ACCOUNTING, FINANCE AND COMPUTATIONAL INTELLIGENCE

Comparison of the Accuracy and Efficiency of Linear Regression and Nonlinear Logit and Probit Models in Stock Price Forecasting

1. Parasto Azami: PhD Student, Department of Financial Engineering, Ro.C., Islamic Azad University, Roudehen, Iran

2. Najmeh Kargar Kamour*: Department of Accounting, Ro.C., Islamic Azad University, Roudehen, Iran. Email: Kargarkamvar@iau.ac.ir (Corresponding Author)

3. Hoda Hemmati: Department of Accounting, Ro.C., Islamic Azad University, Roudehen, Iran

Article history

Received: 25 November 2025

Revised: 04 April 2026

Accepted: 12 April 2026

Initial Publish: 25 May 2026

Final Publish: 21 April 2027

Abstract:

This study aims to compare the accuracy and efficiency of linear regression with nonlinear Logit and Probit models in short-term stock price forecasting. The study uses daily stock data of Isfahan Steel Company from January 1, 2024, to November 19, 2025. After eliminating multicollinear variables, the closing price was considered the dependent variable, while the number of trades, price-to-earnings ratio (P/E), net percentage of retail trading, and per capita buy and sell values of individual investors were selected as independent variables. An out-of-sample forecasting approach was applied, reserving the last 30 trading days as the test set. The results indicate that although linear regression shows acceptable explanatory power, it underperforms in short-term forecasting. The Logit model achieved the lowest MSE, RMSE, and MAE values, followed by the Probit model. The findings confirm the nonlinear nature of stock price dynamics and highlight the superior predictive performance of probability-based nonlinear models, particularly Logit.

Keywords: Stock Price Forecasting, Linear Regression, Logit, Probit, Capital Market

Citation: Azami, P., Kargar Kamour, N., & Hemmati, H. (2027). Comparison of the Accuracy and Efficiency of Linear Regression and Nonlinear Logit and Probit Models in Stock Price Forecasting. *Accounting, Finance and Computational Intelligence,* 5(1), 1-16.

Copyright: © 2027 by the authors. Published under the terms and conditions of Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

Extended Abstract

Introduction

Stock price forecasting remains one of the most persistent and theoretically complex challenges in financial economics. Capital markets are characterized by high volatility, nonlinear dynamics, structural breaks, and strong behavioral components, which collectively limit the effectiveness of purely deterministic or linear predictive frameworks. Accurate short-term stock price prediction is particularly important for investors, portfolio managers, and policymakers, as it directly influences trading strategies, risk management decisions, and capital allocation efficiency. Empirical evidence suggests that stock price movements are driven not only by firm-level fundamentals but also by macroeconomic uncertainty, investor sentiment, liquidity conditions, and informational shocks, all of which interact in nonlinear ways (Adebayo et al., 2022; Salisu et al., 2022; Yang, 2023).

Traditional econometric models—especially linear regression—have long served as the foundation of empirical stock market analysis due to their interpretability, simplicity, and relatively modest data requirements. Canonical econometric literature emphasizes that linear regression remains a valuable benchmark model, particularly when theoretical relationships are assumed to be stable and additive (Greene, 2003; Gujarati, 2004). However, financial time-series data frequently violate the assumptions underlying linear models, exhibiting heteroskedasticity, heavy tails, asymmetric distributions, and regime-dependent behavior. These characteristics often result in declining predictive accuracy, especially in short-term forecasting horizons (Guo & Wang, 2023; Zakamulin, 2024).

In response to these limitations, a growing body of research has explored nonlinear and probabilistic modeling approaches. Logit and Probit models, which transform the forecasting problem into a probability-based framework, allow for greater flexibility in capturing threshold effects and regime shifts in stock price behavior. Empirical studies indicate that probability-based models often outperform linear specifications in volatile environments and when investor behavior plays a dominant role (Alizadeh Chamazkoti et al., 2024; Omogoroye et al., 2023). These findings align with behavioral finance theories that emphasize investor overreaction, sentiment-driven trading, and nonlinear adjustment processes (Boons et al., 2023; Zakamulin, 2024).

Parallel to this development, advances in machine learning and artificial intelligence have introduced sophisticated forecasting techniques capable of capturing complex patterns in large datasets. While such methods have demonstrated strong predictive performance in some contexts, their high computational cost, data intensity, and limited interpretability constrain their practical adoption in many financial institutions (Kumbure et al., 2022; Masini et al., 2023; Thakkar & Chaudhari, 2021). Consequently, there remains a clear research gap for systematic comparisons between classical linear models and interpretable nonlinear econometric alternatives.

This gap is particularly pronounced in emerging markets, where informational inefficiencies, dominance of retail investors, and market microstructure frictions amplify nonlinear price behavior. Empirical evidence from emerging economies—including Iran—suggests that traditional linear models often fail to capture short-term price dynamics, while nonlinear approaches offer superior predictive performance (Arman et al., 2022; Nourvosh et al., 2020; Setayesh & Kazemnezhad, 2019; Shirzad et al., 2021). Despite these insights, limited research has directly compared linear regression with Logit and Probit models using out-of-sample forecasting and error-based evaluation criteria.

Accordingly, this study addresses this gap by providing a structured comparison of linear regression, Logit, and Probit models in short-term stock price forecasting, contributing to both the methodological literature and practical investment analysis (Akhbari et al., 2024; Papageorgiou et al., 2024; Tsai et al., 2024).

Methods and Materials

This study employs daily stock market data from Isfahan Steel Company over the period from January 1, 2024, to November 19, 2025, covering all available trading days within the sample interval. The dataset includes a broad range of price-based, transactional, and behavioral variables. To mitigate multicollinearity and ensure model stability, overlapping variables were systematically removed through diagnostic screening.

The final specification includes the closing stock price as the dependent variable. Independent variables consist of the number of trades, the price-to-earnings ratio (P/E), the net percentage of retail (individual) trading, and per capita retail buy and sell values. For nonlinear models, the dependent variable was transformed into a binary indicator based on whether the closing price exceeded its sample median.

Three models were estimated: ordinary least squares linear regression, Logit regression, and Probit regression. To evaluate predictive performance, an out-of-sample forecasting framework was implemented. The final 30 trading days were reserved as a test set, while the remaining observations were used for model estimation.

Forecast accuracy was assessed using three standard error metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). These criteria allow for a comprehensive comparison of both average deviation and sensitivity to large forecasting errors.

Findings

Descriptive statistics reveal that the stock price and trading-related variables exhibit substantial volatility, wide dispersion, and non-normal distributional characteristics. Such properties suggest that linear assumptions may be restrictive in capturing short-term price movements.

Estimation results from the linear regression model indicate statistically significant relationships between the independent variables and the closing price. The price-to-earnings ratio demonstrates a strong positive effect, while retail trading intensity variables exhibit negative effects, reflecting potential behavioral trading patterns. Despite acceptable explanatory power, the linear model shows relatively large forecasting errors in the out-of-sample period.

The Logit model yields superior predictive accuracy across all evaluation metrics. It records the lowest MSE, RMSE, and MAE values among the three models, indicating the closest alignment between predicted and actual prices during the test period. The Probit model performs better than linear regression but remains less accurate than the Logit specification.

Overall, the ranking of predictive performance is Logit first, Probit second, and linear regression last. These results confirm that probabilistic nonlinear models provide a more accurate representation of short-term stock price dynamics in volatile market conditions.

Discussion and Conclusion

The results of this study provide clear empirical evidence that short-term stock price dynamics are fundamentally nonlinear and probability-driven. While linear regression remains useful for interpretive analysis and baseline modeling, its predictive limitations become pronounced in environments characterized by volatility, asymmetric information, and behavioral trading.

The superior performance of the Logit model highlights the advantages of transforming price prediction into a probabilistic classification problem. By focusing on the likelihood of price movement regimes rather than precise point estimates, the Logit model effectively captures threshold effects and nonlinear adjustments in investor behavior. The Probit model, although conceptually similar, appears more sensitive to distributional assumptions, which may reduce its flexibility under real-market conditions.

These findings underscore the importance of model selection in financial forecasting. More complex models are not inherently superior; rather, models that align closely with the statistical and behavioral properties of the data yield the most reliable predictions. In this regard, interpretable nonlinear econometric models offer a valuable middle ground between overly simplistic linear approaches and opaque machine learning algorithms.

