Gasoline Price Forecasting

DSA 9 Team 7

by Asadullah Qamar, Lucas Moy, Sidi Zainul, Nurul Afeeqah

11:30AM, 11th August 2022

Profit from Gasoline can be so much more!!

PETRONAS Gas Bhd's (PetGas) net profit for the first quarter (1Q) ended March 31, 2022, slipped 20% year-on-year (YoY) to RM410.6 million from RM516.4 million profit posted a year ago for the same period due to lower utility margins following higher fuel gas price and higher operating costs at its gas processing, transportation and regasification segments.

Revenue for the quarter increased 9% YoY to RM1.45 billion mainly driven by higher revenue from the utilities segment as a result of higher product prices and higher electricity sales volumes recorded, PetGas stated in a filing to Bursa Malaysia today.

Earning per share for the quarter was 20.75 sen and the company declared a first interim dividend of 16 sen per ordinary share.

Based on the headline

1. PetGas profit decreased by **RM105.9 Million** first quarter 2022 comparing first quarter 2021.
2. Mainly due to cost of price increase in higher fuel gas price and higher operating costs.
3. Can we avoid this losses if we **anticipate** early high fuel price changes?

How might we improve profit from gasoline?

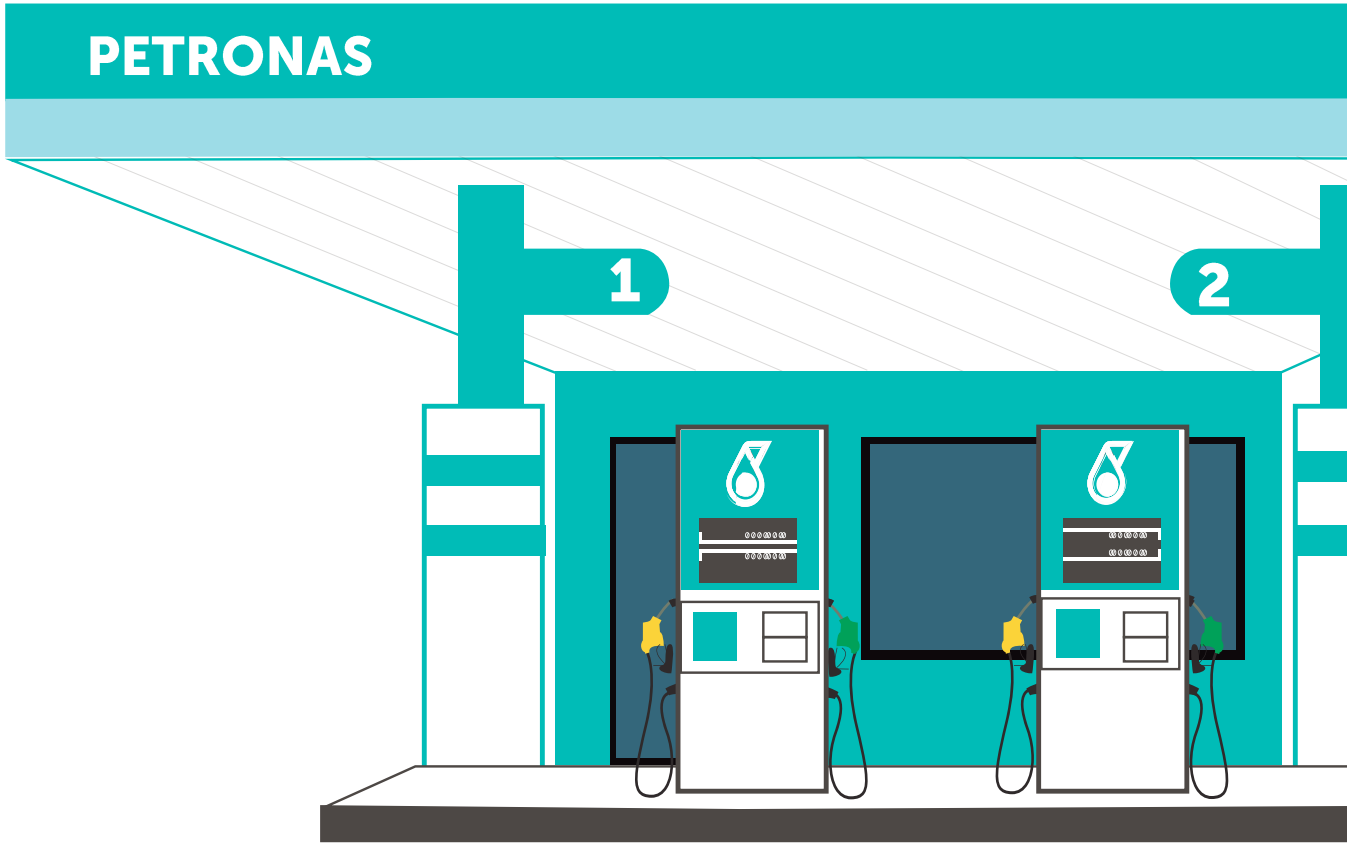
- Planning for price changes.
- Keeping customers satisfied.

How we can help.

1. Predict the prices accurately for at least the first 7 days.
2. Predict further prices after that with a reasonable accuracy.
3. Convey which sentiments to look out for and how they can influence the price.
4. Suggest what times could be difficult and require mitigation plans.

