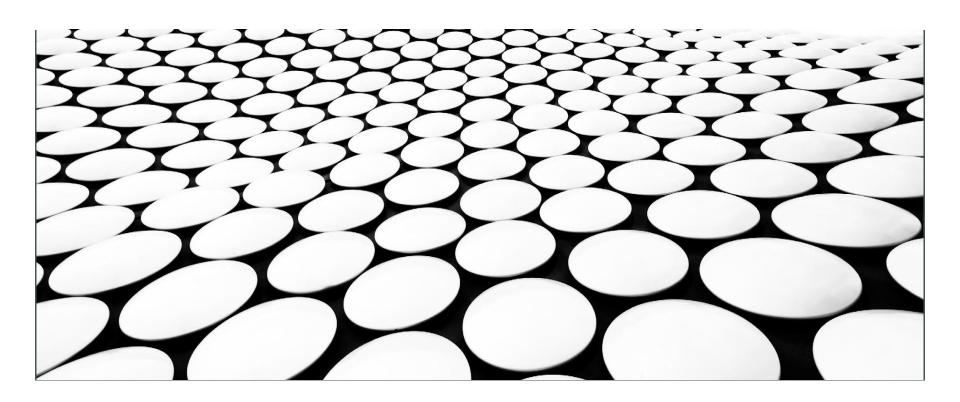
CRUDE OIL PRICE PREDICTION

MENTORS: KARTHIK MUSKULA &

DHANYAPRIYA SOMASUNDARAM



Members

na Vishnu u Nandhini dindla Divya Naveen Kumar Pradeep Kumar t Patil



Business Problem:



With a global trading price of 88\$ per barrel, oil prices have hit an all-time high in the last seven years. Due to rise in oil price the economy are getting affected vary badly.

Objective:

Oil is a product that goes completely in a different direction for a single market event as the oil prices are rarely based on real-time data, instead, it is driven by externalities making our attempt to forecast it even more challenging

As the economy will be highly affected by oil prices our model will help to understand the pattern in prices to help the customers and businesses to make smart decisions.

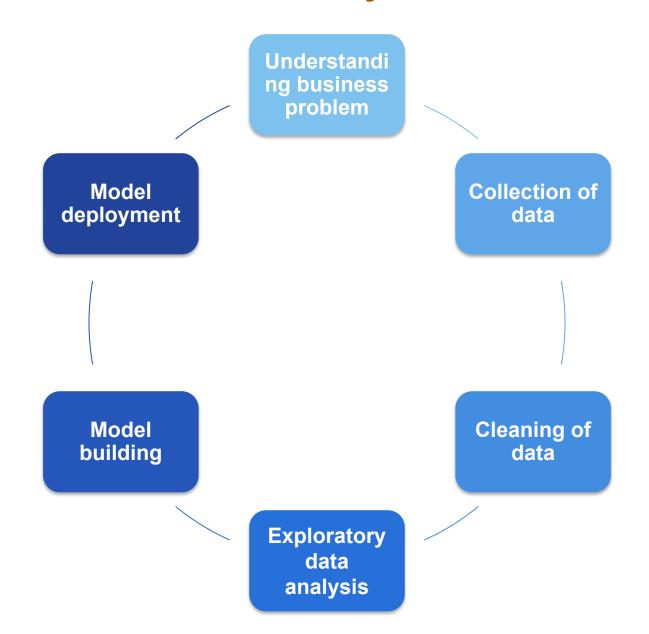
Factors That Affect the Price of Oil



affected by most of the factors, but the one's that affect the most ar and supply



Project Architecture / Project Flow



taset

wnloaded the crude oil price dataset from NASDAQ website.

aset has total 9345 rows and 2 columns which includes the crude oil price on 02-01-1986 to 21-12-2022.



Exploratory Data Analysis (EDA)

Dataset Details

Dataset contains 9345 rows and 2 columns

Datatypes are DataTime and Float Data ranges from 02-01-1986 to 21-12-2022.