From a practical perspective, the results suggest that investors and financial analysts can improve short-term forecasting accuracy by incorporating probability-based nonlinear models into their analytical frameworks. Such models can enhance decision-making, risk assessment, and tactical trading strategies without imposing excessive computational or data burdens.

In conclusion, this study demonstrates that Logit and Probit models—particularly Logit—provide meaningful improvements over linear regression in short-term stock price forecasting. The findings contribute to the empirical finance literature by reinforcing the relevance of nonlinear econometric modeling and offering actionable insights for market participants operating in volatile and emerging market environments.

Authors' Contributions

Authors equally contributed to this article.

Acknowledgments

Authors thank all participants who participate in this study.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

مقایسه دقت و کارایی رگرسیون خطی با مدل‌های غیرخطی لاجیت و پروبیت در پیش‌بینی قیمت سهام

۱. پرستو اعظمی: دانشجوی دکتری، گروه مهندسی مالی، واحد رودهن، دانشگاه آزاد اسلامی، رودهن، ایران

۲. نجمه کارگر کامور*: گروه حسابداری، واحد رودهن، دانشگاه آزاد اسلامی، رودهن، ایران. ایمیل: Kargarkamvar@iau.ac.ir (نویسنده مسئول)

۳. هدی همتی: گروه حسابداری، واحد رودهن، دانشگاه آزاد اسلامی، رودهن، ایران

تاریخچه مقاله

تاریخ دریافت: ۴ آذر ۱۴۰۴

تاریخ بازنگری: ۱۵ فروردین ۱۴۰۵

تاریخ پذیرش: ۲۳ فروردین ۱۴۰۵

تاریخ چاپ اولیه: ۴ خرداد ۱۴۰۵

تاریخ چاپ نهایی: ۱ اردیبهشت ۱۴۰۶

چکیده

هدف این پژوهش مقایسه دقت و کارایی مدل رگرسیون خطی با مدل‌های غیرخطی لاجیت و پروبیت در پیش‌بینی کوتاه‌مدت قیمت سهام شرکت فولاد اصفهان است. این پژوهش از داده‌های روزانه سهام شرکت فولاد اصفهان در بازه زمانی ۱ ژانویه ۲۰۲۴ تا ۱۹ نوامبر ۲۰۲۵ استفاده می‌کند. پس از حذف متغیرهای دارای همخطی، قیمت پایانی به‌عنوان متغیر وابسته و تعداد معاملات، نسبت قیمت به سود (P/E)، درصد خالص معاملات حقیقی و سرانه خرید و فروش حقیقی به‌عنوان متغیرهای مستقل انتخاب شدند. برای ارزیابی توان پیش‌بینی، از رویکرد پیش‌بینی خارج از نمونه استفاده شد و ۳۰ روز پایانی به‌عنوان داده آزمون در نظر گرفته شد. نتایج برآوردها نشان داد هرچند رگرسیون خطی از قدرت تبیینی قابل قبول برخوردار است، اما در پیش‌بینی کوتاه‌مدت عملکرد ضعیف‌تری دارد. مدل لاجیت با کمترین مقادیر خطا در معیارهای MSE، RMSE و MAE بهترین عملکرد پیش‌بینی را نشان داد و مدل پروبیت در رتبه بعدی قرار گرفت. یافته‌ها نشان می‌دهد روابط میان متغیرهای معاملاتی و قیمت سهام ماهیتی غیرخطی دارد و استفاده از مدل‌های مبتنی بر احتمال، به‌ویژه لاجیت، می‌تواند دقت پیش‌بینی قیمت سهام را به‌طور معناداری افزایش دهد.

کلیدواژگان: پیش‌بینی قیمت سهام، رگرسیون خطی، لاجیت، پروبیت، بازار سرمایه

شیوه استناددهی: اعظمی، پرستو.، کارگر کامور، نجمه.، و همتی، هدی. (۱۴۰۶). مقایسه دقت و کارایی رگرسیون خطی با مدل‌های غیرخطی لاجیت و پروبیت در پیش‌بینی قیمت سهام. *حسابداری، امور مالی و هوش محاسباتی*، ۵(۱)، ۱۶-۱.

مجوز و حق نشر: © ۱۴۰۶ تمامی حقوق انتشار این مقاله متعلق به نویسنده است. انتشار این مقاله به‌صورت دسترسی آزاد مطابق با گواهی (CC BY-NC 4.0) صورت گرفته است.

مقدمه

پیش‌بینی قیمت و بازده سهام همواره یکی از محوری‌ترین و در عین حال چالش‌برانگیزترین مسائل در ادبیات مالی و اقتصاد مالی بوده است. بازارهای سرمایه به‌دلیل ماهیت پویا، نوسانی و متأثر از عوامل متنوع اقتصادی، سیاسی، رفتاری و اطلاعاتی، محیطی پیچیده برای تحلیل و تصمیم‌گیری فراهم می‌کنند؛ به‌گونه‌ای که حتی تغییرات جزئی در انتظارات سرمایه‌گذاران یا شوک‌های اطلاعاتی می‌تواند به نوسانات شدید قیمتی منجر شود. از این‌رو، توسعه مدل‌های دقیق و کارآمد برای پیش‌بینی قیمت سهام، نه‌تنها از منظر دانشگاهی، بلکه برای سرمایه‌گذاران، مدیران دارایی و سیاست‌گذاران مالی اهمیتی بنیادین دارد. مطالعات متعدد نشان می‌دهند که دقت پیش‌بینی قیمت سهام نقش تعیین‌کننده‌ای در بهینه‌سازی تخصیص منابع، مدیریت ریسک و افزایش بازده سبدهای سرمایه‌گذاری ایفا می‌کند (Adebayo et al., 2022; Salisu et al., 2022).

در دهه‌های گذشته، روش‌های متنوعی برای پیش‌بینی قیمت و بازده سهام توسعه یافته‌اند که از مدل‌های کلاسیک اقتصادسنجی تا رویکردهای نوین مبتنی بر یادگیری ماشین و هوش مصنوعی را دربر می‌گیرند. در میان این روش‌ها، مدل‌های خطی به‌ویژه رگرسیون خطی، به‌دلیل سادگی، شفافیت تفسیری و قابلیت پیاده‌سازی آسان، جایگاه ویژه‌ای در پژوهش‌های مالی داشته‌اند. متون کلاسیک اقتصادسنجی نشان می‌دهند که رگرسیون خطی همچنان به‌عنوان نقطه شروع بسیاری از تحلیل‌های تجربی در بازار سرمایه مورد استفاده قرار می‌گیرد (Greene, 2003; Gujarati, 2004). با این حال، فرض اساسی خطی‌بودن روابط میان متغیرها، در بسیاری از موارد با واقعیت‌های بازارهای مالی سازگار نیست؛ زیرا داده‌های مالی اغلب با نوسانات شدید، توزیع‌های نامتقارن، چولگی، کشیدگی بالا و رفتارهای غیرخطی همراه هستند (Boons et al., 2023; Zakamulin, 2024).

در واکنش به این محدودیت‌ها، پژوهشگران به‌تدریج به سمت توسعه و به‌کارگیری مدل‌های غیرخطی حرکت کرده‌اند. این مدل‌ها با هدف انعکاس بهتر پیچیدگی‌های ذاتی بازار سهام، توانایی بیشتری در شناسایی الگوهای پنهان و روابط غیرخطی میان متغیرها دارند. در این میان، مدل‌های مبتنی بر احتمال مانند لاجیت و پروبیت، به‌عنوان ابزارهایی مهم در تحلیل داده‌های مالی مطرح شده‌اند. این مدل‌ها به‌ویژه در شرایطی که رفتار قیمت سهام به‌صورت رژیمی، آستانه‌ای یا وابسته به حالات مختلف بازار بروز می‌کند، عملکرد مناسبی از خود نشان می‌دهند (Guo & Wang, 2023; Omogoroye et al., 2023). شواهد تجربی نشان می‌دهد که در محیط‌های پرنوسان، مدل‌های غیرخطی می‌توانند دقت پیش‌بینی را نسبت به مدل‌های خطی به‌طور معناداری افزایش دهند (Papageorgiou et al., 2024).

