FINAL PROJECT

AI accelerator course - Technion

Stav Aizik

Step 1

Date	Close	High	Low	Open	Volume
2021-09-01	244.696671	247.330002	243.756668	244.693329	39612900
2021-09-02	244.130005	246.990005	243.513336	244.833328	38331900
2021-09-03	244.523331	244.666672	241.399994	244.083328	45738300
2021-09-07	250.973328	253.399994	246.419998	246.666672	60119400
2021-09-08	251.289993	254.816666	246.923340	253.860001	56379000
...	...	...	...	...	...
2022-08-25	296.070007	302.959991	291.600006	302.359985	53230000
2022-08-26	288.089996	302.000000	287.470001	297.429993	57163900
2022-08-29	284.820007	287.739990	280.700012	282.829987	41864700
2022-08-30	277.700012	288.480011	272.649994	287.869995	50541800
2022-08-31	275.609985	281.250000	271.809998	280.619995	52107300

252 rows × 5 columns

	Date Created	Tweets
80	2022-08-31 15:39:27+00:00	@ChrisZheng001 I intend to do so as soon as po...
81	2022-08-31 15:16:57+00:00	@teslaownersSV
82	2022-08-31 15:13:23+00:00	@SciGuySpace Yeah, aiming for up to 100 flight...
83	2022-08-31 15:06:53+00:00	Now launching every ~5 days
84	2022-08-30 22:52:58+00:00	@WholeMarsBlog @jonbbc @aelluswamy 10.69.2 goe...
...	...	...
3737	2021-09-01 16:46:52+00:00	@DJSnM @thesheetztweetz @johnkrausphotos @Lyla...
3738	2021-09-01 16:42:32+00:00	@DJSnM @thesheetztweetz @johnkrausphotos @Lyla...
3739	2021-09-01 16:21:41+00:00	@thesheetztweetz Filing legal actions against ...
3740	2021-09-01 16:13:33+00:00	@SciGuySpace He should consider spending some ...
3741	2021-09-01 02:44:40+00:00	@MarcusHouse @BocaChicaGal @NASASpaceflight Yup

