Journal of Capital Market Analysis

Volume 3 Issue 2 Online ISSN: 2783-3488

Page: 59-83 https://jcma.aliabad.iau.ir

# Stock Trading Strategy Based on Regression Learning Algorithms

**Nasser Heydari**
Department of Finance, Arak Branch, Islamic Azad University, Arak, Iran.
**Majid Zanjirdar (**Corresponding author**)**
Department of Finance, Arak Branch, Islamic Azad University, Arak, Iran.
m-zanjirdar@iau-arak.ac.ir
**Ali Lal Bar**
Department of Finance, Arak Branch, Islamic Azad University, Arak, Iran.

**Article Info**

**Article type:**
Research Article

**Article history:**
Received: 06 May 2023
Received in revised form: 21 Aug 2023
Accepted: 14 Sep 2023

**Keywords**:
Trading Strategy, Machine Learning, Regression Algorithms, Stock Exchange.

**Abstract**

The aim of this study is to develop a stock trading strategy using regression learning algorithms. The researcher utilized the Yahoo Finance database to collect the necessary data using Python programming. Key technical analysis indicators and oscillators were calculated and incorporated into the model. The performance of the regression algorithms was evaluated using indicators such as determination coefficient, mean error of the mean, and square root of the error. Advanced statistical methods and software including Python, Spider, SPSS, and Excel were employed to analyze the differences between the evaluation indices of the designed algorithms. The Kruskal-Wallis test was used for meaningful comparison. Additionally, a diversified research sample consisting of companies from various sectors was chosen to generalize the findings. The selected companies were actively traded on the New York Stock Exchange with an average volume greater than 1 million and a market value larger than 200 trillion dollars. The sample was determined using a filter writing method on ۲۰۲۱/۰۶/۲۸ equal to 41 numbers as the sample of this research. The research was completed by the end of February 2023, and the random forest trading strategy model was identified as the most suitable approach.

© The Author(s). Publisher: Islamic Azad Univercity of Aliabad Katoul Branch.

فصلنامه تحلیل بازار سرمایه

شاپا الکترونیک: ۲۷۸۳-۳۴۸۸ سال ۳ شماره۲

https://jcma.aliabad.iau.ir صفحه: ۸۳-۵۹

# راهبرد معاملاتی خرید و فروش سهام بر اساس الگوریتم های یادگیری رگرسیونی

**ناصر حیدری**

گروه مالی، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران.

**مجید زنجیردار** (مسئول مکاتبات)

گروه مالی، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران.

m-zanjirdar@iau-arak.ac.ir

**علی لعل بار**

گروه مالی، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران.

**اطلاعات مقاله**

**نوع مقاله:**

پژوهشی

**تاریخچه مقاله:**

تاریخ دریافت: ۱۶ اردیبهشت ۱۴۰۲

تاریخ ارسال بازنگری: ۳۰ مرداد ۱۴۰۲

تاریخ پذیرش: ۲۳ شهریورماه ۱۴۰۲

**واژگان کلیدی:**

راهبرد معاملاتی، یادگیری ماشین، الگوریتم های رگرسیونی، بورس اوراق بهادار.

**چکیده**

هدف این پژوهش طراحی راهبرد معاملات سهام بر اساس الگوریتم های یادگیری رگرسیونی می باشد. پژوهشگر بوسیله بهره‌برداری از پایگاه اطلاع رسانی یاهو فاینانس، داده های مورد نیاز پژوهش را با استفاده از برنامه نویسی در محیط نرم افزار پایتون استخراج و بر اساس آن شاخص ها و نوسان سازهای تحلیل تکنیکال را محاسبه و به عنوان مولفه های اولیه وارد مدل نموده است. در راستای ارزشیابی الگوریتم های رگرسیونی از شاخص های (ضریب تعیین، خطای میانگین مربعات و ریشه میانگین مربعات خطا) استفاده شده است. پژوهشگر از روش های پیشرفته آماری و نرم افزارهای پایتون، اسپایدر، اس پی اس اس و اکسل استفاده نموده و تفاوت بین شاخص های ارزشیابی الگوریتم های طراحی شده را به لحاظ معنی داری از طریق آزمون کروسکال والیس بررسی و راهبرد خرید و فروش سهام را معرفی می‌نماید. در راستای تنوع بخشی نمونه تحقیق جهت تعمیم پذیری نتایج، شرکت های فعال در حوزه حمایت مصرف کننده، خدمات ارتباطی، مراقبت های بهداشتی، خدمات مالی، تکنولوژی، چرخه مصرف کننده و انرژی در بورس نیویورک با میانگین حجم معامله بزرگتر از ۱ میلیون و ارزش بازار بزرگ تر از ۲۰۰ تریلیون دلار از طریق روش فیلترنویسی در تاریخ ۱۴۰۰/۰۴/۰۷ برابر با ۴۱ عدد به عنوان نمونه این پژوهش انتخاب و انجام تحقیق تا پایان بهمن ماه ۱۴۰۱ به اتمام رسید و مدل راهبرد معاملاتی جنگل تصادفی به عنوان مدل مناسب معرفی گردید.

**ناشر:** دانشگاه آزاد اسلامی واحد علی آباد کتول. ©نویسندگان.

CC BY NC ND

**۱. مقدمه**

بازار بورس نیویورک[1] با ارزش بازار حدود ۲۶ تریلون دلار در ماه می ۲۰۲۱ اولین و بزرگترین بازار بورس کشور آمریکا می باشد که در سال ۱۷۹۲ در خیابان وال شهر نیویورک آغاز به کار نموده است. تعداد شرکت های پذیرش شده در بورس نیویورک بیش از ۳۵۰۰ عدد از مشهورترین و فعالترین شرکت ها می باشند. لازم به ذکر است اقتصاد آمریکا با تولید ناخالص حدود ۲۲,۹۳۹ تریلیون دلار، حدود ۲۴ درصد از اقتصاد کل دنیا را به خود اختصاص داده است. در بازار سهام ایالات متحده و بسیاری از بازارهای مالی توسعه یافته دیگر، حدود ۸۰ درصد از حجم معاملات از طریق معاملات الگوریتمی انجام می شود و از طرفی بازار سرمایه نقش مهمی در تخصیص کارایی وجوه سرمایه ایفا می کند و برای رونق بخشیدن به اقتصاد کشورها که از یک سو با حجم عظیم نقدینگی و از سوی دیگر با کمبود امکانات سرمایه گذاری مولد مواجه است، به عنوان نهادی برای تجهیز منابع مالی شرکت ها از اهمیت ویژه‌ای برخوردار است(صمدی و همکاران،۲۰۲۳). استفاده از الگوریتم های رایانه‌ای و هوش مصنوعی از اوایل قرن گذشته در بازارهای سهام ایالات متحده در حال افزایش است. در سال ۲۰۰۳ ، معاملات الگوریتمی فقط ۱۵ درصد از حجم بازار را تشکیل می داد، اما بین سالهای ۲۰۰۹ و ۲۰۱۰ ، بیش از ۷۰ درصد از معاملات ایالات متحده از طریق الگوریتم های رایانه ای انجام شده است. جذابیت و اهمیت بازار سرمایه برای فعالان حقیقی و حقوقی مانند شرکت های سرمایه گذاری، سبدگردانی، مشاوره سرمایه گذاری، صندوق های سرمایه گذاری در سهام، صندوق های سرمایه گذاری مختلط، صندوق های سرمایه گذاری اهرمی، صندوق های سرمایه گذاری با درآمد ثابت، شرکت های پردازش اطلاعات مالی و سایرنهادهای مالی، پژوهشگران را به بررسی و طراحی راهبرد ها و الگوریتم های مناسب خرید و فروش سهام سوق داده است. زهنگ و همکاران[2](۲۰۲۱) در پژوهش خود نمایش می دهد شبکه های عصبی بیزی می تواند پیش بینی معقول تری از قیمت های آتی ارائه نماید، سوعود الاتابی[3](۲۰۲۱) نشان می دهد مدل های ترکیبی از دقت بالاتری در

1 . New York Stock Exchange
2 . Zhang et al
3 . Alotaibi

پیش بینی قیمت برخوردار می باشند، میرزایی و همکاران(۲۰۲۱) بر اساس یافته های پژوهش خود اظهار می کند پیش بینی شاخص بورس اوراق بهادار بر اساس الگوریتم شبکه های عصبی و بهینه سازی آن دارای دقت مناسبی می باشد، ایساک کوفی[۴](۲۰۲۰) نشان داده است مدل بردار پشتیبان دارای دقت مناسب تری بوده است، نتایج پژوهش کرولا رویا و گویجارو[۵](۲۰۲۰)حاکی از آن است که یادگیری بر اساس الگوریتم جنگل تصادفی دارای دقت بالاتری می باشد، ایساک کوفی آنتی[۶](۲۰۲۰)نشان می دهد شبکه های عصبی در مقایسه با ماشین بردار پشتیبان و درخت تصمیم دارای خطای کمتری می باشند، زهنگ و همکاران(۲۰۱۸) در یافته های پژوهش خود نمایش می دهند دقت الگوریتم ماشین بردار پشتیبان نسب به سایر مدل ها بزرگ تر می باشد.

