Ensemble Machine Learning Approach To Stock Market Prediction

Ku Jing Hao | 1201203310

Prof. Dr. Ting Choo Yee | Dr. Goh Hui Ngo

FYP02-DS-T2510-0032

MULTIMEDIA UNIVERSITY

FACULTY OF COMPUTING & INFORMATICS

Abstract

Stock price prediction remains challenging due to the complex, nonlinear dynamics of financial markets that traditional technical analysis methods often fail to capture. This research introduces an **ensemble machine learning framework** that integrates multiple models using an **averaging ensemble method** to improve stock price predictions. The framework combines linear and nonlinear modeling approaches to address individual model limitations, such as poor generalization and sensitivity to market volatility. Comprehensive evaluations were conducted using historical stock data across **hourly and daily intervals**. The results revealed a counterintuitive finding: while **Linear Regression** consistently outperformed the best-performing **LR + LSTM** ensemble in **predictive accuracy** in hourly scenarios, and **LSTM** outperformed the **LR + LSTM + GRU** ensemble in daily intervals. However, **ensemble methods consistently surpassed individual models** in **real-world trading scenarios**. Specifically, the **LR + LSTM** ensemble outperformed **Linear Regression** in **hourly trading**, while the **LR + LSTM + GRU** ensemble outperformed **LSTM** in **daily trading**. This suggests that ensemble methods offer superior risk management and more consistent returns through improved signal filtration and reduced susceptibility to market noise, demonstrating that predictive accuracy does not directly translate to trading profitability. **Real-time trading evaluations** further validated the ensemble approach, particularly the **LR + LSTM** ensemble in **hourly** intervals, which exhibited **exceptional capital preservation** during volatile market conditions. These findings present a robust framework for stock price prediction and trading strategy development, with practical implications for algorithmic trading and risk management.

Problem Statements

1. **Limitations of Individual Models :** Individual machine learning models often fail to capture the stock market's non-linear patterns, suffer from overfitting, and lack generalization across different market conditions.
2. **Limitations of Conventional Technical Indicators in Trading Strategies :** Rule-based trading methods using technical indicators struggle with noisy data, rely on lagging signals, and are unable to adapt dynamically to market changes.
3. **Lack of Practical Implementation of Machine Learning Models :** Most machine learning-driven stock prediction models focus on accuracy but lack assessment in real-world environments, missing robust evaluation through backtesting and live simulations in actual trading scenarios.

Objective

1. To propose an ensemble machine learning approach for stock price prediction through the enhancement of technical indicators in a trading strategy.
2. To compare the performance of individual machine learning models against ensemble machine learning approaches for stock price prediction.
3. To assess the real-world viability and benchmark conventional statistical methods with the proposed ensemble machine learning approach through backtesting and real-time trading simulations.

