

ML Engineer Intern – Project Report

Submitted by- Shruti Shukla

1. Introduction

This project involves developing machine learning models for financial forecasting using real-world stock market and cryptocurrency datasets. The goal is to preprocess the data, engineer relevant features, train predictive models, evaluate them, and deploy the final system in an interactive user interface.

The project uses two datasets processed separately:

- **stocks.csv** → Stock High Price Prediction
- **cryptocurrency.csv** → Crypto Closing Price Prediction

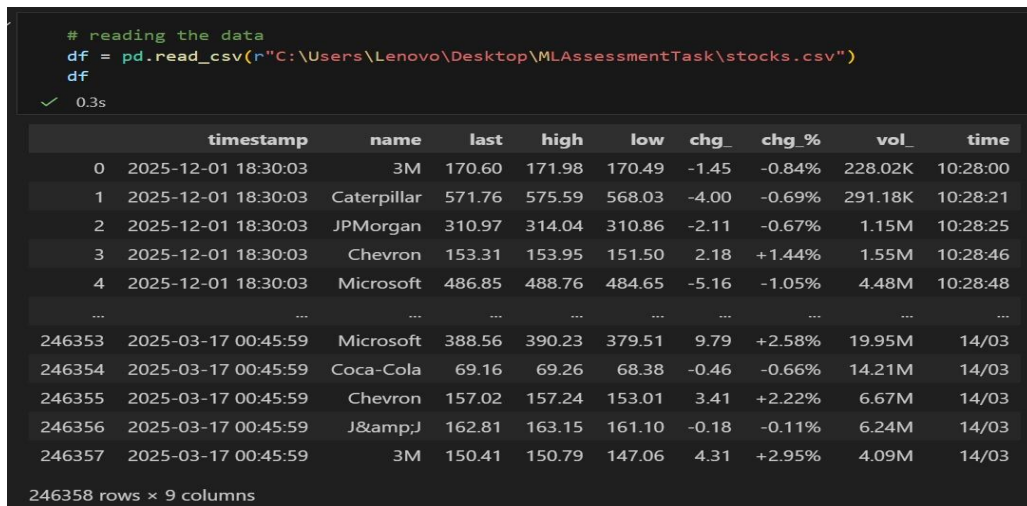
SECTION A – STOCK MARKET PREDICTION (stocks.csv)

2. Dataset Overview – Stocks

The stocks.csv dataset contains historical records for multiple stock assets, including:

- Date
- Open, High, Low, Close
- Volume
- Asset Name

This dataset was filtered to select specific stock tickers required for forecasting the next-day High Price.



```
# reading the data
df = pd.read_csv(r"C:\Users\Lenovo\Desktop\MLAssessmentTask\stocks.csv")
df
```

✓ 0.3s

	timestamp	name	last	high	low	chg_	chg_%	vol_	time
0	2025-12-01 18:30:03	3M	170.60	171.98	170.49	-1.45	-0.84%	228.02K	10:28:00
1	2025-12-01 18:30:03	Caterpillar	571.76	575.59	568.03	-4.00	-0.69%	291.18K	10:28:21
2	2025-12-01 18:30:03	JPMorgan	310.97	314.04	310.86	-2.11	-0.67%	1.15M	10:28:25
3	2025-12-01 18:30:03	Chevron	153.31	153.95	151.50	2.18	+1.44%	1.55M	10:28:46
4	2025-12-01 18:30:03	Microsoft	486.85	488.76	484.65	-5.16	-1.05%	4.48M	10:28:48
...
246353	2025-03-17 00:45:59	Microsoft	388.56	390.23	379.51	9.79	+2.58%	19.95M	14/03
246354	2025-03-17 00:45:59	Coca-Cola	69.16	69.26	68.38	-0.46	-0.66%	14.21M	14/03
246355	2025-03-17 00:45:59	Chevron	157.02	157.24	153.01	3.41	+2.22%	6.67M	14/03
246356	2025-03-17 00:45:59	J&J	162.81	163.15	161.10	-0.18	-0.11%	6.24M	14/03
246357	2025-03-17 00:45:59	3M	150.41	150.79	147.06	4.31	+2.95%	4.09M	14/03

246358 rows × 9 columns

3. Data Preprocessing – Stocks

The following preprocessing steps were executed:

3.1 Dataset Cleaning

- Converted Date column to proper datetime format.
- Handled missing values using forward fill.
- Filtered required stock tickers.