تضاد نتایج مطالعات ارائه شده در خصوص عملکرد الگوریتم های یادگیری پژوهشگر را بر آن داشت، راهبرد معاملاتی خود مبتنی بر مدل های متنوع رگرسیونی یادگیری طراحی و عملکرد آن را بر اساس شاخص های آماری بررسی نماید. با عنایت به خلأها و تضادهای پژوهش های موجود و اهمیت راهبرد معاملاتی برای سرمایه گزاران حقیقی و حقوقی، پژوهشگر قصد دارد شاخص ها[۷] و نوسان سازهای[۸] تحلیل تکنیکال را به عنوان مولفه های ورودی مدل مدل نظر قرار داده و بر اساس الگوریتم های یادگیری رگرسیونی راهبرد معاملاتی سهام شرکت های پذیرش شده در بورس نیویورک را طراحی نماید. افزایش سودآوری و کاهش ریسک، افزایش حجم داده‌ها و خطای محاسباتی در تحلیل آن، اهمیت دقت و سرعت در زمان خرید و فروش، بررسی راهبرد های براساس الگوریتم های رگرسیونی و ارزش بازار بورس نیویورک (حدود ۲۵ تریلون دلار در ماه ژوئن ۲۰۲۱) حاکی از اهمیت و ضرورت هدف پژوهش می باشد. سرمایه گذاران حقیقی و حقوقی مانند شرکت های تامین سرمایه، شرکت های مادر(هلدینگ)، سرمایه گذاری، مشاور سرمایه گذاری، خدمات بازار سرمایه، سبدگردان، کارگزاری، صندوق های سرمایه گذاری در سهام، مختلط و با درآمد ثابت، صندوق های بازارگردان و شرکت های پردازش اطلاعات مالی قادر

4 . Nti et al
5 . Cervelló-Royo et al
6 . Nti, I. K et al
7 .indicator
8 .oscillator

خواهند بود از نتایج این تحقیق بهره مند گردند. بهره‌برداری از پایگاه اطلاع رسانی یاهوفاینانس و کتابخانه های متنوع نرم افزار پایتون به منظور جمع آوری داده‌ها، کاربرد شاخص ها و نوسان سازهای تجزیه و تحلیل تکنیکال مختلف به عنوان داده‌های ورودی راهبرد، استفاده از الگوریتم های رگرسیونی پیچیده یادگیری مانند ماشین بردار پشتیبان، جنگل تصادفی، نزدیک ترین همسایه، شبکه های عصبی چند لایه و رگرسیون لوجستیک و ارزشیابی الگوریتم های با شاخص های آماری مانند ضریب تعیین، خطای میانگین مربعات و ریشه میانگین مربعات خطا، بهره برداری از فن آوری اطلاعات و علوم داده در مالی را می توان از نوآوری های پژوهش نام برد.

**۲. مبانی نظری و پیشینه پژوهش**

چارلز داو[9] معتقد بود، قیمت منعکس کننده تمام اطلاعات گذشته می باشد. برخی مفاهیم در تحلیل تکنیکال از نوشته‌های ایشان در مورد تئوری‌های بازار برگرفته شده است. او باور داشت که بازار همه چیز را کمرنگ می‌کند، به عبارتی تمام اطلاعاتی که ما اکنون داریم قبلا بر روی قیمت‌ها تاثیر خود را گذاشته‌اند. برای مثال اگر انتظار برود که یک کمپانی گزارش درآمد بسیار خوب و مثبتش را بدهد، این موضوع قبل از اینکه اتفاق بیفتد تاثیرش را روی بازار گذاشته است. در این شرایط درخواست و تقاضا برای سهام آن‌ها قبل از انتشار گزارش، بالا می‌رود و ممکن است که بعد از بیرون آمدن آن گزارش درآمد مثبت، قیمت تغییرات چندانی را تجربه نکند. داو در برخی مواقع متوجه شد که قیمت سهام یک کمپانی در بورس بعد از آمدن یک خبر خوب کاهش پیدا می‌کند؛ زیرا این خبر آن طور که انتظار می‌رفت، خوب نبوده است! این قوانین از نظر بسیاری از معامله‌گران و سرمایه‌گذاران، به خصوص آن‌هایی که به طور گسترده از ابزارهای تحلیل تکنیکال استفاده می‌کنند، هنوز هم درست است. در سال ۱۹۹۱، نظریه داو توسط مایکل اوهیگینز[10] و جان داونز[11] که معمولاً به عنوان نظریه سگ های داو شناخته می شود به عنوان یکی از شناخته شده ترین استراتژی های سرمایه گذاری در ایالات متحده با هدف عملکرد بهتر از بازار معرفی شد(کیو و همکاران،۲۰۱۳).

9 . Charles Dow
10 . Michael O'Higgins
11 . John Downes

یادگیری ماشین شاخه نوین هوش مصنوعی است که به مطالعه الگوریتم های به منظور حل مسائل می پردازد و یک فرایند یادگیری است، که با شناسایی حوزه یادگیری شروع می شود و با تست و به کارگیری نتایج به دست آمده مسائل را حل می نماید(پروج و همکاران، ۲۰۱۲). بسیاری از الگوریتم های یادگیری ماشین برای پیش بینی بازار سهام تهیه و استفاده شده اند. نظریه یادگیری ماشینی، همچنین به عنوان نظریه یادگیری محاسباتی شناخته می شود، هدف آن است که اصول اساسی یادگیری را به عنوان یک فرایند محاسباتی درک کند. اهداف این نظریه نه تنها کمک به طراحی روش‌های یادگیری خودکار بهتر، بلکه درک مسائل اساسی در خود فرایند یادگیری است. یادگیری ماشینی ابزار مهمی برای استفاده از فناوری‌های پیرامون هوش مصنوعی است. به دلیل توانایی‌های یادگیری و تصمیم‌گیری، یادگیری ماشین اغلب به عنوان هوش مصنوعی شناخته می‌شود، اگرچه، زیرشاخه‌ای از هوش مصنوعی است. یادگیری ماشینی به یک ابزار بسیار مهم برای رایانش ابری و تجارت الکترونیک تبدیل شده است و در انواع فناوری‌های پیشرفته مورد استفاده قرار می‌گیرد. یادگیری ماشینی به معنای امروزی با نام روانشناس فرانک روزنبلات[۱۲] از دانشگاه کرنل بر اساس سیستم عصبی انسان، ایجاد شده است. گروهی که بر اساس این ایده ماشینی را در سال های ۱۹۵۷ تا ۱۹۶۰ برای تشخیص حروف الفبا ساختند و آن ماشین را، پرسپترون نامیدند. این ماشین اولین نمونه مدرن شبکه عصبی مصنوعی است که به مدل حیوانی و انسانی نزدیک می باشد. در سال ۱۹۶۲ نویکف[۱۳] یک شرایط همگرایی را به مدل پرپسترون اضافه کرد که الگوریتم یادگیری در تعدادی از مراحل محدود و به شناخت بهتر آن کمک نمود. تاریخچه این تئوری نشان میدهد که در طول چند دهه اول آن رویکردها به هم نزدیک بودن و همکاری ثمربخشی بین آنها مشاهده می شود. با این حال، توسعه چشمگیر یادگیری ماشین طی دو دهه اخیر منجر به واگرایی در این تئوری شده است که از آن می توان به یادگیری بریستو و همکاران[۱۴] (۲۰۰۶) و یادگیری تقویتی رچ[۱۵](۲۰۱۸) اشاره کرد(فرادکو،۲۰۲۰).

12 . Frank Rosenblatt
13 . Novikoff
14 . Bristow et al
15 . Recht

روش‌های رگرسیونی خطی، سعی دارند که با ساختن یک ابرسطح (که عبارت است از یک معادله خطی)، داده‌ها را از هم تفکیک کنند. روش رگرسیونی ماشین بردار پشتیبان[16] که یکی از روش‌های رگرسیونی خطی است، بهترین ابرسطحی را پیدا می‌کند که با حداکثر فاصله، داده‌های مربوط به دو طبقه را از هم تفکیک کند. الگوریتم ماشین بردار پشتیبان در سال ۱۹۶۳ اولین با توسط ولادمیر، وپنیک و الکسی مطرح شد. در سال ۱۹۹۲ برنهارد بوسر، ایزابل گویان، ولادمیر و پنیک راهکاری جهت ایجاد طبقه بندی غیر خطی با استفاده از تابع کرنل[17] پیشنهاد کردند(بوسر و همکاران،۱۹۹۲).