A1	<u> </u>	fx Date								
	Α	В	С	D	E	F	G	Н	1	J
1	Date	Price								
2	02-01-1986	25.56								
3	03-01-1986	26								
4	06-01-1986	26.53								
5	07-01-1986	25.85								
6	08-01-1986	25.87								
7	09-01-1986	26.03								
8	10-01-1986	25.65								
9	13-01-1986	25.08								
10	14-01-1986	24.97								
11	15-01-1986	25.18								
12	16-01-1986	23.98								
13	17-01-1986	23.63								
14	20-01-1986	21.33								
15	21-01-1986	20.61								
16	22-01-1986	20.25								
17	23-01-1986	19.93								



Exploratory Data Analysis (EDA)

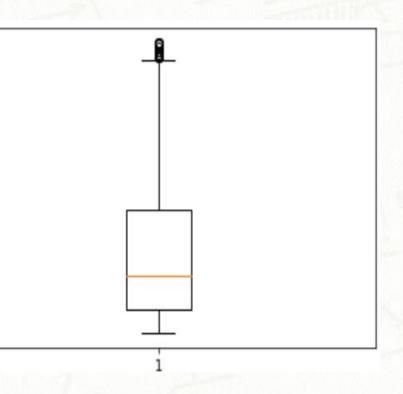
- Dataset has total 9345 rows and 2 columns which includes the crude oil price data from 02-01-1986 to 21-12-2022.
- Calculated Mean, Median, Variance, Skewness, Kurtosis. Mean = 46.028511, Median = 36.060000, Variance = 871.16243, Kurtosis = -0.553953 (Flat peak or Platykurtic), Skewness = 0.761338 (Positive Skewness or Right skewed).
- We found that the dataset has 6 null values. We applied mean imputation technique to remove the null values from the data.
- Calculated the correlation between Price and Year. Correlation = 0.723501 (Positive correlation)
- We used outlier detection technique (Histogram, Boxplot, Describe function) and found that the data has some outliers.
- We used boxplot formula to calculate the upper extreme and lower extreme values and removed the outliers using drop function.
- Finally, null values and outliers have been removed from the data.