Content

1. Problem Statement & Use Cases
2. What We Found
3. What We Did
4. How We Did It
5. What We Learn
6. What Next

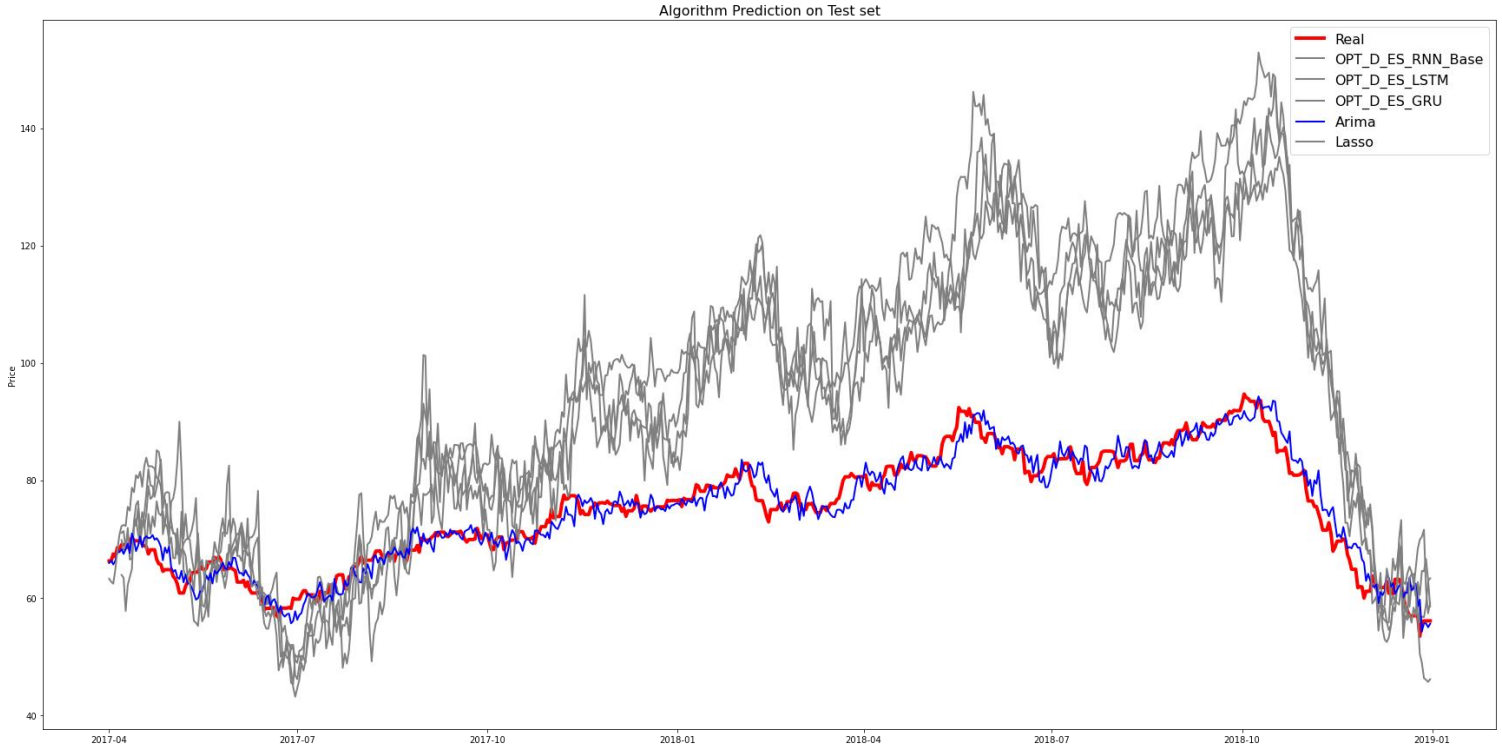


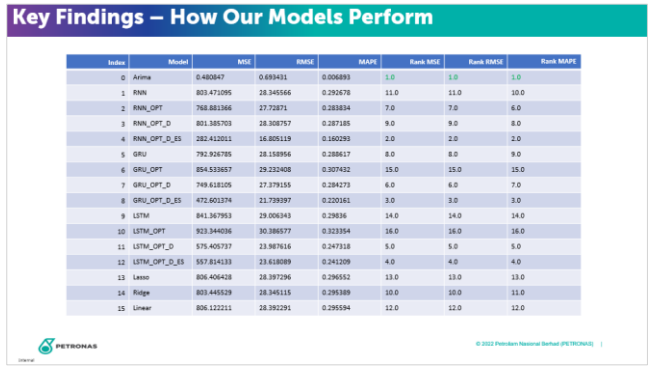
What We Found

PETRONAS

Key Findings – Our Models

Comparing all our models performance





Index	Model	MSE	RMSE	MAPE	Rank MSE	Rank RMSE	Rank MAPE
0	Arima	0.480847	0.693431	0.006893	1.0	1.0	1.0
1	RNN	803.471095	28.345566	0.292678	11.0	11.0	10.0
2	RNN_OPT	768.881366	27.72871	0.283834	7.0	7.0	6.0
3	RNN_OPT_D	801.385703	28.308757	0.287185	9.0	9.0	8.0
4	RNN_OPT_D_ES	282.412011	16.805119	0.160293	2.0	2.0	2.0
5	GRU	792.926785	28.158956	0.288617	8.0	8.0	9.0
6	GRU_OPT	854.533657	29.232408	0.307432	15.0	15.0	15.0
7	GRU_OPT_D	749.618105	27.379155	0.284273	6.0	6.0	7.0
8	GRU_OPT_D_ES	472.601374	21.739397	0.220161	3.0	3.0	3.0
9	LSTM	841.367953	29.006343	0.29636	14.0	14.0	14.0
10	LSTM_OPT	923.344036	30.386577	0.323354	16.0	16.0	16.0
11	LSTM_OPT_D	575.405737	23.987616	0.247318	5.0	5.0	5.0
12	LSTM_OPT_D_ES	557.814133	23.618089	0.241209	4.0	4.0	4.0
13	Lasso	806.406428	28.397296	0.296552	13.0	13.0	13.0
14	Ridge	803.445529	28.345115	0.295389	10.0	10.0	11.0
15	Linear	806.122211	28.392291	0.295594	12.0	12.0	12.0

Key Findings – Our Value Proposition

First Week of 2019

Date	ARIMAX
01/01/2019	58.5117
02/01/2019	59.6563
03/01/2019	54.1998
04/01/2019	55.6266
05/01/2019	55.6243
06/01/2019	54.9596
07/01/2019	55.5461