3.2 Feature Engineering

multiple financial indicators:

- Moving Averages (5-day, 10-day, 20-day)
- Lag Features (1–3 day lags of High Price)
- Daily Returns
- Price Range (High – Low)
- Normalized Features using StandardScalar.

	timestamp	name	last	high	low	chg	chg_%	vol_	time	MA7	MA30	last_lag1	last_lag2	high_lag1	low_lag1
245487	2025-03-18 05:46:00	Amazon.com	-0.710820	-0.702786	-0.700737	-2.21	-1.12%	-0.188625	15:59:59	-0.710634	-0.697770	-0.710752	-0.710684	-0.702720	-0.700737
245428	2025-03-18 06:46:00	Amazon.com	-0.710820	-0.702786	-0.700737	-2.21	-1.12%	-0.188625	15:59:59	-0.710634	-0.698412	-0.710752	-0.710684	-0.702720	-0.700737
245398	2025-03-18 07:46:01	Amazon.com	-0.710820	-0.702786	-0.700737	-2.21	-1.12%	-0.188625	17/03	-0.710634	-0.699053	-0.710752	-0.710684	-0.702720	-0.700737
245386	2025-03-18 08:46:00	Amazon.com	-0.710820	-0.702786	-0.700737	-2.21	-1.12%	-0.188625	17/03	-0.710634	-0.699695	-0.710752	-0.710684	-0.702720	-0.700737
245353	2025-03-18 09:46:00	Amazon.com	-0.710820	-0.702786	-0.700737	-2.21	-1.12%	-0.188625	17/03	-0.710634	-0.700337	-0.710752	-0.710684	-0.702720	-0.700737
...
158	2025-12-01 16:00:03	NVIDIA	-0.874070	-0.873611	-0.856665	-3.26	-1.81%	0.782758	28/11	-0.873892	-0.873204	-0.874003	-0.873936	-0.873545	-0.856665
121	2025-12-01 16:30:04	NVIDIA	-0.874070	-0.873611	-0.856665	-3.26	-1.81%	0.782758	28/11	-0.873892	-0.873204	-0.874003	-0.873936	-0.873545	-0.856665
76	2025-12-01 17:00:02	NVIDIA	-0.874070	-0.873611	-0.856665	-3.26	-1.81%	0.782758	28/11	-0.873892	-0.873204	-0.874003	-0.873936	-0.873545	-0.856665
48	2025-12-01 17:30:02	NVIDIA	-0.874070	-0.873611	-0.856665	-3.26	-1.81%	0.782758	28/11	-0.873892	-0.873204	-0.874003	-0.873936	-0.873545	-0.856665
32	2025-12-01 18:00:03	NVIDIA	-0.866666	-0.878811	-0.881340	0.85	+0.48%	-0.309014	9:59:10	-0.872834	-0.872957	-0.874003	-0.873936	-0.873545	-0.856665

4. Model Development – Stocks

Multiple ML models were trained to predict the next-day High price:

- Random Forest Regressor
- Gradient Boosting Regressor
- XGBoost Regressor
- MLP Regressor

```
# initializing the models
models = {
    "Random Forest": RandomForestRegressor(n_estimators=100, random_state=42),
    "Gradient Boosting": GradientBoostingRegressor(n_estimators=100, learning_rate=0.1, random_state=42),
    "XGBoost": xgb.XGBRegressor(n_estimators=100, learning_rate=0.1, random_state=42),
    "MLP Regressor": MLPRegressor(hidden_layer_sizes=(100,100), max_iter=500, random_state=42)
}
```

✓ 0.0s

4.1 Model Selection

RandomForest delivered the best performance based on validation metrics.