افزون بر این، تحولات فناورانه و گسترش داده‌های کلان، زمینه‌ساز رشد سریع رویکردهای یادگیری ماشین و هوش مصنوعی در پیش‌بینی بازار سهام شده است. مطالعات مروری نشان می‌دهند که الگوریتم‌های یادگیری عمیق، شبکه‌های عصبی و مدل‌های ترکیبی توانسته‌اند در برخی شرایط عملکرد برتری نسبت به مدل‌های سنتی ارائه دهند (Ghallabi et al., 2025; Kumbure et al., 2022; Masini et al., 2023). با این حال، پیچیدگی محاسباتی بالا، نیاز به داده‌های گسترده و دشواری تفسیر نتایج، استفاده عملی از این روش‌ها را برای بسیاری از فعالان بازار محدود کرده است. به همین دلیل، همچنان نیاز به مقایسه نظام‌مند و دقیق مدل‌های اقتصادسنجی کلاسیک و نیمه‌کلاسیک، به‌ویژه در بستر بازارهای نوظهور، احساس می‌شود (Thakkar & Chaudhari, 2021; Tomar & Periyasamy, 2023).

در ادبیات بین‌المللی، مطالعات متعددی به بررسی عوامل مؤثر بر بازده و قیمت سهام پرداخته‌اند. برخی پژوهش‌ها بر نقش عوامل کلان اقتصادی و ریسک‌های سیستماتیک تمرکز داشته‌اند و نشان داده‌اند که عدم‌قطعیت سیاست اقتصادی، ریسک ژئوپلیتیک و شوک‌های کلان می‌توانند پیش‌بینی‌پذیری بازار سهام را تحت تأثیر قرار دهند (Adebayo et al., 2022; Salisu & Vo, 2020). گروهی دیگر از پژوهش‌ها، اهمیت اطلاعات شرکتی، پیش‌بینی‌های مدیریت و متغیرهای بنیادی را در شکل‌دهی انتظارات سرمایه‌گذاران برجسته کرده‌اند (Al-Zubaidi et al., 2025; Kitagawa & Shuto, 2024; Tsai et al., 2024). در کنار این عوامل، مطالعات رفتاری نشان می‌دهند که واکنش‌های بیش‌ازحد یا کمتر از حد سرمایه‌گذاران در بازارهای صعودی و نزولی می‌تواند به انحراف قیمت‌ها از ارزش‌های بنیادی منجر شود (Boons et al., 2023; Zakamulin, 2024).

در بازارهای نوظهور، از جمله ایران، پیچیدگی‌های نهادی، محدودیت‌های اطلاعاتی و نقش پررنگ سرمایه‌گذاران حقیقی، اهمیت انتخاب مدل مناسب برای پیش‌بینی قیمت سهام را دوچندان می‌کند. پژوهش‌های داخلی نشان داده‌اند که متغیرهایی نظیر نقدشوندگی، حباب‌های قیمتی، شرایط مالی کلان و ساختار شبکه‌ای بازار سرمایه نقش مهمی در

پیش‌بینی بازده و قیمت سهام دارند (Arman et al., 2022; Nourvosh et al., 2020; Taghizadeh & Abdzadeh Kanafi, 2023). همچنین نتایج برخی مطالعات حاکی از آن است که روش‌های غیرخطی و ترکیبی در مقایسه با رگرسیون خطی، توان پیش‌بینی بالاتری در بازار سرمایه ایران ارائه می‌دهند (Akhbari et al., 2024; Setayesh & Kazemnezhad, 2019; Shirzad et al., 2021).

در سطح بین‌المللی نیز، پژوهش‌هایی که به مقایسه مستقیم مدل‌های مختلف پیش‌بینی پرداخته‌اند، نتایج ناهمگونی ارائه کرده‌اند. برخی مطالعات برتری روش‌های غیرخطی و یادگیری ماشین را گزارش می‌کنند (Nabipour et al., 2020; Singh et al., 2022; Toochaei & Moeini, 2023)، در حالی که برخی دیگر نشان می‌دهند که در افق‌های زمانی کوتاه یا در شرایط خاص بازار، مدل‌های ساده‌تر می‌توانند عملکردی قابل رقابت داشته باشند (Ma et al., 2021; Yang, 2023). این ناهمگونی نتایج، ضرورت انجام مطالعات مقایسه‌ای دقیق و زمینه‌محور را برجسته می‌سازد.

نکته مهم دیگر، تأثیر کیفیت داده‌ها و ویژگی‌های آماری آن‌ها بر عملکرد مدل‌های پیش‌بینی است. داده‌های مالی اغلب با نویز بالا، همبستگی‌های پیچیده و تغییرپذیری شدید همراه هستند که می‌تواند دقت مدل‌های خطی را کاهش دهد (Guo & Wang, 2023). در چنین شرایطی، مدل‌های غیرخطی مبتنی بر احتمال، به‌ویژه لاجیت و پروبیت، به‌دلیل انعطاف‌پذیری بیشتر در مواجهه با توزیع‌های غیرنرمال و رفتارهای نامتقارن، گزینه‌ای مناسب برای تحلیل محسوب می‌شوند (Greene, 2003; Omogoroye et al., 2023). افزون بر این، شواهد اخیر نشان می‌دهد که متغیرهای رفتاری و اطلاعاتی نظیر احساسات بازار و اخبار مالی می‌توانند نوسانات قیمت سهام را تشدید کرده و ضرورت استفاده از مدل‌های منعطف‌تر را افزایش دهند (Nuriksan & Nikmah, 2025; Shami et al., 2025).

با وجود گستردگی پژوهش‌های انجام‌شده، هنوز خلأ قابل توجهی در زمینه مقایسه نظام‌مند عملکرد مدل‌های خطی و غیرخطی کلاسیک، به‌ویژه لاجیت و پروبیت، در پیش‌بینی قیمت سهام با استفاده از داده‌های واقعی و آزمون‌های خارج از نمونه وجود دارد. بسیاری از مطالعات یا بر یک روش خاص تمرکز کرده‌اند یا از داده‌های محدود و افق‌های زمانی کوتاه استفاده نموده‌اند (Alizadeh Chamazkoti et al., 2024). از این‌رو، انجام پژوهشی که به‌صورت تجربی و مقایسه‌ای، دقت و کارایی این مدل‌ها را در یک چارچوب واحد ارزیابی کند، می‌تواند به غنای ادبیات مالی و بهبود تصمیم‌گیری‌های سرمایه‌گذاری کمک شایانی نماید. بر این اساس، هدف این پژوهش مقایسه دقت و کارایی مدل رگرسیون خطی با مدل‌های غیرخطی لاجیت و پروبیت در پیش‌بینی کوتاه‌مدت قیمت سهام است.

روش پژوهش و مواد

داده‌های مورد استفاده دراین تحقیق، داده‌های روزانه سهام شرکت «فولاد اصفهان» در بازه زمانی ۲۰۲۴/۰۱/۰۱ الی ۲۰۲۵/۱۱/۱۹ شامل هر روز معاملاتی در این بازه می‌باشد. در طول بازه مذکور صرفاً تعداد بسیار محدودی از روزهایی که بعضی از داده‌ها موجود نبوده بطورکامل مورد استفاده قرار گرفته است. متغیرهای بکار رفته به شرح ذیل است:

قیمت بازگشایی: قیمت اولین معامله در آن روز معاملاتی برحسب ریال.

بالاترین: بالاترین قیمت معامله شده در طول روز برحسب ریال.

پایین‌ترین: پایین‌ترین قیمت در طول روز برحسب ریال.

آخرین معامله یا قیمت پایانی: قیمت آخرین معامله در روز یا قیمت پایانی گزارش‌شده؛ در بورس ایران معمولاً «پایانی» به عنوان قیمت مرجع پایان روز در نظر گرفته می‌شود برحسب ریال.

قیمت تعدیلی: قیمت پایانی تعدیل‌شده برای اثرات تقسیم سود یا افزایش سرمایه برحسب ریال.

میزان تغییر: تغییر مطلق قیمت نسبت به روز قبل: $Change_t = Close_t - Close_{t-1}$ برحسب ریال.

درصد تغییر: $Percent\ Change_t = \frac{Close_t - Close_{t-1}}{Close_{t-1}} * 100$ برحسب درصد.

ارزش بازار: تعداد سهام درگردش × قیمت پایانی؛ برحسب ریال.

تعداد معاملات: تعداد معاملات ثبت‌شده در آن روز (تعداد سفارشات اجرایی).

حجم: تعداد برگه سهم معامله‌شده در روز برحسب سهم.

ارزش معاملات: مجموع ارزش همه معاملات در روز؛ برحسب ریال.

سهام شناور: درصد سهام قابل معامله توسط سهامداران خرد برحسب درصد یا سهم.

