

ارائه ی یک روش پیش بینی قیمت سهام بورس به وسیله ی الگوریتم های بهبود یافته ی ترکیبی یادگیری ماشین

آرش خسروی (نویسنده مسئول)^۱، سید محمدجواد امیرحسینی^۲
^۱د گروه مهندسی کامپیوتر، دانشکده مهندسی، مرکز آموزش عالی محلات، محلات

۱ - مقدمه

گسترش اینترنت باعث شد تا خرید و فروش سهام کارخانه ها هم به صورت آنلاین و لحظه ای شود و بسیاری از مردم به دلیل آسانی و سرعت بالا به آن روی آوردند. اما خرید و فروش بدون حساب و کتاب ضرر و زیان بسیاری برای افراد به همراه دارد، پس باید حساب شده یک سهام را در بهترین زمان ممکن معامله کرد تا به بیشترین میزان سود یا حتی کمترین میزان ضرر برسیم.

یکی از روش هایی که می توان معامله ی خوبی انجام دهیم تحلیل گذشته ی سهام یک شرکت و پیش بینی آن است که کاری بسیار زمان بر و دشوار است. اما با گسترش علم و دانش رایانه ای می توان این محاسبات را به رایانه ها سپرد و با استفاده از یادگیری ماشین به رایانه ها یاد دهیم که قیمت یک سهام را در آینده پیش بینی کنند و درصد خطای انسانی در محاسبات را حد بسیار زیادی پایین آورند.

بیان مسئله

زمانی که خرید و فروش ها به صورت آنلاین شد، تعداد خرید و فروش در لحظه بالا رفت و در حال حاضر در لحظه سهم های زیادی را معامله کنیم. اما ریسک خرید و فروش بسیار بالا است چون قیمت ها لحظه ای در حال تغییر هستند و ما باید دنبال معیاری برای خرید و یا فروش در نظر بگیریم. بهترین معیار برای به دست آوردن بهترین زمان معامله تحلیل روند صعودی یا نزولی بودن سهم در گذشته به وسیله ی فرمول های خاصی است تا بتوانیم به کمک آنها آینده را پیش بینی کنیم و به ماشین نیز یاد دهیم تا بتواند برای ما معامله کند. پس ما به دنبال این هستیم که چگونه ماشین به کمک تاریخچه یک سهم در آینده به ما اطلاع دهد که چه کاری کنیم.

اهمیت و ضرورت تحقیق

در بازار های مالی معامله ها ریسک بسیار بالایی دارند و تاثیر مستقیم بر روی سرمایه ی اصلی معامله کننده می گذارد، به این صورت که یک معامله ی خوب می تواند سرمایه ی یک فرد را چندین برابر کند اما یک معامله ی بد امکان این را دارد که

چکیده

امروزه که بیشتر مردم در حال سرمایه گذاری در بازارهای مالی آنلاین هستند، تحلیل و بررسی و پیش بینی قیمت سهام برای خرید و فروش با سود بیشتر از اهمیت بسیار زیادی برخوردار است. لذا درصدد آن برآمدهیم تا بتوانیم برنامه ای تهیه و تنظیم کنیم که بتواند قیمت سهام های را پیش بینی کند و به ما اطلاعات بدهد که چه زمانی بهترین زمان برای خرید یا فروش است. ما به کمک اندیکاتور ها طبق اتفاقاتی که در گذشته برای یک سهم افتاده است می توانیم به کمک یادگیری ماشین به فرمولی دست پیدا کنیم که قیمت سهام را در آینده پیش بینی کند و به ما اطلاع دهد که الان باید یک سهم را بخریم، بفروشیم یا هیچ کاری انجام ندهیم. اما این کار را بسیار پیشرفته تر یعنی با ترکیب چند اندیکاتور انجام داده ایم و سپس با الگوریتم های یادگیری ماشین امتحان کردیم. طبق نتایج بدست آمده ما در بعضی سهم ها ما می توانیم بسیار با دقت بالا قیمت را پیش بینی کنیم و سود بالایی از معاملات خود در بازار های مالی آنلاین بگیریم.

کلمات کلیدی: اندیکاتور، یادگیری ماشین، الگوریتم های طبقه بندی

تاریخچه مقاله:

تاریخ ارسال: ۱۴۰۲/۰۲/۱۱

تاریخ اصلاحات: ۱۴۰۲/۰۳/۱۹

تاریخ پذیرش: ۱۴۰۲/۰۵/۲۱

تاریخ انتشار: ۱۴۰۲/۰۶/۱۰

ایمیل نویسنده مسئول: Khosravi.280@gmail.com

شرکت است و در سود و زیان آن شرکت سهیم است. امروزه خرید و فروش سهام شرکت‌های مختلف به راحتی از طریق اینترنت و به واسطه‌ی کارگزاری‌ها انجام می‌شود. اما باید برای سود بیشتر در این بازار سهام شرکت‌هایی را خریداری کنیم که سود بیشتری دارند. داده‌های بورسی به منظور دیتای خام در نظر گرفته می‌شود که آنها را تحلیل و بررسی می‌کنند. منظور از داده‌های بورسی، چارت‌های گذشته‌ی یک سهم است که در گذشته میزان ارزش یک سهم را ثبت کرده است.

تحلیل تکنیکال با اندیکاتور ها

تحلیل: هدف اصلی از سرمایه‌گذاری در سازمان بورس سودبیشتر از مبلغ پولی است که در آن درگیر کردیم. با توجه به این نکته که سهام و ارزش شرکت‌های بورسی در نوسان قرار دارد و برای سود بیشتر باید در خرید و فروش سهام ها دقت بالایی داشته باشیم.

به این منظور برای گرفتن بیشترین سود و یا کمترین ضرر باید بتوانیم ارزش یک سهام را در آینده پیش‌بینی کنیم و یکی از معروف ترین و پر کاربرد ترین روش‌های پیش‌بینی، استفاده از تحلیل و بررسی قیمت یک سهم در گذشته است که به صورت جداولی وجود دارد که باید برای پیش‌بینی قیمت آینده‌ی یک سهم به گذشته‌ی آن نگاه کنیم.

برای موفقیت و در بازارهای مالی روش‌هایی برای تحلیل وجود دارد و بر اساس مدل تحلیل داده‌های گذشته است که شامل ۳ مدل تحلیل می‌شود:

۱) تحلیل تکنیکال^۱:

در تحلیل تکنیکال تحلیلگر تلاش می‌کند تا با بررسی نمودارهای گذشته‌ی یک سهم میزان ارزش آن در آینده را تخمین بزند و با معامله‌ی به موقع سود بیشتری را به دست آورد.

تحلیل تکنیکال دارای ۳ اصل اساسی است که شامل

- همه‌ی جوانب در قیمت لحاظ شده است
- قیمت‌ها بر اساس روند خاصی حرکت می‌کنند
- حتماً گذشته دوباره تکرار می‌شود

کل سرمایه‌ی فرد را بسوزاند پس سرمایه‌گزاران بازارهای مالی باید بسیار دقت کنند و بفهمند که بهترین زمان برای خرید و یا فروش چه زمانی است.

برای به دست آوردن بهترین نقطه‌ی خرید یا فروش باید رفتار گذشته‌ی سهم را بر اساس تحلیل تکنیکال بررسی کنیم و بتوانیم حدس بزنیم که الان بهترین کار ممکن چیست. حالا فرض کنید ما بتوانیم به کمک الگوریتم‌های فرا ابتکاری و یادگیری ماشین رفتار گذشته‌ی سهم را بررسی و به رایانه یاد دهیم که این سهم چگونه رفتاری در بازار داشته و یادگیری ماشین بتواند تحلیل کند و به ما بگوید که بهترین نقطه برای خرید و فروش کدام است. این یعنی ما به کمک رایانه‌ها می‌توانیم تا حد بسیار زیادی خطای انسانی را کاهش دهیم.

ما قصد داریم که ابتدا سهم‌هایی را از بازار بورس تهران انتخاب کنیم و سپس با اندیکاتورهای مختلف مورد بررسی و نتیجه و قرار دهیم و بعد از اینکه یک به یک با اندیکاتورهای مختلف بررسی کردیم اندیکاتورهای منتخب را با هم ترکیب کنیم تا یک نتیجه‌ی واحد به دست آید. نتیجه‌ی به دست آمده را با الگوریتم‌های یادگیری ماشین مقایسه می‌کنیم و بر اساس دقت به دست آمده ۳ سیستم یادگیری ماشین را انتخاب می‌کنیم و در نهایت زمانی که رای اکثریت صادر شد نتیجه‌ی نهایی ما خواهد شد.