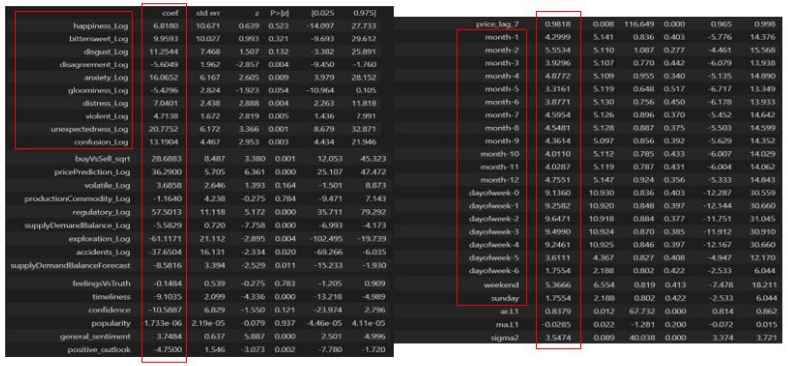
Margin of Error: +- 1.7157

First Month of 2019

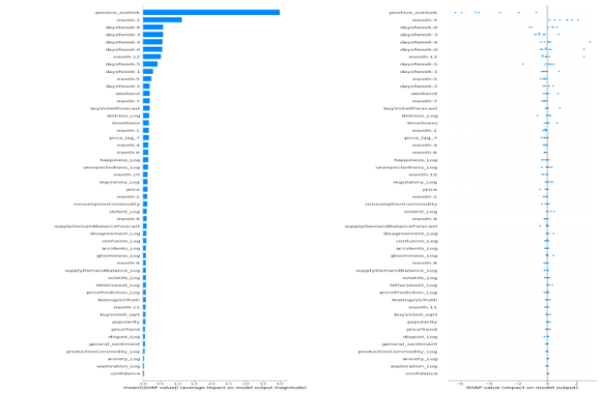
Date	ARIMAX
01/01/2019	64.053914
02/01/2019	62.903297
03/01/2019	62.137733
.	.
.	.
.	.
31/01/2019	55.546106

Margin of Error: +- 2.0826

Interpreting the Feature Significance



Key Findings – What some features that algorithm favours

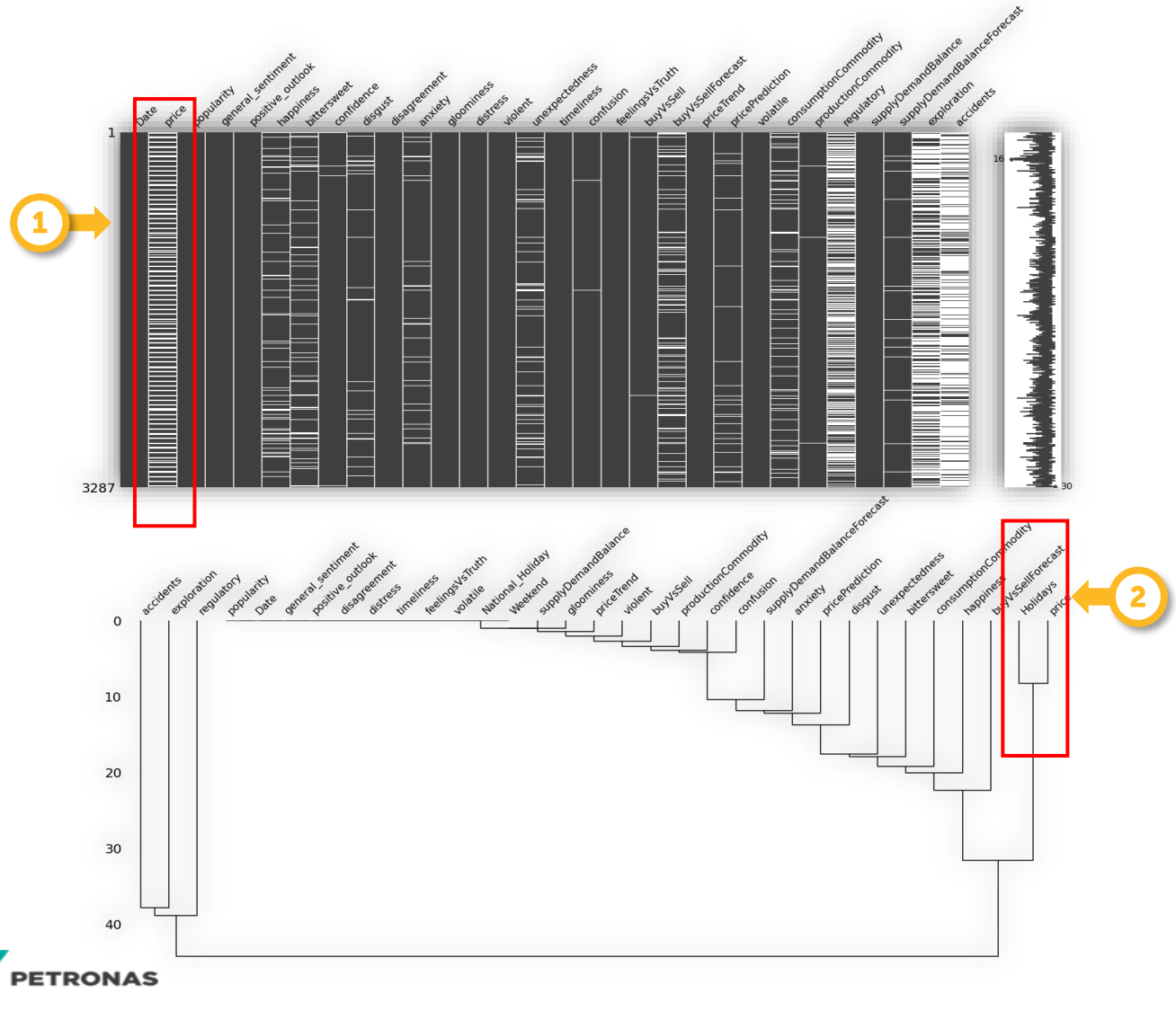


1. DL Model learn the most from *positive outlook* variable
2. Positive outlook sentiment have negative impacts toward price as in when price increase.