الگوریتم جنگل تصادفی[18]یک الگوریتم یادگیری ماشین با قابلیت استفاده آسان است که اغلب اوقات نتایج بسیار خوبی را حتی بدون تنظیم فراپارامترهای آن، فراهم می‌کند. این الگوریتم، جنگلی را به طور تصادفی می‌سازد. «جنگل» ساخته شده، در واقع گروهی از درخت‌های تصمیم[19] است. کار ساخت جنگل با استفاده از درخت‌ها اغلب اوقات به روش کیسه‌گذار[20] انجام می‌شود، به بیان ساده، جنگل تصادفی چندین درخت تصمیم ساخته و آن‌ها را با یکدیگر ادغام می‌کند تا پیش‌بینی‌های صحیح‌تر و پایدارتری حاصل شوند. این روش که توسط بریمن[21] (۲۰۰۱) ابداع شد. جنگل تصادفی یک الگوریتم یادگیری ماشین بر اساس درختان تصمیم و کیسه بندی است. درختان تصمیم، الگوریتم هایی هستند که بر اساس یک سری از نقش ها پیش بینی می کنند. قوانین پیش بینی و طبقه بندی توسط گره ها و شاخه ها ارائه شده و پیش بینی و طبقه بندی نهایی در چهارچوب برگ درختان نمایش داده می شود(بریمن،۲۰۰۱).

روش نزدیک‌ترین همسایه[22] یک روش یادگیری موردی است و از جمله ساده‌ترین الگوریتم‌های یادگیری ماشین می‌باشد. الگوریتم نزدیکترین همسایه برای اولین بار توسط کاور و هارت در سال ۱۹۶۷ بیان شد. در این الگوریتم مقدار آماره K و نزدیک ترین همسایه دو موضوع

16 . SVM (Support Vector Machine).
17 . Kernel Function
18 . Random Forest
19 . Decision Trees
20 . Bagging
21 . Breiman
22 . K-Nearest Neighbors

حیاتی می باشند. از مهمترین توابعی که برای اندازه گیری فاصله نزدیکترین همسایه استفاده می شود عبارتند از اقلیدسی[۲۳]، ماهالانوبیس[۲۴]، فاصله و کوسینوس زاویه منهتن[۲۵] (زهنگ و همکاران،۲۰۲۱).

الگوریتم شبکه های عصبی چندلایه پیشخور[۲۶] یک شبکه عصبی پیش‌خور با دو لایه یا بیشتر و تابع فعال‌سازی خطی[۲۷]، مشابه رگرسیون خطی چند متغیره است. نرون‌های ورودی همان متغیرهای مستقل یا رگرسورها هستند و نرون خروجی همان برآورد متغیر وابسته است. وزن‌های مختلف شبکه نیز، مشابه پارامترهای رگرسیون و جمله اریب نیز همان عرض از مبدا یا جمله ثابت در رگرسیون است. در صورتی که وقفه‌های وابسته را به مجموعه ورودی‌ها اضافه کنیم، در آن صورت به شبکه‌ای مشابه با راهبرد اتورگرسیون[۲۸] خطی دست می‌یابیم. گره‌ها در این نوع شبکه، در لایه‌های متوالی قرار گرفته‌اند و ارتباط یک طرفه است و زمانی که یک الگوی ورودی به شبکه اعمال می‌شود، اولین لایه، مقدار خروجی خود را محاسبه کرده و در اختیار لایه بعدی قرار می‌دهد. لایه بعدی این مقادیر را به عنوان ورودی دریافت کرده و مقادیر خروجی را به لایه بعدی منتقل می‌کند و هر گروه فقط به گره‌های لایه بعد، سیگنال منتقل می‌کنند. شبکه‌های پرسپترون چند لایه جزء این نوع شبکه‌ها می‌باشند.

شبکه های عصبی چند لایه پیشخور یک الگوریتم شبکه عصبی مصنوعی شامل چند لایه ورودی، لایه های پنهان، که با استفاده از تابع غیر خطی سیگموئید[۲۹] و با استفاده از آموزش انجام شده بوسیله تطبیق داده های ورودی و خروجی و نودهای موجود، بر اساس میزان عملکرد به خود وزن تخصیص داده و در نهایت داده های خروجی را پیش بینی می کند(برک آکایا و همکاران[۳۰]، ۲۰۱۹) .

23 . euclidean distance
24 . Mahalanobis distance
25 . Manhattan distance and angle cosine distance
26 . Multi Layer Perceptron
27 . linear activation function
28 . autoregression
29 . sigmoidal nonlinear function
30 . Berke Akkaya et al

رگرسیون لجستیک نوعی از رگرسیون است که متغیرهای پیش بین (مستقل) می‌تواند هم در مقیاس کمی و هم مقوله‌ای باشد؛ ولی متغیر وابسته مقوله‌ای دو سطحی است. این دو مقوله به گونه‌ای معمول به عضویت یا عدم عضویت در یک گروه اشاره دارد.

یکی از مزایای رگرسیون لجستیک بی نیازی آن به مفروضات محدود کننده آماری در رابطه با متغیرهاست. مسأله اصلی که در این راهبرد وجود دارد این است که تغییرات یکسان وضعیت موجود همیشه تغییرات یکسانی را در احتمال به دست آمده به دنبال ندارد. این موضوع زمانی که احتمال نزدیک مقادیر ۱ یا ۰/۵ باشد اهمیت بیشتری پیدا می‌کند(برک آکایا و همکاران[۳۱]، ۲۰۱۹).

**۳. مروری بر ادبیات و پیشینه پژوهش**

به منظور طراحی راهبرد معاملاتی پژوهشگران خارجی و داخلی از روش‌های یادگیری، شاخص-های ارزشیابی و داده‌های مختلف استفاده نموده‌اند که خلاصه پژوهش و یافته های آن در جدول شماره ۱ نشان داده شده است.

**جدول (۱):** خلاصه پیشینه پژوهش ها ی داخلی و خارجی

| پژوهشگر | سال | روش یادگیری | شاخص ارزشیابی مدل | داده ای استفاده شده | نتیجه گیری |
|---|---|---|---|---|---|
| زهنگ و همکاران(2021) | 2021 | یادگیری عمیق شبکه عصبی بیزی | میانگین مربع خطا | قیمت سهام | در این پژوهش چهار شرکت از کشورهای ایالت متحده، چین، آلمان و استرالیا انتخاب شده است و نتایج نشان می‌دهد که به دلیل نوسانات بالای قیمت سهام در طول دوره کرونا، ارائه پیش‌بینی چالش‌برانگیزتر است. با این حال، یافته ها نشان می دهد که شبکه‌های عصبی بیزی می‌توانند پیش‌بینی‌های معقولی را ارائه دهند. |
| ساعود الاتابی(2021) | 2021 | جنگل تصادفی، بردار پشتیبان و شبکه عصبی | خطای نسبی ریشه میانگین مربعات، میانگین درصد خطای مطلق، میانگین خطای نسبی مطلق و ریشه میانگین درصد مربع خطا | میانگین محدوده واقعی ، میانگین متحرک نمایی، شاخص قدرت نسبی و نرخ تغییر | نتایج تحقق نشان می دهد مدل ترکیبی پیشنهادی دارای دقت حدود ۹۸ درصد برای پیش بینی قیمت سهام بورس عربستان صعودی می باشد |

31 . Berke Akkaya et al

| پژوهشگر | سال | روش یادگیری | شاخص ارزشیابی مدل | داده ای استفاده شده | نتیجه گیری |
|---|---|---|---|---|---|
| هادیان و همکاران(2021) | ۲۰۲۱ | شبکه های عصبی، منطق فازی و الگوریتم ژنتیک | بازده | شاخص قدرت نسبی، شاخص جهت قیمت، استوکاستیک، قیمت پایانی و میانگین متحرک | یافته نشان می دهد بازده سرمایه گذاری با روش فعال بر اساس الگوریتم های تحقیق نسبت به روش خرید و نگهداری بیشتر می باشد |
| توکلی و دوستی(۲۰۲۱) | 2021 | شبکه های عصبی، ماشین بردار پشتیبان و الگوریتم ژنتیک | میانگین مجذورات خطا و ریشه میانگین مجذورات خطا | قیمت | نشان می دهد شبکه های عصبی با بهینه سازی از طریق الگوریتم ژنتیک دارای دقت بالایی می باشد |
| شریف فر و همکاران(2021) | 2021 | الگوریتم حافظه کوتاه مدت ماندگار | مجذور میانگین مربعات خطا و میانگین قدر مطلق درصد خطا | میانگین متحرک وزن دار ۱۰روزه، مومنتوم، استوکاستیک، شاخص قدرت نسبی، میانگین متحرک همگرایی واگرایی، شاخص ویلیامز، شاخص حجم و شاخص کانال کالایی و قیمت باز و بسته شدن و بالا و پایین سهام، ارزش معاملات، سرانه حجم خرید و فروش سهامداران حقیقی و حقوقی و تغییر مالکیت حقیقی به حقوقی | در این مطالعه خطای مدل کمتر از ۵ درصد نمایش داده شده است |
| کیانی زاده(2021) | 2021 | شبکه عصبی، بردار پشتیبان و رگرسیون لجستیک، الگوریتم نزدیک ترین همسایه و اعتبارسنجی متقابل | ضریب تعیین، خطای نسبی ریشه میانگین مربعات و خطای میانگین مربعات | قیمت باز و بسته شدن و بالا و پایین سهام | نتایج حاصل از این آزمایش ها نشان دهنده این است که الگوریتم تخمین ماشین بردار پشتیبان مدل بهینه در پیش بینی قیمت سهام بورس اوراق بهادار تهران می باشد و الگوریتم شبکه عصبی پرسپترون چند لایه در رتبه دوم قرار دارد. رگرسیون لجستیک دارای رتبه سوم در پیش بینی قیمت سهام می باشد و روش های اعتبارسنجی |

