تو این بخش میخواهیم ارزش سهام یک شرکت را به عنوان نمونه پیش بینی کنیم .

ما برای این کار از اطلاعات شرکت تسلا که به صورت زیر است ، استفاده کردیم :

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import seaborn as sns
import matplotlib.pyplot as plt

data = pd.read_csv("/content/gdrive/MyDrive/stocks/tsla.us.txt")

#Let's examine a few examples from our data.
data.head()
```

که تعدادی از سطر و ستون های آن به این صورت است :

	Date	Open	High	Low	Close	Volume	OpenInt
0	2010-06-28	17.00	17.00	17.00	17.00	0	0
1	2010-06-29	19.00	25.00	17.54	23.89	18783276	0
2	2010-06-30	25.79	30.42	23.30	23.83	17194394	0
3	2010-07-01	25.00	25.92	20.27	21.96	8229863	0
4	2010-07-02	23.00	23.10	18.71	19.20	5141807	0

این دیتاست دارای ۱۸۵۸ سطر و ۷ ستون است :

```
print("Data Shape -->", data.shape)

Data Shape --> (1858, 7)
```

به عنوان پیش پردازش یک سری اطلاعات مانند میانگین و میانه و ... از این دیتاست استخراج می کنیم :

	Open	High	Low	Close	Volume	OpenInt
count	1858.000000	1858.000000	1858.000000	1858.000000	1.858000e+03	1858.0
mean	150.389741	152.898737	147.688064	150.355047	4.416508e+06	0.0
std	107.071675	108.490099	105.481665	107.023737	4.244294e+06	0.0
min	16.140000	16.630000	8.030000	15.800000	0.000000e+00	0.0
25%	31.002500	31.732500	30.285000	31.112500	1.283324e+06	0.0
50%	184.440000	188.660000	181.450000	184.850000	3.421026e+06	0.0
75%	231.477500	235.375000	227.772500	230.920000	5.917672e+06	0.0
max	386.690000	389.610000	379.345000	385.000000	3.714989e+07	0.0

سپس بررسی می کنیم که مقدار Null در این دیتاست وجود دارد یا نه :

```
Do you have a null column?
 Date       0
Open       0
High       0
Low        0
Close      0
Volume     0
OpenInt    0
dtype: int64
```

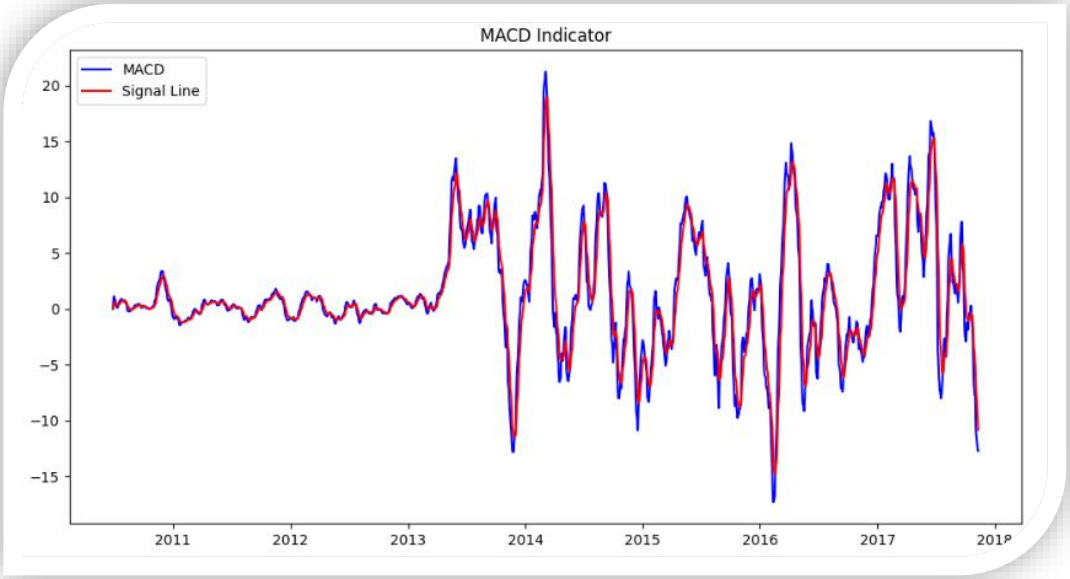
همانطور که واضح است ، مقدار Null در هیچ یک از ستون ها وجود ندارد .

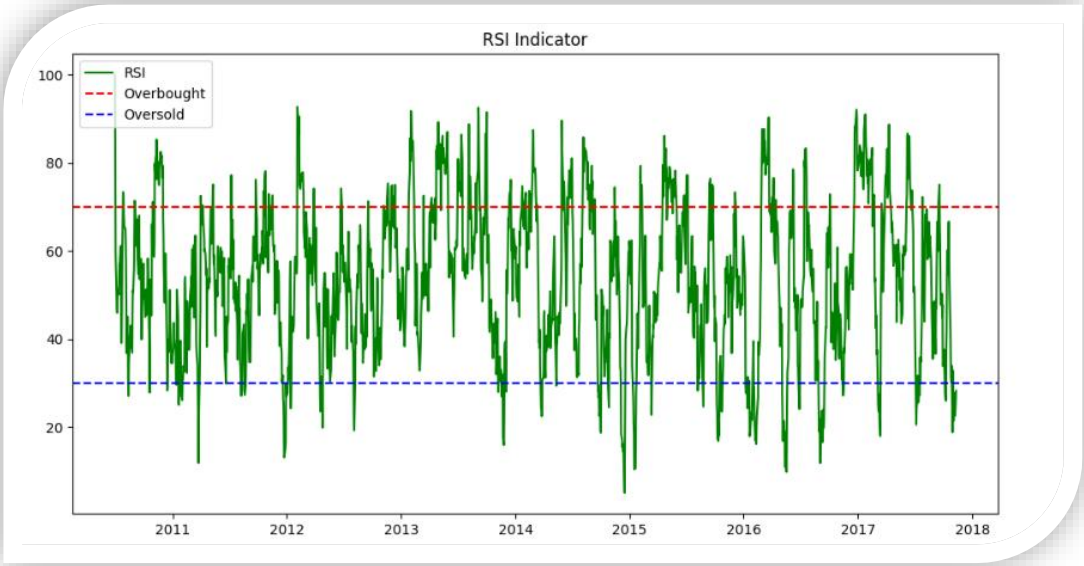
در قدم بعدی از معیار های محاسباتی دیگر مرتبط با بازار های بورس مانند MACD و RSI نیز استفاده می کنیم :