سوالات و فرضیه‌های تحقیق

آیا با ترکیب چند اندیکاتور به نتیجه‌ی بهتری می‌رسیم؟

از چه زبان‌های ماشینی برای به دست آوردن نتیجه‌ی بهتر استفاده کنیم؟

آیا امکان دارد الگوریتم‌های یادگیری ماشین را با هم ترکیب کنیم؟ در صورت آری بودن نتیجه‌ی چه خواهد بود؟

۲- پیشینه پژوهش

بورس

در واقع بورس یک بازار دیجیتالی است که در آن اوراق بهادار مانند سهام معامله می‌شود. تاریخچه‌ی بورس به قرن پانزدهم میلادی بر می‌گردد و بر اساس به حراج گذاشتن کالاها شکل گرفت از دیگر مفاهیم بورس می‌تواند به شراکت و تقسیم سود اشاره کرد. فرایند بورس به این صورت است که فردی با خرید سهام یک شرکت به میزان تعداد سهمی که خریده مالک آن

¹ Technical analysis

اندیکاتور RSI^5 : یک اندیکاتور بازگشتی است که برای به دست آوردن نتیجه به میانگین سود و زیان در طی یک دوره ی گذشته نیاز دارد. این الگوریتم دو بخشی است و برای انجام بخش دوم الگوریتم نیازمند این است که فرمول اول را برای ۱۴ دوره بدست آورد.

اندیکاتور $BBands^6$: از سه خط تشکیل شده است که یکی از رایج ترین محاسبات آن با استفاده از میانگین متحرک ۲۰ روزه است. خط بالایی دو برابر انحراف استاندارد روزانه بر اساس خط میانی است و خط پایینی نیز منهای دو برابر انحراف استاندارد روزانه محاسبه می شود.

هوش مصنوعی^۷

به شاخه از علوم کامپیوتر گفته می شود که در آن قرار است یک سیستم کتمپیوتر بتواند مانند مغز انسان فکر کند و تصمیم بگیرد. یعنی در واقع هوش مصنوعی به زبان ساده تر هوش مصنوعی با عت می شود که کامپیوتر ها هم در مواجه با رفتاری قابلیت فکر کردن و پاسخ دادن متناسب با آن رفتار را داشته باشند. یک مثال در این زمینه این است که اگر یک انسان در حال پیام دادن به یک رایانه است آن سیستم رایانه به گونه ای پاسخ او را بدهد که انسان متوجه نشود با یک رایانه در حال صحبت است.

یادگیری ماشین^۸

یکی از زیر مجموعه های هوش مصنوعی می باشد که به جای استفاده از دستور العمل های واضح، از الگو ها و استنباط ها برای انجام وظیفه عمل می کند. به این صورت که مثلا زمانی که می خواهیم به یک خردسال یاد بدهیم که گربه چیست، به محض دیدن گربه به کودک می گوئیم که این گربه است و کودک بعد از چند بار تکرار و دیدن گربه های بیشتر متوجه می شود که این گربه است و آن را از بقیه ی چهار پایان جدا می کند.

در الگوریتم های یادگیری ماشین از الگوریتم های ریاضی استفاده می شود تا با تکرار و آموزش و بررسی داده های قبلی به رایانه یاد بگیرد که چه چیزی پیشنهاد دهد.

تحلیل گران در این سبک تحلیل با توجه به سه اصل بالا و بر اساس نمودار های گذشته تصمیم گیری می کنند.

۲) تحلیل بنیادی:

در این سبک از تحلیل تحلیل گران برای هر سهم قیمت ذاتی ای در نظر می گیرند و آن را با قیمت فعلی سهم مقایسه می کنند و اگر قیمت روز یک سهم کمتر از ارزش ذاتی آن را خریداری می کنند اما اگر قیمت فعلی یک سهم از ارزش ذاتی آن بیشتر باشد اقدام به فروش آن می کنند

۳) تابلو خوانی:

با توجه به اطلاعات موجود در تابلوی اطلاعات یک سهم مانند حجم خرید و... تحلیلگران می توانند اقدام به خرید و فروش کنند

اندیکاتور: در تحلیل تکنیکال تحلیل گران نیازمند ابزار هایی برای به دست آوردن بهترین الگو از رفتار های گذشته ی یک سهم هستند. برای این منظور به وسیله ی فرمول های ریاضی و جداول گذشته به فرمول های خاصی دسترسی پیدا می کنند که با نام اندیکاتور ها شناخته می شود. اندیکاتور های مختلفی وجود دارد که هر کدام ویژگی خاص خود را دارند که در ادامه برخی از آنها را مورد بررسی قرار می دهیم

اندیکاتور SMA^2 : به میانگین متحرک ساده معروف است، که این اندیکاتور قیمت پایانی هر روز را با هم جمع می کند و تقسیم بر تعداد کل روز ها می کند.

اندیکاتور $MACD^3$: با عنوان اندیکاتور همگرایی و واگرایی شناخته می شود. در واقع در این اندیکاتور، دو میانگین با بازه ی زمانی مختلف از هم دیگر ولی از رفتار یک سهم گرفته می شود و سپس خط مکدی را بر اساس میانگین آن دو کشیده می شود.

اندیکاتور SAR^4 : این اندیکاتور بیشتر برای تایین صعودی یا نزولی روند استفاده می شود و شکل کلی آن به صورت نقطه چین است و برای به دست آوردن کف قیمت در خرید ها استفاده می شود

⁶ Bollinger Bands

⁷ Artificial intelligence

⁸ Machine learning

² Simple Moving Average

³ Moving Average Convergence Divergence

⁴ Stop And Reverse

⁵ Relative Strength Index

نام مشتری	میزان حقوق	تعداد فرزندان	خانه دارد؟	مجرد است؟	برچسب
علی رستمی	۳,۰۰۰,۰۰۰	۰	۰	۱	۰
مجتبی رضایی	۵,۰۰۰,۰۰۰	۱	۱	۰	۱
حسن کلظمی	۱۱,۰۰۰,۰۰۰	۳	۰	۰	۱

جدول ۱- یادگیری ماشین

الگوریتم‌های طبقه‌بندی به ما کمک می‌کند که بتوانیم با بررسی اطلاعات گذشته‌ی یک مشتری و بررسی برچسب آن بتوانیم در صورت اضافه شدن مشتری جدید تخمین بزنیم که برچسب آن چگونه است و آیا این مشتری می‌تواند وام را پس دهد یا نه.

الگوریتم‌های یادگیری ماشین در دسته‌ی طبقه‌بندی بسیار زیاد هستند ولی ما نیاز داریم الگوریتم‌هایی را پیدا کنیم که بازدهی بالایی داشته باشند.

الگوریتم KNN^{۱۴}

این الگوریتم که به الگوریتم (کا نزدیک ترین همسایه) نیز معروف است، یکی از الگوریتم‌های یادگیری ماشین است که در دسته‌ی یادگیری ماشین طبقه‌بندی شده قرار می‌گیرد و می‌تواند به ما کمک کند که بتوانیم تشخیص دهیم که یک سطر جدید با مشخصات داده شده چه برچسبی می‌گیرد که با چند مثال برای درک بهتر مطلب موضوع را توضیح می‌دهیم. فرض کنید شما فروشنده‌ی یک فروشگاه هستید و مشتریان خود را به ۲ دسته‌ی کسانی که بالای ۱۰۰ هزار تومان خرید می‌کنند و دسته‌ی بعدی کسانی که زیر ۱۰۰ هزار تومان خرید می‌کنند تقسیم می‌کنید. شما بر اساس تجربه‌ای که در چند سال به دست آوردید متوجه شدید که خانم‌هایی که بالای ۳ سال سن دارند و با اتومبیل به فروشگاه می‌آیند بالای ۱۰۰ هزار تومان خرید می‌کنند و مردان زیر ۲۰ سال که بدون اتومبیل که به فروشگاه می‌آیند زیر ۱۰۰ هزار تومان خرید می‌کنند. حالا فرض کنید که یک خانم میانسال با اتومبیل خود به فروشگاه می‌آید، شما قطعاً از او توقع خرید بالای ۱۰۰ هزار تومان را دارید.