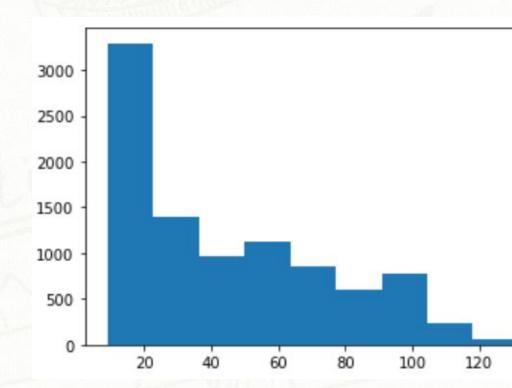
13.439	15.6333	16.15	15.0363	19.1557	13.2835	14.5815	13.7973	15.2005	14.846	14.9617	14.787
18.8643	19.5481	17.5276	18.1105	18.9514	20.8278	19.8295	18.9876	19.4565	18.8989	19.7178	19.3948
17.3271	15.7048	16.4755	16.2289	16.4938	15.6395	16.0159	16.2386	16.7777	14.9782	14.714	15.0232
20.736	18.9659	20.8915	18.6126	18.6071	19.32	19.8117	19.7109	20.059	19.8064	19.7155	19.4067
20.4016	27.0722	26.1405	22.8005	23.5425	21.5638	19.5568	21.6855	20.8739	28.7705	30.7838	30.8862
21.4832	21.9565	20.0257	20.4879	23.1718	21.8438	20.5235	20.711	21.0791	21.8839	22.9082	22.0915
20.3148	21.1471	20.004	19.5419	19.5627	21.36	21.5557	19.7174	20.9168	20.3226	21.0492	21.3109
19.8133	18.2943	16.2214	19.2772	19.0118	18.03	18.5765	19.5265	19.5	17.4132	18.0505	17.7477
16.799	17.4205	17.2376	15.7128	16.1657	18.4391	18.7464	15.6832	17.9142	17.3665	17.2948	17.3059
19.5635	18.3382	18.938	18.5961	18.3895	17.7455	18.1892	18.5017	19.1265	18.198	17.8018	18.1348
23.0885	21.9352	23.8605	20.605	19.4114	21.156	21.006	22.2457	21.2917	23.4252	24.0345	23.1676
20.0738	20.1595	18.9225	21.2532	23.3648	20.2809	20.015	21.1686	20.9505	19.8268	20.8624	20.238
14.9532	13.8905	12.611	15.354	15.9095	14.3159	14.1005	15.019	14.834	12.8861	14.0027	15.0443
17.7073	20.2165	22.8636	15.2941	15.1824	20.0105	18.792	16.8371	17.371	22.8284	21.4933	22.4804
27.0039	31.5736	28.616	29.9761	29.1965	29.6561	31.757	28.9545	29.4165	32.9295	31.9741	32.6205
26.8833	27.0543	22.8229	27.659	28.9668	26.1881	26.7732	26.3348	27.5735	21.9435	23.5152	26.0919
25.8281	27.6722	28.7871	23.2333	21.6795	26.201	25.7805	25.725	26.4691	25.9852	27.6159	28.675
29.275	30.719	32.5163	33.8242	32.323	31.0252	30.8515	31.247	29.9295	31.8647	30.8892	28.7905
38.6119	43.7205	42.4555	37 9579	37.5914	41.19	38.359	38.5033	40.659	46.0862	48.9081	44.3365
52.9043	62.5891	58.6752	52.7622	52.2767	57.2952	57.4582	54.461	51.8574	56.4025	58.9524	62.3364
69.2118	69.4323	63.7285	62.7428	67.002	71.3368	68.5946	64.0065	68.9435	62.24	61.5545	63.5152
66.5237	70.7413	83.7177	64.607	60.1174	74.3295	70.7152	64.0567	67.3204		82.4614	76.0933
107.827	109.745	57.7405	99.2642	92,4742	115.614	119.658	100.461	116.075	72.6985	77.1074	103.654
54 1368	67.4814	68.7643	49.9845	49.1905	63.3573	64.9067	54.919	60.47	70.8753	72.9382	65,4335
82.3941	76.6274		78.7647	77.2224	78.0986	78.5295	81.3	74.8595	81 4895	81.1429	77.3343
103.818		96.2133			96.7674		99.8155	97.4621	96.7476	90.7471	87.9355
100.246			101.037	97.7189		85.9841	102.065		89.16		94.6738
92.9071	103.166	98.4495	95.9144	96,2995	102.397	97.124	95.7445	95.7835	95.5241	99.1295	102.922
98.5535	94.5805	69.7045	98.9568	94.955	100.583	102.21	97.2478	99.0914	82.782	86.1348	93.2265
53.6523	45.0505	41.4686	50.0975	48.0055	49.4468	55.7062	47.533	55.7535	44.1563	46.7591	46.9339
42.107	44.9325	48.8373	36.3706	36.1818	43.8527	47.1255	39.999	46.1379	44.1018	47.279	44.4395
50.3458	48.5432	54.6609	52.6326	52.2595	48.9314	47.3667	49.5021	49.8676	54.1318	51.6295	50.5629
66.9379	66.7609	54.7359	63.1704	63.5041	67.0171	67.1536	63.7009	67.8746	58.3633	67.6657	68.8052
60.3875	55.687	59.1786	55.9839	54.5352	55.8905	56.112	57.8543	58.7652	56.8609	55.7786	57.503
25.561	41.3143	44.5735	46.4655	48.6224	40.9483	39.6681	31.9814	34.9779	41.2379	41.0378	39.7164
63.8982	67.3909	70.9143	64.8541	59.329	69.9595	70.379	63.559	65.394	72.4784	77.4433	71.0843
100.332	93.675	83 8773	94.4072	90.0338	96.5196	104.761	102.021	104.611	87.2086	89.7267	87.115
Apr	Aug	Dec	Feb	Jan	Jul	Jun	Mar	May	Nov	Oct	Sep