3662 rows × 3 columns

STOCK

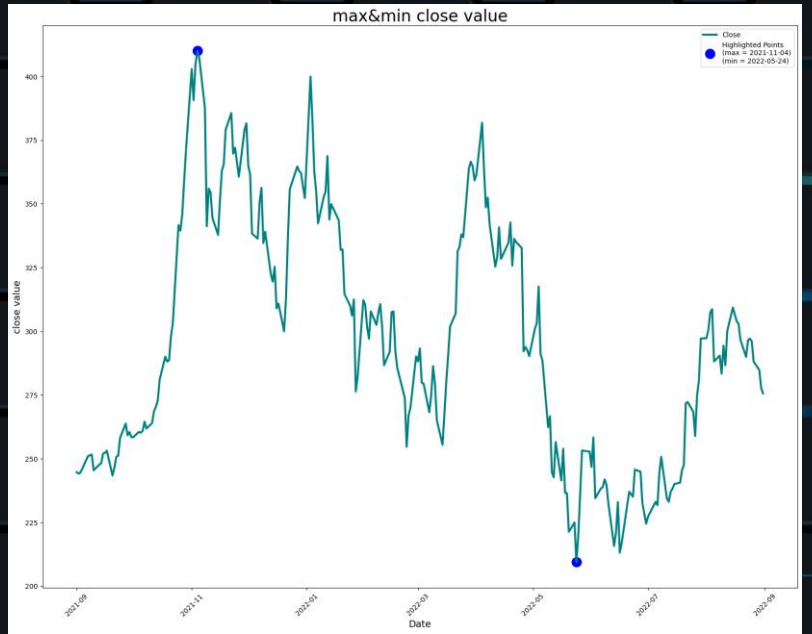
TESLA

DATASET

Tweets – Elon Musk

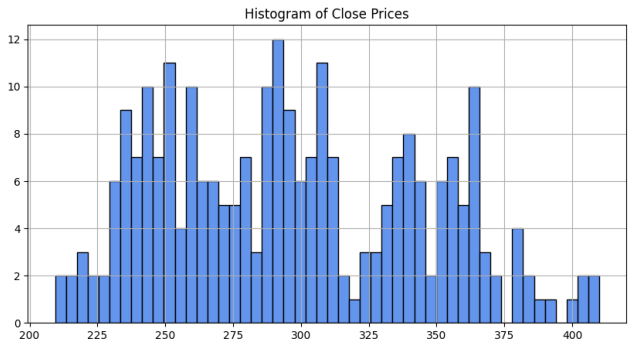
TIME

1/9/2021-01/09/2022



Step 2- Tesla statistics

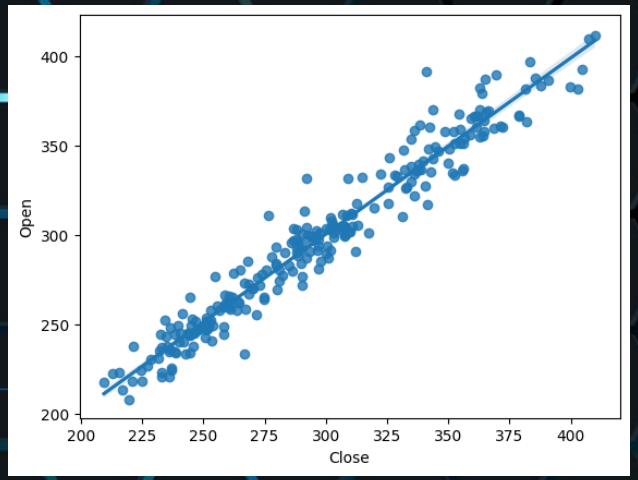
Histogram



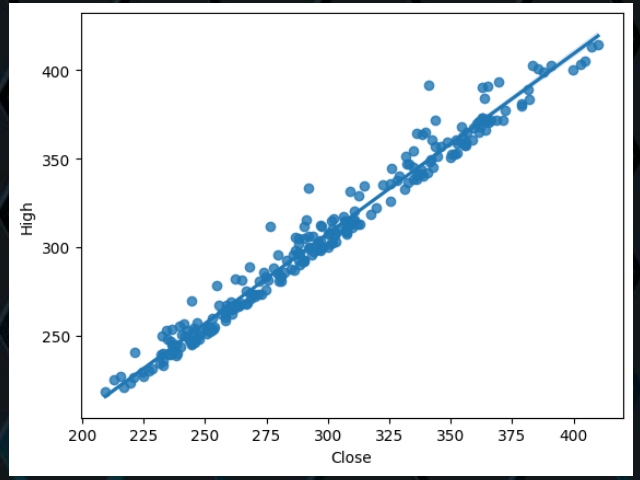
Describe

	Close	High	Low	Open	Volume
count	252.000000	252.000000	252.000000	252.000000	2.520000e+02
mean	296.773651	304.358889	289.330211	297.211839	8.012612e+07
std	47.546290	48.915248	46.225322	47.989156	2.608064e+07
min	209.386673	217.973328	206.856674	207.949997	3.504270e+07
25%	255.262497	261.730003	249.826668	255.566669	6.192960e+07
50%	292.128342	301.179993	285.813339	295.166672	7.710210e+07
75%	336.472504	344.957512	327.908340	335.700005	9.351338e+07
max	409.970001	414.496674	405.666656	411.470001	1.885563e+08