What We Did

PETRONAS

Missing Values



PETRONAS

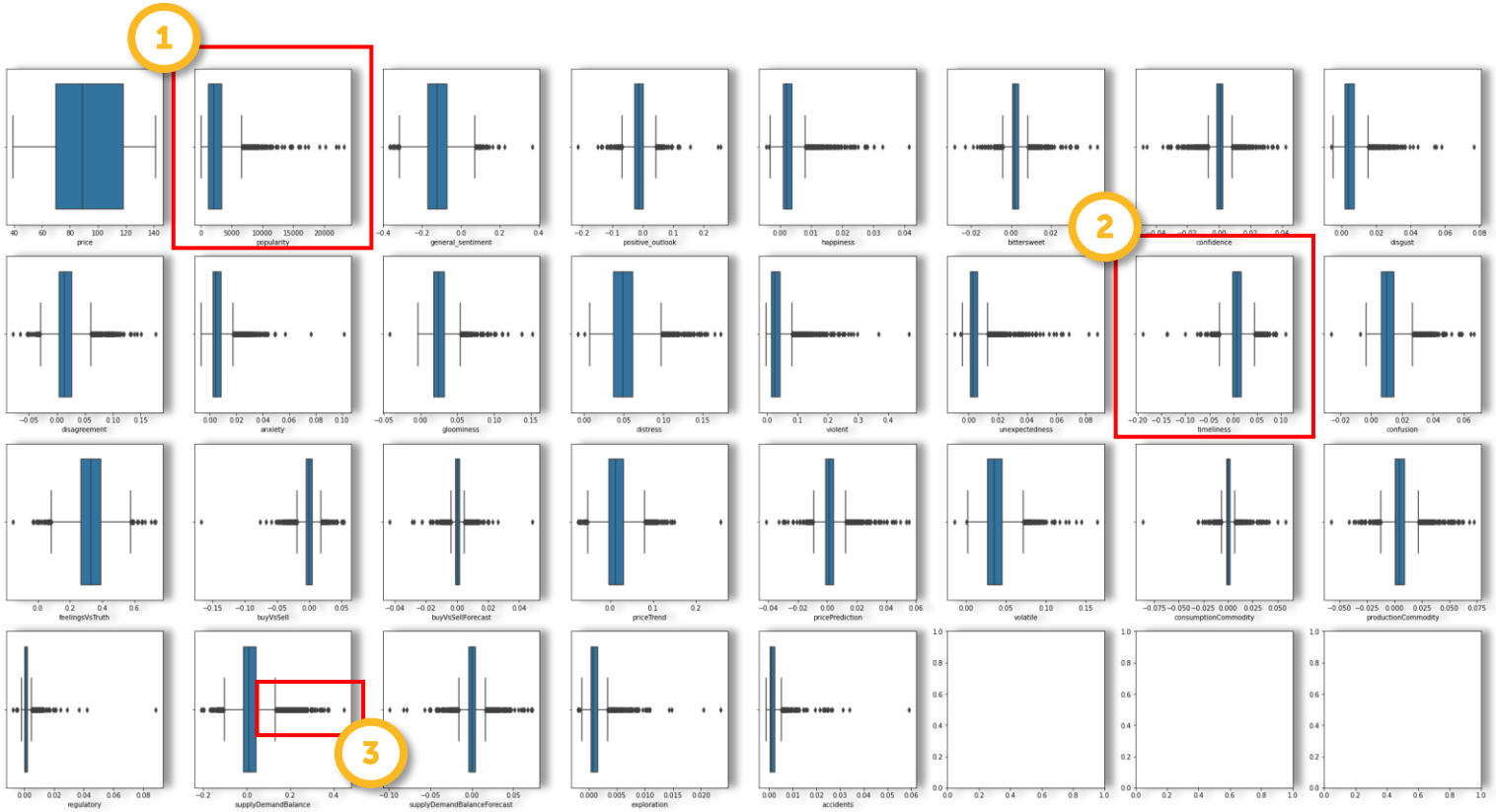
KEY INSIGHTS

1. There appears to be a pattern to the missing values in price!

2. Missing values in 'price' column corresponds to Malaysian holidays and weekends (non-trading days)

Imputed using forward fill – price fixed according to the last trading day for non-trading days