Inferences:

Since Heatmap provide us easy tool to understand the correlation between two entities, we used the heatmap to represent the correlation between the two variables year & month.

tlier Detection

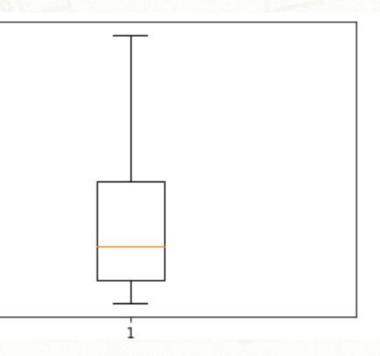


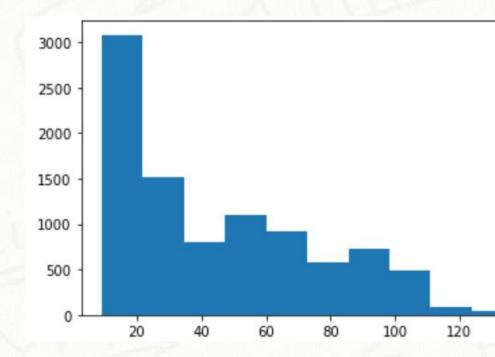


ices:

ere are outliers above the positive upper extreme whisker.
ere are no outliers below the negative side of the lower whisker.
e the data doesn't follow Normal Distribution. Right skewed data.

oing outliers

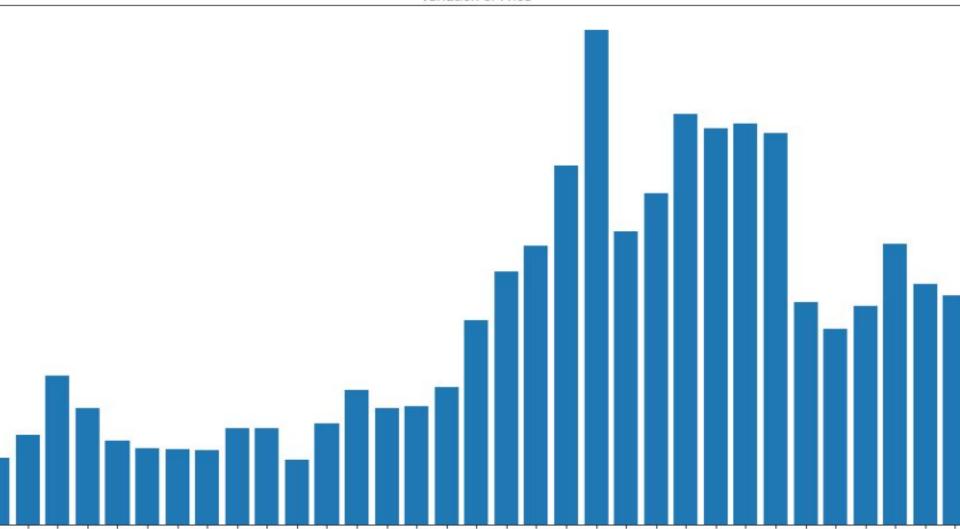




e no outliers above the positive upper whisker.

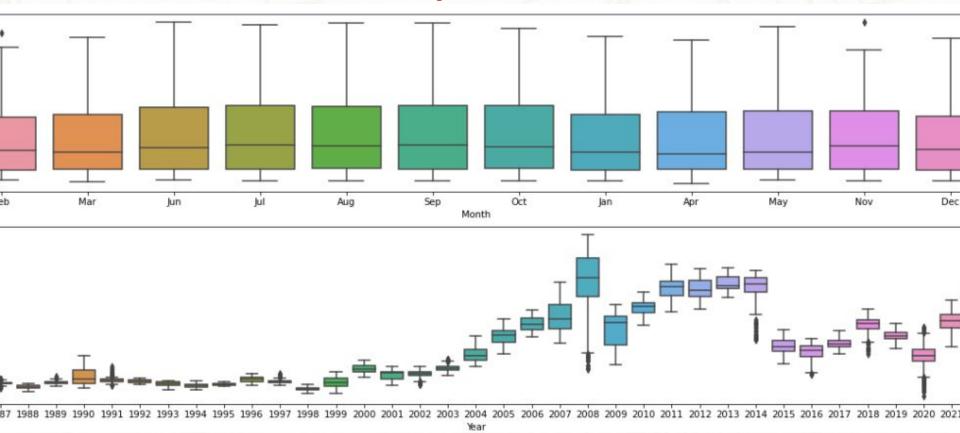
Data VisualizationBar Plot of Variation of Price over Years