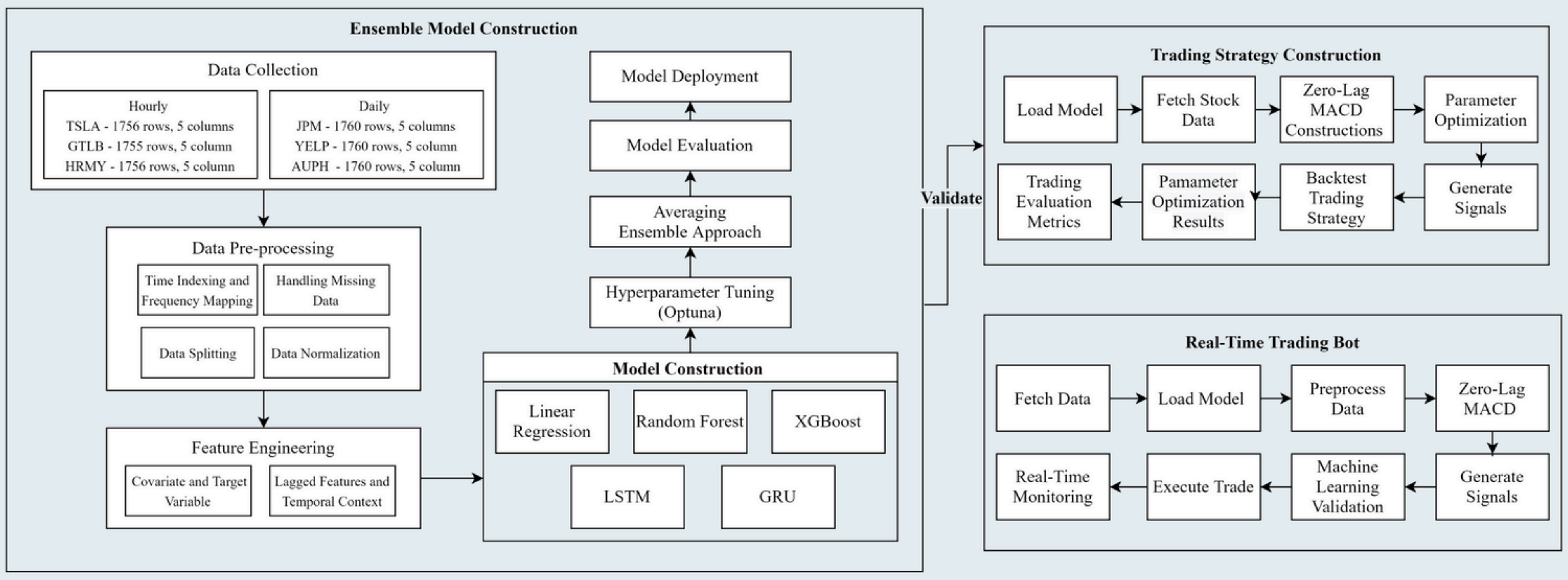
Literature Review

Author	Algorithm	Contribution	Findings
Saud & Shakya (2024)	MACD, DMI, KST LSTM, GRU	Developed intelligent trading strategies combining technical indicators with RNNs	GRU outperformed LSTM; MACD-GRU most profitable; intelligent models beat classical methods
Saifan et al. (2020)	Gradient Boosting, Random Forest, Extra Trees	Evaluated Machine learning models in trading via Quantopian simulations	All models outperformed benchmark; GB had best risk-adjusted returns
Shi et al. (2022)	MLP, RNN, LSTM, CNN, FCNN, ResNet	Enhanced technical indicators by classifying buy/sell points using neural networks	Neural networks improved trading precision; LSTM and CNN showed strong results
Sui et al. (2024)	Ensemble of LSTM, GRU, LightGBM, LR, Keras DNN	Built a hybrid ensemble combining deep learning and time series models	Ensemble achieved higher prediction accuracy compared to individual models
Zhao et al. (2024)	Weight-training ensemble (LSTM, SVR, LightGBM, LR)	Built a model that learns optimal combining weights from past data to forecast stock prices	Outperformed individual models and simple averaging
Sonkavde et al. (2023)	Random Forest + XG-Boost + LSTM	Reviewed many ML and DL models; built ensemble for two Indian stocks; compared with single models	Ensemble had best RMSE & R^2; showed ensemble & hyperparameter tuning improve accuracy

1. Saud & Shakya (2024), Saifan et al. (2020), Shi et al. (2022): Although they integrated machine learning models with technical indicators to improve trading strategies for real-world trading scenarios, they focused only on individual models.
2. Sui et al. (2024), Das et al. (2024), Sonkavde et al. (2023): Demonstrated ensemble model performance but lacked practical trading evaluation and real-time testing.

Research Methodology

The methodology flowchart outlines the process of developing an **ensemble machine learning framework** for stock price prediction, with a focus on **real-world applicability** to enhance trading strategies. The process extends beyond model development to include the deployment of models for **constructing trading strategies**. It integrates a **Zero-Lag MACD** and signal generation with parameter optimization as a supplementary mechanism to **validate the execution** of actionable signals. **Backtesting** is employed to iteratively refine and optimize the trading strategy. Additionally, a **real-time trading bot** is implemented to further evaluate the proposed methodology by actively monitoring and validating trades in a live environment, ensuring its practical effectiveness.



