



Developing a Stock Price Prediction Model Using a Convolutional Neural Network and Long Short-Term Memory

Mostafa Shaygani

MSc., Department of Financial Management, Faculty of Management and Accounting, Farabi Campus, University of Tehran, Qom, Iran. E-mail: shaygani@ut.ac.ir

Asgar Noorbakhsh *

*Corresponding Author, Assistant Prof., Department of Financial Management, Faculty of Management and Accounting, Farabi Campus, University of Tehran, Qom, Iran. E-mail: anoorbakhsh@ut.ac.ir

Abstract

Objective

Predicting stock prices remains a significant challenge in the field of finance due to the complex and dynamic factors that influence market behavior, including economic conditions, financial news, and market fluctuations. Consequently, the development of accurate and robust prediction models has become increasingly important for supporting investment decisions and risk management. In this study, deep learning techniques are employed to develop a model for stock price prediction.

Methods

In this research, which is pragmatically oriented in terms of research objectives and quantitative in terms of data type, the focus is on learning models. The data used covers 21 years from 2001 to 2022. Initial data preprocessing was carried out using Excel software. Subsequently, for data preprocessing, the Anaconda distribution and Jupyter Notebook interpreter were employed. Pandas, NumPy, and scikit-learn libraries were used for data analysis and preprocessing, and the matplotlib library was utilized for data visualization. For price prediction, the Keras and TensorFlow libraries were employed. This study employed two approaches for stock price prediction: single models and a hybrid model. The single-model approach utilized either a Convolutional Neural Network (CNN) or a Long Short-

Citation: Shaygani, Mostafa & Noorbakhsh, Asgar (2026). Developing a Stock Price Prediction Model Using a Convolutional Neural Network and Long Short-Term Memory. *Financial Research Journal*, 28(3), 693-715. <https://doi.org/10.22059/frj.2025.369854.1007546> (in Persian)



Term Memory (LSTM) network, with each model comprising three layers. The hybrid model consisted of four layers, combining a CNN for feature extraction with an LSTM for sequential prediction, thereby leveraging the strengths of both architectures to improve prediction performance. The research process involves importing data, scaling the data, dividing it into training and testing sets (80% for training and the rest for testing), and transforming the data from time series to supervised learning. Finally, the data are fed into the models. Three evaluation metrics, R-Square, MAE, and RMSE, are used to assess the models.

Results

The results obtained from the evaluation criteria indicate that the hybrid model and the convolutional neural network (CNN) model yield acceptable results, while the outcomes from the short-term memory neural network model are less desirable. These findings underscore the importance of not only selecting the correct independent variables and defining and configuring the model correctly, which are essential prerequisites for deep learning models, but also highlight the significant impact of feature extraction on model prediction. The results demonstrate that the CNN model outperforms the short-term memory neural network model, while recurrent models such as the short-term memory neural network have often shown favorable outcomes in predicting time series. Ultimately, the results suggest that combining these two models leads to better performance compared to each individual model.

Conclusion

The presence of factors such as the range of fluctuations in model accuracy has a significant impact and increases the uncertainty of the model. However, in this research, which exclusively utilizes technical variables, the results indicate that, to achieve desirable outcomes, the choice of neural network architecture is crucial. Additionally, the number of layers in each model also proves to be influential in the results, as models that are overly simplistic or excessively complex can lead to issues such as underfitting or overfitting. Furthermore, the selection of features (independent variables) will also have a considerable impact on the results.

Keywords: Stock price prediction, Convolutional neural network, Long short-term memory neural network, Artificial neural network.

ارائه مدلی برای پیش‌بینی قیمت سهام مبتنی بر شبکه عصبی پیچشی و شبکه عصبی حافظه کوتاه‌مدت ماندگار

مصطفی شایگانی

کارشناس ارشد، گروه مدیریت مالی، دانشکده مدیریت و حسابداری، دانشکده‌گان فارابی، دانشگاه تهران، قم، ایران. رایانامه: shaygani@ut.ac.ir

عسگر نوربخش

* نویسنده مسئول، استادیار، گروه مدیریت مالی، دانشکده مدیریت و حسابداری، دانشکده‌گان فارابی، دانشگاه تهران، قم، ایران. رایانامه: anoorbakhsh@ut.ac.ir

چکیده

هدف: در دهه‌های اخیر، پیش‌بینی قیمت سهام به‌عنوان یکی از چالش‌های مهم در حوزه مالی مورد توجه قرار گرفته است. از آنجا که قیمت سهام تحت تأثیر عوامل گوناگونی از جمله شرایط اقتصادی، اخبار رویدادها و تغییرات در بازار قرار می‌گیرد، توسعه مدل‌های دقیق و پیشرفته برای پیش‌بینی آن از اهمیت ویژه‌ای برخوردار است. در این پژوهش برای پیش‌بینی قیمت سهام از مدل‌های یادگیری عمیق استفاده شده است.

روش: این پژوهش از منظر هدف، کاربردی و از نظر نوع داده‌ها، کمی است. تمرکز اصلی در این پژوهش، روی مدل‌های یادگیری است. داده‌های پژوهش در بازه زمانی سال‌های ۱۳۸۰ تا ۱۴۰۰ (در مجموع ۲۱ سال) جمع‌آوری شده است. پالایش اولیه داده‌ها با استفاده از نرم‌افزار اکسل صورت پذیرفته است. در ادامه، برای پیش‌پردازش داده‌ها، از توزیع‌کننده آناکوندا و مفسر ژوپیتر بوت استفاده شده است. برای تحلیل و پیش‌پردازش داده‌ها، از کتابخانه‌های Numpy، Pandas و scikit-learn و برای نمایش داده‌ها، از کتابخانه Matplotlib و برای پیش‌بینی قیمت، از کتابخانه‌های keras و Tensorflow استفاده شده است. در این پژوهش از دو مدل به‌صورت مستقل و یک مدل ترکیبی برای پیش‌بینی قیمت بهره‌برده شده است. مدل‌های مستقل از سه لایه تشکیل شده‌اند و عبارت‌اند از: مدل شبکه عصبی پیچشی و مدل شبکه عصبی حافظه کوتاه‌مدت ماندگار. مدل ترکیبی نیز از چهار لایه تشکیل شده؛ یعنی برای استخراج ویژگی‌ها، از شبکه عصبی پیچشی و برای پیش‌بینی داده‌ها، از شبکه عصبی حافظه کوتاه‌مدت ماندگار استفاده شده است. مراحل انجام پژوهش به این صورت بوده است که ابتدا داده‌ها وارد نرم‌افزار شده، سپس در مقیاس داده‌ها تغییر ایجاد شد. در ادامه، داده‌ها به دو دسته آموزش و آزمون تقسیم شدند؛ به‌گونه‌ای که ۸۰ درصد داده‌ها برای آموزش مدل و مابقی برای آزمون استفاده شدند. در نهایت، داده‌ها از حالت سری زمانی به‌صورت نظارت شده تبدیل و وارد مدل شدند. برای ارزیابی مدل‌ها، از سه معیار ارزیابی که شامل ضریب تعیین، خطای میانگین مطلق و ریشه میانگین مربعات خطا استفاده شد.

استناد: شایگانی، مصطفی و نوربخش، عسگر (۱۴۰۵). ارائه مدلی برای پیش‌بینی قیمت سهام مبتنی بر شبکه عصبی پیچشی و شبکه عصبی حافظه کوتاه‌مدت ماندگار. *تحقیقات مالی*، ۳(۳)، ۶۹۳-۷۱۵.

تاریخ دریافت: ۱۴۰۲/۰۹/۲۸

تاریخ ویرایش: ۱۴۰۴/۰۲/۱۹

تاریخ پذیرش: ۱۴۰۴/۰۶/۲۱

تاریخ انتشار: ۱۴۰۵/۰۷/۰۹

doi: <https://doi.org/10.22059/frj.2025.369854.1007546>

تحقیقات مالی، ۱۴۰۵، دوره ۲۸، شماره ۳، صص. ۶۹۳-۷۱۵

ناشر: دانشکده‌گان مدیریت دانشگاه تهران

نوع مقاله: علمی پژوهشی

© نویسندگان

یافته‌ها: نتایج به‌دست‌آمده از معیارهای ارزیابی، نشان می‌دهد که مدل ترکیبی و مدل شبکه عصبی پیچشی نتایج قابل قبولی دارند؛ اما نتایج حاصل از مدل شبکه عصبی حافظه کوتاه‌مدت ماندگار، مطلوب نیست. این نتایج نشان می‌دهد که علاوه بر انتخاب صحیح متغیرهای مستقل و همچنین، تعریف و پیکربندی درست مدل که برای مدل‌های یادگیری عمیق بسیار لازم است، استخراج ویژگی‌ها، تأثیر بسزایی در پیش‌بینی مدل دارد؛ زیرا نتایج نشان می‌دهد که مدل شبکه عصبی پیچشی، نسبت به مدل شبکه عصبی حافظه کوتاه‌مدت ماندگار عملکرد بهتری دارد، در حالی که مدل‌های بازگشتی مانند شبکه عصبی حافظه کوتاه‌مدت ماندگار، اغلب نتایج مطلوبی در پیش‌بینی سری‌های زمانی از خود نشان داده‌اند. در نهایت، نتایج نشان می‌دهد که ترکیب این دو مدل عملکرد بهتری نسبت هر یک از مدل‌ها دارد.