Variation of Price



38 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 202

Box plot

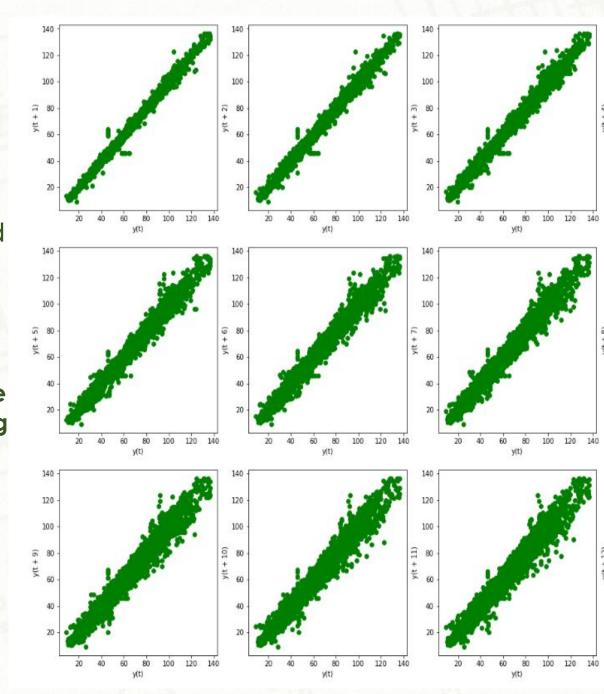


LOT: Inferences: Years 2018 and 2020 show outliers. 2018 The price of oil dropped in November 2018 becaus umber of factors, including "rising petro-nations' oil production, the U.S. shale oil boom, and swelling in North can oil inventories," according to Market Watch. 2020 As rising Covid cases prompted fears of demand own, due to that the price of the oil dropped.

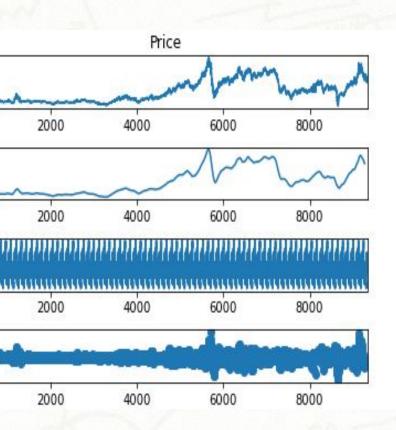
ation arly)

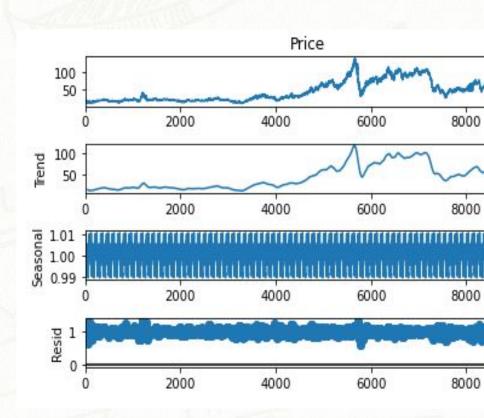
pecial type of scatter plot in represents the dataset with behind or ahead as compared e difference between these ed lag or lagged and it is

series is showing autoregressive good for time series modelling lives a Positive correction ace they can be modelled.