الگوریتم‌های یادگیری ماشین در ۳ دسته‌ی اصلی قرار می‌گیرند که در ادامه آن‌ها را معرفی کرده و توضیح می‌دهیم:

۱) یادگیری با نظارت^۹: در این روش دیتاست (مجموعه

ی داده) تمرینی ما دارای ورودی و همچنین خروجی مناسب است. در طول جلسه یادگیری ماشین، مدلی ایجاد میشود که متغیرهای آن را برای نمایش ورودی به خروجی متناظر تنظیم می‌کند.

۲) یادگیری بدون نظارت^{۱۰}: در این دسته بندی هدفی

برای تولید خروجی وجود ندارد بلکه الگوریتم‌های استفاده شده، داده‌های ما را به گروه‌های مختلف دسته بندی میکنند.

۳) یادگیری تقویتی^{۱۱}: این الگوریتم‌ها براساس تصمیم

گیری آموزش دیده اند. بنابراین بسته به تصمیمات، الگوریتم‌ها خود را برای تولید خروجی موفقیت آمیز و یا شکست آموزش میدهند. در نهایت این الگوریتم دارای تجربه‌ای است که قادر به ارائه پیش بینی‌های خوب در یک موضوع خواهد بود.

یادگیری ماشین در دو مدل (طبقه بندی^{۱۲}) و (خوشه بندی^{۱۳}) است که به مطالعه‌ی آن می‌پردازیم

الگوریتم‌های یادگیری ماشین

یکی از مهم‌ترین شاخه‌های یادگیری ماشین است که به این صورت کار می‌کند که یک ماتریس را به عنوان ورودی می‌گیرد و با یادگیری فیلد‌های قبلی جدول و بررسی برچسب آن در صورتی که یک داده‌ی جدید وارد جدول شود می‌تواند روی آن برچسب بزند.

بگذارید با یک مثال بهتر به عمق درک مطلب بپردازیم. فرض کنید که ما اطلاعات وامی مشتریان یک بانک را داریم که نشان می‌دهد یک مشتری می‌تواند وام گرفته را پس دهد یا نه جدول ۱. حالا فرض کنید که یک مشتری جدید به لیست اضافه شود، برچسب آن چه خواهد بود؟

¹² Classification

¹³ Clustering

¹⁴ K-Nearest Neighbors

⁹ Supervised Algorithms

¹⁰ Unsupervised Algorithms

¹¹ Reinforcement Algorithms

الگوریتم LR^{۱۵}

یکی دیگر از الگوریتم های یادگیری ماشین است که می تواند بین دو طبقه از داده های تمایز قائل شود برای درک بهتر مطلب مثال زیر را در نظر بگیرید.

می خواهیم به کمک یک نرم افزار که از الگوریتم Logistic Regression استفاده می کند تشخیص دهیم که یک پیامک جدید تبلیغاتی است یا عادی و معیار ما برای این کار تعداد کاراکتر های آن پیامک است، که این یعنی دو ستون داریم. یکی ستون برچسب که مشخص می کند پیامک تبلیغاتی است یا خیر و دیگری که تعداد کاراکتر های پیامک است. برای این کار از فرمول معادله خط فرمول ۱،۲ که در دبیرستان یاد گرفتیم استفاده می کنیم.

$$x = \theta \cdot \text{count} + b$$

فرمول ۱

سپس به سراغ تابع تعیین احتمال سیگموند فرمول ۲ می رویم که احتمال وقوع به رخداد را برای ما مشخص می کند و در آن متغیر (X) از فرمول بالا به دست می آید.

$$\text{sigmond } x = \frac{1}{1 + e^{-x}}$$

فرمول ۲

با استفاده این دو تابع می توانیم خطوطی موربی رسم کنیم تا بر روی داده های ما به طورت مناسب قرار گیرند.

الگوریتم درخت تصمیم

در اصول کامپیوتر درخت ها جایگاه ویژه ای دارند و به دلیل باینری بودن آنها می توانیم در همه ی مباحث رایانه ای از آنها کمک بگیریم. در اصول طبقه بندی یادگیری ماشین هم ما از الگوریتم درخت تصمیم استفاده می کنیم تا بتوانیم لیبل داده ی جدید را محاسبه و پیشبینی کنیم. لطفا جدول زیر جدول ۴ را در نظر بگیرید که به دانشجویانی مربوط است که می خواهیم بدانیم دکترا قبول شده اند یا خیر و ما به کمک الگوریتم های یادگیری ماشین و سوابق دانشجویان گذشته می خواهیم به این نتیجه برسیم.

با یک مثال علمی تر می توان بهتر موضوع را درک کرد. جدول ۲ زیر را فرض کنید که می خواهیم در آن متوجه شویم که یک مشتری می تواند وام را پس دهد یا خیر؟

شماره مشتری	خانه از خود دارد؟ (بعد ۱)	چند سال مشتری است؟ (بعد ۲)	وام را پس داده؟ (برچسب)
۱	۱	۳	بله
۲	۰	۱	خیر
۳	۱	۱/۵	بله
۴	۱	۵	بله
۵	۰	۲	خیر

جدول ۲- الگوریتم KNN

حالا فرض کنید که مشتری جدید داریم که خانه دارد و ۲/۵ سال است که مشتری ما است. می خواهیم ببینیم که این دو نفر می توانند وام را پس دهند یا خیر. باید فاصله ی منهن را یک به یک با مشتریان قدیم حساب کنیم. یعنی بعد ۱ را از بعد ۱ کم کنیم و سپس بعد ۲ را از بعد ۲ کم کنیم و حاصل آن را با هم جمع کنیم و فقط یک نکته که برای اینکه عدد ما منفی نشود باید برای آنها قدر مطلق قرار دهیم که به صورت جدول ۳ زیر است که برای مشتری جدید اول حساب کرده ایم.

شماره مشتری	خانه از خود دارد؟	چند سال مشتری است؟	وام را پس داده؟	مشتری اول
۱	۱	۳	بله	$ 1-1 + 2/5-3 = 0.5$
۲	۰	۱	خیر	$ 1-0 + 2/5-1 = 2/5$
۳	۱	۱/۵	بله	$ 1-1 + 2/5-1/5 = 1$
۴	۱	۵	بله	$ 1-1 + 2/5-5 = 2/5$
۵	۰	۲	خیر	$ 1-0 + 2/5-2 = 1/5$

جدول ۳- الگوریتم KNN

با فرض اینکه مقدار $K=3$ است ما ۳ نفر از نزدیکترین افراد لیست را انتخاب می کنیم و میدانیم که ۲ نفر از آنها وام را پس داده اند و یکی نتوانسه وام را پس دهد. پس احتمالا مشتری جدید هم می تواند وام را پس دهد. سپس به مشتری جدید برچسب بله را می زنیم

¹⁵ Logistic Regression

الگوریتم جنگل تصادفی^{۱۶}

یکی از الگوریتم‌های است که به تکنیک یادگیری نظارت شده تعلق دارد. به این صورت کار می‌کند که همانطور از نامش پیدا است از چند درخت تصمیم برای نتیجه‌گیری استفاده می‌کند و مبنای کار آن مانند درخت تصمیم بر اساس درست یا غلط بودن یا بله و خیر است که در نهایت بعد از پاسخ دادن به تک تک سوالات جنگل به نتیجه می‌رسد.

الگوریتم XGBoost

این الگوریتم بر اساس بهبود و تقویت درخت تصمیم ایجاد شده است و یک کتابخانه‌ی متن باز است که به سرعت بالا معروف است و قابلیت تقویت گرادیان آن در بین دیگر الگوریتم‌ها بی نظیر و فوق العاده است. فرمول نهایی این الگوریتم به شرح زیر است.

$$f(x) = f_{(m)}(x) = \sum_{m=0}^M f_m(x)$$

فرمول ۳

الگوریتم ماشین بردار پشتیبان^{۱۷}

مبنای کار ماشین بردار پشتیبان دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌کنیم ابر صفحه‌ای را انتخاب کنیم که حاشیه اطمینان بیشتری داشته باشد. وظیفه‌ی آن پیدا کردن خط بهینه با استفاده از تابع phi است و می‌توانیم برای مسائل با پیچیدگی بالا از آن استفاده کنیم.