```python
# Ensure the 'Date' column is converted to datetime format
data['Date'] = pd.to_datetime(data['Date'])

# Calculate EMA (Exponential Moving Average) for MACD
short_window = 12
long_window = 26
signal_window = 9

data['EMA12'] = data['Close'].ewm(span=short_window, adjust=False).mean()
data['EMA26'] = data['Close'].ewm(span=long_window, adjust=False).mean()

# Calculate MACD and Signal Line
data['MACD'] = data['EMA12'] - data['EMA26']
data['Signal_Line'] = data['MACD'].ewm(span=signal_window, adjust=False).mean()

# Calculate RSI (Relative Strength Index)
window_length = 14

delta = data['Close'].diff(1)
gain = delta.where(delta > 0, 0)
loss = -delta.where(delta < 0, 0)

avg_gain = gain.rolling(window=window_length, min_periods=1).mean()
avg_loss = loss.rolling(window=window_length, min_periods=1).mean()

rs = avg_gain / avg_loss
data['RSI'] = 100 - (100 / (1 + rs))

# Display the data with calculated indicators
print(data[['Date', 'Close', 'MACD', 'Signal_Line', 'RSI']])

# Plot the MACD indicator
plt.figure(figsize=(12, 6))
plt.plot(data['Date'], data['MACD'], label='MACD', color='blue')
plt.plot(data['Date'], data['Signal_Line'], label='Signal Line', color='red')
plt.legend(loc='upper left')
plt.title('MACD Indicator')
plt.show()

# Plot the RSI indicator
plt.figure(figsize=(12, 6))
plt.plot(data['Date'], data['RSI'], label='RSI', color='green')
plt.axhline(70, linestyle='--', color='red', label='Overbought')
plt.axhline(30, linestyle='--', color='blue', label='Oversold')
plt.legend(loc='upper left')
plt.title('RSI Indicator')
plt.show()
```

که نتیجه ی اجرای این کد به صورت زیر است :

	Date	Close	MACD	Signal_Line	RSI
0	2010-06-28	17.00	0.000000	0.000000	NaN
1	2010-06-29	23.89	0.549630	0.109926	100.000000
2	2010-06-30	23.83	0.969201	0.281781	99.136691
3	2010-07-01	21.96	1.137707	0.452966	78.117914
4	2010-07-02	19.20	1.036590	0.569691	59.499136
...	...	...	...	...	...
1853	2017-11-06	302.78	-11.999531	-8.274504	23.147115
1854	2017-11-07	306.05	-12.311572	-9.081918	22.653061
1855	2017-11-08	304.31	-12.554550	-9.776444	24.156692
1856	2017-11-09	302.99	-12.707144	-10.362584	25.661773
1857	2017-11-10	302.99	-12.681887	-10.826445	28.305495





MACD: خط MACD تفاوت میان EMA کوتاه‌مدت (۱۲ روزه) و EMA بلندمدت (۲۶ روزه) است.

RSI: برای اندازه‌گیری شدت حرکات قیمت و تشخیص مناطق اشباع خرید یا فروش استفاده می‌شود.

در بخش بعدی از مدل های آماده برای پیش بینی استفاده می کنیم . در این قسمت از مدل DistilBert به دلیل زمان زیاد پردازش مدل Bert استفاده کردیم . مدل پیش‌فرض bert شامل ۱۱۰ میلیون پارامتر است. چون داده‌ها کوچک هستند، می‌توانیم از مدل‌های کوچک‌تر مثل distilbert استفاده کنیم . **DistilBERT** حدود ۴۰ درصد سبک‌تر از BERT است.

همچنین از فریز کردن وزن ها هم استفاده کردیم . برای استفاده از **فریز کردن وزن‌ها** (freezing weights)، لایه‌های اصلی مدل (به جز لایه‌های جدیدی که برای وظیفه پیش‌بینی اضافه شده‌اند) ثابت نگه داشته می‌شوند. این باعث کاهش زمان آموزش می‌شود، زیرا فقط لایه‌های اضافه‌شده آموزش می‌بینند.

همچنین برای کاهش زمان پردازش ، از ۵۰ درصد داده های دیتاست و ۳ ایپاک برای آموزش مدل استفاده کردیم .

با توجه به این فرضیات مدل را برای پیش بینی ارزش سهام آموزش دادیم و به نتایج زیر رسیدیم :

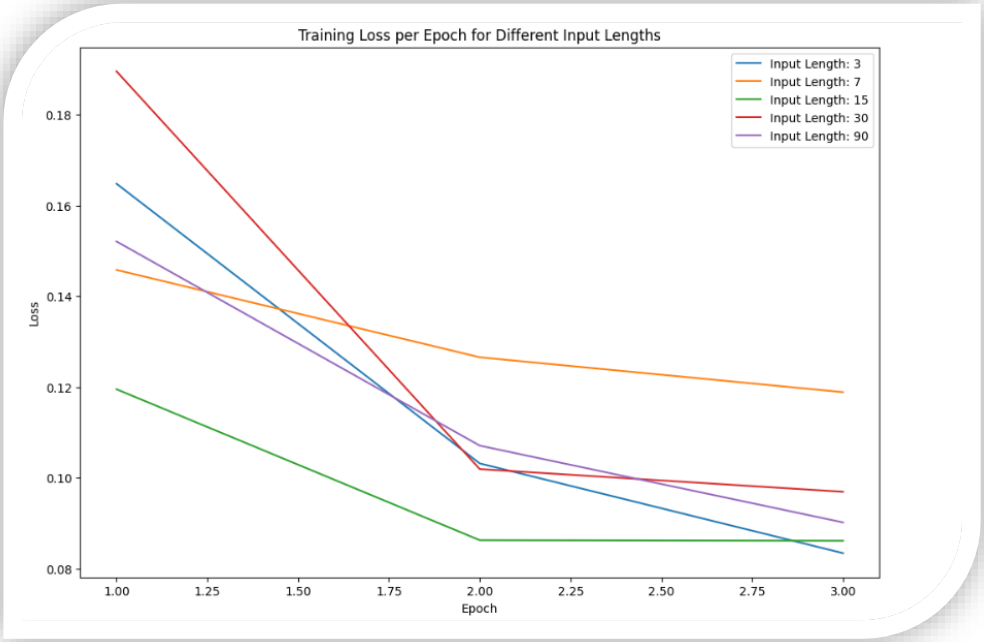
```
Input Length: 3, Epoch: 1, Loss: 0.16481662392616273
Input Length: 3, Epoch: 2, Loss: 0.10315789133310319
Input Length: 3, Epoch: 3, Loss: 0.08338293880224228
```