4.2 Evaluation Metrics (Stocks)

Metric	Result
MAE	0.000365
RMSE	0.001604
R ² Score	0.999998

```
# Training and evaluating each model
results = {}

for name, model in models.items():
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)

    mse = mean_squared_error(y_test, y_pred)
    rmse = np.sqrt(mse)
    mae = mean_absolute_error(y_test, y_pred)
    r2 = r2_score(y_test, y_pred)

    results[name] = {
        "RMSE": rmse,
        "MAE": mae,
        "R2 Score": r2
    }

# Display results
results_df = pd.DataFrame(results).T
results_df = results_df.sort_values(by="R2 Score", ascending=False)
print(results_df)
```

✓ 1.0s

	RMSE	MAE	R2 Score
Random Forest	0.001604	0.000365	0.999998
Gradient Boosting	0.002029	0.000395	0.999996
XGBoost	0.004035	0.000649	0.999986
MLP Regressor	0.024804	0.009939	0.999453

SECTION B – CRYPTOCURRENCY PREDICTION (cryptocurrency.csv)

6. Dataset Overview – Crypto

The cryptocurrency.csv dataset contains historical values for multiple cryptocurrencies, including:

- Date
- Open, High, Low, Close
- Volume
- Asset Name

The task focuses on forecasting the Future Closing Price of selected crypto assets.

```
# Import dataset
df = pd.read_csv("cryptocurrency.csv")
✓ 0.1s

# Inspect dataset
df.head()
✓ 0.0s
```

	timestamp	name	symbol	price_usd	vol_24h	total_vol	chg_24h	chg_7d	market_cap
0	2025-12-01 23:26:07	USDS	usds	\$1.00	\$12,927,922.00	0.00%	+0.00%	+0.01%	\$9,431,355,896
1	2025-12-01 23:26:07	LEO Token	leo	\$9.85	\$795,012.00	0.00%	+0.15%	+3.35%	\$9,084,160,660
2	2025-12-01 23:26:07	sUSDS	susds	\$1.08	\$2,825,118.00	0.00%	+0.17%	+0.09%	\$4,102,962,214
3	2025-12-01 23:26:07	Figure Heloc	figr_heloc	\$1.03	\$18,622,760.00	0.01%	+1.15%	+0.32%	\$14,359,098,283
4	2025-12-01 23:26:07	Wrapped Beacon ETH	wbeth	\$2,988.07	\$23,282,537.00	0.01%	-9.14%	-6.59%	\$9,766,387,527

7. Data Preprocessing – Crypto

7.1 Cleaning

- Converted Date to datetime format
- Sorted chronologically
- Addressed missing data

7.2 Feature Engineering

Generated indicators commonly used in financial forecasting:

- Moving Averages (5, 10, 50 days)
- Lag Features (1–5 days of Close Price)
- Volatility Indicators
- Daily Returns

Normalization

Scaling applied based on model type (especially for LSTM).

	MA7	MA30	price_lag1	daily_return	price_range_proxy
0	0.326112	0.330028	0.325759	0.681885	0.123430
1	0.326378	0.329995	0.325867	0.691256	0.123548
2	0.326506	0.329951	0.326296	0.675120	0.122970
3	0.326539	0.329911	0.326171	0.677720	0.123748
4	0.326606	0.329844	0.326135	0.679801	0.123859

8. Model Development – Cryptocurrency

Models trained for Closing Price prediction:

- Linear Regression
- Random Forest
- XGBoost
- LSTM Neural Network

8.1 Model Selection

The **LSTM model** produced the most stable and accurate trend predictions.