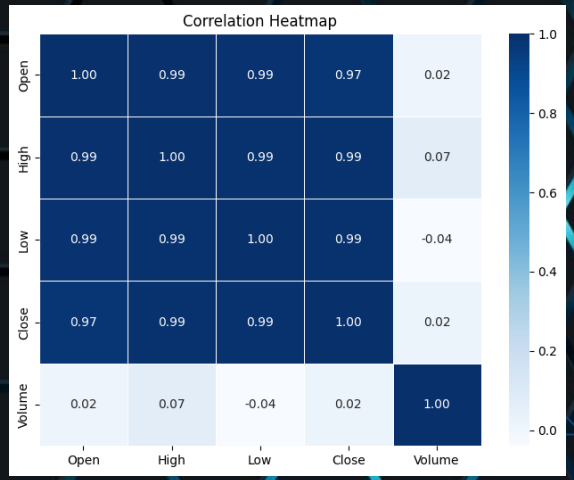
Close - Open



Close - High



HeatMap

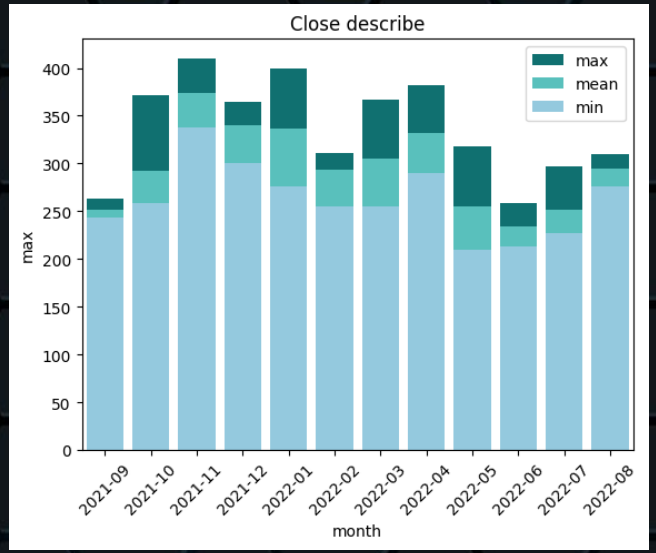


Step 2- Tesla statistics

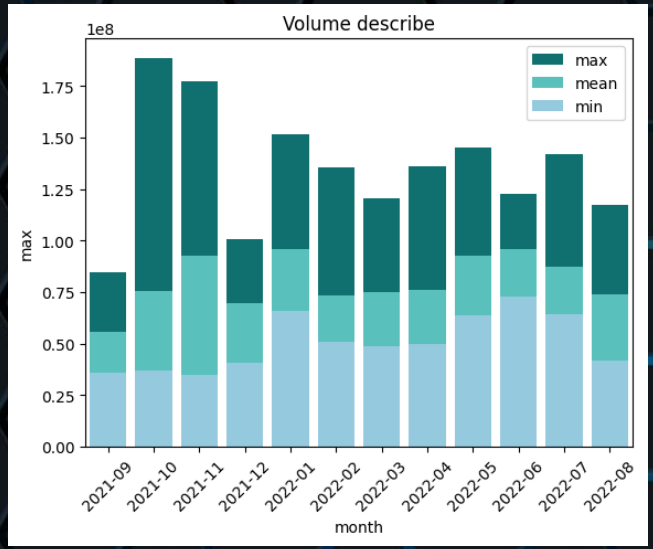
Describe By Month - Close

	Close							
	count	mean	std	min	25%	50%	75%	max
month								
2021-09	21.0	251.317463	5.874000	243.389999	246.460007	251.213333	253.163330	263.786682
2021-10	21.0	292.783017	36.500997	258.406677	263.980011	281.010010	303.226654	371.333344
2021-11	21.0	373.554125	22.071917	337.796661	355.983337	372.000000	387.646667	409.970001
2021-12	22.0	339.629999	20.612209	299.980011	322.934990	338.666672	356.665001	365.000000
2022-01	20.0	336.722830	31.675523	276.366669	312.412498	342.911667	354.824989	399.926666
2022-02	19.0	292.961577	16.457416	254.679993	286.163330	297.046661	307.404999	310.666656
2022-03	23.0	304.793186	36.075913	255.456665	279.431656	290.533325	334.958344	366.523346
2022-04	20.0	332.462503	24.950996	290.253326	325.627502	334.889999	344.184174	381.816681
2022-05	21.0	255.223333	29.855613	209.386673	236.473328	252.753326	266.679993	317.540009
2022-06	21.0	234.025873	11.247435	213.100006	228.490005	235.070007	239.706665	258.333344
2022-07	20.0	251.394666	19.678442	227.263336	236.365829	245.036667	269.251656	297.149994
2022-08	23.0	294.869857	9.581290	275.609985	288.130005	296.453339	301.728333	309.320007