```
Input Length: 7, Epoch: 1, Loss: 0.14582319036126137
Input Length: 7, Epoch: 2, Loss: 0.1265606015920639
Input Length: 7, Epoch: 3, Loss: 0.11887033879756928
```

```
Input Length: 15, Epoch: 1, Loss: 0.11951293349266053
Input Length: 15, Epoch: 2, Loss: 0.08623470216989518
Input Length: 15, Epoch: 3, Loss: 0.08611071109771729
```

```
Input Length: 30, Epoch: 1, Loss: 0.18959926441311836
Input Length: 30, Epoch: 2, Loss: 0.10190499015152454
Input Length: 30, Epoch: 3, Loss: 0.09691385179758072
```

```
Input Length: 90, Epoch: 1, Loss: 0.1520886073509852
Input Length: 90, Epoch: 2, Loss: 0.10708119223515193
Input Length: 90, Epoch: 3, Loss: 0.09015892694393794
```




```
Input Length: 3, MSE: 0.09624351086635183
Input Length: 7, MSE: 0.10181001541538885
Input Length: 15, MSE: 0.1143583076751993
Input Length: 30, MSE: 0.12232927836197852
Input Length: 90, MSE: 0.11207982255776443
```

این نتایج برای پیش بینی ۱ روز آینده است .

برای پیش بینی ۷ روز آینده به نتایج زیر رسیدیم :

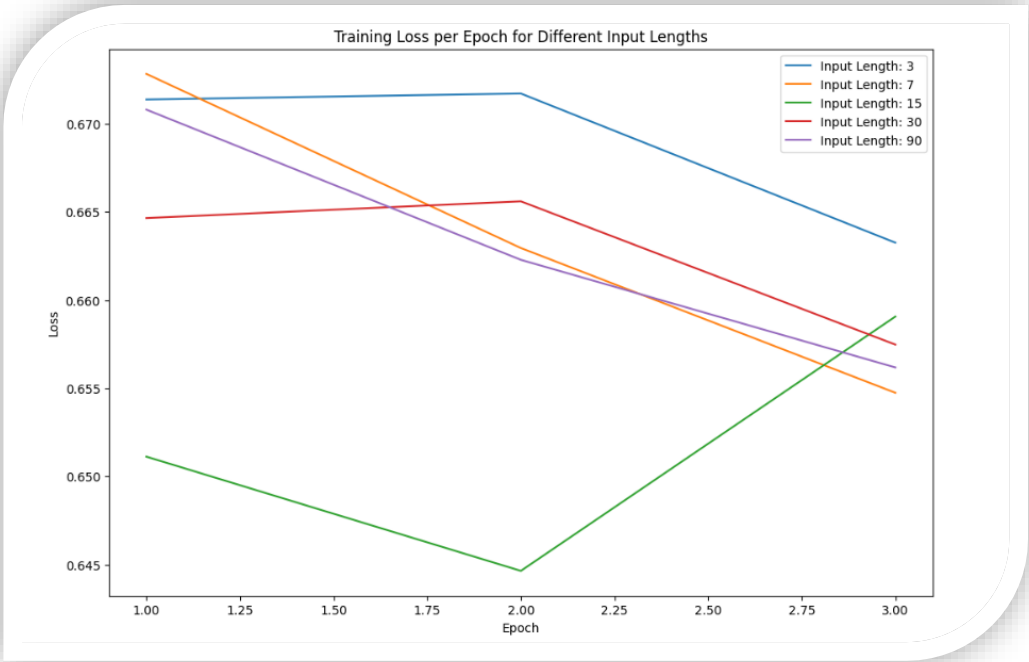
```
Input Length: 3, Epoch: 1, Loss: 0.6713655352592468
Input Length: 3, Epoch: 2, Loss: 0.6717054009437561
Input Length: 3, Epoch: 3, Loss: 0.6632513880729676
```