نتیجه‌گیری: وجود عواملی مانند دامنه نوسان در دقت مدل‌ها اثرگذاری زیادی دارند و ارباب مدل را افزایش می‌دهند؛ اما در این پژوهش که از متغیرهای فنی استفاده شده است، نتایج پژوهش نشان می‌دهد که برای رسیدن به نتایج مطلوب، علاوه بر انتخاب نوع شبکه عصبی، تعداد لایه‌های هر مدل نیز در نتایج مؤثر است؛ زیرا مدل‌های بسیار ساده یا بسیار پیچیده، باعث ایجاد مشکلاتی اعم از کم برآزش یا بیش برآزش خواهد شد. همچنین انتخاب ویژگی‌ها (متغیرهای مستقل) نیز اثر شایان توجهی در نتایج خواهد داشت.

کلیدواژه‌ها: پیش‌بینی قیمت سهام، شبکه عصبی پیچشی، شبکه عصبی حافظه کوتاه‌مدت ماندگار، شبکه عصبی مصنوعی.

مقدمه

بازارهای مالی یکی از جذاب‌ترین ابداعات عصر حاضر است. در این نوع بازارها، موفقیت یک سرمایه‌گذار به کیفیت اطلاعات و سرعت تصمیم‌گیری وی بستگی دارد. بنابراین پیش‌بینی آن‌ها برای سرمایه‌گذاران بسیار مهم است؛ زیرا پیش‌بینی دقیق به آن‌ها این امکان را می‌دهد تا آگاهانه درباره سرمایه‌گذاری‌ها تصمیم‌گیری کنند و ریسک‌ها را کاهش دهند. طبق نظریه‌های گشت تصادفی و بازار مالی کارا، هیچ سرمایه‌گذاری با استفاده از داده‌های تاریخی و فعلی نمی‌تواند سودی بیشتر از نرخ متوسط بازار به‌دست آورد؛ بنابراین پیش‌بینی امکان‌پذیر نیست (اوسار و دریفوس^۱، ۲۰۰۰) با این حال، پژوهشگران و سرمایه‌گذاران با فرضیه بازار کارآمد، هم از لحاظ نظری و هم از لحاظ تجربی مخالف و معتقدند که متغیرهای مالی مانند قیمت سهام، ارزش شاخص‌های بازار سهام و قیمت مشتقات مالی تا حدی پیش‌بینی‌پذیر هستند (چانگ، هان و پارک^۲، ۲۰۱۷). بنابراین به ارائه و آزمایش مدل‌های مربوط به رفتار بازار سهام علاقه و توجه نشان داده و تئوری‌های متعددی در مورد بازار سهام مطرح کرده‌اند. در این میان، محبوبیت شبکه‌های عصبی، به دلیل رفتار بی‌نظم، بی‌ثبات، نویزی و ناپارامتریک بازارهای مالی رو به افزایش است؛ زیرا از توانایی ذاتی در تخمین روابط غیرخطی با دقت زیاد، تحلیل‌های چندمتغیره بدون هیچ پیش‌فرضی و سهولت تعمیم برخوردارند که این امر باعث موفقیت نسبی آن‌ها در مدل‌سازی و پیش‌بینی بخش‌های متعدد بازارهای مالی شده است (لی^۳، ۲۰۰۹).

روند تغییر قیمت سهام، همواره به عنوان مشکل بسیار مهم در حوزه اقتصادی شناسایی شده است (وانگا و اسلوکا^۴، ۲۰۲۰). قیمت سهام تحت تأثیر عوامل مختلف داخلی و خارجی از جمله محیط اقتصادی داخلی و خارجی، وضعیت بین‌المللی، چشم‌انداز صنعت، داده‌های مالی شرکت‌های فهرست‌شده در بورس و عملکرد بازار سهام است (ژانگ و کیم^۵، ۲۰۱۹).

روش تحلیلی سنتی مبتنی بر اقتصاد و امور مالی است که به‌طور عمده از روش تحلیل بنیادی و روش تحلیل تکنیکال استفاده می‌کند. از یک سو، روش تحلیل بنیادی به ارزش ذاتی سهام بیشتر توجه می‌کند و عوامل خارجی مؤثر بر سهام مانند نرخ بهره، نرخ ارز، تورم، سیاست‌های صنعت، امور مالی شرکت‌های بورسی، روابط بین‌الملل را به صورت کیفی تحلیل می‌کند و از سوی دیگر، روش تحلیل تکنیکال بیشتر بر جهت قیمت سهام، حجم معاملات و انتظارات روان‌شناختی سرمایه‌گذاران متمرکز است. در حال حاضر، تحلیل بنیادی و تحلیل تکنیکال، رایج‌ترین روش‌های مورد استفاده برای بسیاری از سازمان‌ها و سرمایه‌گذاران فردی است (سوسا، مونته ویچی و میراندا^۶، ۲۰۱۹).

شبکه‌های عصبی مصنوعی با تحلیل‌های چندمتغیره، از ویژگی‌های غیرخطی داده‌های استنتاجی و سهولت تعمیم الگوریتم‌های پیش‌تاز و اصلی در استفاده از روش‌های یادگیری ماشین در مطالعات مالی برخوردارند و به ابزاری عامه‌پسند

1. Oussar & Dreyfus
2. Chong, Han & Park
3. Lee
4. Vanaga & Sloka
5. Zhang & Kim
6. Sousa, Montevecchi & Miranda

و غالب در امور مالی و اقتصاد تبدیل شده‌اند. عملکرد شبکه‌های عصبی، به‌طور مستقیم با کیفیت ویژگی‌ها مرتبط است و قراردادن تعداد بیش از اندازه متغیر مستقل در رگرسیون، یکی از موانع بسیار بزرگ در کاربرد علمی آن‌هاست؛ به همین دلیل، انتخاب ویژگی‌های مفید در پیش‌بینی نقش مهمی دارد.

با توجه به اینکه پیش‌بینی و مدل‌سازی قیمت همواره امری پُرچالش است و جزء جدایی‌ناپذیر تحلیل‌های سرمایه‌گذاران به‌شمار می‌رود و از طرفی، استفاده از شبکه‌های عصبی در امور مالی، گسترشی روز افزون داشته است، استفاده از مدل‌های مختلف شبکه عصبی، به‌خصوص مدل‌های ترکیبی، می‌تواند باعث ایجاد مدل‌های قدرتمند از هر لحاظ در این زمینه شود.

پیش‌بینی قیمت با استفاده از مدل‌های جدید، دقت در تحلیل و پیش‌بینی را افزایش دهد که این موضوع عملکرد سرمایه‌گذاری را بهبود می‌دهد. همچنین افزایش دقت پیش‌بینی قیمت، می‌تواند به کمک سازمان‌های نظارتی بیاید و تحرکات غیرقانونی قیمت را شناسایی کنند. در این پژوهش ما علاقه‌مند بودیم تا دقت پیش‌بینی روش‌های مختلف یادگیری عمیق را در بازار سهام تهران بررسی کنیم.

پیشینه پژوهش

با توجه به اهمیت زیاد پیش‌بینی قیمت‌ها در بازار سهام، سرمایه‌گذاران از روش‌های متنوعی برای این کار استفاده می‌کنند. عمده این روش‌ها را می‌توان به دو دسته طبقه‌بندی کرد: برخی از آن‌ها گذشته نگرد و با پیروی از الگوهای قیمتی، تلاش می‌کنند که بازارها را بهتر پیش‌بینی کنند. برخی از روش‌های دیگر، به‌صورت آینده نگر سعی می‌کنند تا با تحلیل سناریوهای مختلف، این مهم را انجام دهند.

خاره، دارکار، گوپتا و عطار^۱ (۲۰۱۷) طی مقاله‌ای تحت عنوان «پیش‌بینی کوتاه‌مدت قیمت سهام با استفاده از یادگیری عمیق» از داده‌های یک‌ساله ۱۰ شرکت از بورس اوراق بهادار نیویورک، به‌صورت دقیقه‌ای (تایم فریم یک دقیقه) استفاده کرده است. در این مقاله از دو نوع شبکه عصبی استفاده شده است: یک شبکه عصبی پیش‌خور با عنوان شبکه عصبی چندلایه پرسپترون و یک شبکه عصبی بازگشتی با عنوان شبکه عصبی کوتاه‌مدت ماندگار. برای مقایسه این دو مدل، از معیار خطای میانگین مربعات ریشه استفاده شده است. نتایج نشان می‌دهد که مدل شبکه عصبی چندلایه پرسپترون عملکرد بهتری داشته است.

لی و سو^۲ (۲۰۱۷) طی مقاله‌ای تحت عنوان «پیش‌بینی قیمت سهام با اخبار مالی مبتنی بر شبکه‌های عصبی پیچشی بازگشتی» از ترکیب یک شبکه عصبی پیچشی و یک شبکه عصبی کوتاه‌مدت ماندگار و الگوریتم تبدیل کلمه به بردار استفاده کردند. داده‌های ورودی، قیمت‌های تاریخی و عناوین اخبار آن سهم بود. نتایج به‌دست آمده نشان داد که مدل ترکیبی از مدل شبکه عصبی کوتاه‌مدت ماندگار عملکرد بهتری دارد.

کائو، لی و لی^۱ (۲۰۱۸) طی مقاله‌ای تحت عنوان «مدل پیش‌بینی سری زمانی مالی بر اساس LSTM-CEEMDAN» روی چهار شاخص DAX^۲، HSI^۳، 500S&P و SSE^۴ تمرکز کردند و از قیمت بسته شدن روزانه برای پیش‌بینی استفاده کردند. نتایج این مدل ترکیبی نشان می‌دهد که خطای به‌دست آمده از این مدل، براساس توابع MAE^۵، MAPE^۶، RMSE^۷، نسبت به مدل‌های LSTM، SVM CEEMDAN-SVM و CEEMDAN-MLP عدد کمتری را نشان می‌دهد. همچنین این مدل قابلیت بهینه‌تر شدن را دارد؛ زیرا برای ورودی فقط از قیمت بسته‌شدن استفاده شده است، در حالی که می‌توان تعداد ورودی‌ها را افزایش داد و از بیشترین قیمت و کمترین قیمت و حجم معاملات و سایر ویژگی‌های قیمت استفاده کرد.