| پژوهشگر | سال | روش یادگیری | شاخص ارزشیابی مدل | داده ای استفاده شده | نتیجه گیری |
|---|---|---|---|---|---|
| | | | | | به ترتیب K متقابل و نزدیک ترین همسایه های رتبه های چهارم و پنجم را در پیش بینی قیمت سهام به دست آورده اند |
| ایساک کوف(2020) | ۲۰۲۰ | ماشین بردار پشتیبان، جنگل تصادفی، درخت تصمیم، الگوریتم ژنتیک و شبکه عصبی چند لایه | ناحیه زیر منحنی، خطای نسبی ریشه میانگین مربعات، میانگین درصد خطای مطلق، انحراف استاندارد، دقت و درستی | قیمت باز و بسته شدن سهام، پایین ترین و بالاترین قیمت سهام در یک سال گذشته، میانگین متحرک ساده، میانگین متحرک نمایی، میانگین متحرک همگرایی و واگرایی، شاخص قدرت نسبی، شاخص حجم، استوکاستیک | در این تحقیق دقت مدل بردار پشتیبان برابر با حدود ۹۴ درصد گزارش شده است و یافته های تحقیق نشان می دهد دقت این مدل نسبت به سایر مدلها بیشتر می باشد |
| ایسا کوفی انتی(۲۰۲۰) | 2020 | شبکه عصبی، ماشین‌های بردار پشتیبان و درخت‌های تصمیم | خطای نسبی ریشه میانگین مربعات، میانگین خطای نسبی و خطای میانگین مربعات | میانگین متحرک ساده، میانگین متحرک نمایی ، میانگین متحرک همگرایی واگرایی، شاخص قدرت نسبی و نرخ تغییر | نشان می‌دهد که شبکه‌های عصبی مصنوعی برای پیش‌بینی بازار سهام در مقایسه با ماشین‌های بردار پشتیبان و درخت‌های تصمیم، بر اساس خطای نسبی ریشه میانگین مربعات ، میانگین خطای نسبی و خطای میانگین مربعات بسیار مناسب هستند |
| زهرمن و همکاران(۲۰۱۹) | 2019 | الگوریتم بردار پشتیبان رگرسیونی | خطای میانگین استاندارد شده، خطای نسبی ریشه میانگین مربعات و میانگین درصد خطای مطلق | مانند میانگین متحرک ساده، میانگین متحرک نمایی، میانگین متحرک واگرایی همگرایی، شاخص قدرت نسبی، شاخص کانال کالایی، نرخ تغییرات و شاخص استوکاستیک | یافته های تحقیق نشان می دهد استفاده از داده های ورودی کمتر به دلیل ساده نمودن مدل و از طرفی استفاده از نرخ یادگیری کمتر به نرخ کاهش خطای میانگین استاندارد کمک می کند. |
| دوسدوگر و همکاران (۲۱۰۸) | 2018 | مدل ترکیبی شبکه های عصبی و فراابتکاری | خطای نسبی ریشه میانگین مربعات، میانگین خطای نسبی و | شاخص تراکم/توزیع، نوسان ساز چانکن، | در این تحقیق از مدل‌های شبکه عصبی مختلف برای بهبود دقت پیش‌بینی‌های بازار سهام استفاده |

| پژوهشگر | سال | روش یادگیری | شاخص ارزشیابی مدل | داده ای استفاده شده | نتیجه گیری |
|---|---|---|---|---|---|
| | | | خطای میانگین مربعات | میانگین متحرک همگرایی واگرایی، شاخص حجم منفی، شاخص قدرت نسبی، باندهای بولینگر، شاخص تغییرات قیمت، شاخص تغییرات حجم، مومنتوم | می‌کنیم و ترکیب فراابتکاری و مدل شبکه عصبی دارای مزایای ارزشمندی است. |
| زهنگ و همکاران (۲۰۱۸) | 2018 | از حافظه کوتاه مدت و شبکه عصبی کانولوشنال اقتصادسنجی آرچ، قارچ و آریما و مدل هوش مصنوعی شبکه های عصبی و بردار پشتیبان | خطای نسبی ریشه میانگین مربعات و دقت جهت پیش بینی | بالا، پایین، باز و بسته شدن قیمت سهام و از نوسان و شاخص سازهای تحلیل تکنیکال مانند شاخص قدرت نسبی، باندهای بولینگر، حجم معاملات، گردش، چولگی، مومنتوم، میانگین متحرک نمایی، استوکستیک، میانگین متحرک ساده و میانگین متحرک همگرایی واگرایی | یافت های تحقیق نشان می‌دهد که رویکرد پیشنهادی این مطالعه می‌تواند به طور موثر دقت پیش‌بینی جهت قیمت سهام را بهبود بخشد و خطای پیش‌بینی را کاهش دهد |
| تهانه و همکاران(2018) | 2018 | روش رگرسیون چندگانه | آر اسکوئر خطای درصد مطلق پیش بینی | قیمت سهام، شاخص های کلان اقتصادی مانند شاخص قیمت مصرف کننده، قیمت طلا، نرخ تسعیر، گردش مالی صادارت و واردات، شاخص صنعتی تولید و شاخص قیمت دلار، نرخ بهره وام بلند مدت و نرخ | یافته های تحقیق نشان میدهد مدل ارائه شده برای پیش بینی قیمت سهام کشور ویتنام از دقت بالایی برخوردار است |

| پژوهشگر | سال | روش یادگیری | شاخص ارزشیابی مدل | داده ای استفاده شده | نتیجه گیری |
|---|---|---|---|---|---|
| | | | | بهره وام سه ماه | |
| قاسمی و همکاران(2017) | 2017 | مدل ترکیبی شبکه های عصبی و فراابتکاری | میانگین مربعات خطا، ریشه میانگین مربعات خطا و میانگین خطای کامل | میانگین ساده، میانگین نمایی، شاخص قدرت نسبی، باندهای بولینگر، شاخص توزیع و تراکم، مومنتوم | یافته های تحلیل های آماری نشان می دهد که ترکیبی از شبکه های عصبی مصنوعی و فراابتکاری نقش بسزایی در پیش بینی دقیق قیمت سهام بورس اوراق بهادار تهران دارند.نتایج نشان می دهد مدل هوش مصنوعی و عملکرد برتری در پیش بینی PSO بهینه سازی قیمت سهام دارند |
| زهو و همکاران (۲۰۱۸) | 2018 | از حافظه کوتاه مدت و شبکه عصبی کانولوشنال اقتصادسنجی آرچ، قارچ و آریما و مدل هوش مصنوعی شبکه های عصبی و بردار پشتیبان | خطای نسبی ریشه میانگین مربعات و دقت جهت پیش بینی | بالا، پایین، باز و بسته شدن قیمت سهام و از نوسان و شاخص سازهای تحلیل تکنیکال مانند شاخص قدرت نسبی، باندهای بولینگر، حجم معاملات، گردش، چولگی، مومنتوم، میانگین متحرک نمایی، استوکستیک، میانگین متحرک ساده و میانگین متحرک همگرایی واگرایی | یافت های تحقیق نشان می‌دهد که رویکرد پیشنهادی این مطالعه می‌تواند به طور موثر دقت پیش‌بینی جهت قیمت سهام را بهبود بخشد و خطای پیش‌بینی را کاهش دهد |

## ۴. شیوه پژوهش

سوال تحقیق عبارت است از: آیا راهبرد معاملاتی طراحی شده بر اساس الگوریتم‌های یادگیری رگرسیونی برای خرید و فروش سهام شرکت‌های پذیرش شده در بورس نیویورک با داده ها برازش دارد؟

با عنایت به طراحی راهبرد معلاتی سهام و بهره برداری آن توسط سرمایه گذاران حقیقی و حقوقی این پژوهش از نظر هدف در گروه تحقیقات کاربردی طبقه بندی و نظر به اینکه یک بررسی بازنگرانه جهت شناسایی و نقش عوامل، جهت طراحی راهبرد معاملاتی سهام و ارائه پیام خرید، نگهداری و فروش سهام می‌باشد از نظر گردآوری داده ها جزء تحقیقات پس رویدادی است.