۲-۳- کارهای پیشین

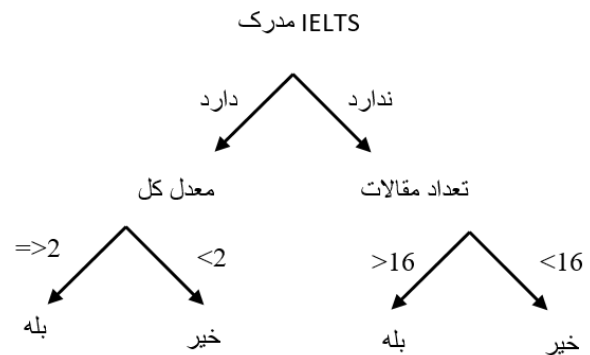
در مقاله‌ی [۱] روش‌های مختلفی بررسی شده است و روی روش یادگیری ماشین تمرکز بیشتری گذاشته شده است و یک چهارچوب ترکیبی از دستگاه بردار پشتیبان برای پیش‌بینی موثر پیشنهاد می‌کند و همچنین یک نظریه‌ی دقیق در مورد SVM با وزن برای طبقه‌بندی داده‌ها ارائه می‌دهد. نتایج آزمایش بر روی بازار سهام چین و دو شاخصه‌ی بورس شانگهای و شنژن تست شده که با مدل پیشنهادی در این مقاله می‌توان به ترتیب در کوتاه مدت، میان مدت و بلند مدت به قابلیت پیش‌بینی بهتری برای شاخص‌های بورس شانگهای و شنژن دست پیدا کرد.

[۲] در مقاله‌ی خود بر سری‌های زمانی تمرکز دارند و پس از توضیحاتی که راجع به سری‌های زمانی می‌دهند، غیر خطی بودن سری‌های زمانی را یک ویژگی بارز می‌دانند. این مقاله با

ردیف	معدل کل	تعداد مقالات	مدرک IELTS	سنوات تحصیلی	دکتری
۱	۱۹.۵	۳	۱	۳	بله
۲	۱۶.۵	۰	۱	۴	خیر
۳	۱۵	۰	۰	۳	خیر
۴	۱۷	۲	۱	۲.۵	بله
۵	۱۸.۵	۲	۰	۲.۵	بله
۶	۱۵.۵	۱	۱	۲.۵	خیر
۷	۱۹	۳	۱	۳	بله

جدول ۴- الگوریتم درخت تصمیم

سپس درخت تصمیم با استفاده از فرمول‌های ریاضی (Entropy) و (Gain) یک مدل پیشنهادی ارائه می‌دهد که می‌توان در شکل زیر شکل ۱ مشاهده کرد که در صورت ورود داده‌ی جدید سریع در ابتدا و راس این درخت قرار می‌گیرد و در نهایت مشخص می‌شود که آیا این دانشجوی جدید می‌تواند دکترا قبول شود یا خیر؟



شکل ۱- الگوریتم درخت تصمیم

که تمامی مراحل کاوش آن به وسیله‌ی دستور (If) انجام می‌شود و می‌توان به راحتی دانشجوی جدید را مورد بررسی قرار داد و به این نتیجه رسید که آیا دکترا قبول می‌شود و این طبقه‌بندی کارکرد الگوریتم درخت تصمیم است.

الگوریتم Voting

یکی دیگر از الگوریتم‌های یادگیری ماشین است که به الگوریتم تجمیعی معروف است که در آن ابتدا داده‌ها را در دو دسته قرار می‌دهد و سپس با ورود داده‌ی جدید آن را با داده‌های گذشته مقایسه می‌کند و در یکی از خوشه‌ها دسته‌بندی قبلی قرار می‌دهد

¹⁶ Random Forest

¹⁷ Support Vector Machine

(MV) پیشنهاد می کند که به طور خاص دو مرحله در این مدل درگیر هستند، پیشبینی سهام و انتخاب پرتفوی در مرحله اول یک مدل ترکیبی از تقویت گرادیان با یک الگوریتم کرم شب تاب برای پیشبینی قیمت سهام پیشنهاد شده است. در مرحله دوم با میانگین واریانس بازده بلقوه ی بالاتر انتخاب می شود. منبع این مقاله بازار بورس شانگهای است و معیار های بدست آمده از نظر بازده و ریسک برتری دارند.

در مقاله ی [۷] به بررسی دقیق ریسک سیستم های مالی با روش یادگیری ماشین و پیشنهاد جهت گیری برای کار های آینده می پردازد و بعد از مطالعه و تحقیق بر روی مقاله های دیگر به روش های طبقه بندی اشیاء تحقیق و سپس طبقه بندی روش ها به سراغ روش های یادگیری ماشین برای ریسک مالی سیستمی می رود. که در ادامه می خواهد روش تحلیل احساسات در بازار های مالی را بررسی کند که برای این کار از الگوریتم های یادگیری ماشین کلاسیک استفاده می کنند. در نهایت به این نتیجه رسیده اند که پاسخ به ریسک سیستمی مالی باید سیاست های تنظیم دقیق تری را با استفاده از روش های هوشمند مبتنی بر داده های بزرگ مالی در آینده ایجاد کند.

مقاله ی [۸] که چهار گروه بازار شامل انواع مالی، نفت، کانی های غیر فلزی و فلزات اساسی از بورس اوراق بهادار تهران را برای ارزیابی تجربی انتخاب کرده و همچنین از ۹ الگوریتم یادگیری ماشین شامل درخت تصمیم، تقویت تطبیقی، حافظه ی کوتاه مدت تطبیقی و... استفاده می کند که نتیجه با دو رویکرد قابل بیان است که در کل تمامی الگوریتم ها به خوبی پیشبینی می کنند زیرا با مقادیر پیوسته که رویکرد اول نیز تا ۶۷ درصد آموزش داده می شوند و با رویکرد باینری یا دوم نیز تا ۸۳ درصد آموزش داده می شوند که باید مقادیر غیر ثابت در رویکرد اول را به مقادیر قطعی روند در رویکرد دوم تبدیل کند.

[۹] یک مطالعه ی تحقیقاتی است که در آن روش های سنتی مانند میانگین متحرک یکپارچه اتو رگرسیون را با روش ها یادگیری ماشین اخیر در زمینه های مالی را مقایسه می کند و به این نتیجه می رسد که این روش ها بهتر از روش های سنتی عمل می کنند و همچنین یک راه حل موثر را برای به دست آوردن نتایج رضایت بخش در مشکلات مالی را با روش هایی مانند AR به دست آورده است.

در مقاله ی [۱۰] برخی سیستم های معاملاتی مانند تجارت جفت، میانگین متحرک، HFT و... را بررسی می کنند و آنها را در ۳ دسته ی تجزیه و تحلیل فنی، تحلیل های متنی و معاملات

هدف ارائه ی روشی برای پیشبینی حداکثر و حداقل قیمت سهام سه شرکت توزیع نیروی برق برزیل، که در سهام بورس سائوپائولو میباشد نوشته شده است. یکی از مهمترین و جدید ترین موضوع این مقاله نسبت به مقاله های گذشته پیشبینی محدوده قیمت بسته شدن سهام شرکت توزیع نیروی برق است که در نتیجه سرمایه گذاران بتوانند ارزش معاملاتی سهام خود را تعیین کنند. پیشبینی واقعی توسط ANN انجام می شود که عملکرد آن با استفاده از MAE، MAPE و RMSE ارزیابی می شود و کسب نتایج فقط به ANN بستگی دارد.

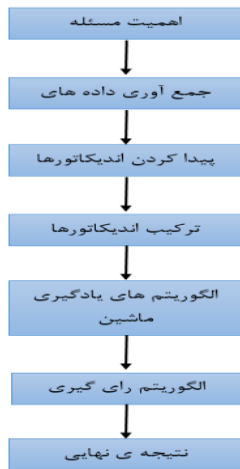
در مقاله ی [۳] از آنجایی که نمی توان GA و HS را با ANN یکپارچه کرد، لذا ما فقط قصد بهینه سازی معماری ANN را داریم و برای بدست آوردن شاخص هایی که بیشترین تاثیر را بر پیشبینی دارند و برای انتخاب متغیر ورودی به کار می روند تلاش می کنیم. در این مقاله تحلیل مولفه های اصلی، با استفاده از 20 متغیر حسابداری پیشبینی پذیری قیمت ها در بازار بورس تهران را بررسی می کند.