هیرانشا، گوپالاکریشنان و سومان^۸ (۲۰۱۸) طی مقاله‌ای تحت عنوان «پیش‌بینی بازار سهام هند با استفاده از مدل‌های یادگیری عمیق» با استفاده از چهار نوع معماری یادگیری عمیق، یعنی MLP^۹، RNN^{۱۰}، LSTM و CNN، به پیش‌بینی قیمت سهام از دو بازار مختلف هند و آمریکا اقدام کردند؛ به‌گونه‌ای که مدل‌ها بر اساس سهام شرکتی از هند آموزش داده شدند و پنج شرکت مختلف از دو بازار هند و آمریکا پیش‌بینی را پیش‌بینی کردند. نتایج نشان می‌دهد که مدل CNN از سایر مدل‌ها در این روش عملکرد بهتری داشته است.

یاماک، یوجیان و گادوسی^{۱۱} (۲۰۱۹) طی مقاله‌ای تحت عنوان «مقایسه بین مدل‌های ARIMA و LSTM و GRU برای پیش‌بینی سری‌های زمانی»، به مقایسه بین مدل‌های اقتصادسنجی و یادگیری عمیق پرداختند و از داده‌های ارز دیجیتال بیت‌کوین استفاده کردند. آن‌ها از قیمت باز شدن، بسته شدن، بیشترین قیمت و کمترین قیمت و حجم معامله بیت‌کوین و حجم دلار استفاده کردند. نتایج این پژوهش بر اساس معیار MAPE و RMSE، نشان می‌دهد که مدل ARIMA از دو مدل دیگر عملکرد بهتر و مدل GRU نیز از LSTM عملکرد مطلوب‌تری داشته است. البته، یکی از محدودیت‌های این مقاله آن است که برای یادگیری عمیق، داده‌های کمی انتخاب شده است و محققان این مقاله نیز خود به این موضوع آگاه هستند که اگر داده‌های بیشتری استفاده شود، ممکن است نتایج تغییر کند.

مهتاب، سن و دوتا^{۱۲} (۲۰۲۱) طی مقاله‌ای با عنوان «پیش‌بینی قیمت سهام با استفاده از یادگیری ماشین و مدل‌های یادگیری عمیق مبتنی بر LSTM» از مقادیر شاخص NIFTY50 بورس ملی هند (NSE) در دوره زمانی ۲۹ دسامبر ۲۰۱۴ تا ۳۱ ژوئیه ۲۰۲۰ استفاده کردند. چارچوب پیش‌بینی در این پژوهش، ساخت چهار مدل رگرسیون مبتنی

1. Cao, Li & Li
2. Dax performance-index
3. Hang Seng Index
4. SSE Composite Index
5. Mean Absolute Error
6. Mean Absolute Percentage Error
7. Root Mean Square Error
8. Hiransha, Gopalakrishnan & Soman
9. Multi Layer Perceptron
10. Recurrent Neural Network
11. Yamak, Yujian & Gadosey
12. Mehtab, Sen & Dutta

بر یادگیری عمیق با استفاده از شبکه‌های حافظه بلندمدت و کوتاه‌مدت (LSTM) با رویکرد جدید walk-forward validation است. نتایج به‌وضوح نشان می‌دهد که مدل تک‌متغیره مبتنی بر LSTM که از داده‌های یک هفته قبل به‌عنوان ورودی برای پیش‌بینی قیمت باز شدن هفته آینده سری زمانی NIFTY50 استفاده می‌کند، دقیق‌ترین مدل است.

جینگ، وو و ونگ^۱ (۲۰۲۱) طی مقاله‌ای با عنوان «یک مدل ترکیبی ادغام یادگیری عمیق با تحلیل احساسات سرمایه‌گذار برای پیش‌بینی قیمت سهام»، یک مدل ترکیبی ارائه کردند که رویکرد یادگیری عمیق را با یک مدل تحلیل احساسات برای پیش‌بینی قیمت سهام ترکیب می‌کند. آن‌ها از یک مدل شبکه عصبی پیچشی برای طبقه‌بندی احساسات پنهان سرمایه‌گذاران استفاده کردند که از یک مجموعه داده بزرگ سهام استخراج شده بود. سپس با استفاده از رویکرد شبکه عصبی LSTM برای تجزیه و تحلیل شاخص‌های فنی از بازار سهام و نتایج تحلیل احساسات، یک مدل تحقیقاتی ترکیبی ارائه شده است. علاوه بر این کار، در شش صنعت کلیدی، در سه بازه زمانی در بورس اوراق بهادار شانگهای، آزمایش‌های واقعی انجام دادند تا اثربخشی و کاربرد مدل پیشنهادی را تأیید کنند. نتایج آزمایش نشان می‌دهد که مدل پیشنهادی نسبت به طبقه‌بندی‌کننده‌های پایه در طبقه‌بندی احساسات سرمایه‌گذار به عملکرد بهتری دست یافته است و این رویکرد ترکیبی در پیش‌بینی قیمت سهام در مقایسه با مدل واحد و مدل‌های بدون تحلیل احساسات عملکرد بهتری دارد.

نوربخش و شایگانی^۲ (۲۰۲۴) به مقایسه مدل‌های یادگیری عمیق برای پیش‌بینی قیمت سهام در بورس اوراق بهادار تهران پرداختند و پنج مدل، شامل دو مدل هیبریدی (CNN-LSTM و CNN-GRU) و سه مدل مستقل (CNN، LSTM و GRU) را بررسی کردند. داده‌های این پژوهش، اطلاعات پنج شرکت طی سال‌های ۲۰۰۱ تا ۲۰۲۲ بود. همچنین، از شاخص‌های فنی برای پیش‌بینی استفاده کردند. نتایج نشان داد که مدل هیبریدی CNN-LSTM در چهار شرکت از پنج شرکت، بهترین عملکرد را داشته است، در حالی که مدل مستقل CNN نیز عملکردی قابل مقایسه با مدل‌های هیبریدی از خود نشان داد. این مطالعه، اهمیت انتخاب معماری مناسب شبکه عصبی و تأثیر نوسان‌های بازار بر دقت پیش‌بینی را نشان می‌دهد.

فقیهی نژاد و مینایی (۱۳۹۷) در مقاله خود، به پیش‌بینی رفتار بازار سهام مبتنی بر شبکه‌های عصبی مصنوعی با رویکرد یادگیری جمعی هوشمند پرداختند. مطابق با نتایج پژوهش، مهم‌ترین چالش‌های مدل‌های پیشنهادی در بازار سهام، دقت نتایج و چگونگی افزایش کارآمدی پیش‌بینی بود. تحقیقات در این زمینه توجه کمی به پیش‌بینی جهت حرکت بعدی قیمت سهام داشته است، در حالی که این موضوع در مورد سودآوری بسیار مهم است.

حیدری و امیری (۱۴۰۱) در مقاله‌ای با عنوان «بررسی قدرت مدل‌های مبتنی بر هوش مصنوعی در پیش‌بینی روند قیمت سهام بورس اوراق بهادار تهران» به بررسی قدرت پیش‌بینی مدل‌های مبتنی بر هوش مصنوعی در پیش‌بینی روند قیمت سهام در بورس اوراق بهادار تهران پرداختند. در این پژوهش با استفاده از داده‌های ۱۵۰ شرکت بزرگ در بازه

1. Jing, Wu & Wang

2. Noorbakhsh & Shaygani

زمانی ۱۳۹۰ تا ۱۳۹۹، عملکرد مدل‌های مختلف یادگیری ماشین شامل مدل‌های خطی، شبکه‌های عصبی و مدل‌های حافظه بلندمدت کوتاه‌مدت (LSTM) مورد ارزیابی قرار گرفت. نتایج این پژوهش نشان می‌دهد که مدل‌های یادگیری عمیق، به‌ویژه LSTM، با دقت حدود ۷۰ درصد، نسبت به مدل‌های سنتی عملکرد بهتری دارند. همچنین، این مدل‌ها در پیش‌بینی روندهای نزولی موفق‌تر عمل می‌کنند. با این حال، به نظر نویسندگان این مدل‌ها نتایج خیره‌کننده‌ای برای سرمایه‌گذاران ارائه نمی‌دهند.

اصولیان، نیکمram و کریمی (۱۴۰۴) در پژوهشی به بررسی پیش‌بینی روند شاخص کل بورس اوراق بهادار تهران با استفاده از یک شبکه عصبی هیبریدی مبتنی بر استخراج ویژگی‌های چندمقیاسی زمانی پرداختند. مدل پیشنهادی ترکیبی از شبکه عصبی کانولوشنی (CNN) و شبکه حافظه بلندمدت - کوتاه‌مدت (LSTM) بود. برای تحلیل، داده‌های روزانه شاخص کل از سال ۱۳۷۷ تا ۱۴۰۱ استفاده شده بود. نتایج این پژوهش نشان داد که مدل هیبریدی پیشنهادی در مقایسه با مدل‌های رقیب، به‌ویژه در مقیاس‌های زمانی هفتگی و ماهانه، دقت کمتری دارد و مدل CNN به تنهایی عملکرد بهتری از خود نشان می‌دهد. این یافته‌ها اهمیت بررسی معماری‌های ساده‌تر شبکه‌های عصبی را در پیش‌بینی بازارهای مالی پرنوسان برجسته می‌کند.

