



# Machine Learning and AI Model For Stock Price Prediction

Celal YELGEL

- To improve operations and investment performances in financial institutions more and more investments are made on ML and DL algorithms
- Most of the AI, ML, and DL techniques require such large amounts of data for data-driven finance
- The prediction of stock prices is a very complex and challenging field for research
- ML and DL as the most promising approaches for solving complex problems of classification

36	0.00	5,239,900	11,977	564,285
95	-3.66	5,825,700	2,322,248	18,089
75	+0.88	5,859,400	33,569	4,310
57	+1.95	2,399,100	3,776	3,300
5	-0.99	479,300	2,389	1,313
37	-0.72	182,000	249	2,788
22	0.00	0	0	0.76
76	-1.30	17,288,600	13,224	3,491
79	0.00	249,300	197	368
14.1	+4.44	2,913,100	40,573	3,540
3.42	0.00	0	0	1,051
3.34	0.00	0	0	5.65
5.65	0.00	16,273,100	93,345	3,789
82.5	+0.92	4,700	385	3,108
4.2	+1.45	247,000	1,033	2,640
4.7	0.00	38,265,200	183,197	1,520
20.7	+0.98	3,888,600	81,029	1,214
				7,284

- AI powered robo-advisers can perform real-time analysis on massive datasets and trade securities at an extremely faster rate compared to human traders
- AI-powered trading could potentially reduce risk and maximise returns
- By accurately predicting stock prices, investors can maximise returns and know when to buy or sell securities

36	0.00	5,239,900	11,901	
95	-3.66	5,825,700	2,322,248	564,285
75	+0.88	5,859,400	33,569	18,089
57	+1.95	2,399,100	3,776	4,310
5	-0.99	479,300	2,389	3,300
37	-0.72	182,000	249	1,313
22	0.00	0	0	2,788
76	-1.30	17,288,600	13,224	3,491
79	0.00	249,300	197	368
14.1	+4.44	2,913,100	40,573	3,540
3.42	0.00	0	0	1,051
3.34	0.00	0	0	3,789
82.5	+0.92	4,700	385	3,108
4.2	+1.45	247,000	1,033	2,640
4.7	0.00	38,265,200	183,197	1,520
20.7	+0.98	3,888,600	81,029	7,284

# Methodology

- Prefer exploratory data analysis and visitation
- Prepare the data for the training and AI/ML model
- Data set is divided into 85% for training and 15% for testing
- Training set: used for model training
- Testing set: used for testing trained model. Make sure that the testing dataset has never been seen by trained model before
- Used Ridge regression to avoid overfitting

36	0.00	5,239,900	11,901	564,285
95	-3.66	5,825,700	2,322,248	18,089
75	+0.88	5,859,400	33,569	4,310
57	+1.95	2,399,100	3,776	3,300
5	-0.99	479,300	2,389	2,788
37	-0.72	182,000	249	1,313
22	0.00	0	0	3,491
76	-1.30	17,288,600	13,224	368
79	0.00	249,300	197	3,540
14.1	+4.44	2,913,100	40,573	1,051
3.42	0.00	0	0	3,789
3.34	0.00	0	0	3,108
5.65	0.00	16,273,100	93,345	2,640
82.5	+0.92	4,700	385	1,520
4.2	+1.45	247,000	1,033	7,284
4.7	0.00	38,265,200	183,197	1,214
20.7	+0.98	3,888,600	81,029	

## Original Vs. Prediction



Linear Regression Score: 0.9585799112345615

# Future Work

- **LSTM**
- **XGBOOST**
- **Convolutional Neural Network**
- **Sentiment Analysis**

36	0.00	5,239,900	11,901	564,285
95	-3.66	5,825,700	2,322,248	18,089
75	+0.88	5,859,400	33,569	4,310
57	+1.95	2,399,100	3,776	3,300
5	-0.99	479,300	2,389	1,313
37	-0.72	182,000	249	2,788
22	0.00	0	0	0
76	-1.30	17,288,600	13,224	3,491
79	0.00	249,300	197	368
14.1	+4.44	2,913,100	40,573	3,540
3.42	0.00	0	0	1,051
3.34	0.00	0	0	3,789
5.65	0.00	16,273,100	93,345	3,108
82.5	+0.92	4,700	385	2,640
4.2	+1.45	247,000	1,033	1,520
4.7	0.00	38,265,200	183,197	1,214
20.7	+0.98	3,888,600	81,029	7,284