به منظور انتخاب نمونه پژوهش در ابتدا بر اساس طبقه بندی انجام شده در پایگاه اطلاع رسانی یاهوفاینانس شرکت های فعال پذیرش شده در صنایع متنوع مانند حوزه خدمات مالی، مراقبت های بهداشتی، تکنولوژی، خدمات ارتباطی، چرخه مصرف کننده، حمایت مصرف کننده و انرژی در بورس نیویورک شناسایی و در گام بعدی جهت کاهش ریسک نقدشوندگی، شرکت‌های با ارزش بازار بزرگ تر از ۲۰۰ تریلیون دلار تعیین و در نهایت شرکت های با میانگین حجم معامله سه ماهه بیشتر از ۱ میلیون، بوسیله روش فیلترنویسی در تاریخ ۱۴۰۰/۰۴/۰۷ به عنوان نمونه این پژوهش انتخاب شده اند.

**جدول(۲):** جامعه و نمونه

| ردیف | صنعت | تعداد شرکت | | | تعداد نمونه |
|---|---|---|---|---|---|
| | | پذیرش شده در بورس | با ارزش بازار بزرگتر از ۲۰۰ تریلیون دلار | با میانگین حجم معاملات سه ماهه بالاتر از ۱ میلیون | |
| ۱ | خدمات مالی | ۱۶۰۶ | ۲۳ | ۶ | ۶ |
| ۲ | مراقبت های بهداشتی | ۴۹۳ | ۶ | ۶ | ۶ |
| ۳ | تکنولوژی | ۴۶۷ | ۱۰ | ۹ | ۹ |
| ۴ | خدمات ارتباطی | ۱۸۷ | ۸ | ۸ | ۸ |
| ۵ | چرخه مصرف کننده | ۴۰۱ | ۶ | ۵ | ۵ |
| ۶ | حمایت مصرف کننده | ۱۶۱ | ۵ | ۵ | ۵ |
| ۷ | انرژی | ۲۱۸ | ۲ | ۲ | ۲ |
| **جمع کل نمونه** | | ۳۵۳۳ | ۶۱ | ۴۱ | **۴۱** |

شاخص‌ها و نوسان سازها به عنوان داده های ورودی راهبرد معاملاتی مد نظر قرار گرفته شده و به شرح جدول ذیل محاسبه گردیده است:

جدول(۳): روش های اندازه گیری شاخص ها و نوسان سازهای تحلیل تکنیکال

| ردیف | شاخص اندازه گیری | کد شاخص | روش اندازه گیری |
|---|---|---|---|
| 1 | Simple Moving Average | SMA | $\frac{A_1+A_2+A_3+\cdots+A_n}{N}$ |
| 2 | Exponential Moving Average | EMA | Price(t) × k + EMA(y) × (1 − k) |
| 3 | Moving Average Convergence/Divergence | MACD | $\sum_{i=1}^{n} EMA_k - \sum_{i=1}^{n} EMA_d$ |
| 4 | Relative strength Index | RSI | $100 - [\frac{100}{1+\frac{Average\ gain}{Average\ loss}}]$ |
| 5 | Bollinger Band | BB | BOLU=MA(TP,n)+m∗σ[TP,n]<br>BOLD=MA(TP,n)−m∗σ[TP,n]<br>where:<br>BOLU=Upper Bollinger Band<br>BOLD=Lower Bollinger Band<br>MA=Moving average<br>TP (typical price)=(High+Low+Close)÷3<br>n=Number of days in smoothing period (typically 20)<br>m=Number of standard deviations (typically 2)<br>σ[TP,n]=Standard Deviation over last n periods of TP |
| 6 | Commodity Channel Index | CCI | Commodity Channel Index<br>CCI = (Typical Price - 20-period SMA of TP) / (.015 x Mean Deviation)<br>Typical Price (TP) = (High + Low + Close)/3<br>Constant = .015 |
| 7 | Aroon Up | AU | Aroon Up=$\frac{25-\text{Periods Since 25 period High}}{25}$ ∗100 |
| 8 | Aroon Down | AD | Aroon Down=$\frac{25-\text{Periods Since 25 period Low}}{25}$ ∗100 |
| 9 | Chande Momentum Oscillator | CMO | CMO = $\frac{sH-sL}{sH+sL}$ ∗100<br>sH= sum of higher closes over N periods |

| ردیف | شاخص اندازه گیری | کد شاخص | روش اندازه گیری |
|---|---|---|---|
| | | | sH= sum of lower closes of N periods |
| 10 | Average Directional Index | ADX | $+DI = (\frac{smoothed+DM}{ATR}) * 100$<br>$-DI = (\frac{smoothed-DM}{ATR}) * 100$<br>$DX = (\frac{\lvert +DI-(-DI)\rvert}{\lvert +DI+(-DI)\rvert}) * 100$<br>$ADX = (\frac{(Prior\ ADX*13)+corrent\ ADX}{14}) * 100$ |
| 11 | Momentum | MOM | MOM = Price - Price of n periods ago |
| 12 | Accumulation/Distribution | AD | $AD = (\frac{(Close-Low)-(High-Close))}{(High-Low)}) * Volume$ |

در این پژوهش به منظور بررسی برازش و خطای راهبرد طراحی شده از شاخص های آماری به شرح جدول ذیل استفاده شده است:

جدول (۴): شاخص های ارزشیابی مدل

| ردیف | نام شاخص اندازه گیری | کد شاخص | روش اندازه گیری |
|---|---|---|---|
| ۱ | Mean squared error | MSE | $(y.\hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (y_i - \hat{y}_i)^2$ |
| ۲ | Root Mean squared error | RMSE | $RSME = \frac{\sqrt{\sum_{i=1}^{n}(P_{i-} O_i)}}{n}$ |
| ۳ | $R^2$ score | $R^2$ | $(y.\hat{y}) = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$ |

مباحث نظری و پیشینه پژوهشی از طریق مطالعه منابع، نشریات، کتاب‌های داخلی و خارجی و پایگاه‌های اطلاع رسانی جمع‌آوری شده است. درخصوص منابع داخلی از پایگاه اطلاع رسانی پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)، مجلات علمی و پژوهشی مورد تایید وزارت علوم، تحقیقات و فناوری، پایگاه مرکز اطلاعات علمی جهاد دانشگاهی، پایگاه مجلات تخصصی نور(نورمگز) و پایگاه اطلاع مقالات سیویلیکا و برای مقالات بین المللی از پایگاه گوگل

اسکولار[۳۲]، درگاه تحقیق[۳۳] و پایگاه اطلاع رسانی مجلات تخصصی، کتابخانه الکترونیکی گوگل[۳۴] و مرکز اطلاعات منابع آموزشی[۳۵] بهره برداری شده است.

با عنایت به اینکه جامعه آماری پژوهش شرکت‌های پذیرش شده در بورس نیویورک می-باشد و داده‌های جامعه هدف از طریق منابع و پایگاه های داخلی قابل دستیابی نبوده است، لذا اطلاعات و داده‌های مورد نیاز تحقیق(قیمت باز و بسته شدن، بالاترین و کمترین قیمت در بازه زمانی روزانه[۳۶]) از طریق برنامه نویسی در نرم افزار پایتون و پایگاه اطلاع رسانی یاهو[۳۷] برای سهام نمونه پژوهش استخراج گردیده و با استفاده ازکتابخانه‌های موجود و برنامه نویسی، شاخص و نوسان سازهای مد نظر به عنوان داده‌های ورودی محاسبه گردیده است. برخی از مهمترین کتابخانه های نرم افزاری استفاده شده در این تحقیق عبارتند از: نام پای[۳۸]، پانداس[۳۹]، سی بورن[۴۰]، پای فولیو[۴۱] و مت پلات لیب[۴۲]. هدف از این مطالعه طراحی راهبرد معاملاتی خرید و فروش سهام شرکت های پذیرش شده در بورس نیویورک با استفاده از مولفه های تحلیل تکنیکال بر اساس الگوریتم های یادگیری ماشین طبقه بندی می باشد. ابتدا شاخص ها و نوسان سازهای میانگین متحرک ساده، میانگین متحرک نمایی، میانگین متحرک همگرا و واگرا، شاخص قدرت نسبی، باندهای بولینگر، کانال قیمت، آرون بالا و پایین، مومنتوم، شاخص میانگین جهت دار و شاخص تراکم و توزیع بر اساس قیمت پایین ترین، بالاترین، آغاز و پایان سهام شرکت محاسبه می گردد. سپس، شاخص های محاسبه شده به عنوان داده‌های ورودی در الگوریتم‌های ماشین بردار پشتیبان، جنگل تصادفی و نزدیک ترین همسایه با نرخ یادگیری و آزمون ۶۰ و۴۰ درصد استفاده می شود. در مرحله بعد، در راستای هدف اصلی تحقیق به منظور انتخاب راهبرد معاملاتی از شاخص