12 rows × 40 columns

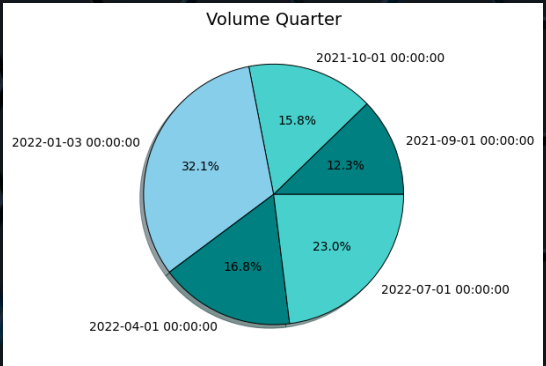


Describe By Month - volume



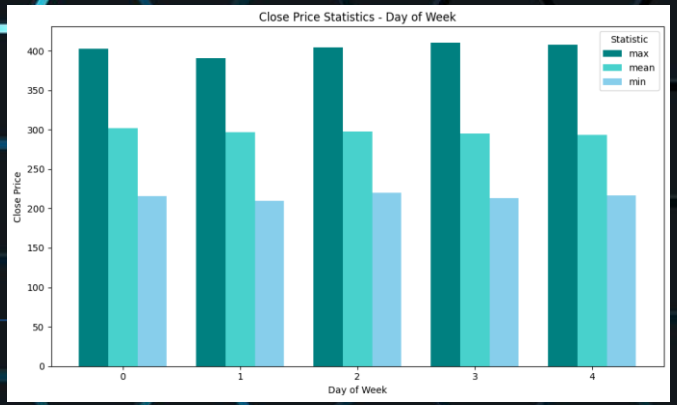
Step 2- Tesla statistics

Describe By Month - volume

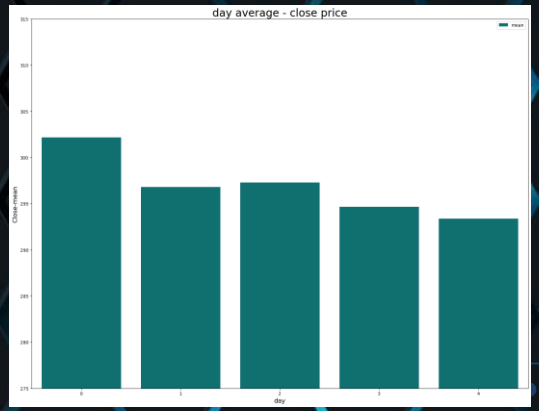


Date	Close	High	Low	Open	Volume	month	day of week	Quarter
2021-09-01	244.696671	247.330002	243.756668	244.693329	39612900	2021-09	2	2021Q3
2021-10-01	258.406677	260.260010	254.529999	259.466675	51094200	2021-10	4	2021Q4
2022-01-03	399.926666	400.356659	378.679993	382.583344	103931400	2022-01	0	2022Q1
2022-04-01	361.529999	364.916656	355.546661	360.383331	54263100	2022-04	4	2022Q2
2022-07-01	227.263336	230.229996	222.119995	227.000000	74460300	2022-07	4	2022Q3

Describe By Day Of Week - Close



day of week	max	min	mean
0	402.863342	215.736664	302.160871
1	390.666656	209.386673	296.813972
2	404.619995	219.600006	297.298490
3	409.970001	213.100006	294.653399
4	407.363342	216.759995	293.381800



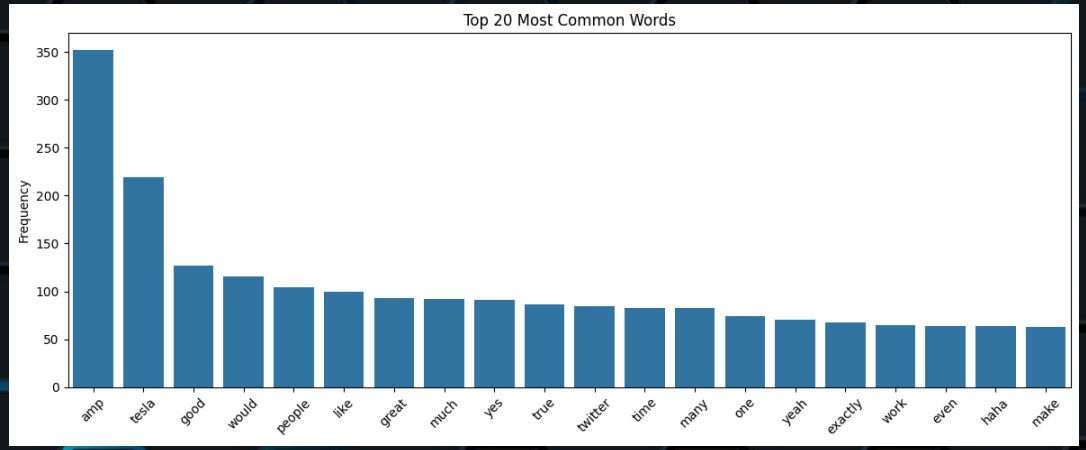
Step 2- Tweets statistics

```
Tweet Frequency Per Day Statistics:
 Mean: 10.83
 Median: 10.0
 Max: 41
 Min: 1
 Std Dev: 7.51
```

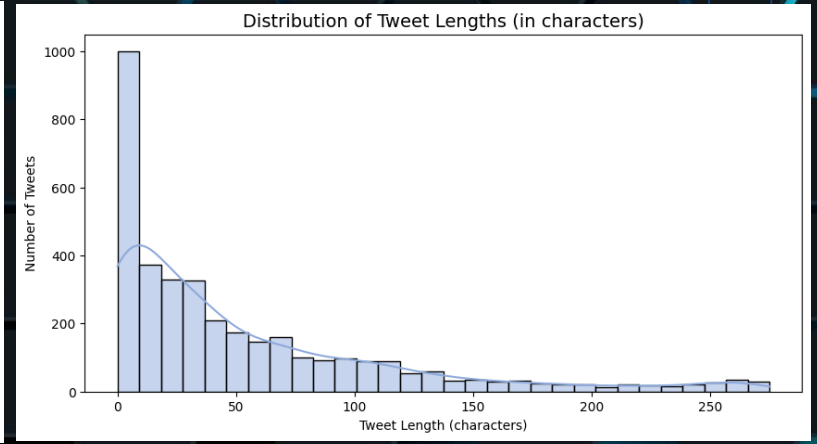
```
df_filtered['tweet_length'].max()

275
```