```
Input Length: 7, Epoch: 1, Loss: 0.6728195428848267
Input Length: 7, Epoch: 2, Loss: 0.6629488706588745
Input Length: 7, Epoch: 3, Loss: 0.6547436237335205
```

```
Input Length: 15, Epoch: 1, Loss: 0.6511232376098632
Input Length: 15, Epoch: 2, Loss: 0.6446486949920655
Input Length: 15, Epoch: 3, Loss: 0.6590629816055298
```

```
Input Length: 30, Epoch: 1, Loss: 0.6646463423967361
Input Length: 30, Epoch: 2, Loss: 0.6655951887369156
Input Length: 30, Epoch: 3, Loss: 0.6574746668338776
```

```
Input Length: 90, Epoch: 1, Loss: 0.6707930763562521
Input Length: 90, Epoch: 2, Loss: 0.6622796058654785
Input Length: 90, Epoch: 3, Loss: 0.6561793684959412
```



برای پیش بینی ۳۰ روز آینده به نتایج زیر رسیدیم :

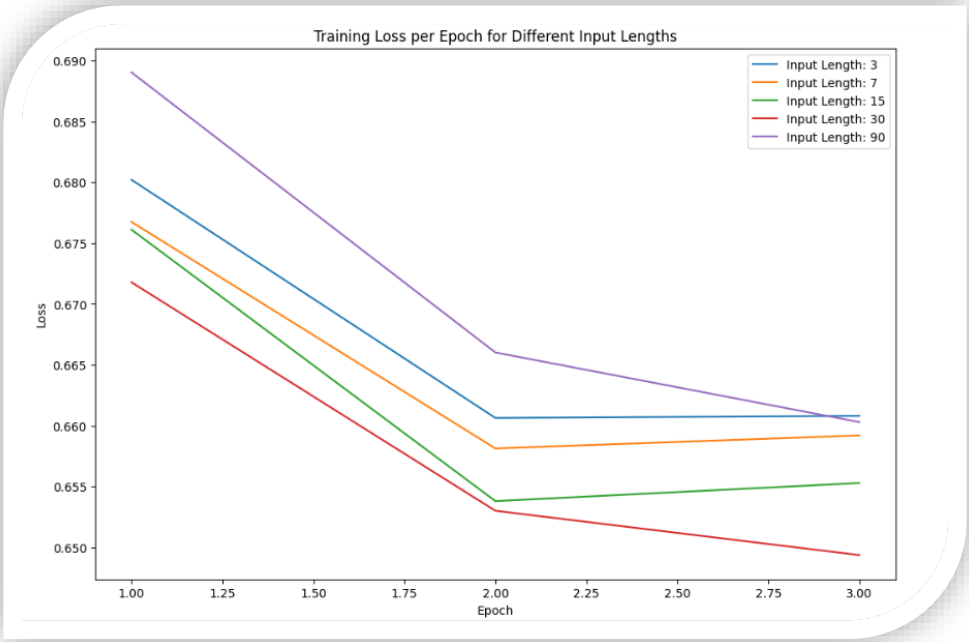
```
Input Length: 3, Epoch: 1, Loss: 0.6802013665437698
Input Length: 3, Epoch: 2, Loss: 0.6606388539075851
Input Length: 3, Epoch: 3, Loss: 0.6608099788427353
```

```
Input Length: 7, Epoch: 1, Loss: 0.6767406463623047
Input Length: 7, Epoch: 2, Loss: 0.6581333875656128
Input Length: 7, Epoch: 3, Loss: 0.6592044979333878
```

```
Input Length: 15, Epoch: 1, Loss: 0.6761045008897781
Input Length: 15, Epoch: 2, Loss: 0.6538086086511612
Input Length: 15, Epoch: 3, Loss: 0.655292809009552
```

```
Input Length: 30, Epoch: 1, Loss: 0.6717807799577713
Input Length: 30, Epoch: 2, Loss: 0.6530149132013321
Input Length: 30, Epoch: 3, Loss: 0.649376168847084
```