محققان پژوهش حاضر با بهره‌گیری از یافته‌های تحقیقات پیشین و با هدف قدرت پیش‌بینی قیمت سهام، در این پژوهش از دو مدل شبکه عصبی به‌صورت جداگانه (یعنی مدل CNN و LSTM) و یک مدل ترکیبی (یعنی مدل LSTM-CNN) برای پیش‌بینی قیمت استفاده کردند. عمده تحقیقاتی که در بالا به آن اشاره شده است، از مدل‌های شبکه‌های عصبی به‌صورت مستقل استفاده کرده‌اند؛ اما در پژوهش حاضر، محققان به دنبال بهبود قدرت پیش‌بینی قیمت سهام با استفاده از مدل ترکیبی شبکه عصبی هستند.

روش‌شناسی پژوهش

داده‌های پژوهش

در این پژوهش به بررسی نتایج حاصل از سه مدل یادگیری عمیق پرداخته شده است. در اینجا از ۲۱ سال (۱۳۸۰ تا ۱۴۰۰) داده شرکت‌های ایران خودرو، نفت بهران، سیمان تهران، توسعه معادن روی ایران و البرز دارو استفاده شده است. در خصوص استفاده از داده‌های ۵ شرکت و دلایل آن توضیحاتی لازم است. در تحقیقات پیش‌بینی قیمت سهام با استفاده از شبکه‌های عصبی، انتخاب معماری مدل (CNN، LSTM و ترکیبی) از تعداد سهام مهم‌تر است؛ اما هر دو عامل به‌صورت مکمل بر کیفیت نتایج تأثیرگذارند. بر اساس مبانی نظری و شواهد تجربی عامل تعیین‌کننده اصلی، معماری مدل است. شواهد نشان می‌دهد که شبکه‌های عصبی پیشرفته مانند LSTM یا Transformer توانایی دارند تا الگوهای پیچیده غیرخطی و وابستگی‌های زمانی را استخراج کنند که حتی با داده‌های محدود نیز عملکرد قابل قبولی ارائه می‌دهند (هوکرایتر و اشمیدوبر^۱، ۱۹۹۷). همچنین مطالعات نشان می‌دهند که بهینه‌سازی هایپرپارامترها (مانند

لایه‌ها، dropout، تنظیم نرخ یادگیری) تأثیر مستقیم‌تری بر دقت پیش‌بینی دارد تا افزایش تعداد سهام (بورویخ، بوت‌ه و اوسترلی^۱، ۲۰۱۷). در پژوهشی که ژانگ و همکاران^۲ (۲۰۱۹) انجام دادند، مشخص شد که یک مدل LSTM بهینه‌شده روی تنها ۳ سهم، به دقت بالاتری نسبت به یک مدل ساده روی ۵۰ سهم دست یافت. با این حال، تعداد سهام تنها زمانی بحرانی است که تنوع رفتاری بازار را پوشش ندهد. برای مثال انتخاب سهام همگی از یک صنعت پرنوسان باشد. طبق تحقیقات ژونگ و انکه^۳ (۲۰۱۹) ۵ تا ۱۰ سهم با ویژگی‌های متمایز (نقدشوندگی، دامنه نوسان، صنعت) برای تعمیم‌پذیری اولیه کافی است. بنابراین در تحقیق حاضر با استناد به نتایج پژوهش‌های بالا و همچنین با در نظر گرفتن تنوع صنعت، نقدشوندگی و حجم معاملات سهم، دامنه نوسان متعارف سهم و اهمیت اقتصادی سهم‌ها، داده‌های این پنج شرکت استفاده شد.

برای انجام این پژوهش، فقط از ویژگی‌های گذشته‌نگر و فنی استفاده شده است که عبارت‌اند از: اولین قیمت، کمترین قیمت، بیشترین قیمت، آخرین قیمت، حجم معاملات، ارزش معاملات، تغییرات قیمت (آخرین قیمت)، درصد تغییرات قیمت (آخرین قیمت).

مدل شبکه عصبی حافظه کوتاه‌مدت ماندگار

شبکه عصبی حافظه کوتاه‌مدت ماندگار^۴، نخستین بار توسط هوکرایتر و اشمیدوبر در سال ۱۹۹۷ معرفی شد. البته در سال‌های بعد، بسیاری از محققان آن را بهبود دادند. شبکه LSTM یک شبکه نسبتاً قدیمی است؛ اما در طیف وسیعی از مسائل استفاده می‌شود و هنوز محبوبیت زیادی دارد.

شبکه عصبی حافظه کوتاه‌مدت ماندگار، نوعی از معماری شبکه‌های عصبی عمیق است که به‌طور خاص برای مدل‌سازی داده‌های دنباله‌ای مانند متون، گفتار، سیگنال‌های زمانی و داده‌های دنباله‌ای دیگر مورد استفاده قرار می‌گیرد. عناصر اصلی یک واحد شبکه عصبی حافظه کوتاه‌مدت ماندگار عبارت‌اند از:

۱. حافظه کوتاه‌مدت^۵: در این بخش، اطلاعاتی که در گام‌های گذشته از دنباله تصمیم‌گیری نهایی نیاز ندارند، حذف می‌شوند.

۲. حافظه بلندمدت^۶: این بخش اطلاعات مهم و معنادار از دنباله را ذخیره می‌کند تا بتواند از آن‌ها در آینده استفاده کند.

۳. دروازه‌ها^۷: دروازه‌هایی مانند دروازه ورودی^۸، دروازه فراموشی^۹ و دروازه خروجی^{۱۰} در یک واحد LSTM وجود

1. Borovykh, Bohte & Oosterlee

2. Zhang et al.

3. Zhong & Enke

4. Long short-term memory (LSTM)

5. Short-term memory

6. Long-term memory

7. Gates

8. Input gate

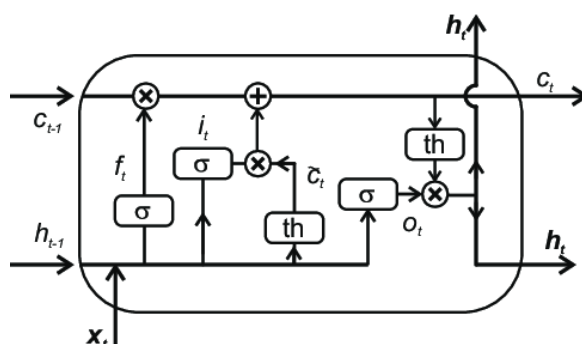
9. Forget gate

10. Output gate

دارد. این دروازه‌ها کنترل می‌کنند که کدام اطلاعات وارد حافظه کوتاه‌مدت شوند، کدام اطلاعات حذف شوند و کدام اطلاعات از حافظه درازمدت فراخوانده شوند.

۴. توابع فعال‌سازی^۱: توابع فعال‌سازی مانند تانژانت هایپربولیک^۲ برای محاسبه و کنترل جریان اطلاعات در واحد LSTM استفاده می‌شوند.

معماری شبکه عصبی LSTM در شکل ۱ نشان داده شده است.



شکل ۱. معماری شبکه عصبی LSTM

مدل شبکه عصبی پیچشی

شبکه عصبی کانولوشن^۳، برای اولین بار در حدود دهه ۱۹۸۰ ساخته شد و مورد استفاده قرار گرفت. بیشترین کاری که شبکه عصبی پیچشی در آن زمان می‌توانست انجام دهد، شناسایی ارقام دست‌نویس بود. این مورد بیشتر در بخش‌های پستی برای خواندن کد پستی، پین کد و غیره مورد استفاده قرار می‌گرفت. نکته مهمی که در مورد هر مدل یادگیری عمیق وجود دارد، این است که برای آموزش مقدار زیادی داده لازم است و همچنین، به منابع محاسباتی زیادی نیاز دارد. این مدل تا مدت‌ها، فقط به استفاده در بخش‌های پستی محدود بود؛ ولی با ظهور سخت‌افزارهای قوی و افزایش قدرت محاسباتی ماشین‌ها، در همه زمینه‌ها، به‌خصوص بینایی ماشین، از آن بسیار استقبال شد.

شبکه عصبی پیچشی، یکی از مهم‌ترین شبکه‌های عصبی عمیق است که معماری آن از الگویی مشابه اتصال نورون‌های مغز به یکدیگر پیروی می‌کند و از بخش Visual Cortex (بخش متعلق به بینایی مغز) الگوبرداری شده است. دسته‌های کوچکی از نورون‌های بینایی در هر قسمت به تنهایی مشغول هستند. ترکیب این نورون‌ها و ایجاد شبکه‌های به‌هم‌پیوسته، باعث می‌شود تا انسان منطقه‌ای را ببیند. شبکه عصبی پیچشی یک بُعدی، نوعی از معماری شبکه‌های عصبی عمیق است که برای پردازش داده‌های دنباله‌ای با یک بُعد، مانند سیگنال‌های زمانی و داده‌های سری زمانی، استفاده می‌شود. همچنین در گاهی برای پردازش متون نیز استفاده می‌شود. این معماری به تشخیص الگوها و ویژگی‌های مهم در داده‌های دنباله‌ای کمک می‌کند.