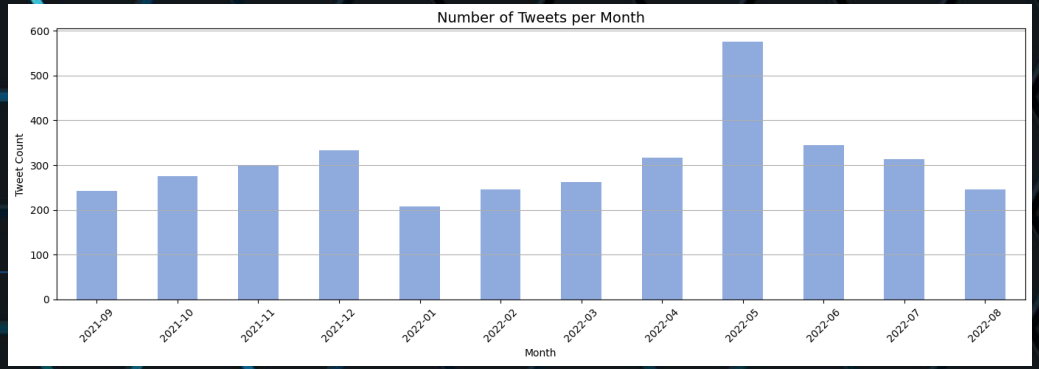
TOP WORDS



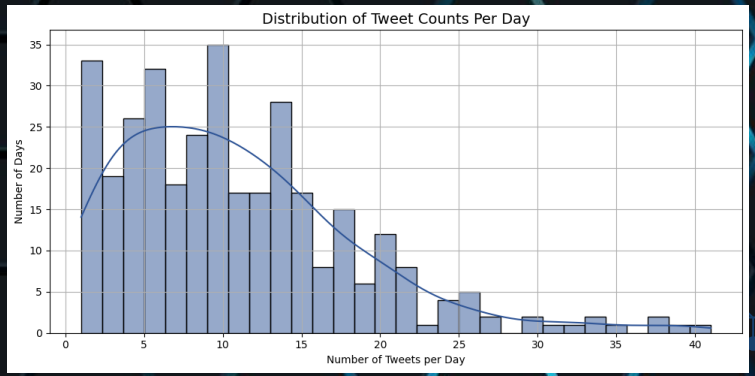
Histogram - distribution of tweets lengths



Number Of Tweets - Month



Histogram – tweets per day



Step 3- Classic Models

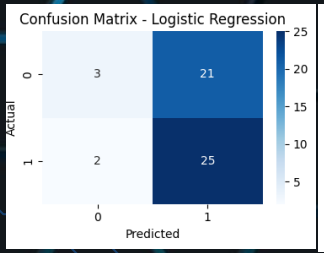
Pre- Processing

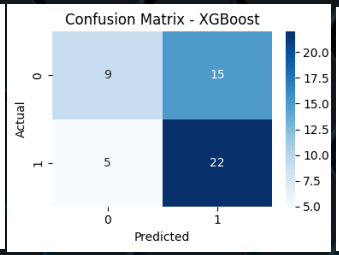
Date	Close	High	Low	Open	Volume	Change	label
2021-09-01	244.696671	247.330002	243.756668	244.693329	39612900	-0.566666	0.0
2021-09-02	244.130005	246.990005	243.513336	244.833328	38331900	0.393326	1.0
2021-09-03	244.523331	244.666672	241.399994	244.083328	45738300	6.449997	1.0
2021-09-07	250.973328	253.399994	246.419998	246.666672	60119400	0.316666	1.0
2021-09-08	251.289993	254.816666	246.923340	253.860001	56379000	0.330002	1.0
...	...	...	...	...	...	...	...
2022-08-25	296.070007	302.959991	291.600006	302.359985	53230000	-7.980011	0.0
2022-08-26	288.089996	302.000000	287.470001	297.429993	57163900	-3.269989	0.0
2022-08-29	284.820007	287.739990	280.700012	282.829987	41864700	-7.119995	0.0
2022-08-30	277.700012	288.480011	272.649994	287.869995	50541800	-2.090027	0.0
2022-08-31	275.609985	281.250000	271.809998	280.619995	52107300	NaN	NaN

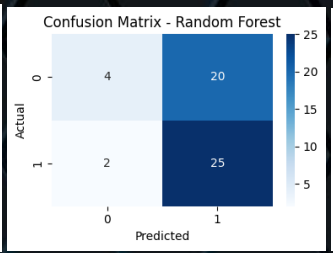
252 rows × 7 columns

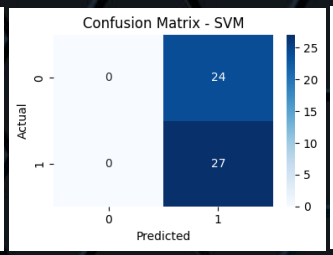
Models

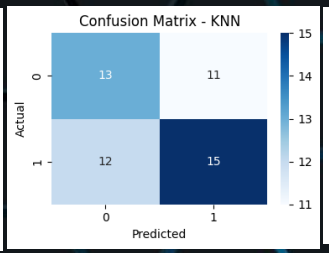
	Model	Accuracy	Precision	Recall	F1-score
1	Random Forest	0.568627	0.555556	0.925926	0.694444
3	SVM	0.529412	0.529412	1.000000	0.692308
5	XGBoost	0.607843	0.594595	0.814815	0.687500
0	Logistic Regression	0.549020	0.543478	0.925926	0.684932
2	Decision Tree	0.549020	0.576923	0.555556	0.566038
4	KNN	0.549020	0.576923	0.555556	0.566038