Implementation & Evaluation

Stock	Interval	Metrics	Individual Model: Linear Regression	Random Forest	XGBoost	LSTM	GRU	Ensemble Model: LR + RF + XGB + LSTM + GRU	LR + RF + XGB	LSTM + GRU	LR + LSTM	LR + RF + XGB + LSTM	LR + RF + LSTM + GRU	LR + RF + LSTM	LR + LSTM + GRU
TSLA	1H	RMSE	**1.4573**	1.5180	1.5666	1.4757	1.6234	1.4814	1.4787	1.5400	**1.4579**	1.4703	1.4788	**1.4609**	1.4920
		MSE	**2.1238**	2.3042	2.4542	2.1776	2.6354	2.1947	2.1864	2.3717	**2.1256**	2.1617	2.1867	**2.1341**	2.2260
		MAE	**0.3801**	0.5922	0.6936	0.5501	0.8671	0.5416	0.4986	0.7119	**0.4504**	0.4864	0.5412	**0.4524**	0.5873
		MAPE	**0.18%**	0.28%	0.33%	0.27%	0.44%	0.26%	0.24%	0.36%	**0.22%**	0.23%	0.27%	**0.22%**	0.29%
		R^2	**0.9978**	0.9976	0.9975	**0.9978**	0.9973	0.9977	**0.9978**	0.9976	**0.9978**	**0.9978**	**0.9978**	**0.9978**	0.9977
		CV	**0.7072**	0.7366	0.7602	0.7161	0.7877	0.7189	0.7175	0.7473	**0.7074**	0.7134	0.7176	**0.7089**	0.7240
GTLB	1H	RMSE	**0.2696**	0.2957	0.3291	0.3407	0.4585	0.2789	0.2823	0.2937	**0.2726**	0.2777	0.2750	**0.2725**	0.2802
		MSE	**0.0727**	0.0874	0.1083	0.1160	0.2102	0.0778	0.0797	0.0862	**0.0743**	0.0771	0.0756	**0.0743**	0.0785
		MAE	**0.0813**	0.1534	0.1696	0.2300	0.3973	0.1192	0.1240	0.1575	**0.1031**	0.1147	0.1143	**0.1046**	0.1281
		MAPE	**0.16%**	0.31%	0.34%	0.45%	0.81%	0.24%	0.25%	0.31%	**0.21%**	0.23%	0.23%	**0.21%**	0.25%
		R^2	**0.9964**	0.9957	0.9946	0.9942	0.9896	0.9961	0.9960	0.9957	**0.9963**	0.9962	0.9962	**0.9963**	0.9961
		CV	**0.5412**	0.5936	0.6607	0.6839	0.9205	0.5599	0.5668	0.5895	**0.5472**	0.5574	0.5520	**0.5471**	0.5625
HRMY	1H	RMSE	0.1403	0.1506	0.1507	0.1448	**0.1398**	0.1406	0.1430	**0.1399**	0.1409	0.1422	0.1401	0.1415	**0.1395**
		MSE	0.0197	0.0227	0.0227	0.0210	**0.0195**	0.0198	0.0204	**0.0196**	0.0199	0.0202	**0.0196**	0.0200	**0.0195**
		MAE	**0.0496**	0.0793	0.0766	0.0732	0.0586	0.0584	0.0651	0.0556	0.0590	0.0638	**0.0552**	0.0611	**0.0520**
		MAPE	**0.16%**	0.25%	0.25%	0.24%	0.19%	0.19%	0.21%	**0.18%**	0.19%	0.21%	**0.18%**	0.20%	**0.17%**
		R^2	**0.9937**	0.9927	0.9933	**0.9937**	**0.9937**	**0.9937**	0.9934	**0.9937**	0.9936	0.9935	**0.9937**	0.9936	**0.9938**
		CV	**0.4515**	0.4847	0.4851	0.4661	0.4499	0.4527	0.4602	**0.4503**	0.4536	0.4577	0.4509	0.4556	**0.4492**

The evaluation of individual and ensemble models for **hourly interval stock price prediction** showed that individual models, particularly **Linear Regression**, outperformed all the ensemble approaches in **predictive accuracy** metrics. Linear Regression achieved a slightly better performance with an example **RMSE of 1.4573** and **MAPE of 0.18%**, compared to the best-performing ensemble **(LR + LSTM)**, which had an **RMSE of 1.4579** and **MAPE of 0.22%** for **TSLA** stocks. This suggests that Linear Regression captures hourly trends slightly more effectively. However, ensemble models like **LR + LSTM** demonstrated improved robustness, although their predictive accuracy lagged slightly behind individual models.