در مقاله ی [۴] یک مدل داده کاوی بر اساس متد روزانه ی ETF و بر اساس 60 ویژگی مالی و اقتصادی ارائه می دهد. برای این کار از 3 تکنیک (PCA) و (FRPCA) و (KPCA) استفاده می کند که منجر به ساده سازی و تنظیم مجدد ساختار داده اصلی در کل مجموعه داده می شود، در نهایت به دوازده مجموعه داده جدید می رسند که از هر یک از 3 روش بالا به دست می آید. در نهایت شبکه عصبی مصنوعی را با 36 مجموعه ی داده ی جدید استفاده می شود

در مقاله ی [۵] به غیر خطی بودن و نواسانات بالای سری های زمانی مالی که پیشبینی بازار سهام را دشوار کرده اشاره می کند. سپس چند روش را بررسی می کند روش هایی از جمله مدل حافظه ی کوتاه مدت و بلند مدت (LSTM)، شبکه های عصبی کانولوشنال (CNN) و الگوریتم های تجزیه ی حالت تجربی مجموعه ی کامل (EMD) و (CEEMD) را با هم ترکیب می کند به روش های جدیدی مثل (CEEMD-CNN) و (LSTM) و (EMD-CNN-LSTM) را پیشنهاد می کند که برای پیشبینی یک گام جلو تر عمل می کند و نتیجه هم همین را نشان می دهد که ترکیب الگوریتم های فوق بهبودی در پیشبینی ایجاد کرده.

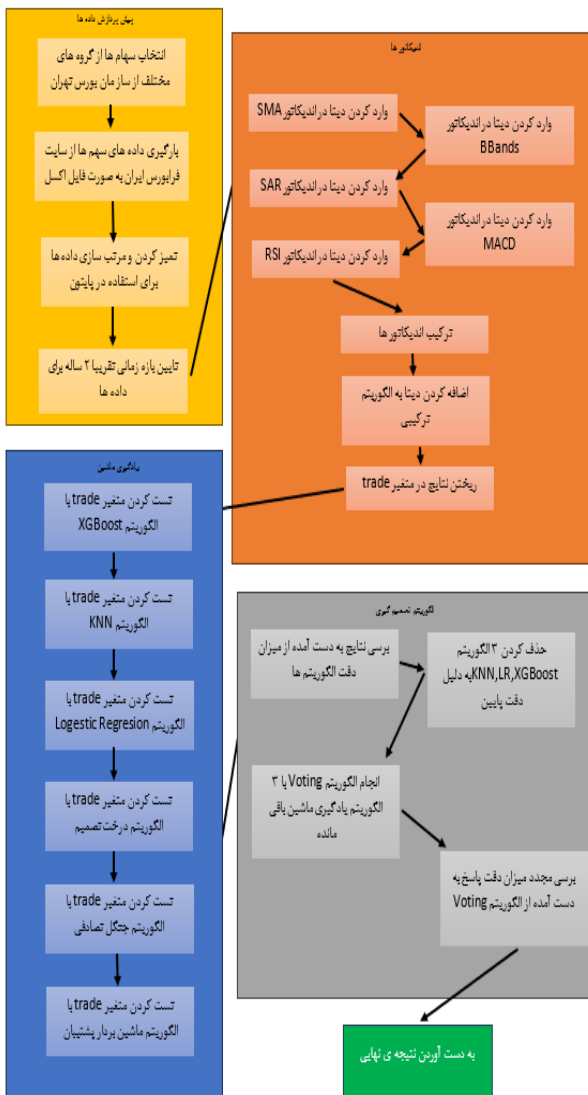
در مقاله ی [۶] به این موضوع اشاره می کند که نمی توان یک مدل واحد برای پیشبینی بازار سهام در نظر گرفت. این مدل جدیدی را مبتنی بر یادگیری ماشین و مدل میانگین واریانس

۳- روش پیشنهادی



شکل ۲ قدم های متودولوژی

مدلی از فرایند کار



شکل ۳ قدم های متودولوژی

با فرکانس بالا قرار می دهند و پس از مقایسه به این نتیجه می رسند که روش آماری PST، یادگیری ماشین و تحلیل متنی از بقیه ی روش ها بهتر بودند. اما در ادامه به این نکته اشاره کردند که این یک مبحث بسیار پیچیده است و نیاز به تحقیق و بررسی بیشتر دارد و هنوز در مراحل اولیه است ولی تا به این مرحله پیشرفت خوبی حاصل شده است.

در مقاله ی [۱۱] بر روی سهام های آمازون، آپل و گوگل از الگوریتم های یادگیری ماشین مانند رگرسیون خطی، میانگین متحرک سه ماهه، هموار سازی نمایی و پیشبینی سری زمانی استفاده کرد تا بتواند میزان درستی و دقت این الگوریتم ها را بررسی کند. در نهایت بعد از بررسی تمامی الگوریتم های یادگیری ماشین، به این نتیجه رسیده اند که نتایج پیشبینی هموار سازی نمایی خطای کمتر و دقت بیشتری را نسبت به بقیه دارد.

در مقاله ی [۱۲] بر روی داده های بزرگ کار می کنند و از الگوریتم های یادگیری ماشین قدرتمند مانند شبکه عصبی عمیق برای انجام تحلیل ها استفاده می کنند. سپس یک فرایند جامع تجزیه و تحلیل داده های بزرگ را انجام می دهد و با شبکه عصبی مصنوعی سنتی مقایسه می کند که مبنای ۳۰۰ روزه را در نظر می گیرد و در نتیجه بهبود و دقت بالایی در DNN را به دست می آورد و برای آینده نیز می خواهد تعداد روز ها را به ۵۰۰ برساند.

مقاله ی [۱۳] بر روی داده های ۳ ارز مجازی اصلی یعنی بیت کوین، اتریوم و لایت کوین مانور می دهد و از سود آور ترین الگوریتم های روز دنیا مانند مدل های خطی، ماشین بردار پشتیبان و جنگل تصادفی برای بررسی استفاده می کند. نتیجه نشان داد که دقت پیشبینی در مدل و ارز های مختلف متفاوت است و الگوی خاصی وجود ندارد اما به طور کل به این نتیجه رسیدند که مدل های فردی نسبت به سایر مدل ها از دقت کمتری برخوردار است.

در مقاله ی [۱۴] بر اساس قابلیت ردیابی بلاکچین و همچنین استفاده از یادگیری ماشین یک مکانیزم برای پاسخ به چالش بین خریدار داده و مالک داده را ایجاد می کند و چالش های این مسیر را بررسی و به آنها پاسخ می دهد. در واقع در این مقاله یک مدل قرار داد هوشمند را به وسیله ی یادگیری ماشین دنبال می کند تا داده هایی بین خریدار و مالک انجام شود و در آینده نیز بر روی جلوگیری از فروش مجدد خریدار به یک فروشنده را به عنوان پروژه ی بعدی انجام خواهد داد.

اهمیت مسئله

در بازار های مالی معامله ها ریسک بسیار بالایی دارند و تاثیر مستقیم بر روی سرمایه ی اصلی معامله کننده می گذارد، به این صورت که یک معامله ی خوب می تواند سرمایه ی یک فرد را چندین برابر کند اما یک معامله ی بد امکان این را دارد که کل سرمایه ی فرد را بسوزاند پس سرمایه گزاران بازار های مالی باید بسیار دقت کنند و بفمنند که بهترین زمان برای خرید و یا فروش چه زمانی است.

برای به دست آوردن بهترین نقطه ی خرید یا فروش باید رفتار گذشته ی سهم را بر اساس تحلیل تکنیکال بررسی کنیم و بتوانیم حدس بزنیم که الان بهترین کار ممکن چیست. حالا فرض کنید ما بتوانیم به کمک الگوریتم های فرا ابتکاری و یادگیری ماشین رفتار گذشته ی سهم را بررسی و به رایانه یاد دهیم که این سهم چگونه رفتاری در بازار داشته و یادگیری ماشین بتواند تحلیل کند و به ما بگوید که بهترین نقطه برای خرید و فروش کدام است. این یعنی ما به کمک رایانه ها می توانیم تا حد بسیار زیادی خطای انسانی را کاهش دهیم.

ما قصد داریم که ابتدا سهم هایی را از بازار بورس تهران انتخاب کنیم و سپس با اندیکاتور های مختلف مورد بررسی و نتیجه و قرار دهیم و بعد از اینکه یک به یک با اندیکاتور های مختلف بررسی کردیم اندیکاتور های منتخب را با هم ترکیب کنیم تا یک نتیجه ی واحد به دست آید. نتیجه ی به دست آمده را با الگوریتم های یادگیری ماشین مقایسه می کنیم و بر اساس دقت به دست آمده ۳ سیستم یادگیری ماشین را انتخاب می کنیم و در نهایت زمانی که رای اکثریت صادر شد نتیجه ی نهایی ما خواهد شد.