```
Input Length: 90, Epoch: 1, Loss: 0.6890235841274261
Input Length: 90, Epoch: 2, Loss: 0.6660170555114746
Input Length: 90, Epoch: 3, Loss: 0.6602992713451385
```



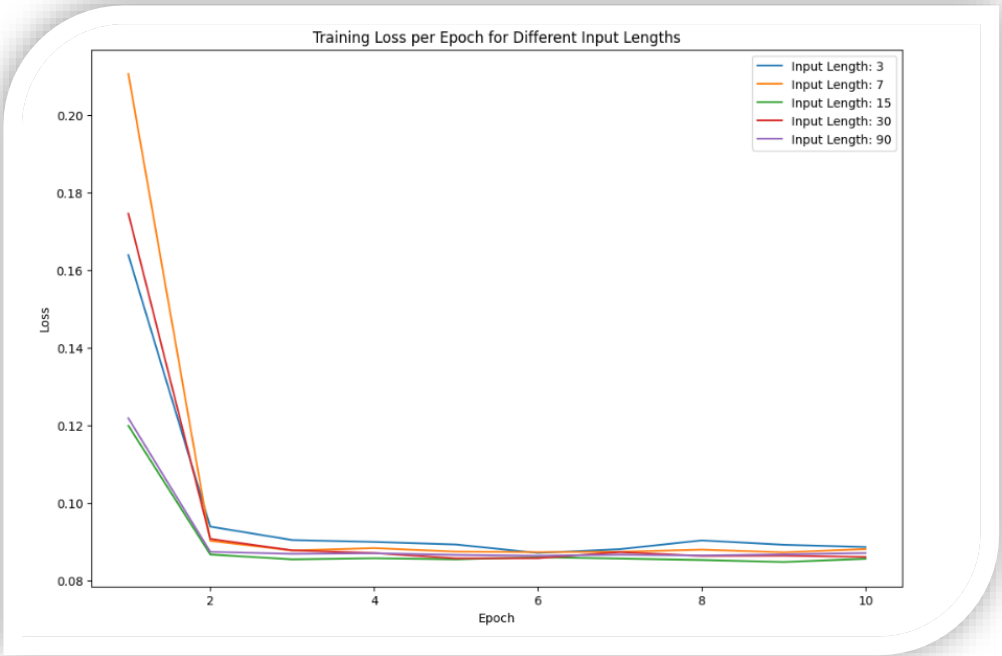

```
Input Length: 3, MSE: 1.1672089834489163
Input Length: 7, MSE: 1.1087089582096068
Input Length: 15, MSE: 1.3429469708387318
Input Length: 30, MSE: 1.3679832527317926
Input Length: 90, MSE: 1.2278182930316581
```

در قدم بعدی بجای مدل BERT از مدل LSTM با همان فرضیات قبلی استفاده کردیم :

```python
# Define the LSTM model
class LSTMModel(nn.Module):
    def __init__(self, input_size, hidden_size, num_layers, output_size):
        super(LSTMModel, self).__init__()
        self.lstm = nn.LSTM(input_size, hidden_size, num_layers, batch_first=True)
        self.fc = nn.Linear(hidden_size, output_size)