Feature Transformation



1. Variable is right skewed and violates the normality assumption.

2. Variable is left skewed and violated the normality assumption.

3. So many outliers!

PETRONAS

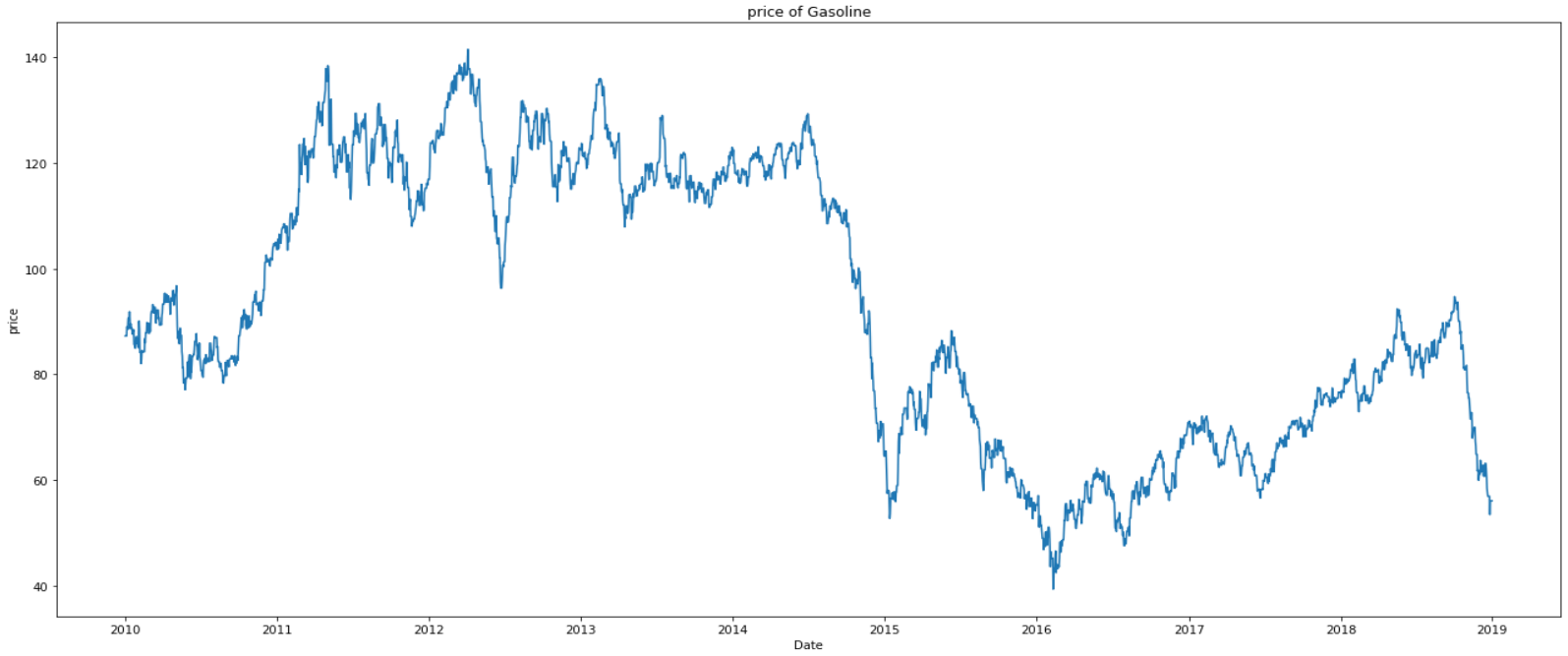
Feature Engineering

No	New Features	Definitions
1	Day of week	Binary: 7 columns for each day of the week. E.g., indicates whether it is Monday or not.
2	Weekend	Binary: Indicates whether it is a weekend or not.
3	Sunday	Binary: Indicates whether it is a Sunday or not.
4	Holidays	Binary: Indicates whether it is a holiday or not.
5	Month	Binary: 12 columns for each month of the year. E.g., indicates whether it is January or not

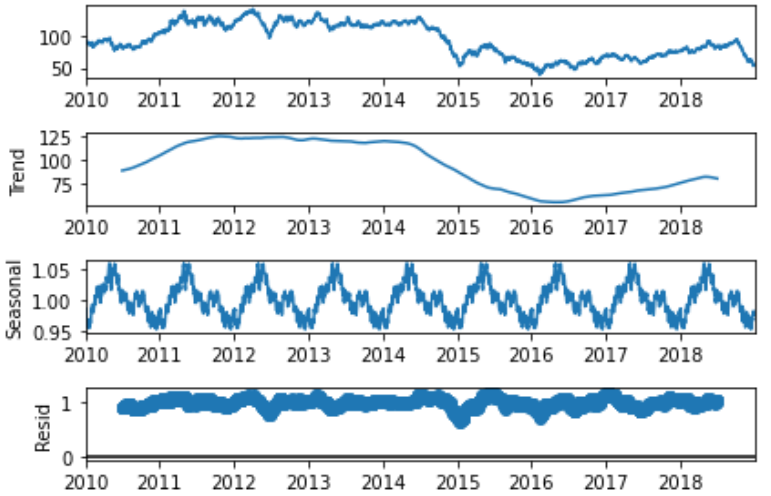
PETRONAS

Key Findings – Time Series EDA

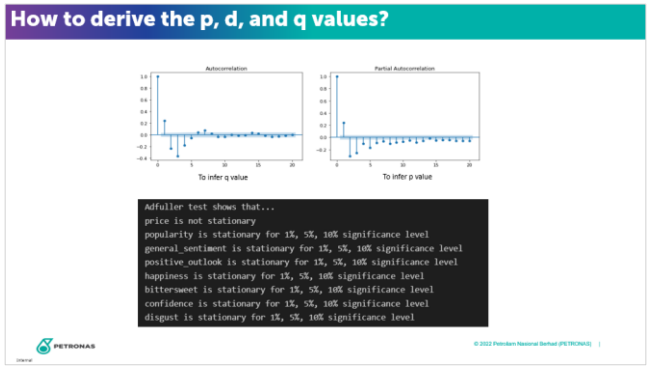
Price against Time



Multiplicative Seasonal Decomposition



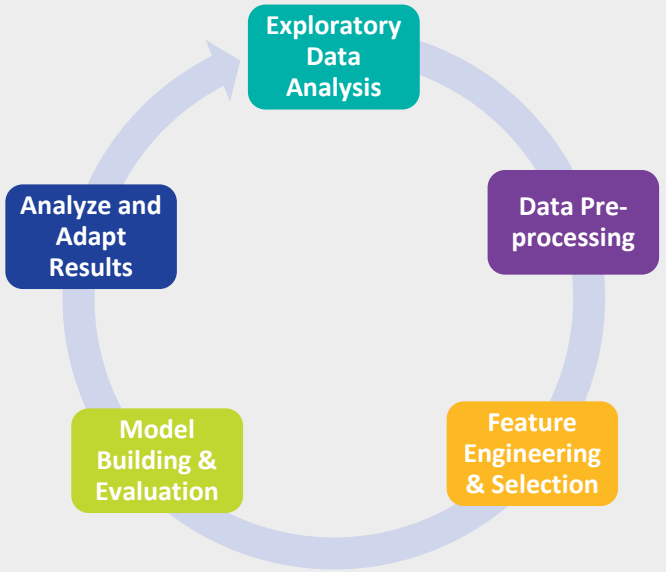
- There is no clear trend.
- The price variable is not stationary.
- The residuals are multiplicative.
- Predicting price is difficult without a robust algorithm.



PETRONAS

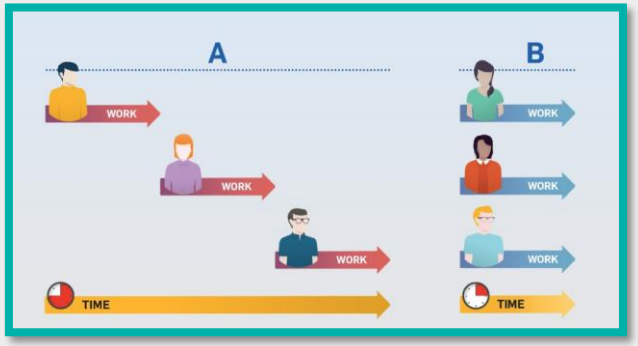
How We Did It

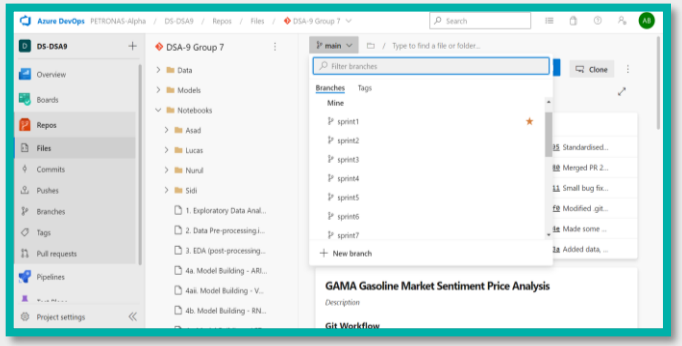
Process Flow



- 8 iterations
- 1 iteration a day
- Constant review with the team
- Discuss regularly with trainers / business

Ways of Working





What We Learn

1. Strengthen data science competencies in EDA, missing value imputation, and feature engineering.
2. Learn the relations between prices and sentiments. Domain knowledge is vital to make decisions.
3. Enhance skillsets on time series and how it is different from regular machine learning.
4. Better understanding on ARIMAX and time series deep learning.
5. Exposure to Azure DevOps, Git, version control, and team collaboration (branching and pull requests).