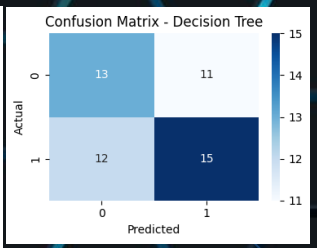




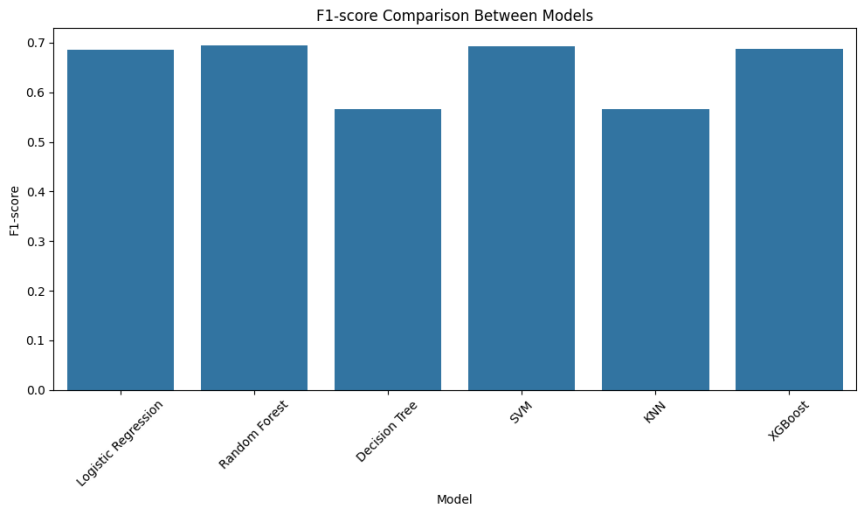








Step 3- Classic Models



XGBoost achieved the highest accuracy (60.7%)

Random Forest had the best F1-score (0.694)

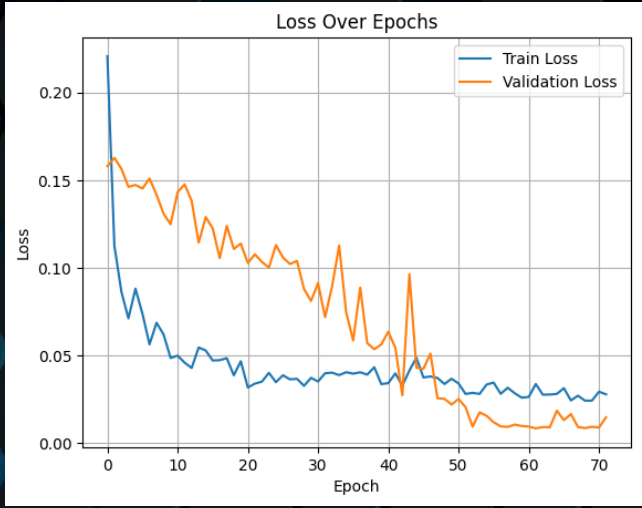
Decision Tree and KNN showed balanced confusion matrices

small data size and high volatility

Recommended to expand the data and add technical and sentiment features to improve prediction.