    def forward(self, x):
        out, _ = self.lstm(x)
        out = self.fc(out[:, -1, :])  # Use the output of the last time step
        return out
```

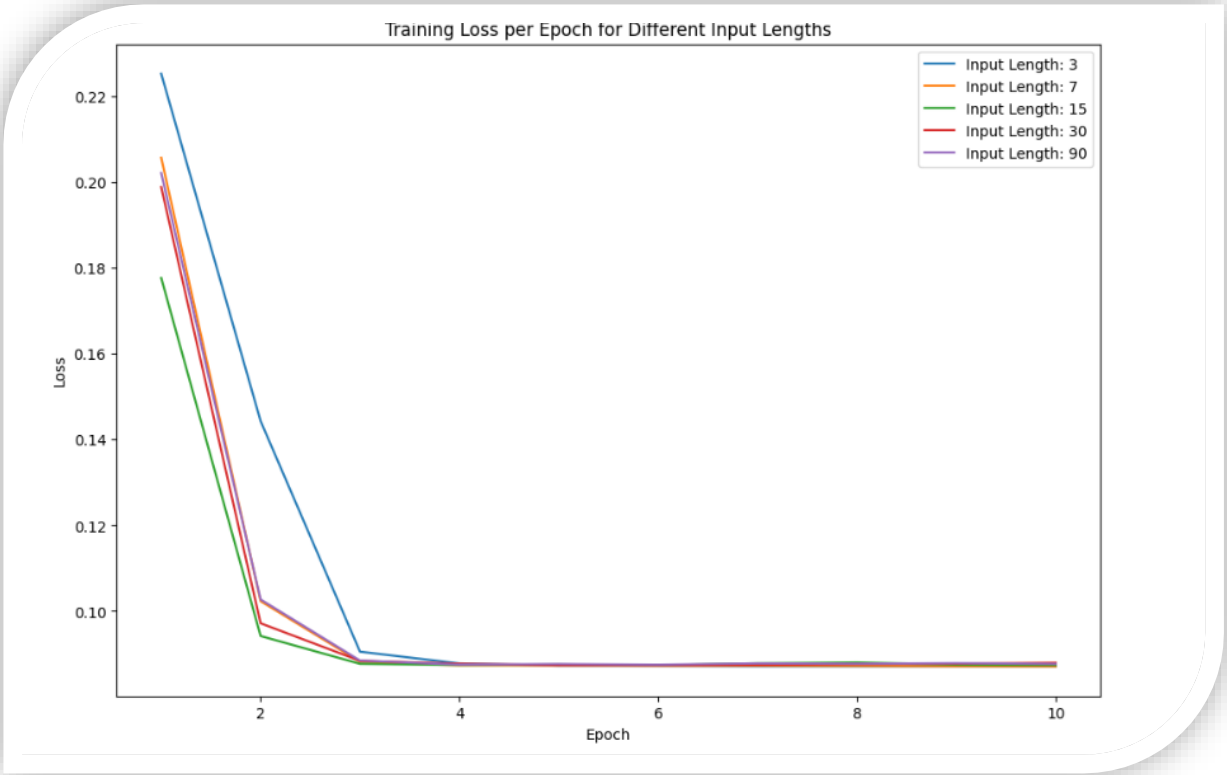
این مدل شامل یک لایه LSTM و یک لایه Fully Connected است .

نتایج به دست آمده برای پیش بینی ۱ روز آینده به صورت زیر است :



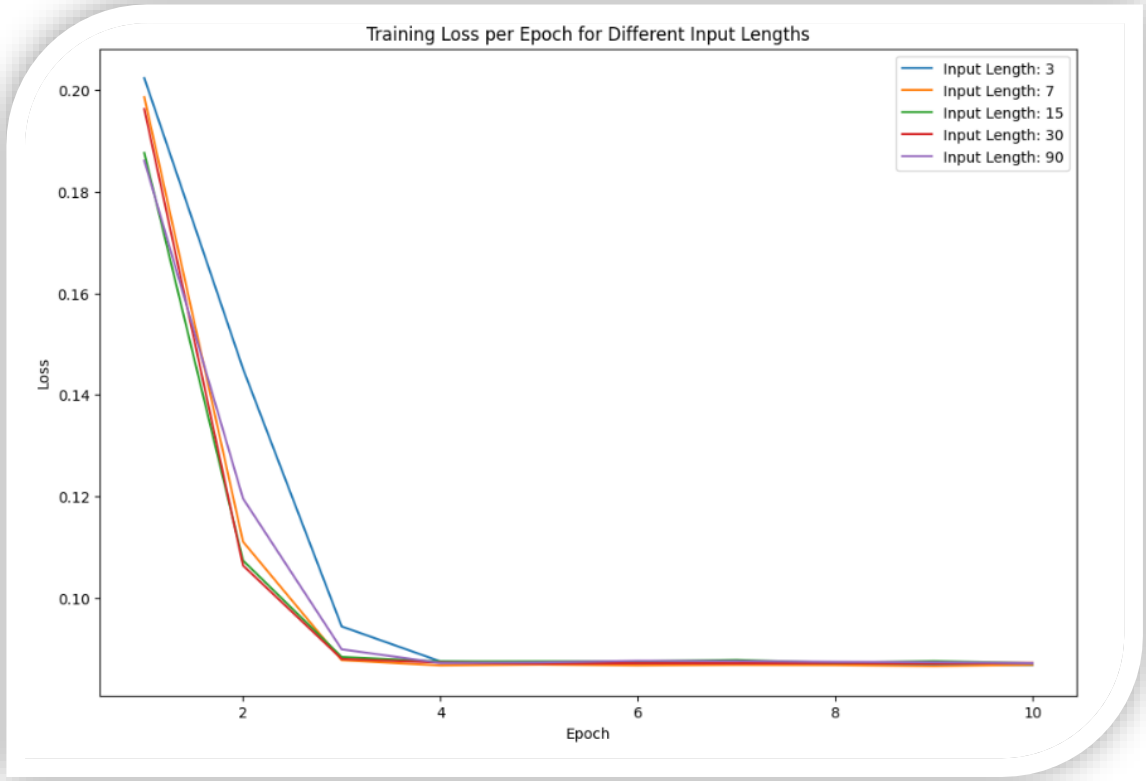

```
Final Results (MSE):
Input Length: 3, MSE: 0.0845
Input Length: 7, MSE: 0.0872
Input Length: 15, MSE: 0.0948
Input Length: 30, MSE: 0.0937
Input Length: 90, MSE: 0.0887
```

نتایج به دست آمده برای پیش بینی ۷ روز آینده به صورت زیر است :