$\frac{P}{E_{ttm}}$: نسبت قیمت به درآمد بر مبنای سود ۱۲ ماه گذشته (*Trailing Twelve Months*).

$\frac{P}{B}$: نسبت قیمت به ارزش دفتری هر سهم.

$\frac{P}{S}$: نسبت قیمت به فروش هر سهم.

قدرت خرید حقیقی: یکی از معیارهای میزانی که نشان می‌دهد خریداران حقیقی در آن روز چه مقدار قدرت خرید خالص داشته‌اند (می‌تواند معیاری مبتنی بر ارزش خرید – فروش حقیقی باشد) برحسب ریال یا سهم.

درصد خالص حقیقی: سهم درصدی از خالص خرید حقیقی نسبت به مجموع معاملات یا سهم.

خالص حقیقی: مجموع خریدهای حقیقی منهای مجموع فروش‌های حقیقی در ارزش یا حجم برحسب ریال یا سهم.

سرانه خرید حقیقی یا سرانه فروش حقیقی: میانگین ارزش خرید–فروش هر حقیقی فعال (مجموع خرید حقیقی تقسیم بر تعداد خریداران حقیقی یا مشابه).

در نهایت با توجه به اینکه بعضی از این متغیرهای همپوشانی دارند (مانند میزان تغییر و درصد تغییر و ...) و این امر موجب همخطی شدید[1] می‌گردد، لذا متغیرهایی که منجر به این امر می‌شود را حذف و نهایتاً متغیرهای باقیمانده شامل قیمت پایانی، تعداد معاملات، $\frac{P}{E_{ttm}}$، درصد خالص حقیقی، سرانه خرید حقیقی و سرانه فروش حقیقی می‌باشند.

در نهایت با توجه به داده‌های مذکور و حجم بالای آن‌ها که می‌تواند کمک شایانی به تخمین بهتر داشته باشد، به پیش‌بینی قیمت سهام مذکور پرداخته می‌شود. در ادامه به بررسی روش‌های مورد استفاده در این تحقیق پرداخته می‌شود:

رگرسیون خطی

با گسترش اهمیت روزافزون تحلیل داده‌ها در علوم مختلف، استفاده از روش‌های آماری برای مدل‌سازی و پیش‌بینی رفتار پدیده‌ها اهمیت ویژه‌ای یافته است. یکی از ساده‌ترین و در عین حال پرکاربردترین این روش‌ها، رگرسیون خطی (*Linear Regression*) است. این روش به‌طور گسترده در حوزه‌هایی نظیر اقتصاد، مهندسی، علوم اجتماعی، پزشکی و یادگیری ماشین برای پیش‌بینی و تحلیل روابط بین متغیرها به کار می‌رود. رگرسیون خطی یکی از ابزارهای پایه‌ای و مؤثر در تحلیل داده‌ها و پیش‌بینی است که علی‌رغم سادگی، در بسیاری از مسائل واقعی عملکرد مناسبی دارد. رابطه کلی آن به صورت زیر تعریف می‌شود:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

رگرسیون لاجیت

رگرسیون لاجیت یک تکنیک آماری است که برای پیش‌بینی احتمال وقوع یک رویداد دوتایی استفاده می‌شود، به‌ویژه زمانی که متغیر وابسته به صورت دوتایی تعریف شده باشد. این روش به کمک تابع لجستیک (*Logistic Function*) که به‌طور خاص نوعی تابع غیرخطی است، می‌تواند به‌خوبی مدل‌سازی کند که چگونه تعدادی از متغیرهای مستقل در پیش‌بینی احتمال وقوع یا عدم وقوع یک پدیده تاثیر می‌گذارند. به این ترتیب، رگرسیون لاجیت یکی از ابزارهای کلیدی در تحلیل داده‌های تجربی و داده‌های وابسته است و می‌تواند در بسیاری از زمینه‌ها مانند علم پزشکی، اقتصاد و علوم اجتماعی به کار گرفته شود.

مهم‌ترین مزیت رگرسیون لاجیت این است که خروجی آن دارای محدودیت بین ۰ و ۱ است، بنابراین می‌توان به‌راحتی تفسیر کرد. در بازاریابی، این مدل برای تحلیل و شناسایی اینکه آیا یک مشتری احتمال خرید یک محصول خاص را دارد یا خیر، مفید است. همچنین، رگرسیون لاجیت به تحلیل‌گران اقتصادی در پیش‌بینی نتایج اقتصادی مختلف

[1] Multicollinearity

کمک می‌کند و در نهایت به مدیران کسب‌وکارها در اتخاذ تصمیمات استراتژیک کمک می‌کند. با توجه به پیشرفت‌های محاسباتی و داده‌های کلان، این روش توانسته است به‌عنوان ابزاری جامع در تحلیل داده‌ها و پیش‌بینی در دنیای مدرن محسوب شود. در شرایطی که داده‌ها به طور ایده‌آل توزیع نرمال نیستند، رگرسیون لاجیت همچنان قادر به ارائه نتایج معنادار و معتبر است. این قابلیت به‌ویژه در پردازش داده‌های مالی که ممکن است دارای توزیع‌های دیگر مانند توزیع‌های غیرخطی، چوله یا پخ‌دار باشند، بسیار حائز اهمیت است. مزیت دیگر رگرسیون لاجیت، قابلیت تفسیر آسان ضرایب مدل است.

علاوه بر ویژگی‌های ذکر شده، رگرسیون لاجیت دارای قابلیت‌های کاربردی اضافی است که در پژوهش‌های مالی و اقتصادی مورد توجه قرار می‌گیرد. به عنوان مثال، این روش می‌تواند نتایج دقیقی را حتی در صورت وجود چندین متغیر توضیحی ارائه دهد و تعاملات بین متغیرهای مختلف را نیز شناسایی کند. این قابلیت به محققان این امکان را می‌دهد که مدل‌های پیش‌بینی دقیقی را برای قیمت سهام طراحی کنند و تأثیر عوامل مختلف را بر نتایج تحلیل کنند. علاوه بر این، رگرسیون لاجیت به دلیل عملکرد خوب خود در شرایط داده‌های ناکامل یا موجودیت ناهماهنگ بین متغیرها، به یکی از ابزارهای محبوب برای پیش‌بینی در زمینه‌های مختلف تبدیل شده است. به این ترتیب، استفاده از این متد در حوزه مالی می‌تواند به تصمیم‌سازی‌های بهتر و بهینه‌سازی سبد سرمایه‌گذاری کمک کند. به طور خلاصه، رگرسیون لجستیک ابزاری قدرتمند است که در تحلیل‌های پیش‌بینی و تصمیم‌سازی به کار می‌رود. رگرسیون لاجیت به شکل زیر تعریف می‌شود:

$$logit(p) = ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i}$$

$$p = Pr(y_i = 1) \;\&\; i = 1, \ldots, n$$

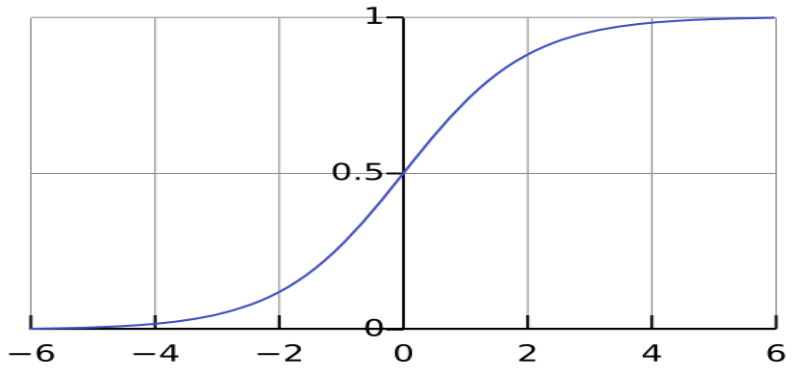
$$p = Pr\left(y_i = 1/\vec{x}_i : \vec{\beta}\right) = \frac{e^{\beta_0+\beta_1 x_{1,i}+\cdots+\beta_k x_{k,i}}}{1+e^{\beta_0+\beta_1 x_{1,i}+\cdots+\beta_k x_{k,i}}} = \frac{1}{1+e^{-(\beta_0+\beta_1 x_{1,i}+\cdots+\beta_k x_{k,i})}}$$