جمع آوری داده ها

داده هایی که ما در این مسئله مورد استفاده قرار می دهیم از سازمان بورس ایران میباشد، اما به دلیل تعداد زیاد سهمی که وجود دارد ما آنها را محدود می کنیم و تعدادی از آنها را بررسی می کنیم و در سال های مشخصی دیتای آن ها را انتخاب می کنیم. برای این کار سهم هایی را از گروه های مختلف انتخاب می کنیم سهم هایی که انتخاب کردیم به شرح زیر است

- (۱) دتماد از گروه مواد و محصولات دارویی
- (۲) های وب از گروه اطلاعات و ارتباطات

- (۳) هرمز از گروه فلزات اساسی
- (۴) خگستر از گروه خودرو و ساخت قطعات
- (۵) کیمیا از گروه فلزات اساسی
- (۶) قمر از گروه قند و شکر
- (۷) ناخت از گروه انبوه سازی، املاک و مستغلات
- (۸) سمزن از گروه سیمان، آهک و گچ
- (۹) شبریز از گروه فراورده های نفتی
- (۱۰) وبصادر از گروه بانک ها و موسسات اعتباری
- (۱۱) وخارزم از گروه سرمایه گذاری ها
- (۱۲) زدشت از گروه زراعت و خدمات وابسته

همان طور که گفته شد از گروه های مختلف سهم هایی را انتخاب کردیم و سپس نیاز داریم که دیتای گذشته ی آنها را به صورت آفلاین^{۱۸} داشته باشیم چراکه الگوریتم ما یک الگوریتم ایستا است و ما برای دقت بالا تر باید دیتای آفلاین گذشته را داشته باشیم. برای این کار از نرم افزار (TseClient) که یک برنامه ی وابسته به مدیریت فناوری بورس تهران است استفاده می کنیم. داده های بورسی را در فایل اکسل^{۱۹} با پسوند (CSV). از این برنامه استخراج می کنیم. داده هایی که ما از بورس استخراج می کنیم به صورت کاملا خام هستند و باید برای استفاده در برنامه ی پایتون کمی تغییر کنند. برای این کار باید جدا کننده ی ستون های آنها کاما(,) باشد و بر اساس تاریخ که به صورت (ماه/ روز /سال) مرتبط شوند و سپس فایل نهایی به صورت شکل ۳ زیر است.

	A	B	C	D	E	F	G	H	I
1	Date	Open	High	Low	Close	Vol	<OPENINT>	<OPENINT>	<OPENINT>
2	1/27/2003	72500	72870	72351	72351	518713	3.7729E+10	638	72500
3	1/28/2003	72351	72358	72351	72351	27929	2020699599	291	72351
4	2/2/2003	71628	71631	71628	71628	35575	2548174992	67	72351
5	2/5/2003	71900	71900	71745	71745	3575	256670782	16	71628
6	2/9/2003	71028	71028	71028	71028	1102	78272856	5	71745
7	2/10/2003	70318	70318	70318	70318	1000	70318000	1	71028
8	2/15/2003	69615	69615	69615	69615	500	34807500	1	70318
9	2/16/2003	68919	68919	68919	68919	2000	137838000	3	69615
10	2/17/2003	68230	68250	68230	68240	36341	2479565729	79	68919
11	2/18/2003	68262	68310	68262	68310	701	47877801	9	68240
12	2/19/2003	68310	68311	68310	68310	1773	121113908	13	68310
13	2/22/2003	68309	68310	68309	68310	3221	220026110	5	68310
14	2/23/2003	68287	68287	68287	68287	140	9560180	3	68310
15	2/24/2003	67605	67605	67605	67605	500	33802500	3	68287
16	2/26/2003	66929	68281	66929	68281	24413	1642269166	53	67605
17	3/1/2003	68281	68963	68281	68963	1467	101105977	9	68281
18	3/2/2003	69652	69652	69652	69652	1000	69652000	4	68963
19	3/3/2003	70348	70348	70348	70348	1000	70348000	2	69652
20	3/5/2003	71051	71051	71051	71051	2016	143238816	2	70348
21	3/8/2003	71761	71761	71761	71761	2000	143522000	5	71051
22	3/9/2003	72478	72478	72476	72478	38925	2821202270	46	71761

شکل ۴ قدم های متودولوژی

همانطور که در شکل بالا می بینید داده ها مرتب شده و در سطر هایی و ستون های مختلفی قرار گرفته اند که سطر ها بر اساس

¹⁹ Excel

¹⁸ Offline

دو سال است که تغییرات اخیر کشور که بر اساس تورم رخ داده است را شامل نمی‌شود و بازار بورس در این بازه ی تاریخی بدون تنشج است.

اندیکاتور هایی که ما در این برنامه استفاده کردیم به شرح زیر است:

- **SMA**: در این اندیکاتور ما باید از دو بازه ی زمانی ۲۰ روزه و ۵۰ روزه استفاده کردیم که بازه ی زمانی ۲۰ روزه را متغیر ($n1$) و بازه ی زمانی ۵۰ روزه را متغیر ($n2$) نام گذاری کرده ایم سپس دستور شرطی گذاشته ایم که اگر ($n1 < n2$) سیگنال خرید بدهد و اگر ($n1 > n2$) بود سیگنال فروش بدهد.
- **BBands**: در این اندیکاتور که ما دو خط بالایی و پایینی داریم در ابتدا خط پایینی را (**Ibbans**) و خط بالایی را (**ubbands**) می‌نامیم. حالا دستور شرطی را می‌گذاریم که به این صورت سیگنال خرید می‌دهد که اگر (**Ibbands**) بزرگ تر از خط میانی شد سیگنال خرید به ما دهد.
- **SAR**: کرانه ی بالا و کرانه ی پایین را در نظر می‌گیریم و دستور شرطی این است که کرانه ی پایین اگر از میانگین متحرک بزرگتر شد سیگنال خرید و اگر از کرانه ی بالا از میانگین متحرک کوچک تر شد سیگنال فروش دهد.
- **MACD**: در این اندیکاتور ما ابتدا به وسیله ی فرمول مکدی میانگین دو خط را به دست می‌آوریم و سپس اگر میانگین مکدی از عدد ۰ کوچکتر شد سیگنال فروش دهد و اگر میانگین مکدی از عدد ۰ بزرگتر شد سیگنال خرید بدهد.

RSI: در ابتدا سیگنال **RSI** را فراخوانی می‌کنیم و سپس شرط عدد ۳۰ را می‌گذاریم که اگر **RSI** از ۳۰ بزرگتر شد سیگنال فروش می‌دهد و اگر کوچکتر بود سیگنال فروش می‌دهد.

ترکیب اندیکاتور ها

بعد از اینکه تک تک اندیکاتور ها را به همراه دیتا ست ها تست کردیم و سیگنال های خرید و فروش گرفتیم، این اندیکاتور ها را با هم ترکیب می‌کنیم چرا که زمانی که می‌خواهیم دقت بالاتری به دست آوریم و همین مسأله زمانی ثابت می‌شود که در ابتدا سیگنال خرید و فروش تک تک اندیکاتور ها را از قبل

روز ها مرتب شده اند و ستون هایی وجود دارد که یک به یک بررسی می‌کنیم:

- **Date**: این ستون مربوط به تاریخ و روز است که یک سهم در آن چه رفتاری دارد
- **Open**: قیمت آغازین یک سهم در آن روز
- **High**: بالاترین قیمت روز
- **Low**: پایین ترین قیمت روز
- **Close**: قیمت پایانی روز

این داده ی اولیه ی ما است که برای استفاده در نرم افزار و زبان پایتون مناسب است و ما تمامی سهم های انتخابی را در پوشه ای ذخیره می‌کنیم و بعد به سراغ اندیکاتور ها می‌رویم تا با استفاده از آن ها و این جدول ها سیگنال خرید و فروش بگیریم.

پیدا کردن اندیکاتور ها

در بازار های مالی معمولاً سرمایه گزاران و افرادی که معامله می‌کنند برای بررسی رفتار یک سهم از اندیکاتور های مختلفی استفاده می‌کنند که همین موضوع باعث شد که ما اندیکاتور های مختلفی را مورد بررسی و آزمایش قرار دهیم و تک تک دیتای سهم هایی که انتخاب کردیم را با اندیکاتور هایی که انتخاب کردیم مورد بررسی و تحلیل قرار می‌دهیم.