```
Final Results (MSE):
Input Length: 3, MSE: 0.0864
Input Length: 7, MSE: 0.0857
Input Length: 15, MSE: 0.0848
Input Length: 30, MSE: 0.0869
Input Length: 90, MSE: 0.0838
```

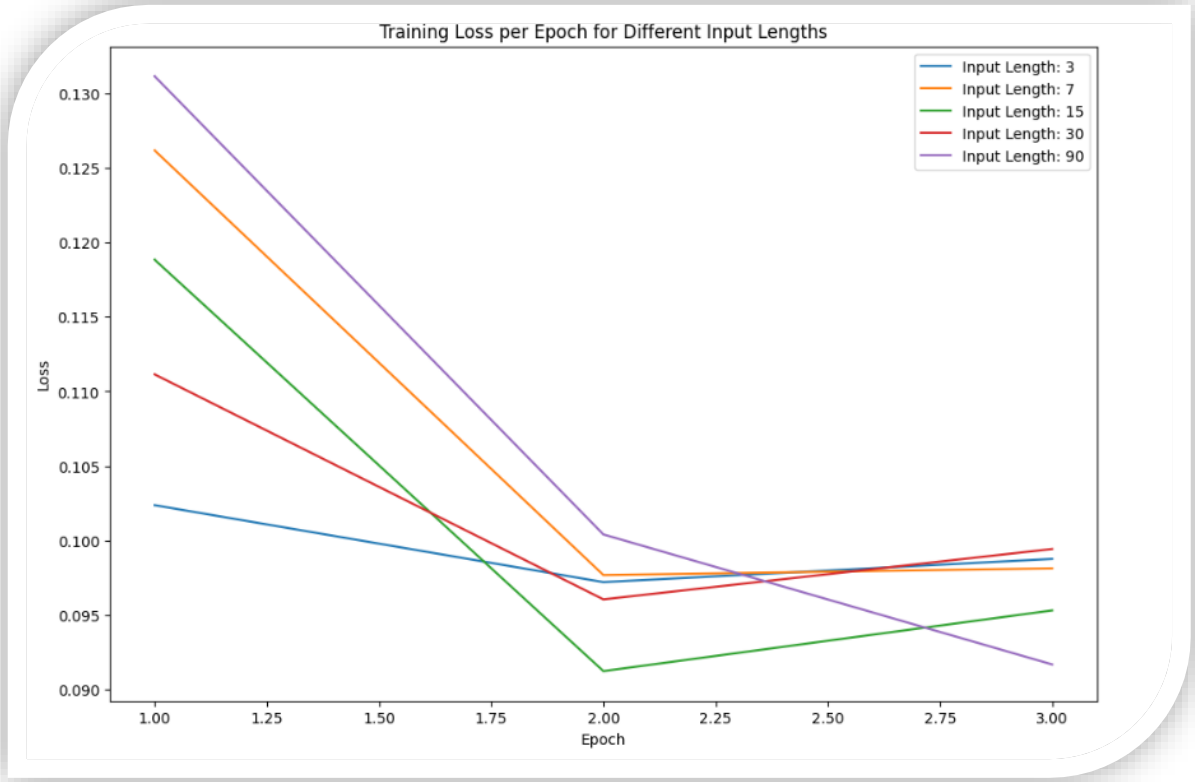
نتایج به دست آمده برای پیش بینی ۳۰ روز آینده به صورت زیر است :