32 .google scholar
33 . ResearchGate
34 .googlebooks
35 . Educational Resources Information Cente(ERIC)
36 . Daily time frame
37 . yahoo finance
38. numpy
39 . pandas
40 . seaborn
41.pyfoloio
42 .matplotlib

ضریب تعیین[۴۳]، میانگین مربعات خطا[۴۴] و ریشه میانگین مربعات خطا[۴۵] جهت بررسی الگوریتم های رگرسیونی استفاده می‌گردد. شاخص های ارزشیابی مدل پس از پردازش در محیط اکسل، وارد محیط اس پی اس اس[۴۶] می گردد. در نهایت به منظور بررسی توزیع داده‌های پژوهش از آزمون‌های آماری کولموگروف اسمیرنوف[۴۷] و شاپیرو ویلک[۴۸] و در راستای سوال تحقیق برای بررسی راهبردهای مختلف از آزمون های جدول توافقی[۴۹]، آزمون کای دو[۵۰]، آزمون کروسکال والیس استفاده شده است.

## ۵. یافته ها و تحلیل

### ۵,۱. سوال تحقیق

کدام راهبرد معاملاتی طراحی شده بر اساس الگوریتم های یادگیری رگرسیونی برای خرید و فروش سهام شرکت های پذیرش شده در بورس نیویورگ بر اساس شاخص های ضریب تعیین، میانگین مربعات خطا و ریشه میانگین مربعات خطا عملکرد مناسب تری دارد؟

جدول(۵): آمار توصیفی شاخص های ارزشیابی الگوریتم های یادگیری رگرسیونی

| شاخص | نمونه | دامنه | کمینه | بیشینه | میانگین | | انحراف معیار | واریانس | چولگی | | کشیدگی | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | مقدار | مقدار | مقدار | مقدار | مقدار | انحراف معیار | مقدار | مقدار | مقدار | انحراف معیار | مقدار | انحراف معیار |
| میانگین مربعات خطا | ۱۸۴۵ | ۱,۸۳ | .۳۴ | ۲,۱۷ | ۱,۱۴ | .۰۰۸۱۴ | .۳۴۹۶ | .۱۲۲ | .۴۱۵ | .۰۵۷ | -.۶۶۲ | .۱۱۴ |
| ریشه میانگین مربعات خطا | ۱۸۴۵ | .۸۹ | .۵۸ | ۱,۴۷ | ۱,۰۵ | .۰۰۳۸۱ | .۱۶۳۵ | .۰۲۷ | .۱۲۰ | .۰۵۷ | -.۶۴۰ | .۱۱۴ |
| ضریب تعیین | ۱۸۴۵ | ۱,۸۱ | -۱,۱۹ | .۶۲ | -.۲۰ | .۰۰۸۱۵ | .۳۴۹۹ | .۱۲۲ | -.۳۷۵ | .۰۵۷ | -.۸۰۹ | .۱۱۴ |

43 .$R^2$
44 .MSE
45 .RMSE
46 .SPSS
47 . Kolmogorov-Smirnov test
48 . Shapiro-Wilk test
49 . crosstab test
50 . Chi-Square tests

فرض آماری در خصوص بررسی نرمال بودن داده های پژوهش:

H: P = P0 (توزیع داده های تحقیق نرمال است)

$H_1$: P ≠ P0 (توزیع داده های تحقیق نرمال نمی باشد)

برای بررسی نرمال بودن داده های پژوهش از دو آزمون کولموگروف اسمیرنوف و آزمون شاپیرو – ویلک استفاده شده و نتایج آن به شرح جدول ذیل گزارش می گردد.

**جدول(۶):** آزمون بررسی نرمال بودن داده ها ی شاخص های ارزشیابی الگوریتم های رگرسیونی

| شاخص | کولموگروف اسمیرنوف | | | شاپیرو – ویلک | | |
|---|---|---|---|---|---|---|
| | آمار | درجه آزادی | سطح معنی داری | آمار | درجه آزادی | سطح معنی داری |
| میانگین مربعات خطا | .۰۹۷ | ۱۸۴۵ | ۰.۰۰۰ | .۹۶۵ | ۱۸۴۵ | ۰.۰۰۰ |
| ریشه میانگین مربعات خطا | .۰۶۷ | ۱۸۴۵ | ۰.۰۰۰ | .۹۸۱ | ۱۸۴۵ | ۰.۰۰۰ |
| ضریب تعیین | .۱۰۷ | ۱۸۴۵ | ۰.۰۰۰ | .۹۶۱ | ۱۸۴۵ | ۰.۰۰۰ |

همانگونه که در جدول فوق الذکر مشاهده می گردد آماره کولموگروف اسمیرنوف برای شاخص های میانگین مربعات خطا، ریشه میانگین مربعات خطا و ضریب تعیین به ترتیب برابر با ۰۹۷. ، ۰۶۷. و ۱۰۷. با درجه آزادی ۱۸۴۵ دارای سطح معنی داری کمتر از ۰۰۰۱. می باشد و به عبارتی سطح معنی داری این آزمون کوچک تر از ۵ درصد بوده و فرض صفر پژوهش مبنی بر اینکه داده های مربوط به سنجه های ارزشیابی آماری(ضریب تعیین، میانگین مربعات خطا و ریشه میانگین مربعات خطا) راهبردهای معاملاتی رگرسیونی دارای توزیع نرمال می باشد رد و فرض خلاف مبنی بر نرمال نبودن داده های پژوهش پذیرفته می شود.

آزمون شاپیرو –ویلک که دارای دقت بیشتری نسبت به آزمون کولموگروف اسمیرنوف می باشد، دارای نتیجه مشابه بوده و آماره این آزمون برای شاخص های خطای میانگین مربعات، خطای جذر میانگین مربعات و ضریب تعیین به ترتیب برابر با ۹۶۵. ، ۹۸۱. و ۹۶۱. با درجه آزادی ۱۸۴۵ دارای سطح معنی داری کمتر از ۰۰۰۱. می باشد به عبارتی سطح معنی داری این آزمون نتایج آزمون کولموگروف اسمیرنوف را تایید می کند.

با عنایت به اینکه داده های مربوط به سنجه های ارزشیابی آماری دارای توزیع نرمال نمی باشد، برای بررسی برتری شاخص های آماری راهبردهای رگرسیونی طراحی شده، نمی توان از آزمون های پارامتریک استفاده نمود و باید از آزمون های غیرپارمتریک استفاده گردد. نظر به سوال تحقیق جهت بررسی شاخص های ارزشیابی مدل های رگرسیونی از آزمون کروسکال والیس استفاده شده است. فرض آماری در خصوص بررسی تفاوت شاخص‌های ضریب تعیین، میانگین مربعات خطا و ریشه میانگین مربعات خطا راهبردهای طراحی شده عبارتست از:

$H_0$: $\mu_1 = \mu_2 = \mu_3$

$H_1$: $\mu_1 \neq \mu_2$

در جدول شماره ۵ شاخص های مرکزی و توزیع فراوانی آمار توصیفی مربوط به سنجه های راهبردهای معاملاتی رگرسیونی(نزدیک ترین همسایه، رگرسیون لوجستیک، شبکه های عصبی چندلایه پیشخور، جنگل تصادفی و ماشین بردار پشتیبان ) نمره میانگین مربعات خطا، ضریب تعیین و میانگین مربعات خطا راهبرد معاملاتی جنگل تصادفی رگرسیونی در مقایسه با سایر راهبردهای معاملاتی طراحی شده رگرسیونی مناسب تر گزارش شده است. لذا به منظور بررسی معنی داری این تفاوت از آزمون غیرپارامتریک کروسکال والیس به شرح جدول ذیل استفاده شده است.