دامنه تابع لجستیک شامل اعداد حقیقی است و برد این تابع در بازه صفر و یک می‌باشد. در ادامه نحوه محاسبه و بازنویسی این تابع نمایش داده شده است:

$$\sigma(t) = \frac{e^t}{1+e^t} = \frac{1}{1+e^{-t}}$$

$$Pr\left(y_i = 1/\vec{x}_i : \vec{\beta}\right) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_{1,i}+\cdots+\beta_k x_{k,i})}} = \sigma(\beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i})$$

با توجه به مدل ریاضی تابع لجستیک، نمایش نموداری این تابع بصورت زیر است:



شکل ۱. تابع لجیستیک

مدل پروبیت

این مدل نیز مانند مدل لاجیت یک مدل غیر خطی است و از تئوری مطلوبیت گرفته شده و توسط مک فادن[1] ارائه شده است. در مدل پروبیت برای تعریف متغیر وابسته از تابع تجمعی نرمال استفاده می‌گردد. در این مدل مقادیر خروجی بازه ۰ تا ۱ است (گرین[2]، ۲۰۰۳). همچنین این مدل دارای توزیع نرمال است و نتایج آن شبیه مدل لاجیت است اما این مدل به علت پیچیدگی محاسبه آن کمتر مورد استفاده می‌شود (گجراتی[3]، ۲۰۰۴). این مدل به صورت زیر است:

$$p = \Phi(x'\beta) = \int_{-\infty}^{x'\beta} \gamma(t)d^{\gamma}$$

که x متغیرهای وابسته بوده و β پارامترهای مدل هستند و $\gamma(z)=(1/2\pi)e^{\left(\frac{-t^2}{2}\right)}$ تابع چگالی نرمال استاندارد است. برای برآورد این مدل نیز از MLE استفاده می‌شود.

یافته‌ها

در ابتدا آمار توصیفی متغیرهای بکار رفته در این تحقیق در جدول زیر (جدول ۱) آمده است:

جدول ۱. آمار توصیفی

متغیر	تعداد	میانگین	انحراف معیار	حداقل	چارک اول	میانه	چارک سوم	حداکثر
قیمت پایانی	۴۰۸	۴۳۸۵.۴۰	۱۱۵۸.۰۴	۲۰۶۴	۳۷۴۷.۲۵	۴۴۶۸.۵۰	۵۲۱۲.۵۰	۶۳۴۰
تعداد معاملات	۴۰۸	۸۲۹۷.۶۸	۵۵۰۱.۰۲	۱۵۵	۴۷۶۷.۵۰	۶۸۵۹.۰۰	۱۰۳۷۹.۰۰	۵۲۹۵۸
$\frac{P}{E_{ttm}}$	۴۰۸	۵.۴۹	۰.۸۶	۳.۴۷	۴.۸۸	۵.۵۱	۶.۲۵	۶.۹۸
درصد خالص حقیقی	۴۰۸	-۰.۰۲۸	۰.۳۱۸	-۰.۸۲۸	-۰.۲۴۱	-۰.۰۴۷	۰.۱۷۱	۰.۸۵۰
سرانه خرید حقیقی	۴۰۸	۳۶۶,۷۸۳,۶۰۰	۴۶۷,۸۴۱,۷۰۰	۴۳,۶۱۰,۲۷۰	۱۵۳,۹۸۴,۱۰۰	۲۳۰,۹۲۹,۸۰۰	۳۶۶,۱۰۸,۲۰۰	۴,۰۶۵,۶۶۶,۰۰۰
سرانه فروش حقیقی	۴۰۸	۴۵۰,۲۱۷,۴۰۰	۲۷۱,۱۷۷,۳۰۰	۱۰۸,۰۵۴,۸۰۰	۲۷۶,۲۶۱,۰۰۰	۳۹۸,۶۹۵,۴۰۰	۵۳۸,۲۲۲,۷۰۰	۳,۲۹۴,۲۲۰,۰۰۰

جمع‌بندی کلی جدول آمار توصیفی نشان می‌دهد که داده‌های مالی مربوط به قیمت و ویژگی‌های معاملاتی سهام فولاد اصفهان دارای نوسانات قابل توجه، چولگی و کشیدگی زیاد و دامنه تغییرات بسیار وسیع هستند. این ویژگی‌ها اهمیت انتخاب مدل‌های غیرخطی مانند پروبیت را افزایش می‌دهد. انحراف معیار حدود ۱۱۴۰ تا ۱۱۷۳ است که بسیار زیاد است و نشان می‌دهد قیمت سهام روزانه نوسان شدیدی داشته است. بیشتر متغیرهای بکار رفته در مدل پایدار نیستند و دارای نوسان زیاد هستند، بنابراین مدل خطی ممکن است عملکرد مناسبی نداشته باشد. بسیاری از متغیرهای مالی دارای نوسانات شدید، کشیدگی بالا و داده‌های پرت هستند. توزیع اکثر داده‌ها نرمال نیست و این موضوع باعث تضعیف مدل خطی می‌شود. مدل پروبیت به دلیل انعطاف‌پذیری بیشتر در مواجهه با داده‌های نامتقارن عملکرد مناسب‌تری دارد. همچنین مطابق با نتایج جدول متغیرهای حجمی، ارزشی، سرانه‌ها و قدرت خرید شدیداً نامتقارن و پرنوسان هستند و نقش مهمی در مدل‌های غیرخطی دارند.

[1] McFadden
[2] Greene
[3] Gujarati

همچنین نتایج برای نرمال بودن داده‌ها با آماره *Jarque–Bera* انجام گرفت و فرض نرمال بودن داده‌ها تأیید شد. در خصوص ناهمسانی واریانس نیز به روش *Breusch–Pagan* نشان داده شد که ناهمسانی واریانس وجود ندارد و در نهایت خودهمبستگی جملات اخلال با آماره *Durbin–Watson* و مقدار ان یعنی ۰.۲۶ نشان داد که خودهمبستگی مثبت وجود دارد که از ویژگی‌های داده‌های مالی روزانه است.

در ادامه نتایج تخمین مدل رگرسیون خطی در جدول زیر (جدول ۲) آمده است:

جدول ۲: نتایج رگرسیون خطی

متغیر	ضریب	خطای معیار	آماره t	P-Value
عرض از مبدأ	۸۲۹.۵۳	۲۹۳.۲۳	۲.۸۳	۰.۰۰۵***
تعداد معاملات	-۰.۰۵۱	۰.۰۰۹	-۵.۴۳	۰.۰۰۰***
$\frac{P}{E_{ttm}}$	۸۳۴.۱۲	۵۳.۸۸	۱۵.۴۸	۰.۰۰۰***
درصد خالص حقیقی	-۶۶۵.۴۵	۲۲۸.۳۲	-۲.۹۱	۰.۰۰۴***
سرانه خرید حقیقی	-۰.۰۰۰۰۲۸۴	-۰.۰۰۰۰۱۱۲	-۲.۵۴	۰.۰۱۱**
سرانه فروش حقیقی	-۰.۰۰۰۱۱۲	-۰.۰۰۰۰۱۹۳	-۵.۸۲	۰.۰۰۰***

***در سطح ۱ درصد معناداری، **در سطح ۵ درصد معناداری، *در سطح ۱۰ درصد معناداری

در نهایت معادله کلی خط رگرسیون به صورت زیر است:

$$Y^=829.53-0.051(TedadeMoamelat)+834.12(P/E)-665.45(DarsadeKhales)-0.0000284(SaraneKharid)-0.000112(SaraneForosh)$$

در نهایت ضرایب متغیرهای مورد بررسی به روشنی نحوه تأثیر و شدت آن را بر روی متغیر وابسته نشان می‌دهند. مطابق با نتایج به دست آمده از جدول ۲، قوی‌ترین عامل بنیادی تأثیرگذار با اثر مثبت بر متغیر وابسته $\frac{P}{E_{ttm}}$ است. در ضمن بایستی یادآور شد که این نتیجه نقش انتظارات سودآوری سرمایه‌گذاران را تأیید می‌کند.