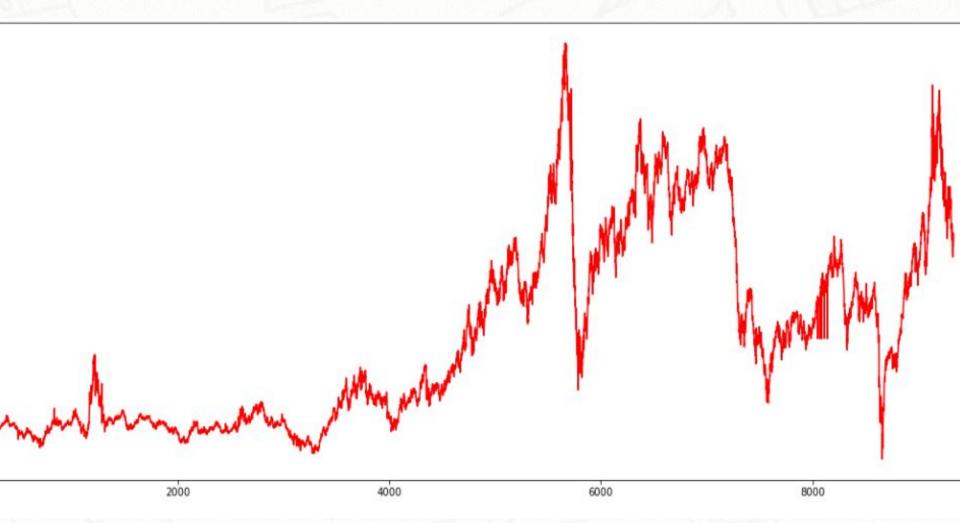


Time series Decomposition Method





s Smoothing



xponential trend in Moving average & it shows orginal Oil Price data.



Model Building

MODEL BASED METHODS

	MODEL	RMSE_Values	MAPE_Values
0	Linear_model	2.918314e+01	48.066785
1	Exp_model	4.420679e+01	92.698980
2	Quad_model	3.435598e+97	150.729517
3	Add_sea_model	4.425814e+01	48.859428
4	Add_sea_quad_model	9.655196e+01	150.898063
5	Mult_sea_model	4.904882e+01	94.686302
6	Mult_add_sea_model	4.423942e+01	92.699036

ARIMA Model

SARIMAX Results

Dep. Variable:	V	No. Observations:	9331
Model:	SARIMAX(0, 0, 2)	Log Likelihood	-16089.679
Date:	Mon, 09 Jan 2023	AIC	32185.357
Time:	15:32:38	BIC	32206.781
mile.	13.32.30	ыс	32200.701
Sample:	0	HQIC	32192.634
	0331		

- 9331

Covariance Type: opg

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.1213	0.003	-40.770	0.000	-0.127	-0.115
ma.L2	-0.0155	0.006	-2.703	0.007	-0.027	-0.004
sigma2	1.8418	0.007	257.533	0.000	1.828	1.856

Ljung-Box (L1) (Q): 0.00 Jarque-Bera (JB): 356642.45

Prob(Q): 0.99 **Prob(JB):** 0.00

Heteroskedasticity (H): 11.12 Skew: -0.67

Prob(H) (two-sided): 0.00 Kurtosis: 33.26

Splitting the data into

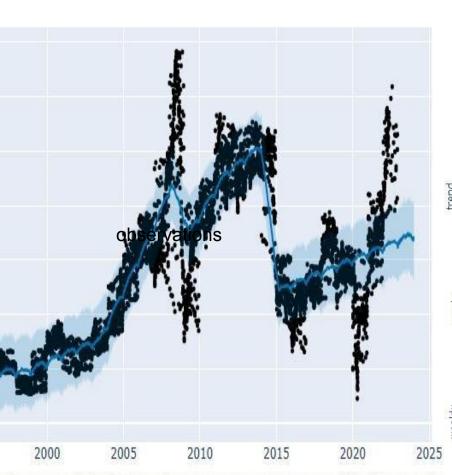
Train data = 66.6%Test data = 33.4%

Model with order

Data Driven Methods

	MODEL	RMSE_Values	MAPE_Values
0	Ses_model	1.147909e+33	37.926884
1	Holt_model	inf	1181.995312
2	Hwe_add_add_model	4.066831e+44	56.706486
3	Hwe_mult_add_model	2.841681e+44	56.403096