جدول(۷): آزمون آماری کروسکال والیس برای شاخص های الگوریتم های رگرسیونی

| شرح | میانگین مربعات خطا | ریشه میانگین مربعات خطا | ضریب تعیین |
|---|---|---|---|
| کای دو | ۱۲۷۶٫۸۸۰ | ۱۲۷۶٫۸۸۰ | ۱۴۱۸٫۸۵۴ |
| درجه آزادی | ۴ | ۴ | ۴ |
| سطح معنی داری | ۰.۰۰۰ | ۰.۰۰۰ | ۰.۰۰۰ |

یافته های آزمون آماری کروسکال والیس بر اساس جدول فوق الذکر نمایش می دهد آماره کای دو برای شاخص های میانگین مربعات خطا ، ریشه میانگین مربعات خطا و ضریب تعیین به ترتیب برابر با ۱۲۷۶٫۸۸۰، ۱۲۷۶٫۸۸۰ و ۱۴۱۸٫۸۵۴ و درجه آزادی ۴ در سطح معنی داری کمتر از ۰۰۰۱. معنی دار می باشد. به عبارتی مقدار آماره کای دو و سطح معنی حاکی از آن است که بین شاخص

های ارزشیابی میانگین مربعات خطا و ریشه میانگین مربعات خطا و ضریب تعیین راهبردهای معاملاتی الگوریتم های رگرسیونی نزدیک ترین همسایه رگرسیونی، رگرسیون لوجستیک، شبکه های عصبی چندلایه پیشخور، جنگل تصادفی رگرسیونی و ماشین بردار پشتیبان رگرسیونی) اختلاف معنی داری وجود دارد. به منظور ارزشیابی راهبردهای معاملاتی رگرسیونی طراحی شده، بر اساس آزمون آماری کروسکال والیس و شاخص مرکزی میانگین راهبردهای طراحی شده به شرح جدول ذیل رتبه بندی می شوند:

جدول (۸): رتبه بندی راهبردهای معاملاتی رگرسیونی بر اساس آزمون کروسکال والیس

| رتبه بر اساس میانگین | میانگین | فراوانی | راهبرد معاملاتی | شاخص |
|---|---|---|---|---|
| ۱ | ۳۰۴,۰۶ | ۳۶۹ | *جنگل تصادفی | میانگین مربعات خطا |
| ۲ | ۷۲۲,۰۱ | ۳۶۹ | ماشین بردار پشتیبان | |
| ۲ | ۷۳۸,۵۴ | ۳۶۹ | نزدیکترین همسایه | |
| ۴ | ۱۳۲۴,۶۸ | ۳۶۹ | شبکه های عصبی | |
| ۵ | ۱۵۲۵,۷۲ | ۳۶۹ | رگرسیون لجستیک | |
| ۱ | ۳۰۴,۰۶ | ۳۶۹ | *جنگل تصادفی | ریشه میانگین مربعات خطا |
| ۲ | ۷۲۲,۰۱ | ۳۶۹ | ماشین بردار پشتیبان | |
| ۳ | ۷۳۸,۵۴ | ۳۶۹ | نزدیکترین همسایه | |
| ۴ | ۱۳۲۴,۶۸ | ۳۶۹ | شبکه های عصبی | |
| ۵ | ۱۵۲۵,۷۲ | ۳۶۹ | رگرسیون لجستیک | |
| ۱ | ۱۵۸۱,۵۹ | ۳۶۹ | *جنگل تصادفی | ضریب تعیین |
| ۲ | ۱۱۳۱,۸۵ | ۳۶۹ | ماشین بردار پشتیبان | |
| ۳ | ۱۱۱۰,۳۶ | ۳۶۹ | نزدیکترین همسایه | |
| ۴ | ۵۰۱,۸۸ | ۳۶۹ | شبکه های عصبی | |
| ۵ | ۲۸۹,۳۳ | ۳۶۹ | رگرسیون لجستیک | |

در جدول شماره ۸ نمایش داده شده است که راهبرد معاملاتی جنگل تصادفی از نظر شاخص های ارزشیابی مدل نسبت به سایر راهبردهای طراحی شده دارای میانگین و رتبه بالاتری می باشد. در جدول ذیل شاخص های مرکزی و توزیع فراوانی راهبرد جنگل تصادفی رگرسیونی به تفکیک صنعت نمایش داده شده است.

### ۶. بحث و نتیجه گیری

نتایج بررسی انجام شده و یافته های پژوهش نشان می‌دهد بر اساس شاخص های آماری میانگین مربعات خطا، راهبرد معاملاتی طراحی شده بر مبنای الگوریتم جنگل تصادفی دارای کمترین مقدار خطا، بر اساس شاخص ریشه میانگین مربعات خطا، راهبرد معاملاتی مبتنی بر جنگل تصادفی دارای کمترین مقدار خطا و بر اساس شاخص ضریب تعیین دارای بیشترین مقدار قدرت تبیین بوده و این اختلاف با سایر الگوریتم های رگرسیونی بر اساس آزمون های آماری معنی دار می باشد. با عنایت به شاخص های آماری فوق الذکر راهبرد معاملاتی طراحی شده بر مبنای الگوریتم جنگل تصادفی به عنوان راهبرد معاملاتی پیشنهاد می شود. شایان ذکر است به منظور پیشگیری از متغیرهای مداخله گر در نتایج تحقیق مانند صدور بخشنامه های جدید سازمان بورس و اوراق بهادار در خصوص تغییر دامنه نوسان، تغییر نسبت سرمایه گذاری صندوق های سرمایه گذاری درآمد ثابت، اعتبار تخصیصی به سرمایه گذاران و نحوه تسویه آن، بستن بلند مدت برخی سهام شرکت های بورسی، صف های خرید و فروش، وضعیت سیاسی کشور و تاثیر پذیری بازار سرمایه از نرخ دلار و ... تلاش شده است در راستای کنترل این موضوع، تحقیق در بازار سرمایه بورس نیویورک و در صنایع متنوع جهت پایداری و افزایش قابلیت تعمیم پذیری یافته های تحقیق انجام شود. نتایج این تحقیق می‌تواند به سرمایه گذاران حقیقی، حقوقی(شرکت های سبدگردان، مشاوره سرمایه گذاری، خدمات مالی، هلدینگ های مالی، سرمایه گذاری و ...) جهت افزایش سودآوری و کاهش ریسک سرمایه گذاری در بورس اوراق بهادار ایران کمک و مورد بهره برداری قرار گیرد. نقدشوندگی و عمق بخشی به بازار سرمایه از دیگر مزایای استفاده از معاملات الگوریتمی می باشد که لازمه توسعه و رشد بورس اوراق بهادار است.

نتایج این پژوهش با نتایج علیزاده و همکاران(۲۰۲۲) که برای پیش بینی روند قیمت از امواج الیوت و الگوریتم یادگیری جنگل تصادفی و روش ارزشابی دقت استفاده نموده، یافته های هادیان و همکاران(۲۰۲۱) که در پژوهش خود با استفاده از الگوریتم ژنتیک، منطق فازی و شبکه های عصبی و استفاده از داده های شاخص قدرت نسبی، استوکاستیک، قیمت پایانی، میانگین متحرک و شاخص جهت قیمت به بررسی تفاوت بین بازده سرمایه گذاران فعال و غیر فعال پرداخته اند و یافته های پژوهش ساعود الاتابی(۲۰۲۱) که در آن از شاخص سازها و نوسان سازهای میانگین محدوده واقعی، میانگین متحرک نمایی، شاخص قدرت نسبی و نرخ تغییر استفاده شده و نتایج پژوهش کرولا رویا و گویجارو(۲۰۲۱) که از شاخص سازها و نوسان سازهای باندهای بولینگر، نوسانات چایکن، نوسانات نزدیک به بسته شدن، شاخص کانال کالایی، شاخص حرکت جهت دار، میانگین متحرک نمایی، شاخص جریان پول، همگرایی واگرایی میانگین متحرک، شاخص حجم، شاخص قدرت نسبی، شاخص مومنتوم استوکاستیک و شاخص پارکینسون برای طراحی راهبرد معاملاتی استفاده نموده است و نتایج پژوهش زهنگ و همکاران (۲۰۲۲)که در این پژوهش از شاخص سازها و نوسان سازهای بالا، پایین، باز و بسته شدن قیمت سهام و از نوسان و شاخص سازهای تحلیل تکنیکال مانند شاخص قدرت نسبی، باندهای بولینگر، حجم معاملات، گردش، چولگی، مومنتوم، میانگین متحرک نمایی، استوکستیک، میانگین متحرک ساده و میانگین متحرک همگرایی واگرایی برای طراحی راهبرد معاملاتی استفاده شده است و نشان می دهند استفاده از الگوریتم های هوش مصنوعی و ماشین بردار پشتیبان به طور موثر دقت پیش‌بینی جهت قیمت سهام را بهبود بخشد و خطای پیش‌بینی را کاهش میدهد، دارای نتایج یکسانی می باشد و با نتایج پژوهش توکلی و دوستی(۲۰۲۱) که از داده های قیمت و روش‌های یادگیری شبکه های عصبی، ماشین بردار پشتیبان و الگوریتم ژنتیک و روش ارزشیابی دقت، ایساک کوفی (۲۰۲۰) که در آن از شاخص سازها و نوسان سازهای قیمت باز و بسته شدن سهام، پایین ترین و بالاترین قیمت سهام در یک سال گذشته، میانگین متحرک ساده، میانگین متحرک نمایی، میانگین متحرک همگرایی و واگرایی، شاخص قدرت نسبی، شاخص حجم، استوکاستیک استفاده شده، دارای نتایج متفاوتی