در اینجا متغیر وابسته دودویی به صورت میانه قیمت‌ها در نظر گرفته شده است. به عبارت دیگر در صورتی که قیمت پایانی از میانه قیمت‌ها بالاتر باشد عدد ۱ و در غیر اینصورت عدد ۰ در نظر گرفته شده است. در ادامه نتایج تخمین مدل لاجیت در جدول زیر (جدول ۳) آمده است:

جدول ۳: نتایج مدل لاجیت

متغیر	ضریب	خطای معیار	آماره z	P-Value	Odds Ratio
عرض از مبدأ	-۷.۱۰	۱.۱۳	-۶.۳۱	۰.۰۰۰***	۰.۰۰۰۸
تعداد معاملات	-۰.۰۰۰۱۹	۰.۰۰۰۰۴	-۴.۳۵	۰.۰۰۰***	۰.۹۹۹۸
$\frac{P}{E_{ttm}}$	۲.۰۱	۰.۲۴	۸.۵۳	۰.۰۰۰***	۷.۴۵
درصد خالص حقیقی	-۲.۱۰	۰.۵۹	-۳.۵۶	۰.۰۰۰***	۰.۱۲
سرانه خرید حقیقی	-۰.۰۰۰۰۰۰۶۵	-۰.۰۰۰۰۰۰۰۳۶	-۱.۸۴	۰.۰۶۷*	۱.۰۰
سرانه فروش حقیقی	-۰.۰۰۰۰۰۰۵۱	-۰.۰۰۰۰۰۰۰۱	-۵.۱۰	۰.۰۰۰***	۱.۰۰

***در سطح ۱ درصد معناداری، **در سطح ۵ درصد معناداری، *در سطح ۱۰ درصد معناداری

در خصوص جدول فوق یادآور می‌گردد که *Odds Ratio* نشان دهنده این است که یک واحد تغییر در متغیر مستقل شانس اینکه متغیر وابسته بیشتر از میانه گردد چقدر تغییر می‌یابد. در اینجا نیز همانند سایر روش رگرسیون خطی قوی‌ترین عامل بنیادی تأثیرگذار با اثر مثبت بر متغیر وابسته $\frac{P}{E_{ttm}}$ است.

در اینجا نیز در صورتی که قیمت پایانی از میانه قیمت‌ها بالاتر باشد عدد ۱ و در غیر اینصورت عدد ۰ در نظر گرفته شده است. مدل پروبیت احتمال وقوع *Y=1* را با استفاده از تابع توزیع نرمال تجمعی برآورد می‌کند. در ادامه نتایج تخمین مدل پروبیت در جدول زیر (جدول ۴) آمده است:

جدول ۴: نتایج مدل پروبیت

متغیر	ضریب	خطای معیار	آماره z	P-Value
عرض از مبدأ	-۴.۴۸	۰.۶۱	-۷.۲۹	۰.۰۰۰***
تعداد معاملات	-۰.۰۰۰۱۴	۰.۰۰۰۰۲	-۶.۵۱	۰.۰۰۰***
$\frac{P}{E_{ttm}}$	۱.۱۸	۰.۱۳	۹.۴۴	۰.۰۰۰***
درصد خالص حقیقی	-۰.۹۲	۰.۳۱	-۲.۹۵	۰.۰۰۳***
سرانه خرید حقیقی	-۰.۰۰۰۰۰۰۰۴۲	-۰.۰۰۰۰۰۰۰۲۰	-۲.۰۸	۰.۰۳۸**
سرانه فروش حقیقی	-۰.۰۰۰۰۰۰۱۸	-۰.۰۰۰۰۰۰۰۲۹	-۶.۰۸	۰.۰۰۰***

***در سطح ۱ درصد معناداری، **در سطح ۵ درصد معناداری، *در سطح ۱۰ درصد معناداری

در اینجا نیز همانند سایر روش‌های پیشین قوی‌ترین عامل بنیادی تأثیرگذار با اثر مثبت بر متغیر وابسته $\frac{P}{E_{ttm}}$ است. همچنین عملکرد ضعیف‌تر مدل پروبیت نسبت به لاجیت را می‌توان به حساسیت بیشتر آن نسبت به فرض نرمال بودن توزیع خطاها نسبت داد. این موضوع سبب می‌شود که در شرایط نوسانات شدید بازار، انعطاف‌پذیری مدل پروبیت کاهش یابد.

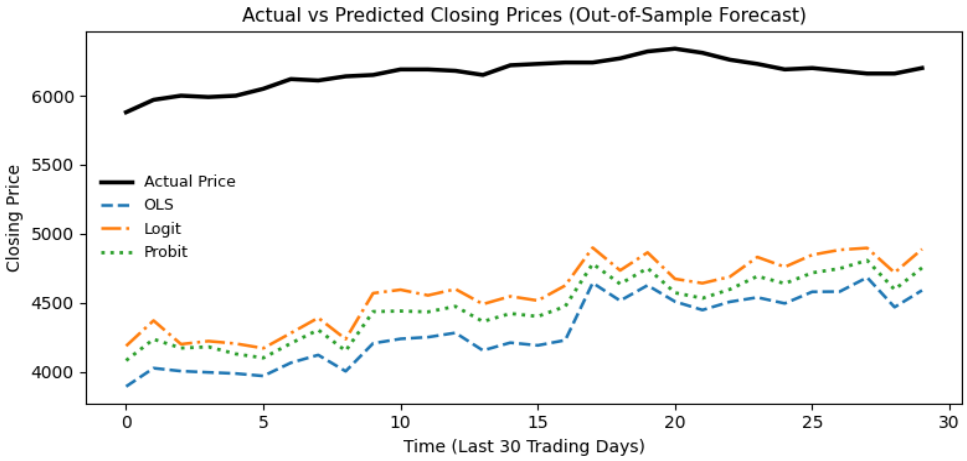
در ادامه به مقایسه دقت پیش‌بینی روش‌های مورد استفاد در این تحقیق پرداخته می‌شود.برای اینکه دقت روش‌های مورد استفاده بررسی شود فرض شده است که داده‌های واقعی سی روز انتهای فایل بعنوان داده‌های واقعی در نظر گرفته شده و سپس با توجه به ضرایب به دست آمده روش‌های قبل و مقادیر آنها به تخمین نتایج پرداخته شد. بعبارت دیگر بهترین انطباق با مقدار واقعی قیمت در ۳۰ روز پایانی در نظر گرفته شده است. نتایج مربوط به دقت پیش بینی بر اساس معیارهای *MSE*[1]، *RMSE*[2] و *MAE*[3] در جدول زیر (جدول ۵) آمده است.

جدول ۵: مقایسه دقت کارایی روش‌های مختلف

مدل	MSE	RMSE	MAE
رگرسیون خطی	۳,۴۸۸,۹۸۵	۱,۸۶۷	۱,۸۵۹
لاجیت	۲,۵۶۲,۷۰۲	۱,۶۰۰	۱,۵۹۱
پروبیت	۲,۹۱۳,۹۰۸	۱,۷۰۷	۱,۶۹۹

نتایج به وضوح نشان می‌دهد که روش لاجیت بهتر از روش‌های دیگر به لحاظ دارا بودن کمترین مقادیر ممکن خطا در هر سه معیار بوده است. همچنین نتایج نشان می‌دهد مدل پروبیت بهتر از رگرسیون خطی بوده و خطاهای کمتری در پیش‌بینی دارد. در ادامه دقت پیش‌بینی روش‌های مورد استفاده به لحاظ نموداری آمده است:

[1] Mean Squared Error
[2] Root Mean Squared Error
[3] Mean Absolute Error



شکل ۲. مقایسه دقت پیش‌بینی مدل‌های مورد استفاده با داده‌های واقعی متغیر وابسته

بحث و نتیجه‌گیری

یافته‌های این پژوهش نشان داد که اگرچه مدل رگرسیون خطی از نظر تبیین رابطه میان متغیرهای معاملاتی و قیمت سهام دارای قدرت توضیحی قابل قبول است، اما در پیش‌بینی کوتاه‌مدت قیمت سهام، به‌ویژه در محیط‌های پرنوسان بازار سرمایه، عملکرد ضعیف‌تری نسبت به مدل‌های غیرخطی از خود نشان می‌دهد. این نتیجه با مبانی نظری اقتصادسنجی همخوان است؛ زیرا رگرسیون خطی بر فرض خطی‌بودن روابط، نرمال‌بودن توزیع خطاها و ثبات واریانس استوار است، در حالی‌که داده‌های مالی روزانه معمولاً دارای نوسانات شدید، چولگی و رفتارهای نامتقارن هستند که این فروض را نقض می‌کنند (Greene, 2003; Gujarati, 2004). بنابراین، ضعف نسبی رگرسیون خطی در پیش‌بینی قیمت سهام را می‌توان ناشی از ناتوانی آن در جذب رفتارهای غیرخطی و شوک‌های کوتاه‌مدت بازار دانست؛ موضوعی که در مطالعات اخیر نیز به آن اشاره شده است (Guo & Wang, 2023; Zakamulin, 2024).