FB Prophet Model



Train2 = data11.iloc[:len(data11)-150] # Total data observations

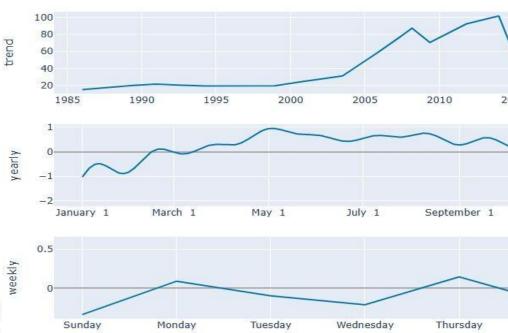
Test2 = data11.iloc[len(data11)-150:] # Only last 1

Result:

rmse: 9.83 MAE: 6.16 MSE: 96.54

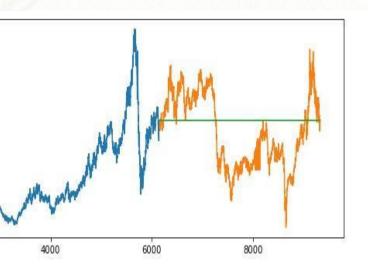
R-squared Score: 0.88

Components



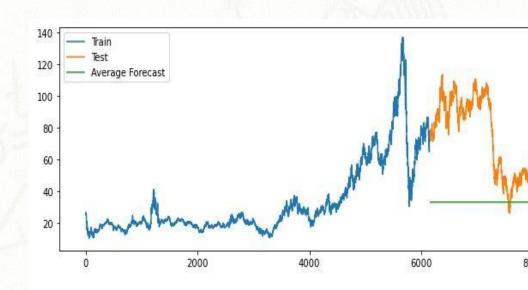
el (Last Sample Method)

ead(6160) I(3172)



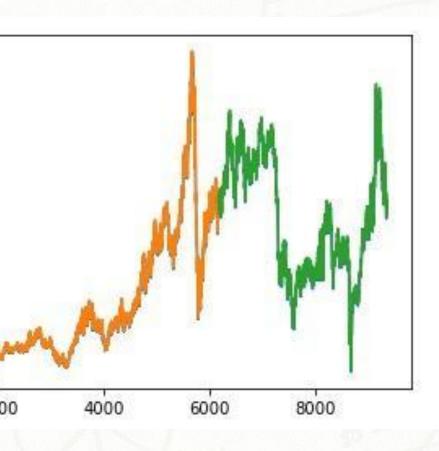
Simple Averages Met

Train: 6159 Test: 3173



rmse:224710588217780.16

Long Short Term Memory (LSTM) Model



Train Score: 1.19 RMSE Test Score: 2.04 RMSE

Train-R2 score: 0.9973721176476127 Test-R2 Score: 0.99215850176137

Model Deployment using Streamlit



Crude Oil Price Forecasting



This data app uses Facebook's open-sources prophet library to automatically general future forecast values from imported dataset. You'll be able to import your data set from a CSV file, visualize trend and feature, analyze forecast performance, and finally download the created forecast



Import Data

Upload here

Drag and drop file here

Limit 200MB per file • CSV

Browse files

Select Forecast Period

How many periods would you like to forecast into the future?