1. Activation functions

2. Tanh

3. Convolutional neural network

عناصر اصلی یک شبکه عصبی پیچشی یک‌بعدی به شرح زیر هستند:

۱. لایه کانولوشن یک‌بعدی: این لایه از یک یا چند فیلتر یک‌بعدی برای اسکن کردن داده‌های ورودی استفاده می‌کند. فیلترها به طول و عرض مشخصی هستند و روی داده حرکت کرده و عملیات کانولوشن را انجام می‌دهند. این عملیات به تشخیص الگوها و ویژگی‌های مهم در داده کمک می‌کند.
۲. لایه‌های ادغام^۱: مانند لایه‌های ادغام در شبکه‌های عصبی پیچشی دوبعدی، لایه‌های ادغام در شبکه عصبی پیچشی برای کاهش ابعاد و تعداد ویژگی‌ها استفاده می‌شوند.

فرایند اجرای پژوهش

قبل از بیان فرایند انجام پژوهش و با توجه به مطالب مطرح شده در بالا، به‌طور خلاصه می‌توان گفت که هدف اصلی پژوهش، ارائه یک مدل ترکیبی مبتنی بر شبکه عصبی پیچشی (CNN) و شبکه عصبی حافظه کوتاه‌مدت ماندگار (LSTM) برای پیش‌بینی قیمت سهام در بورس اوراق بهادار تهران با دقت بیشتر نسبت به مدل‌های مستقل و جداگانه است. پرسش اصلی برای نیل به هدف پژوهش این است که «آیا ترکیب مدل‌های CNN و LSTM می‌تواند دقت پیش‌بینی قیمت سهام را در مقایسه با استفاده جداگانه از این مدل‌ها بهبود بخشد؟»

برای پاسخ به این پرسش، سه فرضیه ضمنی مطرح شده است.

فرضیه اول: مدل ترکیبی CNN-LSTM نسبت به مدل‌های مستقل CNN و LSTM برای پیش‌بینی قیمت سهام عملکرد بهتری دارد.

فرضیه دوم: مدل CNN به‌تنهایی، نسبت به مدل LSTM برای پیش‌بینی قیمت سهام عملکرد بهتری نشان می‌دهد

فرضیه سوم: انتخاب ویژگی‌های فنی مناسب و تنظیم پارامترهای مدل بر دقت پیش‌بینی قیمت سهام تأثیر معناداری دارد.

در ادامه و برای نیل به اهداف پژوهش، فرایند اجرای پژوهش تشریح می‌شود. در هر سه مدل، ابتدا داده با فرمت خاصی وارد نرم‌افزار پایتون شده است.

مرحله اول بعد از ورود داده‌ها، تغییر مقیاس داده‌هاست. برای این کار می‌توان به دو روش نرمال‌سازی و استانداردسازی را انجام داد که در این پژوهش از روش استاندارد سازی استفاده شده است. علت استفاده از این روش وجود ویژگی‌هایی با واحدهای غیریکسان بوده است.

- روش نرمال‌سازی^۲: مقیاس‌بندی Min-Max یا Min-Max Scaler یکی از روش‌های رایج در نرمال‌سازی داده‌ها است که به‌منظور تغییر مقیاس داده‌ها به بازه‌ای مشخص میان دو مقدار حداقل و حداکثر انجام می‌شود. این روش باعث می‌شود تمام داده‌ها به بازه مشخصی مثل [۰, ۱] تغییر مقیاس یابند.

- روش استانداردسازی^۱: استانداردسازی یکی دیگر از روش‌های نرمال‌سازی داده‌هاست که به منظور تبدیل داده‌ها به یک توزیع با میانگین صفر و واریانس یک انجام می‌شود. در این روش، از میانگین و واریانس داده‌ها برای تغییر مقیاس و تبدیل واحدهای داده‌ها استفاده می‌شود. این کار به مدل‌های یادگیری ماشینی کمک می‌کند که بهتر و با دقت‌تر با داده‌ها برخورد کنند.

$$y_i = \frac{x_i - \bar{x}}{\sigma} \quad \text{رابطه (۱)}$$

در رابطه ۱، y_i نشان‌دهنده مقدار استاندارد شده، x_i نشان‌دهنده داده ورودی، $\bar{x}$ نشان‌دهنده مقدار میانگین داده‌های ورودی و σ انحراف معیار است.

در مرحله دوم داده‌ها را به دو دسته آموزش^۲ و آزمون^۳ تقسیم کنیم. در این پژوهش در هر سه مدل، ۸۰ درصد داده‌ها را به عنوان داده‌های آموزش و ۲۰ درصد داده‌ها را به عنوان داده‌های آزمون فرض می‌کنیم. به علت محدود بودن تعداد داده‌ها، از داده‌های آزمون در ارزیابی عملکرد مدل نیز استفاده شده است.

مرحله سوم، تبدیل داده‌های سری زمانی به فرمت نظارت شده است. تبدیل داده‌های سری زمانی به فرمت نظارت شده^۴، مرحله مهم برای استفاده از مدل‌های رگرسیونی در پیش‌بینی و تحلیل داده‌های سری زمانی است. در این فرمت، داده‌های سری زمانی به صورت جفت‌های ورودی و خروجی مرتبط تبدیل می‌شوند. به عبارت دیگر، از نمونه‌های گذشته برای پیش‌بینی مقدار در زمان آینده استفاده می‌شود. در این پژوهش برای تبدیل داده‌ها از سری زمانی به داده‌های نظارت شده، از دستور زیر استفاده می‌کنیم. این کار معمولاً با استفاده از پنجره‌های زمانی (time-step) ویژگی‌های قبلی به عنوان ورودی و مقدار زمان بعدی به عنوان خروجی انجام می‌شود.

ابتدا گام زمانی (time_step) را مشخص می‌کنیم. گام زمانی مشخص می‌کند که برای پیش‌بینی مقدار برچسب‌گذاری از چند گام زمانی قبل استفاده کنیم. در اینجا از گام زمانی ۲۲ استفاده شده است که برابر یک ماه کاری است. در ادامه، ماتریس‌های x_{train} و y_{train} و x_{test} و y_{test} را تشکیل می‌دهیم.

مرحله چهارم برازش و ارزیابی مدل است. برازش مدل^۵، به معنای تطابق مدل با داده‌های آموزشی است. این تطابق می‌تواند به سه صورت مختلف اتفاق بیفتد: برازش خوب^۶، بیش برازش^۷ و کم برازش^۸. این سه وضعیت به کیفیت آموزش مدل نسبت به داده‌های آموزشی و داده‌های جدید اشاره دارد.

1. Standardization
2. Train
3. Test
4. Supervised
5. Model Fitting
6. Goodfit
7. Overfitting
8. Underfitting

- برازش خوب^۱: در این وضعیت، مدل با دقت مناسبی به داده‌های آموزشی پاسخ می‌دهد و همچنین توانایی خوبی در تخمین داده‌های جدید دارد. مدل نه به‌طور زیادی به داده‌های آموزشی پاسخ می‌دهد و نه به اندازه‌ای که بتواند اطلاعات کلی مسئله را به‌دست آورد.
 - بیش‌برازش^۲: در این وضعیت، مدل به‌طور زیادی به داده‌های آموزشی پاسخ می‌دهد و در تخمین داده‌های جدید توانایی ضعیفی دارد. مدل به جزئیات کوچک و نویزهای داده‌های آموزشی حساس می‌شود و نمایی از داده‌های آموزشی را تا حد زیادی حفظ می‌کند؛ اما در مواجهه با داده‌های جدید، ممکن است نتایج نادرستی ارائه دهد.
 - کم‌برازش^۳: در این وضعیت، مدل نه به داده‌های آموزشی به‌درستی پاسخ می‌دهد و نه توانایی خوبی در تخمین داده‌های جدید دارد. مدل به سادگی و ناپیوستگی داده‌ها پاسخ می‌دهد و توانایی تعمیم‌پذیری به داده‌های جدید را ندارد.
- برای دستیابی به برازش خوب، نیاز است تا مدل معماری مناسبی داشته باشد، تابع هزینه مناسبی مورد استفاده قرار گیرد و تعداد دوره‌های آموزش بهینه تعیین شود.
- برای حل بیش‌برازش شدن مدل از روش‌های مختلفی می‌توان استفاده کرد که در اینجا ما از روش Regularization استفاده می‌کنیم.
- روش Regularization یک تکنیک کلی در یادگیری ماشین است که با هدف کاهش بیش‌برازش و کم‌برازش مدل‌ها به داده‌های آموزش، از جمله شبکه‌های عصبی، اعمال می‌شود. این تکنیک با اضافه کردن جملات تنظیمی به تابع هزینه مدل، تأثیر وزن‌ها را کنترل کرده و به مدل کمک می‌کند تا اطلاعات مهم و معناداری را از داده‌های آموزش یاد بگیرد.
- دو نوع اصلی از Regularization وجود دارد:
۱. Lasso Regularization: در این نوع، جمله تنظیمی به تابع هزینه اضافه می‌شود که برابر با مقدار مطلق وزن‌ها است. این کار باعث می‌شود که وزن‌هایی که مقدار کمی دارند، به صفر نزدیک شوند و در نتیجه ویژگی‌های مرتبط با آن‌ها از مدل حذف شوند. این کار می‌تواند به مدل کمک کند تا تنها از ویژگی‌های مهم و معنادار ترکیب‌ها انتخاب کند.
 ۲. Ridge Regularization: در این نوع، جمله تنظیمی به تابع هزینه اضافه می‌شود که برابر با مجموع مربعات وزن‌هاست. این کار باعث می‌شود وزن‌ها کمی به مقادیر کوچک‌تر نزدیک شوند و به این ترتیب، تمایل مدل به یادگیری ویژگی‌های مهم و تعمیم‌پذیر تر افزایش می‌یابد.
- استفاده از Regularization به این صورت انجام می‌شود که جمله تنظیمی به تابع هزینه اصلی اضافه می‌شود و

1. Good Fitting

2. Overfitting

3. Underfitting

مقدار یک پارامتر به نام فاکتور تنظیم^۱ که به عنوان یک مقدار هایپرپارامتر تعیین می شود، تعیین می کند که Regularization روی مدل چقدر تأثیر دارد.