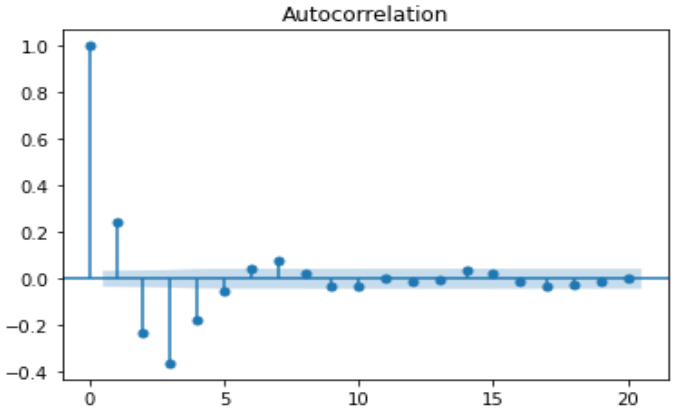
What Next

1. Get more data (historical and more recent data) to train the deep learning models.
2. Explore the other use cases by studying the other features such as supply demand breakdown and how price relates to this.
3. Use cross validation methods to ensure the train-test split is validated for different time windows.
4. Work with software engineers to create a dashboard for users to consume the model through visualisations.

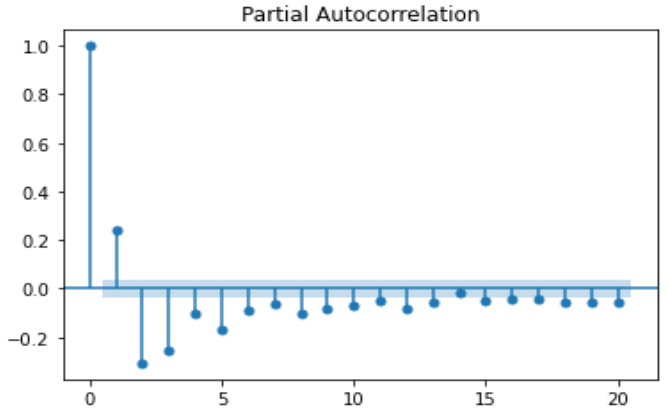
[Open]

Thank you for your passion.

How to derive the p, d, and q values?