روشی که ما برای انجام این مقاله انتخاب کردیم تحلیل تکنیکال است. زیرا تحلیل تکنیکال علمی برای پیشبینی رفتار نمودار ها در آینده با استفاده از اطلاعات گذشته یک سهم است و سرمایه گذاران می‌توانند با استفاده از داده های گذشته روند و الگوریتم یک نماد را شناسایی کنند و آینده را بر اساس آن الگوریتم پیشبینی کنند.

اندیکاتور ها را در بستر گوگل به زبان برنامه نویسی پایتون پیاده سازی می‌کنیم و سپس دیتای سهم ها را روی آن پیاده سازی می‌کنیم تا چگونگی معامله کردن و میزان سود را مشاهده و مورد بررسی قرار دهیم و ما سیگنالی مبنی بر این که بخریم و یا بفروشیم را دریافت کنیم.

اما یک نکته ی بسیار مهم این است که به دلیل نوسانات بسیار زیاد چند سال اخیر در بورس ایران ما نمی‌توانیم از داده های بورسی چند سال اخیر استفاده کنیم و اگر هم بخواهیم از تمام دیتا استفاده کنیم، برنامه ی بسیار سنگین می‌شود و در اجراها دچار مشکل می‌شویم. پس ما برای حل این مشکل اطلاعات سهم ها را در بازه ی تاریخی مشخصی بررسی می‌کنیم و بازه ی تاریخی ای که ما برای این کار برگزیده ایم از تاریخ (۲۰۱۷/۰۵/۰۱) تا تاریخ (۲۰۱۹/۰۳/۰۱) به مدت تقریباً

در انتهای کار ما از یک الگوریتم Voting استفاده می کنیم که این به معنای رای گیری است یعنی به این صورت که بعد از اینکه همه ی الگوریتم های یادگیری ماشینمان برچسب های خود را به جدول دیتا اضافه کردند به دلیل نحوه ی محاسبه و نتایج مختلف ما چند برچسب متفاوت داریم که می خواهیم در این مرحله آنها را یکی کنیم و به یک برچسب واحد برسیم که برچسب (final) نام دارد و به این صورت کار می کند که هر گاه دو الگوریتم یادگیری برچسبی یکسان داشته اند برچسب نهایی ما آن برچسب است ولی چون ما ۳ تا الگوریتم یادگیری ماشین داشتیم و همچنین ۳ مدل برچسب پس اگر هر کدام از آنها برچسب متغیری داشتند برچسب نهایی ما برچسب خنثی است.

۴- آنالیز و نتایج تحقیق

در این فصل ما به تجزیه و تحلیل پیرامون جداول به دست آمده و همچنین نتایج که در هر مرحله به دست آوردیم می پردازیم و در نهایت در فصل بعد بر اساس تجزیه و تحلیل های این فصل نتیجه نهایی را اعلام می کنیم. برای این کار ما باید در ابتدا نتایج به دست آمده از اندیکاتور ها را به دست آوریم و سپس با نتایج به دست آمده از تجمیع اندیکاتور ها مقایسه کنیم و بعد به الگوریتم های یادگیری ماشین بپردازیم.

جداول اندیکاتور ها

در ابتدا جدول و نتایج به دست آمده از تک تک اندیکاتور ها را مورد بررسی قرار می دهیم در جدول ۵ ما نتایج به دست آمده از الگوریتم (SMA) را می بینیم.

نام سهم	تعداد معامله	سود کل	بیشترین سود
دتماد	305	41.31	22.6
های وب	71	35.21	7.59
هرمز	159	55.97	210.5
خ گستر	209	47.84	59.94
کیمیا	137	55.47	61.73
قمر	206	51.94	100.68
ثاخذ	266	48.49	49.22
سمازن	346	14.16	8.8
شیریز	55	52.72	246
وبصادر	56	19.64	8.34
وخارزم	194	47.93	8.28
زدشت	0	0	0

به دست آورده باشیم و سپس با یک دستور شرطی سیگنال خرید و فروش می دهیم.

دستور شرطی ای که برای ترکیب اندیکاتور هایمان استفاده می کنیم به این صورت است که اگر $(MACD < 0)$ و $(lbands)$ از میانگین خط خودش بزرگتر باشد، یا $(RSI < 30)$ و $(sma1 < sma2)$ باشد سیگنال خرید بدهد برای سیگنال فروش هم شرط اینکه اگر (sar) از میانگین خودش کوچکتر باشد و $(ubbands)$ از میانگین خودش بزرگتر باشد را دارد.

برای استفاده از سیگنال ها در سیستم یادگیری ماشین ما باید به سیگنال ها برچسب بزنیم که در این مرحله این کار را انجام می دهیم و زمانی که ترکیب اندیکاتور هایمان سیگنال فروش داد ما برچسب (۰) و اگر سیگنال فروش را دریافت کردیم برچسب (۱) و اگر سیگنالی دریافت نکردیم برچسب (۲) را بر روی جدول بزنیم.

الگوریتم های یادگیری ماشین

ما در این تحقیق از ۳ الگوریتم یادگیری ماشین استفاده می کنیم و بعد میزان دقت هر کدام را مورد بررسی قرار می دهیم تا بتوانیم تشخیص دهیم که الگوریتم ما چند درصد صحت و درستی دارد. برای این کار باید دیتایی را که از قبل با الگوریتم ترکیبی ای که مرحله ی قبل ایجاد کردیم برچسب گذاری کرده باشیم و سپس یک به یک با الگوریتم های یادگیری ماشینی که در قبل معرفی کردیم امتحان کنیم و خروجی ما در این مرحله میزان دقت و درستی الگوریتم های یادگیری ماشین را می سنجیم که به واسطه میزان درصد درستی است. الگوریتم های طبقه بندی ای که ما استفاده کردیم هر کدام در ابتدا سیگنال خرید و فروش را در جدول دیتا وارد می کند و سپس میزان دقت در درستی را نسبت به سیگنالی که از ترکیب اندیکاتور ها به دست آوردیم بررسی می کند و به ما خروجی می دهد.

ما از الگوریتم های یادگیری زیر استفاده کردیم:

- KNN
- Logistic Regression
- درخت تصمیم
- جنگل تصادفی
- ماشین بردار پشتیبان
- XGBoost

الگوریتم های رای گیری

وبصادر	83	30.12	8.8
وخارزم	210	44.76	29.97
زدشت	0	0	0

جدول 8 نتایج اندیکاتور MACD

نام سهم	تعداد معامله	سود کل	بیشترین سود
دتماد	3	33.33	8.04
های وب	1	-1	-1
هرمز	3	66.66	4
خ گستر	1	100	3.81
کیمیا	0	0	0
قمرو	0	0	0
ثاخذ	1	0	-0.7
سمازن	7	0	-0.3
شبریز	2	100	264
وبصادر	0	0	0
وخارزم	1	0	-0.6
زدشت	0	0	0

جدول 9 نتایج اندیکاتور Bbands

همانطور که در جداول بالا مشاهده می‌کنید زمانی که از یکی از اندیکاتورها استفاده می‌کنیم نتایج مختلفی به دست می‌آید که بعضی از آنها خیلی جالب نیست. اما ما با ترکیب این اندیکاتورها به نتایج جدید و متفاوتی دست پیدا کردیم که در جدول ۱۰ می‌توانیم نتایج را مشاهده کنیم و به میزان سودی که از هر سهمی به دست آوریم و تعداد معامله‌های انجام شده دقت کنید که درصد بالایی دارد. (نکته: الگوریتم تجمیع اندیکاتورها را MIX نام‌گذاری کرده ایم).