روابط ۲ و ۳ به ترتیب نشان دهنده Lasso و Ridge هستند.

$$Loss = \text{Error}(y, \hat{y}) + \lambda \sum_{j=1}^p |\beta_j| \quad (\text{رابطه ۲})$$

$$Loss = \text{Error}(y, \hat{y}) + \lambda \sum_{j=1}^p \beta_j^2 \quad (\text{رابطه ۳})$$

در این پژوهش از ترکیب این دو استفاده می شود که به Elastic Net معروف است.

روش Lasso نسبت به داده های پرت حساسیت زیادی ندارد و همچنین، بعضی از وزن ها را به صفر کاهش می دهد؛ ولی ممکن است که بعضی وزن ها همچنان بزرگ باقی بمانند. اما در مدل Ridge که نسبت به داده های پرت حساسیت بیشتری دارد، وزن ها را کوچک می کند ولی هیچ گاه صفر نمی شود. مقادیر لانداها به صورت تجربی و آزمون و خطا به دست آمده است.

ارزیابی پژوهش

برای ارزیابی و سنجش میزان خطا بین مقدار واقعی و مقدار پیش بینی شده از ۵ معیار خطای میانگین خطای مطلق^۲، ریشه میانگین مربعات خطا^۳، ضریب تعیین^۴، میانگین خطای درصدی مطلق متقارن^۵ و ضریب دقت جهت دار^۶ استفاده شده است.

خطای میانگین مطلق یک معیار اندازه گیری است که در زمینه های مختلف، به ویژه در آمار و یادگیری ماشین، برای اندازه گیری میزان تفاوت بین مقادیر پیش بینی شده و مقادیر واقعی یک متغیر استفاده می شود. خطای میانگین مطلق در مقابل خطاهای بزرگ و پرت حساسیت کمتری دارد؛ این معنا که اگر تعدادی از پیش بینی ها خطا داشته باشند، خطای میانگین مطلق کمتر تحت تأثیر قرار می گیرد و مقدار کمی تغییر می کند. خطای میانگین مطلق به سادگی قابل تفسیر است. رابطه ۴ نحوه محاسبه خطای میانگین مطلق را نشان می دهد.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - ty_i| \quad (\text{رابطه ۴})$$

ریشه میانگین مربعات خطا نیز مانند خطای میانگین مطلق، یک معیار اندازه گیری است که در زمینه های مختلف از جمله آمار و یادگیری ماشین، برای اندازه گیری میزان تفاوت بین مقادیر پیش بینی شده و مقادیر واقعی یک متغیر استفاده

^۱. Regularization Factor

^۲. Mean absolute error (MAE)

^۳. Root mean square error (RMSE)

^۴. R- squared (R2)

^۵. Symmetric Mean Absolute Percentage Error (sMAPE)

^۶. Directional Accuracy (DA)

می‌شود. این معیار به‌طور مختصر ریشه میانگین مربعات خطا نامیده می‌شود. به‌عنوان معیاری که تفاوت‌ها را به‌صورت مربعی محاسبه می‌کند، ریشه میانگین مربعات خطا به خطاهای بزرگ‌تر و کوچک‌تر اهمیت می‌دهد که به‌معنای دقت در پیش‌بینی است. میزان تفاوت‌ها، حتی اگر کوچک باشند، در نهایت تأثیر خود را بر نتایج دارند. در مواجهه با خطاهای بزرگ‌تر، ریشه میانگین مربعات خطا اهمیت بیشتری از خطای میانگین مطلق به آن‌ها اختصاص می‌دهد. رابطه ۵ نحوه محاسبه ریشه میانگین مربعات خطا را نشان می‌دهد.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad \text{رابطه ۵}$$

ضریب تعیین، معیاری است که میزان تفسیر متغیر وابسته توسط متغیرهای مستقل در یک مدل را اندازه‌گیری می‌کند. این معیار معمولاً در آمار و یادگیری ماشین استفاده می‌شود تا نشان دهد که مدل چه مقدار از تغییرات متغیر وابسته را توسط متغیرهای مستقل توجیه می‌کند. ضریب تعیین معادل میانگین مربعات ضریب هم‌بستگی بین متغیر وابسته و متغیرهای مستقل است. در واقع، ضریب تعیین نشان می‌دهد که چه مقدار از تغییرات متغیر وابسته توسط تغییرات متغیرهای مستقل توجیه می‌شود. مقدار ضریب تعیین بین ۰ تا ۱ قرار می‌گیرد، که صفر به‌معنای عدم توجیه تغییرات متغیر وابسته توسط متغیرهای مستقل است و ۱ توجیه کامل تغییرات را بیان می‌کند. مقدار ضریب تعیین به شما ایده می‌دهد که مدل شما چه مقدار از تغییرات واقعی را توجیه می‌کند. اگر مقدار ضریب تعیین بالا باشد، این نشان می‌دهد که مدل‌تان با داده‌ها به‌خوبی مطابقت دارد. رابطه ۶ نحوه محاسبه ضریب تعیین را نشان می‌دهد.

$$R^2 = 1 - \frac{(\sum_{i=1}^n (y_i - \hat{y}_i)^2)/n}{(\sum_{i=1}^n (y_i - \bar{y}_i)^2)/n} \quad \text{رابطه ۶}$$

میانگین خطای درصدی مطلق (MAPE) میزان خطای نسبی پیش‌بینی را به‌صورت درصد اندازه‌گیری می‌کند. این معیار برای مقایسه عملکرد مدل‌ها در مقیاس‌های مختلف مفید است. در زمانی که داده‌های با نوسان بالا یا مقادیر نزدیک به صفر داشته باشیم بهتر است به جای معیار MAPE از معیار میانگین خطای درصدی مطلق متقارن (sMAPE) استفاده شود. رابطه ۷ نحوه محاسبه این معیار را نشان می‌دهد.

$$sMAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2} \quad \text{رابطه ۷}$$

که در آن y_i مقدار واقعی، $\hat{y}_i$ مقدار پیش‌بینی‌شده و n تعداد نمونه‌ها است. وقتی مقدار این معیار کمتر از ۱۰ درصد باشد، دقت عالی و وقتی بین ۱۰ تا ۲۰ درصد باشد، دقت خوب و وقتی بیشتر از ۲۰ درصد باشد، نیاز به بهبود مدل را نشان می‌دهد.

ضریب دقت جهت‌دار (DA) درصد پیش‌بینی‌های صحیح جهت تغییرات قیمت (افزایش/کاهش) را محاسبه می‌کند. این معیار برای سرمایه‌گذاران که به روند کلی قیمت اهمیت می‌دهند، بسیار کاربردی است. رابطه ۸ نحوه محاسبه این معیار را نشان می‌دهد.

$$DA = \frac{100\%}{n} \sum_{i=1}^n \mathbb{I}(\text{sign}(y_i - y_{i-1}) = \text{sign}(\hat{y}_i - y_{i-1})) \quad (\text{رابطه ۸})$$

که در آن $\mathbb{I}$ تابع نشانگر (۱ اگر پیش‌بینی و واقعی هم‌جهت و ۰ در غیر این صورت است) و sign علامت تغییرات (مثبت/منفی) است. مقدار بهینه این معیار بالای ۷۰ درصد است.

خلاصه فرایند اجرای پژوهش برای استفاده در بازتولیدپذیری مدل

به‌منظور بازتولیدپذیری مدل، علاوه بر موارد فوق، خلاصه‌ای از فرایند انجام پژوهش بیان می‌شود. در مدل CNN، تعداد فیلترها ۶۴، اندازه کرنل ۳، تابع فعال ساز ReLU، لایه Pooling MaxPooling با اندازه ۲ و تعداد نرون‌های لایه خروجی برابر ۱ با تابع فعال‌سازی خطی مورد استفاده قرار گرفته است. در مدل LSTM، تعداد واحدهای LSTM برابر ۵۰، تابع فعال‌سازی تانژانت هیپربولیک (tanh)، میزان یادگیری ۰/۰۰۱ و بهینه‌ساز Adam استفاده شده است. در مدل ترکیبی LSTM-CNN نیز لایه CNN ۶۴ فیلتر با کرنل ۳ و تابع فعال‌سازی ReLU و در لایه STLM 50 واحد با تابع فعال‌سازی tanh و همچنین لایه خروجی ۱ نرون با تابع فعال‌سازی خطی با نرخ یادگیری ۰/۰۰۱ و بهینه‌ساز Adam استفاده شده است.

همان‌طور که در بخش قبل ذکر شد برای جلوگیری از بیش‌برازش، از تکنیک‌های Regularization شامل L1 و L2 با مقدار لاند ۰/۰۱ استفاده شده است. همچنین، داده‌ها به‌صورت ۸۰ درصد آموزش و ۲۰ درصد آزمون تقسیم شده‌اند و از معیارهای MAE، RMSE، R^2 ، SMAPE و DA برای ارزیابی عملکرد مدل‌ها استفاده شده است. این تنظیمات بر اساس آزمون و خطا و مطالعات پیشین انتخاب شده‌اند تا تعادل مناسبی بین دقت و تعمیم‌پذیری مدل‌ها ایجاد شود.