Step 4- LSTM

Loss

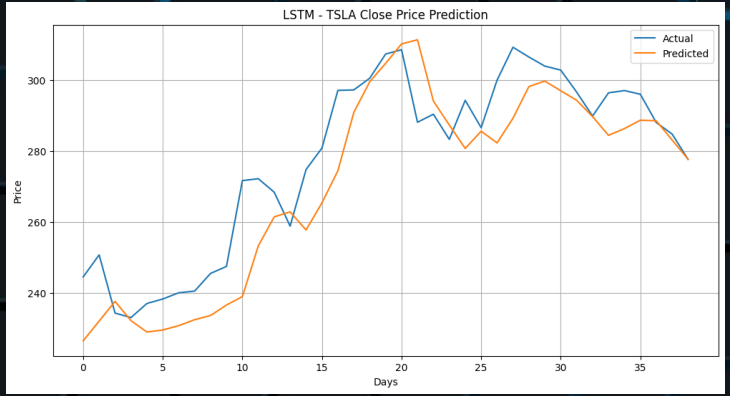


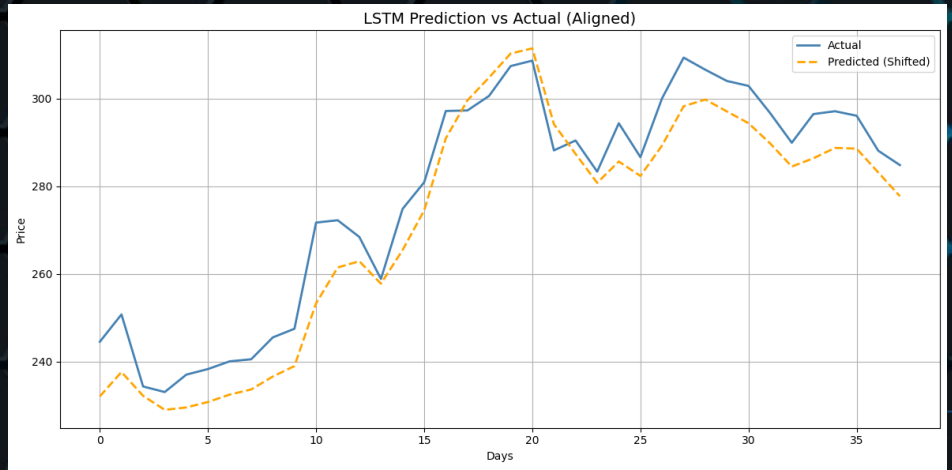
MAE: 9.32

RMSE: 12.13

R^2: 0.75

Models





Step 4- LSTM

strong predictive performance

low error (MAE=9.32, RMSE=12.13) and high explanatory power (R^2=0.75)

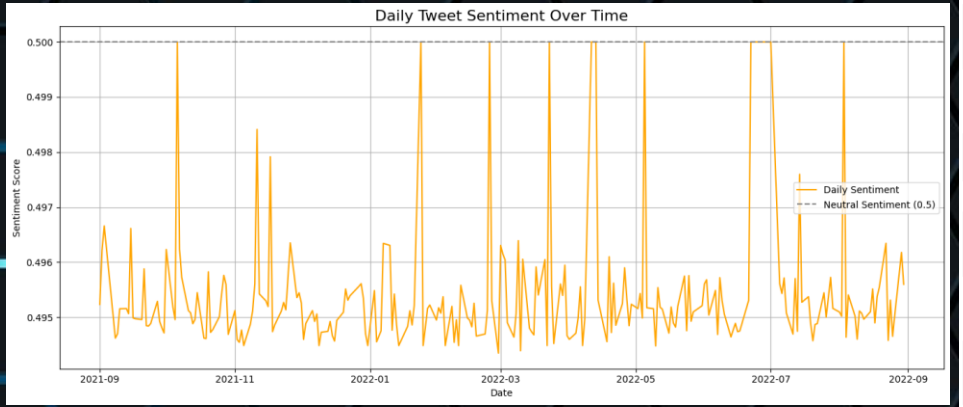
The prediction graph follows actual price trends reasonably well, including peaks and drops

no signs of overfitting

LSTM outperforms classical models with higher accuracy and stability.

Step 5- Sentiment

Graph

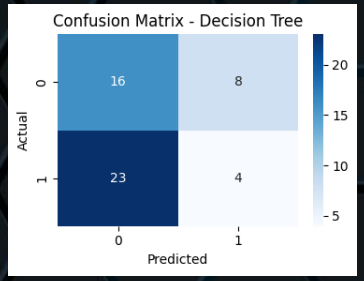


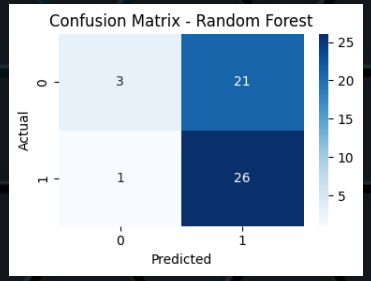
Table

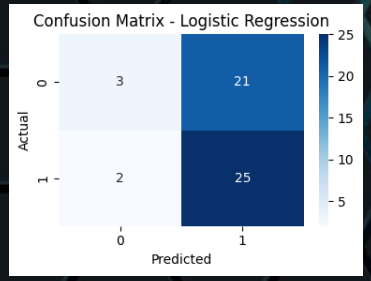
	Close	High	Low	Open	Volume	Change	label	Date	Cleaned_Tweet	sentiment_score
0	244.696671	247.330002	243.756668	244.693329	39612900	-0.566666	0.0	2021-09-01	i was in the berlin paint shop talking to th...	0.495233
1	244.130005	246.990005	243.513336	244.833328	38331900	0.393326	1.0	2021-09-02	im fkn tryin ur welcome pls be nice to the hum...	0.496211
2	244.523331	244.666672	241.399994	244.083328	45738300	6.449997	1.0	2021-09-03	this is obv debatable but id call anything wit...	0.496659
3	250.973328	253.399994	246.419998	246.666672	60119400	0.316666	1.0	2021-09-07	thanks what is your favorite tea and if it doe...	0.495029
4	251.289993	254.816666	246.923340	253.860001	56379000	0.330002	1.0	2021-09-08	that sums it up well	0.494620
...	...	...	...	...	...	...	...	...	...	...
246	297.096680	303.646667	296.500000	297.563324	57259800	-1.026672	0.0	2022-08-24	i would gladly obey their commands but i cant ...	0.495312
247	296.070007	302.959991	291.600006	302.359985	53230000	-7.980011	0.0	2022-08-25	livestream of big news in about an hour releas...	0.494653
248	288.089996	302.000000	287.470001	297.429993	57163900	-3.269989	0.0	2022-08-26	some are indeed sadly antihuman countries shou...	0.495050
249	284.820007	287.739990	280.700012	282.829987	41864700	-7.119995	0.0	2022-08-29	it will grow by at least to meters over time f...	0.496177
250	277.700012	288.480011	272.649994	287.869995	50541800	-2.090027	0.0	2022-08-30	goes out to k owners next week sjm question sp...	0.495598