8.2 Evaluation Metrics (Crypto)

Metric	Result
MAE	0.5911
RMSE	1.5384
R ² Score	1.000

```

super().__init__(**kwargs)
9/9 ————— 0s 38ms/step
Linear Regression --> MAE: 0.5911, RMSE: 1.5384, R2: 1.0000
Random Forest --> MAE: 3.5792, RMSE: 4.0425, R2: 0.9997
XGBoost --> MAE: 10.7149, RMSE: 54.8240, R2: 0.9457
LSTM --> MAE: 135.3056, RMSE: 219.7416, R2: 0.1280

```

SECTION C – DEPLOYMENT

10. UI Application Development

A combined UI was built using Flask .

The application contains two independent modules:

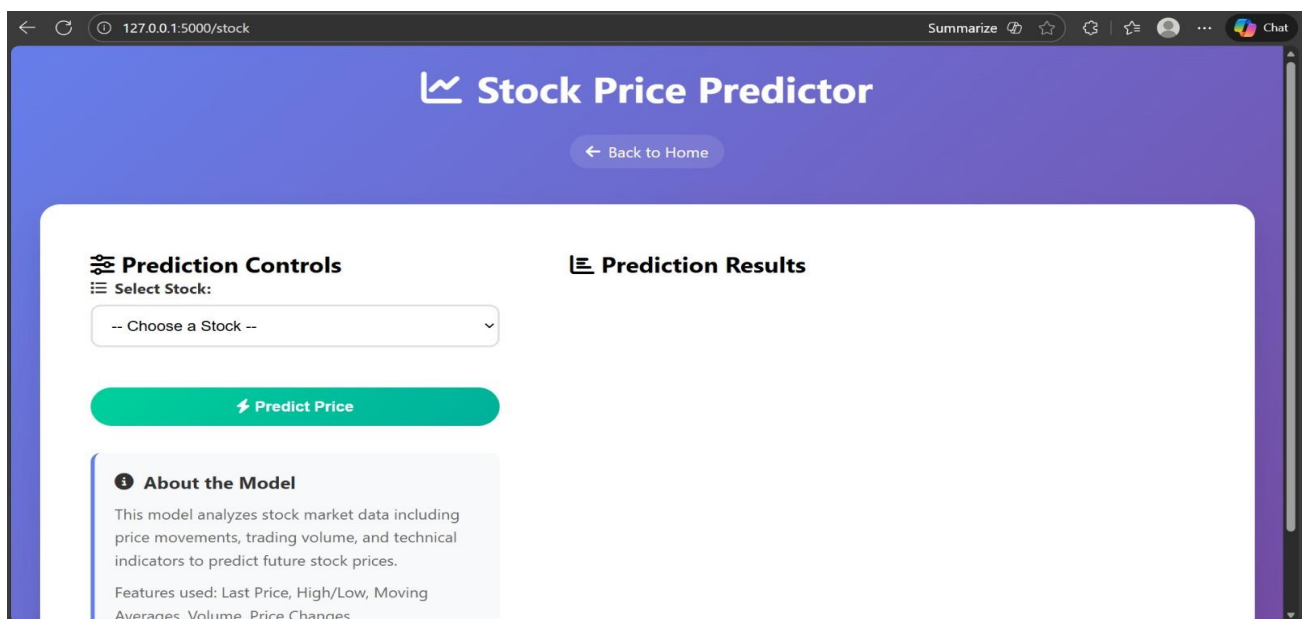
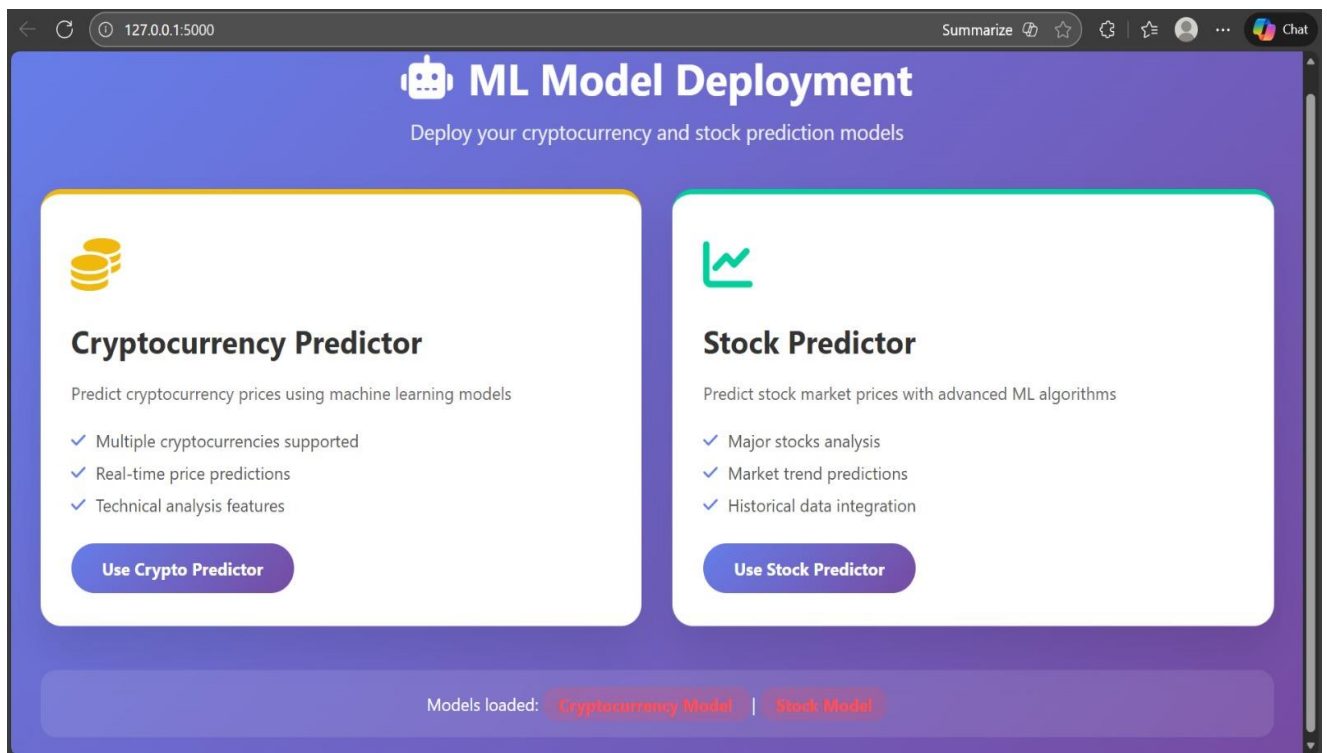
10.1 Cryptocurrency Module

- User selects a cryptocurrency
- Predicts future closing prices
- Displays forecast graphs & comparison charts

10.2 Stock Prediction Module

- User selects a stock ticker
- Predicts next-day High price

14. Attachments Section





Cryptocurrency Price Predictor

[← Back to Home](#)

Prediction Controls

Select Cryptocurrency:

-- Choose a Cryptocurrency --

⚡ Predict Price

About the Model

This model analyzes historical price data, trading volume, and technical indicators to predict future cryptocurrency prices.

Features used: Price, Volume, Moving Averages, Price Changes, Market Cap

Prediction Results

MLASSESSMENTTASK

static

style.css

templates

<> crypto_prediction.html

<> index.html

<> stock_prediction.html

🐍 app.py

≡ best_crypto_model.joblib

≡ best_stock_model.joblib

📄 cryptocurrency.csv

📄 cryptocurrency.ipynb

📖 README.md

≡ requirements.txt

🐍 sample_data.py

📄 stocks.csv

📄 Stocks.ipynb