یافته‌های پژوهش

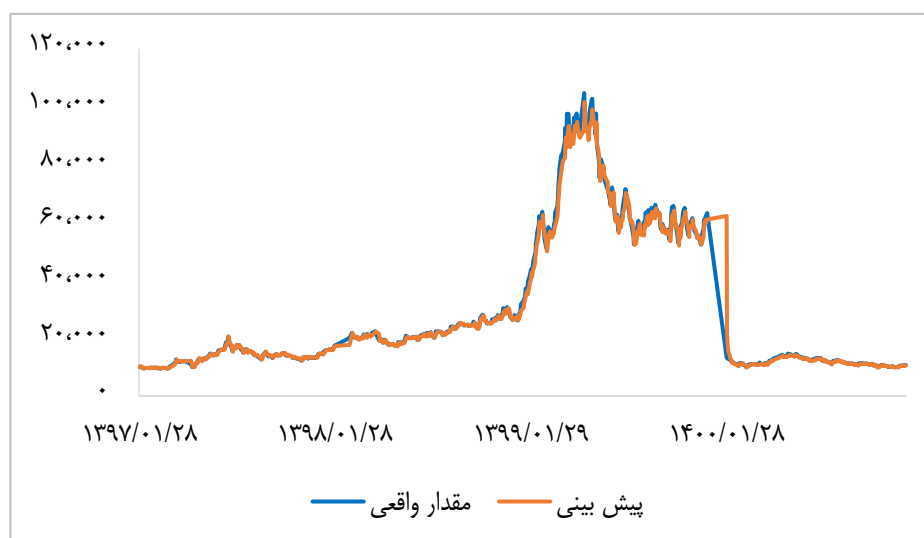
در جدول ۱ نتایج و عملکرد هر سه مدل در هر پنج شرکت قابل مشاهده است. نتایج حاصل از معیارهای ارزیابی نشان می‌دهد که عملکرد مدل‌ها چگونه بوده است.

با توجه به نتایج ۵ معیار ارزیابی استفاده شده، مدل ترکیبی CNN-LSTM از مدل‌های مستقل CNN و LSTM عملکرد بهتری داشته است. همچنین در بین مدل‌های مستقل، مدل CNN عملکرد مطلوب‌تری نسبت به مدل LSTM داشته است.

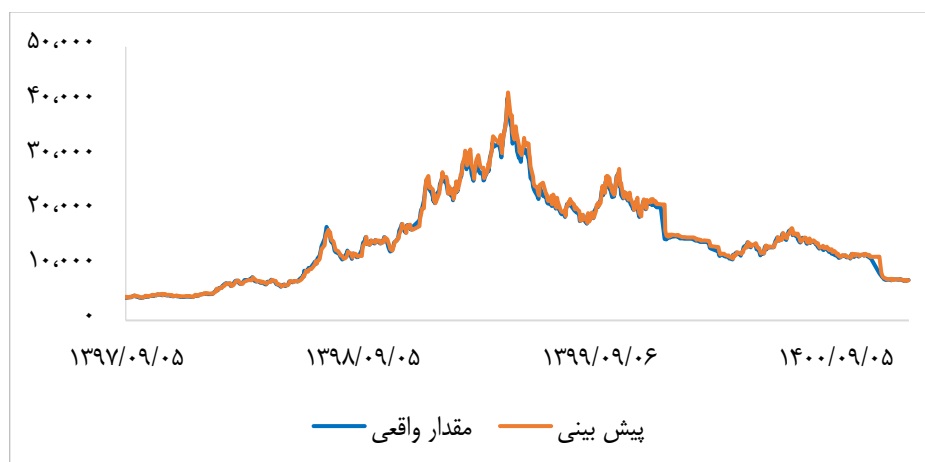
جدول ۱. عملکرد مدل‌های ارزیابی

نماد	مدل	ریشه میانگین مربعات خطا	میانگین خطای مطلق	ضریب تعیین	میانگین خطای درصدی مطلق متقارن	ضریب دقت جهت‌دار (درصد)
شبه‌رن	CNN	۲۵۹۶	۱۰۵۰	۹۸/۷۹	۱۴/۵	۶۵
	LSTM	۶۲۸۳	۲۳۴۴	۹۲/۸۹	۲۲/۳	۵۸/۲
	CNN-LSTM	۲۱۹۹	۸۱۸	۹۹/۱۳	۹/۱	۸۵/۵
دالبر	CNN	۱۲۷۶	۶۹۶	۹۷/۲۴	۱۲/۸	۷۰/۱
	LSTM	۳۳۸۱	۱۶۹۴	۸۰/۶۳	۲۵/۶	۵۵/۷
	CNN-LSTM	۶۲۳	۴۰۹	۹۹/۳۴	۸/۵	۸۳/۲
خودرو	CNN	۴۵۸	۱۵۲	۹۶/۰۵	۱۰/۲	۷۲/۴
	LSTM	۵۹۲	۳۰۴	۹۴/۱۰	۱۸/۹	۶۰/۳
	CNN-LSTM	۴۱۷	۱۲۵	۹۷/۰۸	۷/۸	۸۵
کروی	CNN	۲۶۰۴	۱۴۴۷	۹۶/۶۵	۱۶/۷	۶۳/۸
	LSTM	۶۴۳۲	۲۹۰۳	۷۹/۵۳	۲۸/۴	۵۲/۱
	CNN-LSTM	۱۷۳۳	۱۰۶۲	۹۸/۵۱	۱۱/۲	۸۰/۶
ستران	CNN	۱۲۲۱	۷۳۱	۹۹/۴۵	۱۳/۱	۶۸/۹
	LSTM	۳۰۶۸	۱۱۵۸	۹۶/۵۲	۲۰/۵	۵۷/۴
	CNN-LSTM	۱۱۹۳	۶۶۷	۹۹/۴۷	۸/۹	۸۱/۷

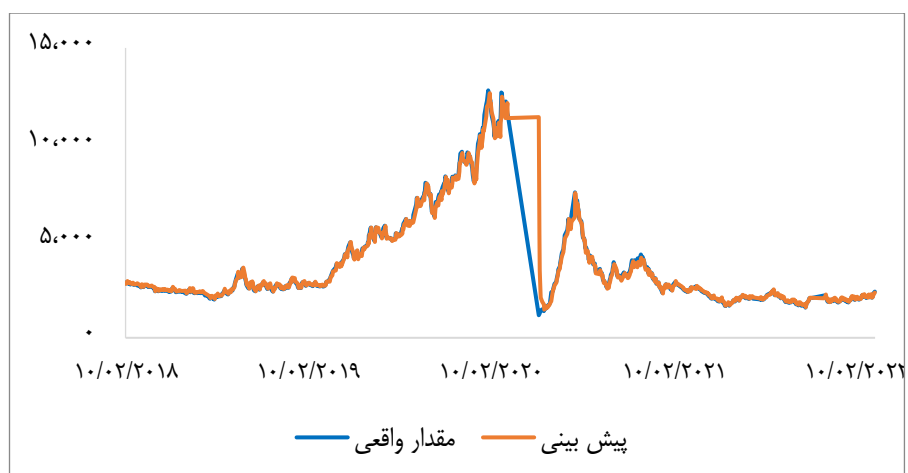
نتایج عملکرد مدل‌ها در شکل‌های ۱ تا ۵ قابل مشاهده است.



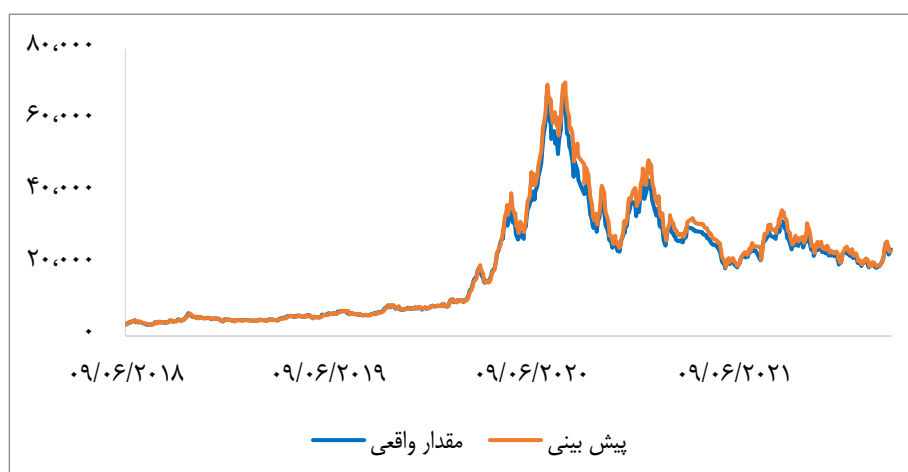
شکل ۲. نمودار شرکت نفت بهران (مدل CNN-LSTM)



شکل ۳. نمودار شرکت البرز دارو (مدل CNN-LSTM)



شکل ۴. نمودار شرکت ایران خودرو (مدل CNN-LSTM)



شکل ۵. نمودار شرکت توسعه معادن روی ایران (مدل CNN-LSTM)



شکل ۶. نمودار شرکت سپان تهران (مدل CNN-LSTM)

مدل ترکیبی معرفی شده در این پژوهش نسبت به مدل‌های مستقل، در پیش‌بینی چرخه‌های مختلف رکود و رونق عملکردی به مراتب دقیق‌تری از خود نشان می‌دهد. مقدار ضریب تعیین به دست آمده در مدل نشان می‌دهد که با استفاده از متغیرهای فنی می‌توان به نتایج مطلوبی در پیش‌بینی قیمت سهام رسید. همان‌طور که بیان شد مدل‌های یادگیری عمیق برای افزایش دقت در پیش‌بینی، به تعداد داده‌های زیادی نیاز دارند؛ اما افزایش تعداد متغیرهای مستقل می‌تواند باعث ایجاد مشکلاتی چون بیش‌برازش مدل شود. در ادامه باید به این موضوع اشاره شود که وجود دامنه نوسان، قدرت پیش‌بینی‌کنندگی این مدل‌ها را کاهش می‌دهد.