در مقابل، نتایج این پژوهش نشان داد که مدل لاجیت در مقایسه با سایر روش‌ها، کمترین میزان خطای پیش‌بینی را بر اساس معیارهای MSE، RMSE و MAE دارد و بیشترین انطباق را با قیمت‌های واقعی سهام ارائه می‌دهد. این یافته تأییدکننده دیدگاهی است که بر ماهیت غیرخطی و رژیمی رفتار قیمت سهام تأکید دارد. مدل لاجیت با تبدیل مسئله پیش‌بینی به یک چارچوب احتمالاتی، قادر است تغییر وضعیت قیمت را بهتر از مدل‌های خطی شناسایی کند و حساسیت بیشتری نسبت به تحولات کوتاه‌مدت بازار نشان دهد (Omogoroye et al., 2023). برتری مدل لاجیت نسبت به رگرسیون خطی در این پژوهش با نتایج مطالعاتی همسو است که نشان می‌دهند مدل‌های مبتنی بر احتمال و طبقه‌بندی، در افق‌های زمانی کوتاه، دقت پیش‌بینی بالاتری نسبت به مدل‌های خطی دارند (Nabipour et al., 2020; Toochaei & Moeini, 2023).

نتایج به‌دست‌آمده درباره عملکرد مدل پروبیت نیز حائز اهمیت است. اگرچه این مدل نسبت به رگرسیون خطی عملکرد بهتری داشت، اما در مقایسه با مدل لاجیت از دقت پیش‌بینی پایین‌تری برخوردار بود. این تفاوت را می‌توان به حساسیت بالاتر مدل پروبیت نسبت به فرض نرمال‌بودن توزیع خطاها نسبت داد؛ به‌گونه‌ای که در شرایط نوسانات شدید و داده‌های نامتقارن، انعطاف‌پذیری آن کاهش می‌یابد (Greene, 2003). مطالعات پیشین نیز نشان داده‌اند که در داده‌های مالی واقعی، که اغلب از توزیع‌های غیرنرمال پیروی می‌کنند، مدل لاجیت معمولاً عملکرد باثبات‌تری نسبت به پروبیت ارائه می‌دهد (Alizadeh Chamazkoti et al., 2024; Gujarati, 2004).

از منظر متغیرهای توضیحی، نتایج این پژوهش نشان داد که نسبت قیمت به سود (P/E) قوی‌ترین متغیر با اثر مثبت بر قیمت سهام در هر سه مدل مورد بررسی است. این یافته نشان‌دهنده نقش کلیدی انتظارات سودآوری سرمایه‌گذاران در شکل‌دهی قیمت سهام است و با مطالعاتی که بر اهمیت اطلاعات بنیادی و پیش‌بینی‌های مدیریت در تعیین بازده آتی سهام تأکید دارند، همخوانی دارد (Al-Zubaidi et al., 2025; Kitagawa & Shuto, 2024; Tsai et al., 2024). در مقابل، متغیرهای رفتاری نظیر درصد خالص معاملات حقیقی و سرانه خرید و فروش حقیقی، عمدتاً اثر منفی و معناداری بر احتمال افزایش قیمت داشتند که می‌تواند بازتاب‌دهنده رفتارهای هیجانی و کوتاه‌مدت

سرمایه‌گذاران حقیقی در بازار سرمایه باشد. این نتیجه با ادبیات مالی رفتاری و شواهد مربوط به واکنش‌های بیش‌ازحد سرمایه‌گذاران همسو است (Boons et al., 2023; Zakamulin, 2024).

برتری مدل لاجیت در پیش‌بینی کوتاه‌مدت قیمت سهام را می‌توان در چارچوب تحولات اخیر ادبیات پیش‌بینی بازار سرمایه نیز تفسیر کرد. پژوهش‌های نوین نشان می‌دهند که حتی با وجود پیشرفت‌های قابل توجه در حوزه یادگیری ماشین و هوش مصنوعی، مدل‌های اقتصادسنجی غیرخطی ساده و قابل تفسیر همچنان می‌توانند عملکردی رقابتی ارائه دهند، به‌ویژه زمانی که داده‌ها محدود یا پرنوسان باشند (Kumbure et al., 2022; Masini et al., 2023; Thakkar & Chaudhari, 2021). در این راستا، نتایج پژوهش حاضر نشان می‌دهد که استفاده از مدل‌های پیچیده‌تر لزوماً به معنای بهبود دقت پیش‌بینی نیست و انتخاب مدل باید متناسب با ماهیت داده‌ها و هدف پیش‌بینی صورت گیرد (Tomar & Periyasamy, 2023).

یافته‌های این پژوهش با نتایج مطالعات انجام‌شده در بازار سرمایه ایران نیز همخوانی قابل توجهی دارد. پژوهش‌های داخلی نشان داده‌اند که روش‌های غیرخطی و ترکیبی، در مقایسه با رگرسیون خطی، توان پیش‌بینی بالاتری در بازارهای نوظهور دارند؛ بازاری که با محدودیت‌های اطلاعاتی، نقش پررنگ سرمایه‌گذاران حقیقی و نوسانات ساختاری مواجه است (Akhbari et al., 2024; Setayesh & Kazemnezhad, 2019; Shirzad et al., 2021). همچنین نقش عوامل کلان و ساختاری نظیر شرایط مالی، نقدشوندگی و شوک‌های قیمتی در کاهش کارایی مدل‌های خطی و افزایش اهمیت مدل‌های غیرخطی مورد تأکید قرار گرفته است (Arman et al., 2022; Nourvosh et al., 2020; Taghizadeh & Abdzadeh Kanafi, 2023).

در سطح بین‌المللی نیز، نتایج این پژوهش با مطالعاتی که به محدودیت‌های پیش‌بینی خطی در شرایط عدم‌قطعیت و شوک‌های کلان اشاره دارند، سازگار است. شواهد نشان می‌دهد که در دوره‌هایی با ریسک سیستماتیک بالا، بحران‌های اقتصادی یا شوک‌های اطلاعاتی، روابط خطی تضعیف شده و مدل‌های غیرخطی یا احتمالاتی عملکرد بهتری از خود نشان می‌دهند (Adebayo et al., 2022; Salisu et al., 2022; Salisu & Vo, 2020). علاوه بر این، پژوهش‌هایی که به نقش احساسات بازار و اخبار مالی پرداخته‌اند، تأکید می‌کنند که این عوامل می‌توانند نوسانات کوتاه‌مدت قیمت را تشدید کرده و ضرورت استفاده از مدل‌های منعطف‌تر را افزایش دهند (Nuriksan & Nikmah, 2025; Shami et al., 2025).

در مجموع، نتایج این پژوهش نشان می‌دهد که روابط میان متغیرهای معاملاتی و قیمت سهام ماهیتی غیرخطی دارد و مدل‌های مبتنی بر احتمال، به‌ویژه لاجیت، می‌توانند ابزار مناسبی برای پیش‌بینی کوتاه‌مدت قیمت سهام باشند. این یافته‌ها نه‌تنها از منظر روش‌شناختی اهمیت دارند، بلکه می‌توانند به بهبود تصمیم‌گیری سرمایه‌گذاران و تحلیل‌گران بازار سرمایه کمک کنند و در عین حال، پلی میان ادبیات کلاسیک اقتصادسنجی و رویکردهای نوین پیش‌بینی ایجاد نمایند (Ma et al., 2021; Papageorgiou et al., 2024; Yang, 2023).

یکی از محدودیت‌های اصلی این پژوهش تمرکز بر داده‌های مربوط به یک شرکت خاص و یک بازه زمانی مشخص است که می‌تواند تعمیم‌پذیری نتایج به سایر شرکت‌ها یا صنایع را محدود کند. همچنین، استفاده از داده‌های روزانه و تمرکز بر افق کوتاه‌مدت پیش‌بینی، امکان بررسی پویایی‌های بلندمدت قیمت سهام را فراهم نمی‌سازد. محدودیت دیگر، عدم ورود متغیرهای کلان اقتصادی و شاخص‌های احساسی بازار به مدل‌هاست که ممکن است بخشی از نوسانات قیمت سهام را توضیح دهند.