To infer q value

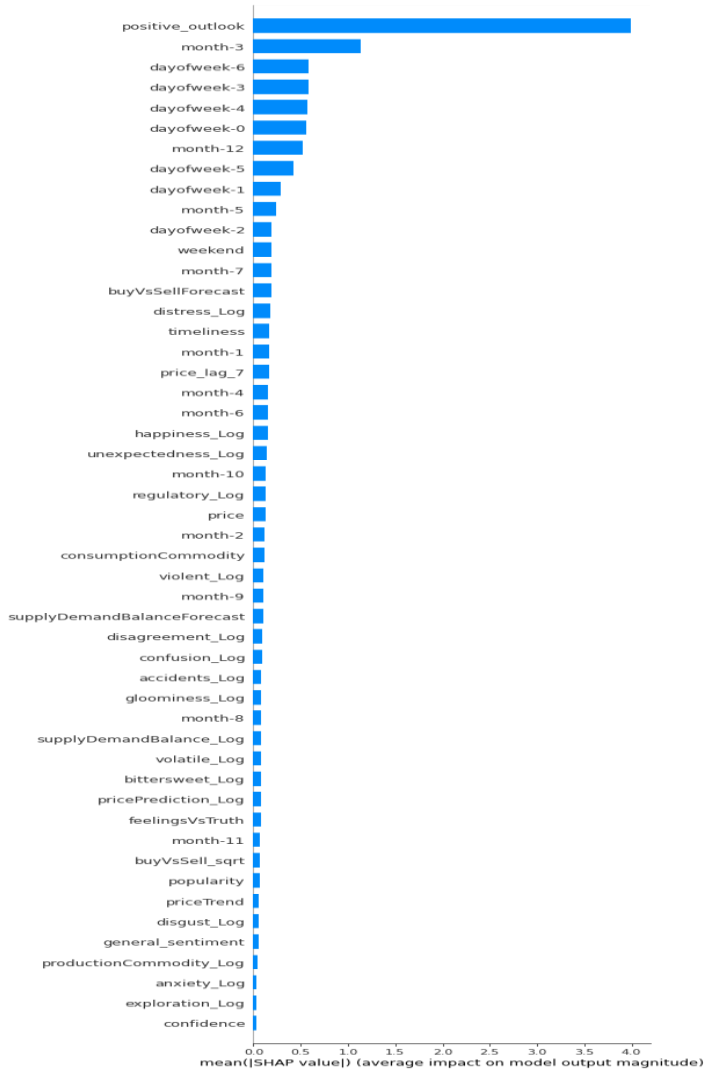


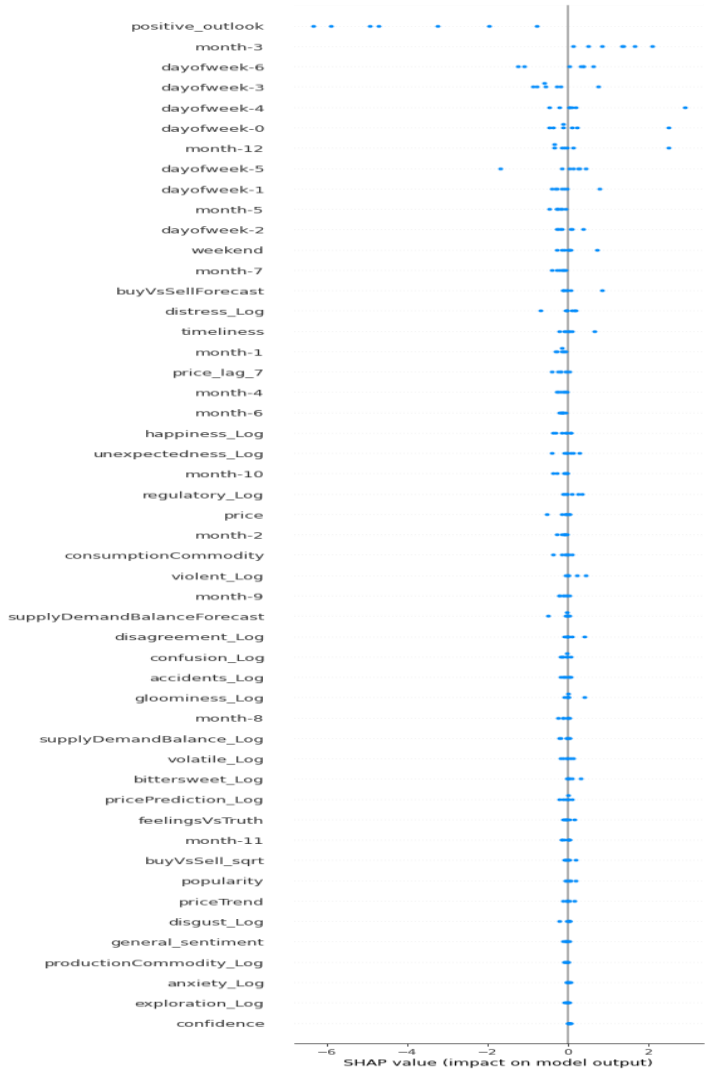
To infer p value

```
Adfuller test shows that...
price is not stationary
popularity is stationary for 1%, 5%, 10% significance level
general_sentiment is stationary for 1%, 5%, 10% significance level
positive_outlook is stationary for 1%, 5%, 10% significance level
happiness is stationary for 1%, 5%, 10% significance level
bittersweet is stationary for 1%, 5%, 10% significance level
confidence is stationary for 1%, 5%, 10% significance level
disgust is stationary for 1%, 5%, 10% significance level
```

PETRONAS

Key Findings – What some features that algorithm favours





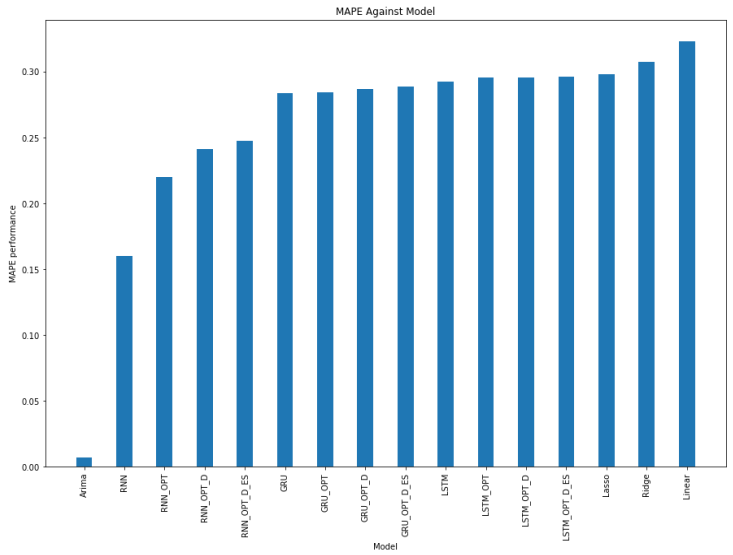
1. DL Model learn the most from *positive outlook* variable
2. Positive outlook sentiment have negative impacts toward price as in when price increase.