Uploading the data

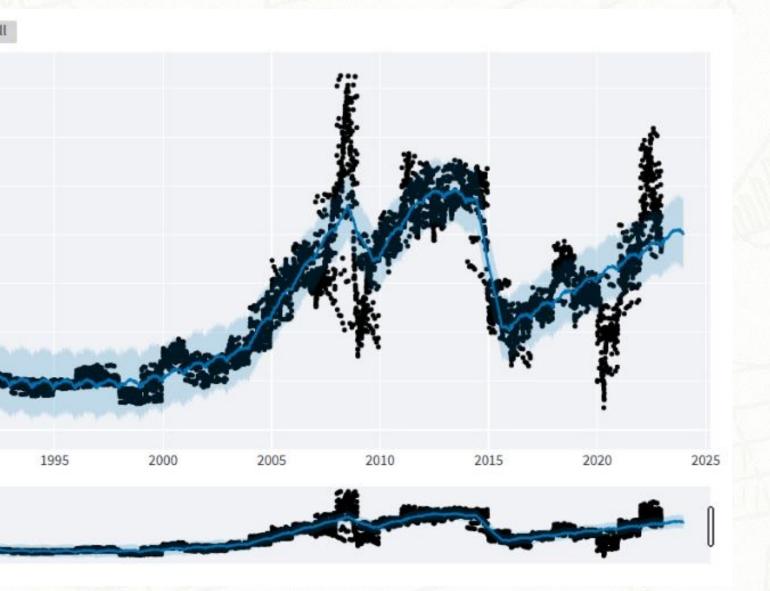
Here, the user should uploa and enter the number of for for which he want the predic

Visualized forecast data

	ds	yhat	yhat_lower	yhat_upper
9700	2023-12-12T00:00:00	80.2908	66.3457	93.8470
9701	2023-12-13T00:00:00	80.0786	66.1881	95.3446
9702	2023-12-14T00:00:00	80.5899	67.2124	94.7726
9703	2023-12-15T00:00:00	80.1658	65.1791	93.7202
9704	2023-12-16T00:00:00	80.9960	66.8745	96.0558
9705	2023-12-17T00:00:00	80.1438	65.8343	94.2432
9706	2023-12-18T00:00:00	80.5524	66.8627	94.5782
9707	2023-12-19T00:00:00	80.3275	65.6418	94.3695
9708	2023-12-20T00:00:00	80.2018	65.6389	94.9570
9709	2023-12-21T00:00:00	80.8024	65.8381	94.8606

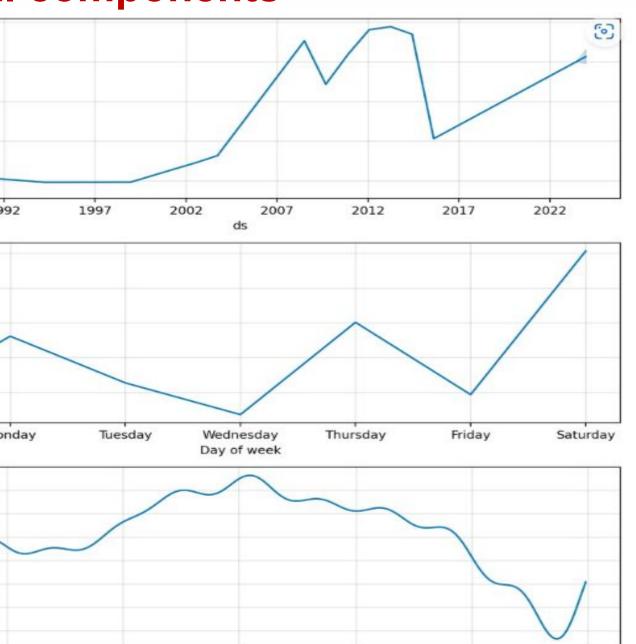
This visual shows future predicted values of the next 365 days. "yhat" is the predicted value, and the upper and lower limits are (by default) 80% confidence intervals.

Visualizations



This visualizes shows the accordance ("Blacker and predicte ("Blue line") time.

I components



This visuals shows a trend of predicted value week trends, and year dataset covers multiputed. The blue shaded are upper and lower continterval



Challenges faced?

The part of the project where we have faced the difficulty was in the model building stage deploying the final model. We ran into many errors in the process.

How did you overcome?

How did you overcome: We finished the deployment for our project on time, even after facing many errors because of our team work and with the help of our mentors and projects team.



Thank you