پژوهش‌های آینده می‌توانند با گسترش دامنه نمونه به چندین شرکت یا صنایع مختلف، قابلیت تعمیم نتایج را افزایش دهند. همچنین، مقایسه مدل‌های خطی و غیرخطی کلاسیک با روش‌های پیشرفته یادگیری ماشین و مدل‌های ترکیبی می‌تواند تصویر جامع‌تری از عملکرد روش‌های پیش‌بینی ارائه دهد. بررسی افق‌های زمانی متفاوت، به‌ویژه میان‌مدت و بلندمدت، و افزودن متغیرهای کلان اقتصادی و شاخص‌های رفتاری نیز می‌تواند مسیرهای پژوهشی ارزشمندی را فراهم آورد.

نتایج این پژوهش می‌تواند برای سرمایه‌گذاران و تحلیل‌گران بازار سرمایه راهنمایی عملی فراهم کند تا در پیش‌بینی کوتاه‌مدت قیمت سهام، صرفاً به مدل‌های خطی اکتفا نکنند و از مدل‌های غیرخطی مبتنی بر احتمال بهره بگیرند. همچنین، نهادهای مالی و شرکت‌های سرمایه‌گذاری می‌توانند با به‌کارگیری این مدل‌ها در سیستم‌های تصمیم‌یار خود، دقت پیش‌بینی و مدیریت ریسک را بهبود بخشند. استفاده ترکیبی از تحلیل بنیادی و مدل‌های غیرخطی نیز می‌تواند به اتخاذ تصمیمات آگاهانه‌تر در بازار سرمایه منجر شود.

مشارکت نویسندگان

در نگارش این مقاله تمامی نویسندگان نقش یکسانی ایفا کردند.

تشکر و قدردانی

از تمامی کسانی که در طی مراحل این پژوهش به ما یاری رساندند تشکر و قدردانی می‌گردد.

تعارض منافع

در انجام مطالعه حاضر، هیچ‌گونه تضاد منافعی وجود ندارد.

حمایت مالی

این پژوهش حامی مالی نداشته است.

موازین اخلاقی

در انجام این پژوهش تمامی موازین و اصول اخلاقی رعایت گردیده است.

References

Adebayo, T. S., Akadiri, S. S., & Rjoub, H. (2022). On the relationship between economic policy uncertainty, geopolitical risk and stock market returns in South Korea: a quantile causality analysis. *Annals of Financial Economics*, *17*(01), 2250008.

Akhbari, H. R., Mohammadzadeh Salteh, H., Baradaran Hasanzadeh, R., & Zeynali, M. (2024). Optimizing risk-based stock return prediction in selected industries of the Tehran Stock Exchange (Data Envelopment Analysis approach). *Financial Research*, *26*(2), 331-354.

Al-Zubaidi, H. S., Abdulhameed, M. A., & Fenjan, F. H. (2025). Evaluating common stock prices through the residual income model: Evidence from Iraq stock exchange. *Journal of Economic and Administrative Sciences*.

Alizadeh Chamazkoti, M., Fathabadi, M., Mahmodzadeh, M., & Ghavidel Doostkouei, S. (2024). The Possibility or Impossibility of Stock Price Prediction: Evidence from the Petrochemical Industry. *Financial Research Journal*, *26*(1), 81-104.

Arman, S. A., Anvari, E., & Raki Kianpour, S. (2022). Modeling the financial conditions dynamic index and examining its impact on the predictability of stock returns in Iran. *Asset Management and Financing*, *10*(1), 47-72.

Boons, M., Ottonello, G., & Valkanov, R. I. (2023). Excess Volatility in Professional Stock Return Forecasts. https://ssrn.com/abstract=4537181

Ghallabi, F., Souissi, B., Du, A. M., & Ali, S. (2025). ESG stock markets and clean energy prices prediction: Insights from advanced machine learning. *International Review of Financial Analysis*, *97*(1), 103889. https://doi.org/10.1016/j.irfa.2024.103889

Greene, W. (2003). *Econometric Analysis* (Fifth Edition ed.). Prentice Hall.

Gujarati, D. (2004). *Basic Econometrics* (Fourth Edition ed.). McGraw-Hill.

Guo, Y., & Wang, W. (2023). A robust adaptive linear regression method for severe noise. *Knowledge and Information Systems*, *65*(11), 4613-4653. https://doi.org/10.1007/s10115-023-01924-4

Kitagawa, N., & Shuto, A. (2024). Unexpected management forecasts and future stock returns. *Journal of Business Finance & Accounting*.

Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, *197*, 116659.

Ma, F., Wang, R., Lu, X., & Wahab, M. I. M. (2021). A comprehensive look at stock return predictability by oil prices using economic constraint approaches. *International Review of Financial Analysis*, *78*, 101899.

Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2023). Machine learning advances for time series forecasting. *Journal of Economic Surveys*, *37*(1), 76-111.

Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., & Mosavi, A. (2020). Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis. *IEEE Access*, *8*, 150199-150212.

Nourvosh, I., Mohseni Dehkalani, N., & Rahimipour, A. (2020). The impact of liquidity shocks and stock bubbles on predicting the stock price index in the Tehran Stock Exchange. *Investment Knowledge*, *9*(33), 259-282.

Nuriksan, S., & Nikmah, N. (2025). Financial News Sentiment and Investor Confidence: Determinants of Stock Price Volatility in IDX30. *Journal of Humanities Social Sciences and Business (Jhssb)*, *4*(3), 493-506. https://doi.org/10.55047/jhssb.v4i3.1648

Omogoroye, O. O., Olaniyi, O. O., Adebiyi, O. O., Oladoyinbo, T. O., & Olaniyi, F. G. (2023). Electricity consumption (kW) forecast for a building of interest based on a time series nonlinear regression model. *Asian Journal of Economics, Business and Accounting*, *23*(21), 197-207.

Papageorgiou, G., Gkaimanis, D., & Tjortjis, C. (2024). Enhancing Stock Market Forecasts with Double Deep Q-Network in Volatile Stock Market Environments. *Electronics*, *13*(9), 1629.

Salisu, A. A., Demirer, R., & Gupta, R. (2022). Financial turbulence, systemic risk and the predictability of stock market volatility. *Global Finance Journal*, *52*, 100699.

Salisu, A. A., & Vo, X. V. (2020). Predicting stock returns in the presence of COVID-19 pandemic: The role of health news. *International Review of Financial Analysis*, *71*, 101546.

Setayesh, M. H., & Kazemnezhad, M. (2019). Investigating the usefulness of variable reduction methods in predicting the stock returns of companies listed on the Tehran Stock Exchange. *Empirical Studies in Financial Accounting*, *16*(63), 83-107.

Shami, A., Taghipourian, Y., Maranjouri, M., & Fallah, R. (2025). Explaining the impact of financial distress risk on stock price crash risk. *Scientific Research Quarterly of Accounting and Management*, *14*(4), 303-312.

Shirzad, L., Kheradyar, S., & Chirani, E. (2021). Analysis of risky and non-risky approaches of investment opportunity levels on the predictability of stock returns. *Financial Knowledge of Securities Analysis (Financial Studies)*, *14*(50), 45-56.

Singh, A., Bhardwaj, G., Srivastava, A. P., Bindra, A., & Chaudhary, P. (2022). Application of neural network to technical analysis of stock market prediction. 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM),

Taghizadeh, R., & Abdzadeh Kanafi, M. (2023). An analysis of the capital market using the network approach. *Financial Research*, *25*(3).

Thakkar, A., & Chaudhari, K. (2021). Fusion in stock market prediction: a decade survey on the necessity, recent developments, and potential future directions. *Information Fusion*, *65*, 95-107.

Tomar, M., & Periyasamy, V. (2023). The role of reference data in financial data analysis: Challenges and opportunities. *Journal of Knowledge Learning and Science Technology*, *1*(1), 90-99.

Toochaei, M. R., & Moeini, F. (2023). Evaluating the performance of ensemble classifiers in stock returns prediction using effective features. *Expert Systems with Applications*, *213*, 119186.

Tsai, H. W., Che, H. C., & Bai, B. (2024). Exploring the Relationship between Patent Forward Citation and Stock Return Rate Using Empirical Data of China Stock Market. *Management*, *12*(2), 67-83.

Yang, J. (2023). Analysis of Influencing Factors of Stock Return Rate. *Highlights in Business, Economics and Management*, *21*, 394-399. https://doi.org/10.54097/hbem.v21i.14495

Zakamulin, V. (2024). Stock price overreaction: evidence from bull and bear markets. *Review of Behavioral Finance*, *16*(6), 998-1011. https://doi.org/10.1108/RBF-03-2024-0088