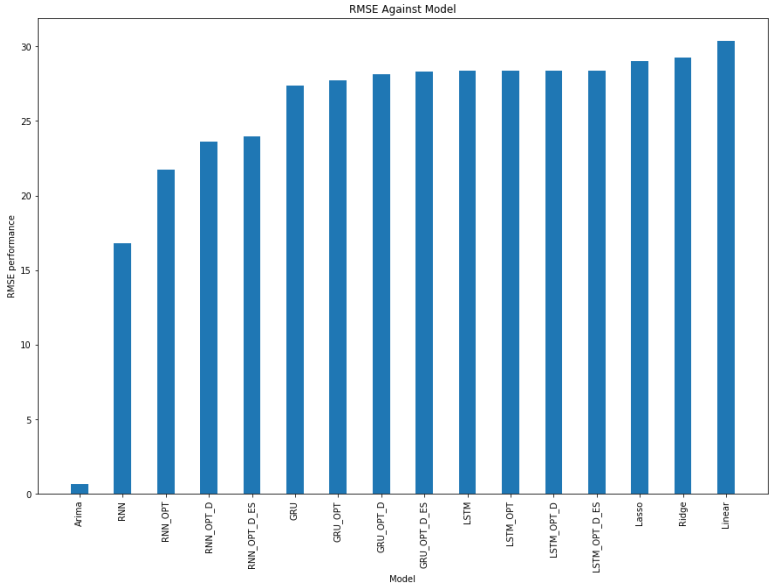
PETRONAS

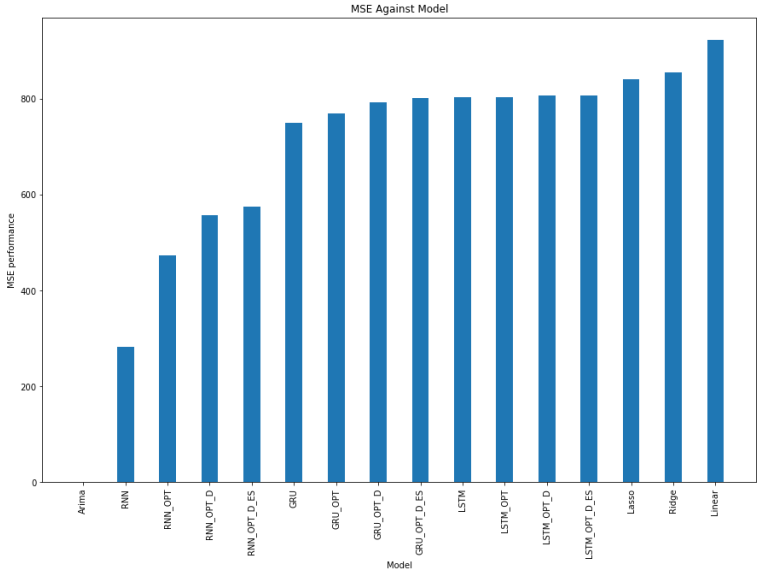
Key Findings – How Our Models Perform

Index	Model	MSE	RMSE	MAPE	Rank MSE	Rank RMSE	Rank MAPE
0	Arima	0.480847	0.693431	0.006893	1.0	1.0	1.0
1	RNN	803.471095	28.345566	0.292678	11.0	11.0	10.0
2	RNN_OPT	768.881366	27.72871	0.283834	7.0	7.0	6.0
3	RNN_OPT_D	801.385703	28.308757	0.287185	9.0	9.0	8.0
4	RNN_OPT_D_ES	282.412011	16.805119	0.160293	2.0	2.0	2.0
5	GRU	792.926785	28.158956	0.288617	8.0	8.0	9.0
6	GRU_OPT	854.533657	29.232408	0.307432	15.0	15.0	15.0
7	GRU_OPT_D	749.618105	27.379155	0.284273	6.0	6.0	7.0
8	GRU_OPT_D_ES	472.601374	21.739397	0.220161	3.0	3.0	3.0
9	LSTM	841.367953	29.006343	0.29836	14.0	14.0	14.0
10	LSTM_OPT	923.344036	30.386577	0.323354	16.0	16.0	16.0
11	LSTM_OPT_D	575.405737	23.987616	0.247318	5.0	5.0	5.0
12	LSTM_OPT_D_ES	557.814133	23.618089	0.241209	4.0	4.0	4.0
13	Lasso	806.406428	28.397296	0.296552	13.0	13.0	13.0
14	Ridge	803.445529	28.345115	0.295389	10.0	10.0	11.0
15	Linear	806.122211	28.392291	0.295594	12.0	12.0	12.0

Key Findings – How Our Models Perform (Visualised)







PETRONAS

Key Findings – Description on the Algorithm

Algorithm	Description
Time Series Models	
Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX)	
Machine Learning Models	
Multiple Linear Regression (LR)	LR is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables
Lasso Regression (Lasso)	Lasso regression is a regression technique which uses L1 regularization technique, that adds a penalty term equal to the absolute value of the magnitude of the coefficient.
Ridge Regression (Ridge)	Ridge regression is a regression technique which uses L2 regularization technique, that introduces a penalty term which is the summed absolute values of the model's parameters multiplied by lambda (regularization rate).
Deep Learning Models	
Recurrent Neural Network (RNN)	RNN is a type of neural network used for temporal or sequential data. RNN is distinguishable from feedforward neural network as they take information from prior inputs to influence current input and output in a sequence.
Long Short Term Memory (LSTM)	LSTM can also be considered an RNN, where the LSTM unit encompasses different gates to help regulate the flow of information better. The gates in an LSTM unit consists of the input, output and forget gate.
Gated Recurrent Unit (GRU)	GRU is also an RNN, where the GRU unit encompasses different gates to regulate the flow of information in the network. The GRU unit consists of a reset gate and an update gate.

Key Findings – Hyperparameter Tuning by Algorithm

Algorithm	Hyperparameter Tuning
Time Series Models	
Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX)	Auto Arima: • p • q • d
Machine Learning Models	
Multiple Linear Regression (LR)	- No parameter tuning
Lasso Regression (Lasso)	Grid Search: • Alpha
Ridge Regression (Ridge)	Grid Search: • Alpha
Deep Learning Models	
Recurrent Neural Network (RNN)	Grid Search: • Epochs • Batch Size • Optimizers
Long Short Term Memory (LSTM)	Grid Search: • Epochs • Batch Size • Optimizers
Gated Recurrent Unit (GRU)	Grid Search: • Epochs • Batch Size • Optimizers

Key Findings – Feature Transformation

```
Skewness for popularity is 2.688565405256412
Skewness for happiness is 3.2513306477618182
Skewness for bittersweet is 1.7272420370757668
Skewness for disgust is 3.0709467314172674
Skewness for disagreement is 1.293105488567805
Skewness for anxiety is 3.3170495437288645
Skewness for gloominess is 1.835446272590075
Skewness for distress is 1.2717534537679729
Skewness for violent is 3.522056002776818
Skewness for unexpectedness is 4.242727724079467
Skewness for confusion is 1.7837124395410207
Skewness for buyVsSell is -1.766697444535481
Skewness for pricePrediction is 1.411985261545558
Skewness for volatile is 1.3899238118231598
Skewness for productionCommodity is 1.1341686875585548
Skewness for regulatory is 11.33927917618782
Skewness for supplyDemandBalance is 1.2828991549923368
Skewness for exploration is 4.529465871754434
Skewness for accidents is 5.8063151839400975
```

Before transformation

```
Skewness for bittersweet_Log is 1.68272289002227
Skewness for disgust_Log is 3.0292443377706912
Skewness for disagreement_Log is 1.155432068942532
Skewness for anxiety_Log is 3.2386127611482878
Skewness for gloominess_Log is 1.6998661256639784
Skewness for distress_Log is 1.1748636642058328
Skewness for violent_Log is 3.095386730229782
Skewness for unexpectedness_Log is 4.20873163291828
Skewness for confusion_Log is 1.7307673934230199
Skewness for buyVsSell_sqrt is -1.9718633590735675
Skewness for pricePrediction_Log is 1.3441092001501687
Skewness for volatile_Log is 1.2891705690549091
Skewness for productionCommodity_Log is 1.0267012003573452
Skewness for regulatory_Log is 13.64291035489982
Skewness for supplyDemandBalance_Log is 0.916574617910244
Skewness for exploration_Log is 5.808360834813212
Skewness for accidents_Log is 10.863739759616594
```

After transformation

PETRONAS