to manage the trade-off between the simplicity imposed by a limited feature set and the risk of overfitting for improved model reliability.

Challenges and Lessons Learned

Lessons Learned

- Model Selection Impact
  - The project highlights the substantial impact of model selection on performance, with models like Recurrent Neural Network (RNN) and Linear Regression showing more promising results compared to CNN. Careful model choice is crucial for favorable outcomes.
- CNN Suitability
  - Convolutional Neural Networks (CNNs) may not be the most suitable choice for datasets with limited features or unclear patterns.
- RNN Suitability
  - Good performance on predicting time series data, but it may not be efficient for datasets characterized by unequally spaced time series data or data that is not inherently ordered.

Conclusion

- The Recurrent Neural Network (RNN) can achieve gasoline price prediction with an error margin of 0.1 USD per gallon.
- The most important feature is the previous price, exhibiting substantial impact and notable results.
- Stakeholders can leverage these predictions to make well-informed decisions regarding activities associated with gasoline prices, including:
  - Business and Investor
    - Informed investment decisions in energy-related sectors.
  - Government and Industry Impact
    - Facilitates economic planning and policy adjustments.
  - Consumer
    - Informed decision-making on optimal times for vehicle refueling.


ILLINOIS