Stock Name	Strategy	Zero-Lag MACD: Frequency	Max Profit ($)	Max Loss ($)	ROI(%)	Linear Regression Enhanced Zero-Lag MACD: Frequency	Max Profit ($)	Max Loss ($)	ROI(%)	Random Forest Enhanced Zero-Lag MACD: Frequency	Max Profit ($)	Max Loss ($)	ROI(%)	LSTM Enhanced Zero-Lag MACD: Frequency	Max Profit ($)	Max Loss ($)	ROI(%)	LR + LSTM Enhanced Zero-Lag MACD: Frequency	Max Profit ($)	Max Loss ($)	ROI(%)	LR + RF + LSTM Enhanced Zero-Lag MACD: Frequency	Max Profit ($)	Max Loss ($)	ROI(%)
TSLA	Buy Above Sell Above	21	251.52	-228.41	11.20	12	237.65	-240.81	55.93	5	252.93	-148.20	74.09	5	256.15	-127.52	88.08	8	237.58	-118.72	121.47	3	256.15	-106.85	56.10
TSLA	Buy Below Sell Above	19	255.69	-217.47	21.67	9	252.43	-184.39	80.18	4	234.17	-91.32	74.01	7	247.56	-142.55	153.44	7	247.56	-141.51	120.87	5	173.60	-111.74	61.91
TSLA	Buy Above Sell Below	4	197.19	-155.13	61.61	1	189.00	189.00	80.37	2	208.22	20.76	96.77	7	234.17	-141.30	108.88	6	201.70	-125.48	108.39	1	237.67	237.67	103.92
TSLA	Buy Below Sell Below	10	219.83	-153.02	19.49	12	194.06	-195.49	54.80	2	170.45	20.01	72.01	2	133.47	13.01	70.67	6	190.63	-119.98	60.91	3	208.30	-24.04	81.48
GTLB	Buy Above Sell Above	12	27.08	-21.66	3.99	5	19.65	-15.07	17.02	2	23.22	-6.95	27.09	2	6.69	0.59	8.48	3	8.51	0.59	12.97	3	13.52	0.59	28.73
GTLB	Buy Below Sell Above	15	29.60	-28.28	12.31	11	33.61	-27.76	39.11	5	20.70	-17.18	63.45	3	23.95	-13.78	25.88	6	33.23	-20.31	49.60	5	29.15	-16.71	44.32
GTLB	Buy Above Sell Below	5	20.53	-15.01	1.84	3	23.28	-13.37	22.23	1	20.97	20.97	39.34	1	4.97	4.97	9.77	3	10.00	-2.74	12.79	2	14.01	3.65	26.78
GTLB	Buy Below Sell Below	17	31.33	-31.41	-23.79	2	30.29	17.92	30.31	3	20.53	-2.56	42.14	1	31.15	31.15	24.20	5	31.12	-17.53	42.94	5	30.51	-15.64	50.54
HRMY	Buy Above Sell Above	5	12.25	-11.69	-3.92	4	12.24	-6.75	8.71	3	6.70	-1.17	7.82	1	5.69	5.69	14.66	9	11.73	-10.62	2.51	3	7.91	-4.14	15.64
HRMY	Buy Below Sell Above	13	9.30	-9.82	-13.97	5	11.60	-3.67	9.16	5	6.57	-4.16	14.63	1	2.92	2.92	6.55	3	5.93	-3.16	7.36	3	7.62	-3.47	9.87
HRMY	Buy Above Sell Below	8	9.31	-8.44	-14.42	4	10.65	-5.78	19.13	2	4.26	-3.83	12.31	2	4.29	0.87	11.16	3	10.78	-8.28	2.81	2	3.85	-3.99	7.73
HRMY	Buy Below Sell Below	5	6.22	-6.45	-17.58	4	6.40	-1.50	15.26	3	4.17	-3.60	12.36	1	5.13	5.13	13.01	2	6.74	-2.81	9.03	4	5.81	-3.97	24.61

In **real-world trading scenarios** assessed through **backtesting (January 1, 2024, to May 1, 2025)**, **ensemble methods outperformed** individual models despite the latter's higher predictive accuracy. Specifically, the **LR + LSTM** ensemble achieved an average **ROI of 45.97%** in **hourly** intervals, surpassing **Linear Regression**'s **36.02%** and **LSTM's 44.57%** across all stocks.

Stock Name	Model Strategy	Frequency	ROI (%)
TSLA	Conventional Zero-Lag MACD	3	-0.99
	LR + LSTM Enhanced Zero-Lag MACD	0	0.00
	LR + RF + LSTM Enhanced Zero-Lag MACD	1	-0.15

Real-time trading evaluation from **June 16 to June 27, 2025**, confirmed the **LR + LSTM** ensemble's **superiority** over the conventional **Zero-Lag MACD**. The conventional strategy incurred a **-0.99% ROI** with 3 trades on **TSLA**, while **LR + LSTM avoided losses** with 0 trades and a **0.00% ROI**, showcasing exceptional capital preservation. The **LR + RF + LSTM** ensemble recorded a minor **-0.15% ROI**, further highlighting ensemble models' **effective signal filtering** in volatile hourly market conditions.

Conclusion

1. **Developed Ensemble Framework**: Successfully integrated Linear Regression and LSTM using an averaging ensemble method for stock price prediction.
2. **Improved Trading Performance:** The LR + LSTM ensemble achieved a 45.97% ROI in hourly trading, surpassing Linear Regression's 36.02% and the conventional Zero-Lag MACD, which had a 5.19% ROI.
3. **Enhanced Risk Management:** Ensemble methods reduced trading frequency, filtered unprofitable signals, and improved capital preservation, especially in volatile markets.
4. **Practical Trading Tool:** Created an ensemble machine learning-enhanced Zero-Lag MACD strategy, providing actionable trading signals for algorithmic trading and risk management.