نتیجه‌گیری و پیشنهادها

با توجه به پیشرفت روز افزون مدل‌های یادگیری عمیق در پیش‌بینی داده‌های مختلف از جمله داده‌های مالی، تلاش کرده‌ایم در این پژوهش به بررسی برخی از مدل‌های یادگیری عمیق برای تخمین و پیش‌بینی در بورس اوراق بهادار تهران بپردازیم.

با توجه به محدودیت‌ها و ویژگی‌های خاص بورس اوراق بهادار تهران، مانند دامنه نوسان، موجب ایجاد علاقه در محققان جهت بررسی و پیش‌بینی این بازار با استفاده از مدل‌های یادگیری ماشین و یادگیری عمیق می‌شود. نتایجی که یافته‌های این پژوهش نشان می‌دهد، این است که مدل‌های یادگیری عمیق در صورتی قابلیت و توانایی پیش‌بینی قیمت با دقت قابل قبول را دارند که ویژگی‌هایی (متغیرهای مستقل) انتخاب شوند که بتوانند بیشترین میزان معناداری را در تفسیر علل فراز و فرودهای قیمت دوره‌های رونق و رکود بازار بیان کنند.

یکی از نتایجی که در این پژوهش می‌توان به آن اشاره کرد، این است که نتایج به دست آمده در مدل ترکیبی (مدل CNN-LSTM) از دو مدل مستقل دیگر عملکرد بهتری داشته؛ اما این عملکرد، نسبت به مدل مستقل CNN، آنچنان عملکرد بهتری نبوده است.

همان طور که نتایج پژوهش نشان داد مدل ترکیبی CNN-LSTM عملکرد بهتری نسبت به مدل‌های مستقل CNN و LSTM داشته است با این حال، برخلاف یافته‌های این پژوهش، برخی پژوهش‌ها (مانند اصولیان و همکاران، ۱۴۰۴) نشان دادند که مدل CNN به تنهایی می‌تواند عملکردی قابل مقایسه با مدل‌های ترکیبی داشته باشد. همچنین، نتایج پژوهش مبنی بر عملکرد ضعیف‌تر LSTM در مقایسه با CNN با یافته‌های لی و همکاران (۲۰۱۷) که ترکیب LSTM و CNN را مؤثرتر دانسته‌اند، تفاوت دارد. این تفاوت‌ها ممکن است ناشی از تفاوت در داده‌ها، ویژگی‌های انتخابی، یا پارامترهای مدل‌سازی باشد.

برتری CNN نسبت به LSTM در این پژوهش را می‌توان به ماهیت پرنویز و غیرخطی داده‌های بورس تهران، محدودیت دامنه نوسان و حجم نسبتاً کم داده‌های آموزشی نسبت داد. CNN با تمرکز بر الگوهای محلی و کاهش حساسیت به نویز، برای این شرایط مناسب‌تر است، در حالی که LSTM به دلیل پیچیدگی ذاتی در مدل‌سازی وابستگی‌های بلندمدت، در محیط پرنوسان بورس تهران عملکرد ضعیف‌تری نشان می‌دهد. البته ترکیب این دو (LSTM-CNN) با بهره‌گیری از مزایای هر دو معماری، بهترین نتایج را ارائه می‌کند.

برای پژوهش‌های آینده می‌توان دو موضوع را پیگیری و بررسی کرد:

۱. قابلیت و دقت پیش‌بینی مدل در دوره‌های زمانی بالاتر چگونه است
۲. متغیرهای مستقل دیگری، از جمله متغیرهای کلان اقتصادی و مالی شرکتی را نیز در تخمین و پیش‌بینی استفاده و نتایج را مقایسه کرد.

محدودیت‌های داده‌ای در این پژوهش، از جمله حذف نقاط پرت و ایستاسازی، عمدتاً با هدف کاهش نویز و افزایش تمرکز مدل بر الگوهای اساسی بازار اعمال شدند. این روش‌ها ممکن است بخشی از اطلاعات را حذف کنند؛ اما با توجه به ماهیت پرنوسان بورس تهران، برای جلوگیری از بیش‌برازش^۱ ضروری بودند. در تحقیقات آتی، می‌توان با استفاده از مدل‌های مقاوم‌تر (مثل Transformerهای مبتنی بر Attention) یا افزودن متغیرهای کلان اقتصادی، این محدودیت‌ها را جبران کرد.

منابع

- اصولیان، محمد؛ نیک‌مرام، علی و کریمی، مهدی (۱۴۰۴). پیش‌بینی روند شاخص کل با استفاده از شبکه‌های عصبی هیبریدی با تمرکز بر استخراج ویژگی مقیاس زمانی چندگانه در بورس اوراق بهادار تهران. *تحقیقات مالی*، ۲۷ (۱)، ۸۵-۱۱۳.
- حیدری، مهدی و امیری، حمیدرضا (۱۴۰۱). بررسی قدرت مدل‌های مبتنی بر هوش مصنوعی در پیش‌بینی روند قیمت سهام بورس اوراق بهادار تهران. *تحقیقات مالی*، ۲۴ (۴)، ۶۰۲-۶۲۳.
- فقیهی نژاد، محمد تقی؛ مینایی، بهروز (۱۳۹۷). پیش‌بینی رفتار بازار سهام بر اساس شبکه‌های عصبی مصنوعی با رویکرد یادگیری جمعی هوشمند. *فصلنامه مدیریت صنعتی*، ۱۰ (۱)، ۳۱۵-۳۳۴.

References

- Borovykh, A., Bohte, S. & Oosterlee, C. W. (2017). *Conditional time series forecasting with convolutional neural networks*. arXiv preprint arXiv:1703.04691.
- Cao, J., Li, Z., & Li, J. (2019). Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A: Statistical mechanics and its applications*, 519, 127-139.
- Chong, E., Han, Ch. & Park F. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, (83), 187-205.
- Faghihi Nezhad, M. T. & Minaei, B. (2018). Prediction of Stock Market Behavior Based on Artificial Neural Networks through Intelligent Ensemble Learning Approach. *Industrial Management Journal*, 10(2), 315-334. doi: 10.22059/imj.2018.255079.1007412 (in Persian)
- Heidari, M. & Amiri, H. (2022). Inspecting the Predictive Power of Artificial Intelligence Models in Predicting the Stock Price Trend in Tehran Stock Exchange. *Financial Research Journal*, 24(4), 602-623. <https://doi.org/10.22059/FRJ.2022.320064.1007149> (in Persian)
- Hiransha, M. E. A. G., Gopalakrishnan, E. A. & Soman, K. P. (2018). NSE stock market prediction using deep-learning models. *Procedia computer science*, 132, 1351-1362.
- Hochreiter, S. & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- Jing, N., Wu, Z. & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications*, 178, 115019.
- Khare, K., Darekar, O., Gupta, P. & Attar, V. Z. (2017, May). Short term stock price prediction using deep learning. In *2017 2nd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT)* (pp. 482-486). IEEE.
- Lee, C. Y. & Soo, V. W. (2017, December). Predict stock price with financial news based on recurrent convolutional neural networks. In *2017 conference on technologies and applications of artificial intelligence (TAAI)* (pp. 160-165). IEEE.
- Lee, M. C. (2009). Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert Systems with Applications*, 36 (8), 10896-10904.
- Mehtab, S., Sen, J. & Dutta, A. (2020, October). Stock price prediction using machine learning and LSTM-based deep learning models. In *Symposium on machine learning and metaheuristics algorithms, and applications* (pp. 88-106). Singapore: Springer Singapore.
- Noorbakhsh, A. & Shaygani, M. (2024). Deep Learning Approach for Stock Price Prediction: Comparing Single and Hybrid Models. *Advances in Industrial Engineering*, 58(1), 237-249.
- Osoolian, M., Nikmaram, A. & Karimi, M. (2025). Predicting Index Trend Using Hybrid Neural Networks with a Focus on Multi-Scale Temporal Feature Extraction in the Tehran Stock

- Exchange. *Financial Research Journal*, 27(1), 85-113.
<https://doi.org/10.22059/FRJ.2024.377816.1007611> (in Persian)
- Oussar, Y. & Dreyfus, G. (2000). Initialization by selection for wavelet network training. *Neurocomputing*, 34 (1), 131-143.
- Sousa, J., Montevechi, J. & Miranda, R. (2019). Economic lot-size using machine learning, parallelism, metaheuristic, and simulation. *Journal of Logistics, Informatics, and Service Science*, 18 (2), 205-216.
- Vanaga, R. & Sloka, B. (2020). Financial and capital market commission financing: Aspects and challenges. *Journal of Logistics*, 7 (1), 17-30.
- Yamak, P. T., Yujian, L. & Gadosey, P. K. (2019, December). A comparison between arima, lstm, and gru for time series forecasting. In *Proceedings of the 2019 2nd international conference on algorithms, computing and artificial intelligence* (pp. 49-55).
- Zhang, L. & Kim, H. (2020). The influence of financial service characteristics on use intention through customer satisfaction with mobile fintech. *Journal of System and Management Sciences*, 10 (2), 82-94.
- Zhang, X., Zhang, Y., Wang, S., Yao, Y., Fang, B. & Yu, P. S. (2019). Improving stock market prediction via heterogeneous information fusion. *Knowledge-Based Systems*, 166, 1-11.
- Zhong, X. & Enke, D. (2019). Predicting the daily return direction of the stock market using hybrid machine learning algorithms. *Financial Innovation*, 5(1), 1-20.