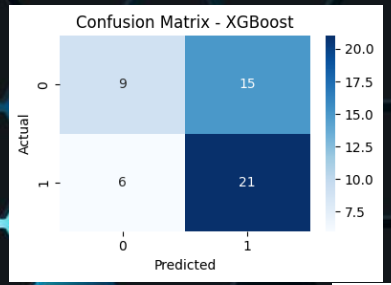
251 rows × 10 columns

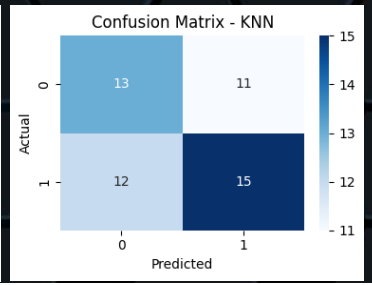
Step 6 – Clasic Models With Sentiment

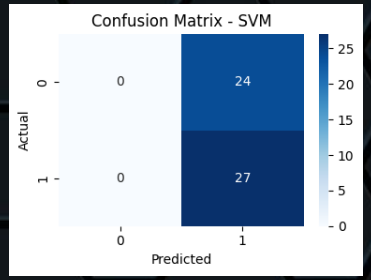


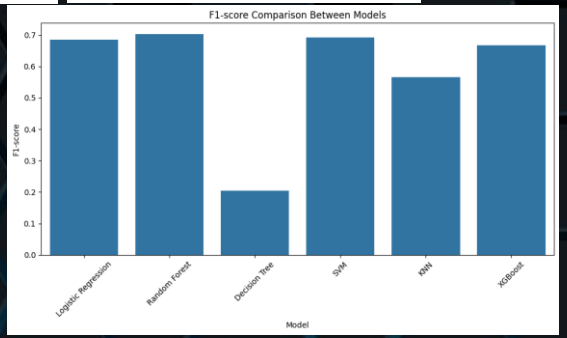












With sentiment

	Model	Accuracy	Precision	Recall	F1-score
1	Random Forest	0.568627	0.553191	0.962963	0.702703
3	SVM	0.529412	0.529412	1.000000	0.692308
0	Logistic Regression	0.549020	0.543478	0.925926	0.684932
5	XGBoost	0.588235	0.583333	0.777778	0.666667
4	KNN	0.549020	0.576923	0.555556	0.566038
2	Decision Tree	0.392157	0.333333	0.148148	0.205128

Without

	Model	Accuracy	Precision	Recall	F1-score
1	Random Forest	0.568627	0.555556	0.925926	0.694444
3	SVM	0.529412	0.529412	1.000000	0.692308
5	XGBoost	0.607843	0.594595	0.814815	0.687500
0	Logistic Regression	0.549020	0.543478	0.925926	0.684932
2	Decision Tree	0.549020	0.576923	0.555556	0.566038
4	KNN	0.549020	0.576923	0.555556	0.566038

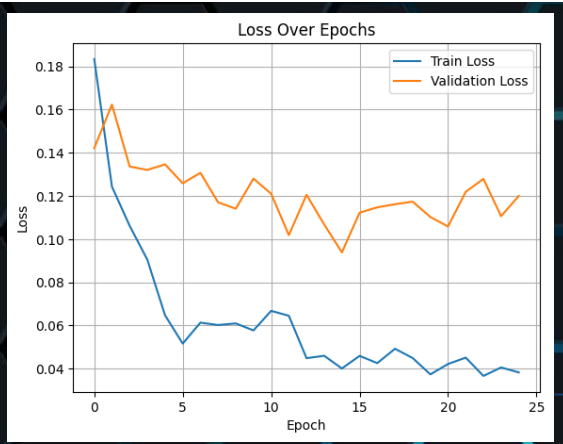
Step 6- LSTM Model With Sentiment

With Sentiment

Clasic models with sentimentic VS without sentimetic
Without - Better

LSTM with sentiment VS without sentiment
Without - Better

LSTM VS Classic models (Both with sentiment)
LSTM – Better



The task conclusions

High variance in stock data affects prediction

LSTM outperformed for trend forecasting

XGBoost was best among classic models

Sentiment model needs improvement

Thank You