بوده است. طراحی راهبرد معاملاتی و استفاده از شاخص سازها و نوسان سازها متنوع به عنوان ورودی مدل، بهره برداری از الگوریتم های پیچیده یادگیری رگرسیونی ماشین بردار پشتیبان، جنگل تصادفی، شبکه های عصبی، رگرسیون لوجستیک و نزدیک ترین همسایه، شاخص های آماری ضریب تعیین، میانگین مربعات خطا و ریشه میانگین مربعات خطا و کاربرد فناوری اطلاعات و علوم داده در حوزه مالی را می توان از نوآوری های پژوهش نام برد. شایان ذکر است حیدری و همکاران (۲۰۲۲) در پژوهش خود نشان دادند از بین راهبردهای طراحی شده برای خرید و فروش سهام بر اساس الگوریتم‌های طبقه بندی ماشین بردار پشتیبان، جنگل تصادفی و نزدیک ترین همسایه با استفاده از شاخص های دقت[۵۱]، فراخوانی[۵۲]و معیار[۵۳]F، راهبرد طراحی شده برای خرید و فروش سهام بر اساس الگوریتم جنگل تصادفی طبقه بندی در مقایسه با سایر الگوریتم های طبقه بندی دارای بیشترین میزان نمره دقت، فراخوانی و F بوده و این اختلاف بر اساس آزمون های آماری معنی دار می باشد.

## منابع


صمدی، فاطمه؛ خسروی، فاطمه؛ اسلامی مفیدآبادی، حسین .(2023) .بهینه سازی سبد سهام با استفاده ازالگوریم‌های جستجوی ممنوعه و فروشنده دوره گرد .*فصلنامه تحلیل بازار سرمایه*، ۳ (1)، ۱۱۰-۱۴۰،

میرزائی، سیداحمد؛ نیکدل، زهرا؛ نیکدل، زکیه .(2021) . بررسی آموزش شبکه عصبی با استفاده از الگوریتم‌های فراابتکاری به‌منظور پیش‌بینی شاخص کل در بورس ایران .*فصلنامه تحلیل بازار سرمایه*،۱ (2)، 188-212،

Alizadeh, H., Zanjirdar, M., & Haji, G. (2022). The Ability of Elliott Waves Theory to Predict the Information Content of Accounting Profit. *Advances in Mathematical Finance & Applications, 7*. doi: 10.22034/AMFA.2022.1950621.1685

Alotaibi, S. S. (2021). Ensemble Technique with Optimal Feature Selection for Saudi Stock Market Prediction: A Novel Hybrid Red Deer-Grey Algorithm. *IEEE Access, 9*, 64929–64944. doi: 10.1109/ACCESS.2021.3073507

Benjamin Recht. (2018). A Tour of Reinforcement Learning: The View from Continuous Control. *Optimization and Control*. doi: https://doi.org/10.48550/arXiv.1806.09460


51 .Accuracy
52 .Recall
53 .F_Measure

Berke AKKAYA, & NURDAN ÇOLAKOĞLU. (2019). Comparison of Multi-class Classification Algorithms on Early Diagnosis of Heart Diseases. *Recent Advances in Data Science and Business Analytics*, 162–172.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory - COLT '92*, 144–152. doi: 10.1145/130385.130401

Breiman, L. (2001). Machine Learning. *Machine Learning*, *45*(1), 5–32. doi: 10.1023/A:1010933404324

Bristow, D. A., & Tharayil, M. (2006). A learning-based method for high-performance tracking control. *IEEE Control Systems Magazine*, *26*(3), 96–114. doi: 10.1109/MCS.2006.1636313

Cervelló-Royo, R., & Guijarro, F. (2020). Forecasting stock market trend: a comparison of machine learning algorithms. *Finance, Markets and Valuation*, *6*(1), 37–49. doi: 10.46503/NLUF8557

DOSDOĞRU, A. T. B. O. R. U. A. M. G. M. Ö. and T. G. (2108). Assessment of hybrid artificial neural networks and metaheuristics for stock market forecasting. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *27*(1), 63–78.

Fradkov, A. L. (2020). Early History of Machine Learning. *IFAC-PapersOnLine*, *53*(2), 1385–1390. doi: https://doi.org/10.1016/j.ifacol.2020.12.1888

Ghasemiyeh, R., Moghdani, R., & Sana, S. S. (2017). A Hybrid Artificial Neural Network with Metaheuristic Algorithms for Predicting Stock Price. *Cybernetics and Systems*, *48*(4), 365–392. doi: 10.1080/01969722.2017.1285162

Hadian, H., Haskuee, M., & Zomorodain, G. (2021). An Algorithmic Trading System Based on Machine Learning in Tehran Stock Exchange. *Advances in Mathematical Finance & Applications*, *6*. doi: doi: 10.22034/amfa.2020.1894049.1380

Heydari. Nasser, Zanjirdar, M., & Lalbar, A. (2022). Designing A Trading Strategy to Buy and sell the stock of Companies Listed on The New York Stock Exchange Based on Classification Learning Algorithms. *Advances in Mathematical Finance & Applications*, *8*. doi: (DOI): 10.22034/AMFA.2022.1967149.1796

Kaur, G., Goyal, S., & Kaur, H. (n.d.). *Brief Review of Various Machine Learning Algorithms*. doi: doi.org/10.2139/ssrn.3747597

kianizadeh, hosain. (2021). *Comparing the accuracy of machine learning models in predicting stock prices and introducing the optimal model*. Islamic Azad University, Kish International Branch.

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020a). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, *53*(4), 3007–3057. doi: 10.1007/s10462-019-09754-z

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020b). Efficient Stock-Market Prediction Using Ensemble Support Vector Machine. *Open Computer Science*, *10*(1), 153–163. doi: 10.1515/comp-2020-0199

Perwej, Y., & Perwej, A. (2012). Prediction of the Bombay Stock Exchange (BSE) Market Returns Using Artificial Neural Network and Genetic Algorithm. *Journal of Intelligent Learning Systems and Applications*, *04*(02), 108–119. doi: 10.4236/jilsa.2012.42010

Qiu, M., Song, Y., & Hasama, M. (2013). *EMPIRICAL ANALYSES OF THE "DOGS OF THE DOW" STRATEGY: JAPANESE EVIDENCE*.

Rustam, Z., & Takbiradzani, K. (2019). Application of Support Vector Regression for Jakarta Stock Composite Index Prediction with Feature Selection Using Laplacian score. *Journal of Theoretical and Applied Information Technology*, *97*(1), 88–97. Retrieved from www.jatit.org

sharifar, A., khaliliaraghi, M., Reisi vanani, I., & Fallah, mir feyz. (2021). The Assessment of the optimal Deep Learning Algorithm on Stock Price Prediction (Long Short-Term Memory Approach). *Engineering Finance and Portfolio Management*, *12*(48). doi: 20.1001.1.22519165.1400.12.48.16.6

Tavakoli, M., & Doosti, H. (2021). Forecasting The Tehran Stock Market by Machine Learning Methods Using a New Loss Function. *Advances in Mathematical Finance & Applications*, *6*. doi: doi: 10.22034/amfa.2020.1896273.1399

Thanh, D. van, Minh Hai, N., & Hieu, D. D. (2018). Building unconditional forecast model of Stock Market Indexes using combined leading indicators and principal components: application to Vietnamese Stock Market. *Indian Journal of Science and Technology*, *11*(2), 1–13. doi: 10.17485/ijst/2018/v11i2/104908

Wang, B., & Zhang, S. (2022). A new locally adaptive K-nearest centroid neighbor classification based on the average distance. *Connection Science*, *34*(1), 2084–2107. doi: 10.1080/09540091.2022.2088695

Zhang, Q., Yang, L., & Zhou, F. (2021). Attention enhanced long short-term memory network with multi-source heterogeneous information fusion: An application to BGI Genomics. *Information Sciences*, *553*, 305–330. doi: 10.1016/j.ins.2020.10.023

Zhang, S., & Li, J. (2021). KNN Classification with One-step Computation. *IEEE Transactions on Knowledge and Data Engineering*, 1–1. doi: 10.1109/TKDE.2021.3119140

Zhang, X., Zhang, Y., Wang, S., Yao, Y., Fang, B., & Yu, P. S. (2018). *Improving Stock Market Prediction via Heterogeneous Information Fusion*. doi: 10.1016/j.knosys.2017.12.025

Zhou, X., Pan, Z., Hu, G., Tang, S., & Zhao, C. (2018). Stock market prediction on high-frequency data using generative adversarial nets. *Mathematical Problems in Engineering*.