```
Final Results (MSE):
Input Length: 3, MSE: 0.0863
Input Length: 7, MSE: 0.0871
Input Length: 15, MSE: 0.0853
Input Length: 30, MSE: 0.0859
Input Length: 90, MSE: 0.0855
```

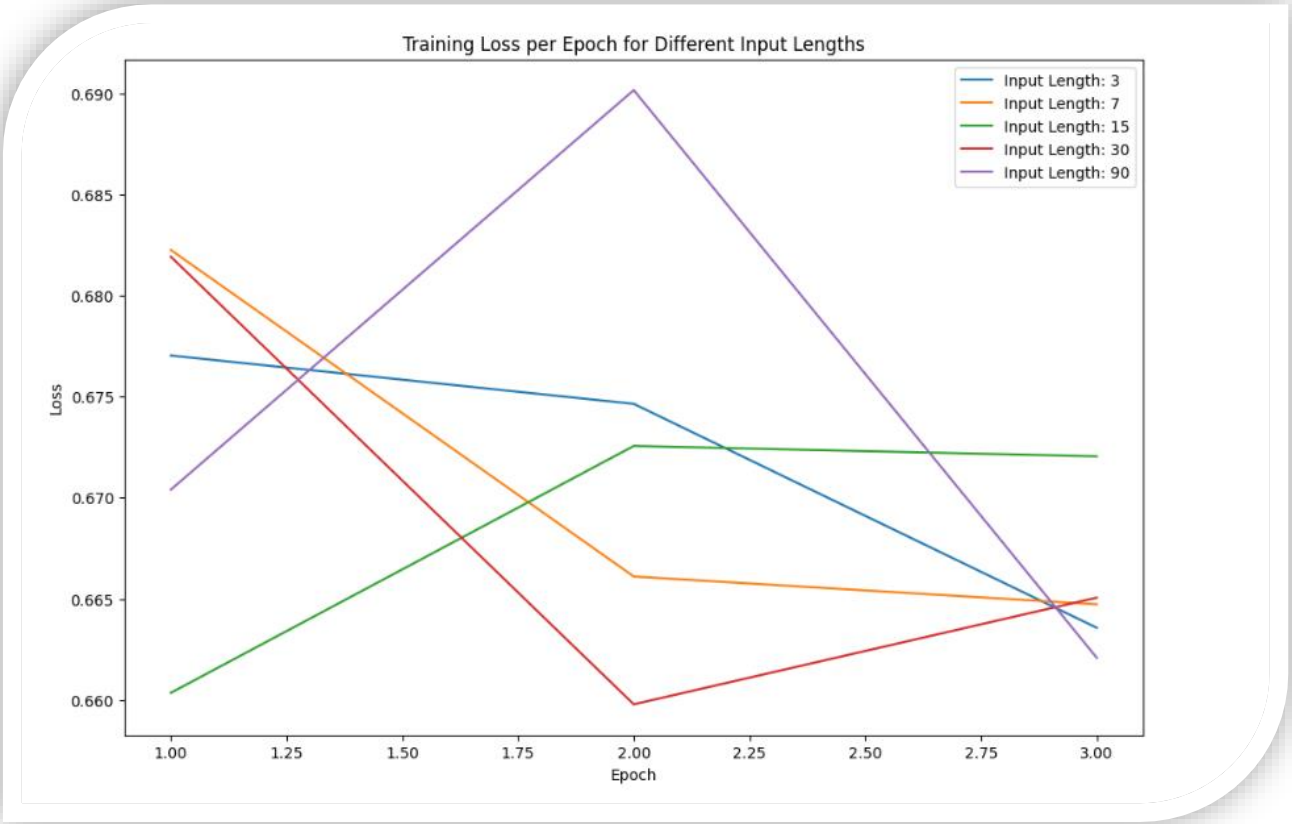
در قسمت بعدی از مدل آماده ی RoBERTa استفاده کردیم . در این مدل هم با همان فرضیات قبلی از جمله فریز کردن وزن ها و استفاده از ۵۰ درصد داده ها برای آموزش و ... برای کاهش زمان پردازش ، مدل را آموزش دادیم .

نتایج به دست آمده برای پیش بینی ۱ روز آینده به صورت زیر است :



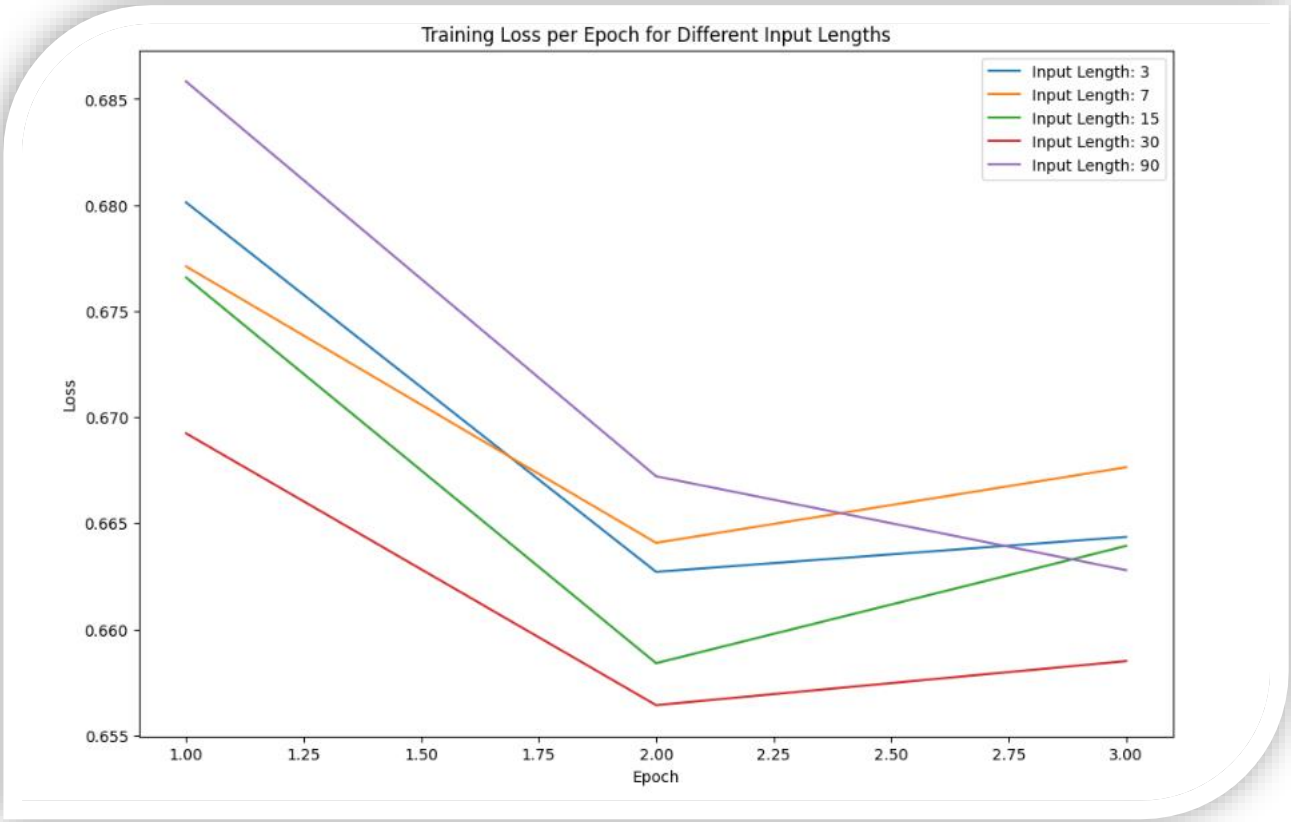

```
Final Results (MSE):
Input Length: 3, MSE: 0.0842404052713824
Input Length: 7, MSE: 0.09154357840785729
Input Length: 15, MSE: 0.09802820393970321
Input Length: 30, MSE: 0.11172072488821153
Input Length: 90, MSE: 0.0885475634154334
```

نتایج به دست آمده برای پیش بینی ۷ روز آینده به صورت زیر است :




```
Final Results (MSE):
Input Length: 3, MSE: 0.7115779135353707
Input Length: 7, MSE: 0.7703081307264299
Input Length: 15, MSE: 0.8736545373487702
Input Length: 30, MSE: 0.9353704754591808
Input Length: 90, MSE: 0.9409487330854713
```

نتایج به دست آمده برای پیش بینی ۳۰ روز آینده به صورت زیر است :




```
Final Results (MSE):
Input Length: 3, MSE: 1.050149863080413
Input Length: 7, MSE: 0.9419036517436097
Input Length: 15, MSE: 1.1778039365611763
Input Length: 30, MSE: 1.233239028775089
Input Length: 90, MSE: 0.9904619682836617
```