نام سهم	تعداد معامله	سود کل	بیشترین سود
دتماد	55	58.18	6.2
های وب	21	33.33	6
هرمز	6	66.66	3.41
خ گستر	24	54.16	11.03
کیمیا	20	60	18.47
قمرو	11	45.45	3.75
ثاخذ	104	39.42	20.22
سمازن	287	6.96	7.42
شبریز	5	40	3.58
وبصادر	49	16.32	4.99
وخارزم	39	30.76	9.48
زدشت	0	0	0

جدول 5 نتایج اندیکاتور SMA

نام سهم	تعداد معامله	سود کل	بیشترین سود
دتماد	111	30.63	22.6
های وب	58	46.55	20.33
هرمز	121	56.19	210
خ گستر	192	47.91	19.5
کیمیا	75	62.66	61.7
قمرو	161	52.79	99.32
ثاخذ	136	45.58	49.22
سمازن	67	35.82	6.71
شبریز	116	52.58	245
وبصادر	99	50.5	35.86
وخارزم	182	46.15	23.2
زدشت	3	66	2.8

جدول 6 نتایج اندیکاتور SAR

نام سهم	تعداد معامله	سود کل	بیشترین سود
دتماد	231	30.3	22
های وب	21	38.09	6
هرمز	13	69.23	9.43
خ گستر	26	50	6.5
کیمیا	20	55	15.63
قمرو	11	45.45	5.5
ثاخذ	134	41	20.22
سمازن	326	6.7	7.4
شبریز	24	58.33	4.58
وبصادر	251	82.86	15.72
وخارزم	63	36.5	6.4
زدشت	74	100	200

جدول 7 نتایج اندیکاتور RSI

نام سهم	تعداد معامله	سود کل	بیشترین سود
دتماد	317	37.22	22
های وب	65	36.99	6.72
هرمز	165	54.54	210
خ گستر	245	48.16	59.94
کیمیا	149	54.36	61.73
قمرو	222	56.3	100
ثاخذ	264	45.45	49.22
سمازن	360	12.22	8
شبریز	61	49.18	245

جدول 10 نتایج تجمیع اندیکاتور ها

Date	Open	High	Low	Close	Vol	SI_OPEN%Tag1	SI_OPEN%Tag_1	SI_OPEN%Tag_2	Trade
2018-08-13 00:00:00	6300	6600.0	6280.0	6620.0	17957.0	311720280.0	117	6590	0
2018-08-14 00:00:00	6300	6600.0	6300.0	6600.0	17957.0	181220280.0	117	6620	0
2018-08-15 00:00:00	7700	7700.0	6618.0	6614.0	44047.0	304001000.0	118	6627	0
2018-08-16 00:00:00	6670	6670.0	6671.0	6660.0	17957.0	179570000.0	66	6661	0
2018-08-18 00:00:00	7800	7800.0	6318.0	6300.0	22280.0	149149715.0	125	6300	0
2018-08-20 00:00:00	6370	6400.0	6350.0	6368.0	49536.0	327134405.0	56	6353	0
2018-08-21 00:00:00	6300	6390.0	6290.0	6300.0	17957.0	207000000.0	66	6300	0
2018-08-22 00:00:00	6700	6700.0	6648.0	6606.0	121625.0	80000381.0	66	6636	0
2018-08-23 00:00:00	6100	6000.0	6370.0	6084.0	214100.0	112000000.0	124	6080	0
2018-08-24 00:00:00	6800	7010.0	6690.0	6811.0	425500.0	307000000.0	706	6804	0
2018-08-27 00:00:00	6870	6750.0	6600.0	6636.0	150002.0	28118400.0	64	6641	0
2018-08-28 00:00:00	6800	6800.0	6800.0	6800.0	17957.0	181220280.0	117	6800	0
2018-08-29 00:00:00	6800	6900.0	6700.0	6700.0	118300.0	77711600.0	75	6900	0
2018-08-30 00:00:00	6800	6800.0	6750.0	6811.0	117800.0	800003475.0	76	6836	0
2018-08-31 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-01 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-02 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-03 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-04 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-05 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-06 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-07 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-08 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-09 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-10 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-11 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-12 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-13 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-14 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-15 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-16 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-17 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-18 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-19 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0
2018-09-20 00:00:00	6800	6800.0	6750.0	6800.0	17957.0	181220280.0	117	6800	0

شکل ۵ دیتای اصلی بعد از برچسب گذاری

جدول نهایی

ما جدول نهایی را جدولی می نامیم که بعد از رای گیری نهایی بین الگوریتم های یادگیری ماشین انجام شده و تمامی نتایج را در یک جدول که جدول نهایی است آورده ایم که شامل دو جدول است که در شکل ۵ ما سیگنال ها را مشاهده می کنیم و در جدول ۱۱ میزان تطبیق پذیری آن ها را به صورت کلی خواهیم دید.

Date	Open	High	Low	Close	Vol	F2	F3	F4	ln	sgboost	rf	shadby	svm	Final
2018-09-04	0	0.0	0.0	1000.0	0	0.000000e+00	0	1000	2	2.0	0.0	2.0	2.0	2.0
2018-09-05	0	0.0	0.0	1000.0	0	0.000000e+00	0	1000	2	2.0	2.0	0.0	0.0	2.0
2018-09-08	0	0.0	0.0	1000.0	0	0.000000e+00	0	1000	2	2.0	0.0	0.0	0.0	2.0
2018-09-09	0	0.0	0.0	1000.0	0	0.000000e+00	0	1000	2	2.0	2.0	0.0	0.0	2.0
2018-09-10	0	0.0	0.0	1000.0	0	0.000000e+00	0	1000	2	2.0	0.0	2.0	2.0	2.0
2019-02-23	3140	3190.0	3000.0	3133.0	4364181	1.367384e+10	1853	3062	1	2.0	0.0	0.0	2.0	3133.0
2019-02-24	3130	3150.0	3052.0	3062.0	4068602	1.253047e+10	2422	3133	1	2.0	1.0	0.0	0.0	3062.0
2019-02-25	3071	3080.0	3024.0	3048.0	3396228	1.035123e+10	2041	3082	1	1.0	0.0	1.0	0.0	3048.0
2019-02-26	3002	3040.0	3001.0	3016.0	2750094	8.293867e+09	1405	3048	1	2.0	0.0	0.0	2.0	3016.0
2019-02-27	3011	3127.0	3010.0	3053.0	3144706	9.694495e+09	1635	3016	1	2.0	0.0	0.0	2.0	3053.0

شکل ۶ جدول نهایی

نام سهم	Final	DT	RF	SVM	KNN	LR	XGBoost
دتماد	50%	47%	47%	35%	51%	44%	44%
های وب	78%	81%	82%	56%	34%	24%	24%
هرمز	65%	66%	65%	47%	33%	23%	24%
خ گستر	54%	54%	54%	39%	42%	34%	33%
کیمیا	64%	66%	64%	49%	32%	22%	22%
قمریو	59%	60%	60%	38%	39%	31%	31%
ثاخذ	50%	49%	49%	35%	50%	42%	43%
سمازن	53%	51%	51%	35%	62%	56%	56%
شبریز	56%	56%	56%	36%	43%	35%	34%
وینار	63%	64%	64%	41%	39%	31%	31%
وخارزم	60%	60%	60%	41%	38%	29%	28%
زدشت	81%	84%	41%	62%	31%	21%	21%
میانگی ن	61%	62%	58%	43%	41%	33%	33%

جدول ۱۱ جدول نهایی میزان تطابق

۵- نتیجه گیری

ما در نهایت بعد از اینکه دیتای ها و سهم های مناسبی از سازمان بورس ایران پیدا کردیم و پس از آنالیز و تمیز کردن، آنها را با اندیکاتور های مختلف مورد بررسی قرار دادیم و پس از ترکیب اندیکاتور ها نتایج را با هم مقایسه کردیم. همان طور که در جدول ۴,۶ مشاهده کردید نتایج خوبی به دست آوردیم.

بعد از اینکه سیگنال خرید و فروش و برچسب گذاری را روی داده ها به وسیله ی تجمیع اندیکاتور ها انجام شد. ما دیتای جدید را با الگوریتم های یادگیری ماشین تست کردیم، سپس ۶ الگوریتم یادگیری ماشین را روی دیتا تست و نتیجه گیری کردیم که نتایج تست و نهایی بسیار با هم تفاوت داشت. اما ما به نتیجه ی نهایی نیاز داشتیم و بعد از اینکه ۶ اندیکاتور میزان تطابق پذیری پذیرنی خود را مشخص کردند ۳ تا از بهترین های آنها را انتخاب و در نهایت عملیات Voting یا رایگیری را بین آن ۳ الگوریتم یادگیری ماشین انجام دادیم و نتیجه ی قابل قبولی را شاهد بودیم.

طبق اطلاعات به دست آمده جامعه هدف ما که برای بهبود عملکرد آنها تلاش می کردیم کلیه سازمان ها و افرادی هستند که در بازار های مالی چه جهانی و چه داخلی فعالیت می کنن و طبق این مقاله ما می توانیم به کمک الگوریتم های یادگیری ماشین و رای گیری به روش Voting به آنها کمک کنیم که قیمت جدید و آینده ی یک سهم را پیشبینی کنند و همچنین بهترین معامله را داشته باشند.

در یک مرحله فراتر می توان از این مقاله به عنوان یک ربات استفاده کرد که به طور خودکار شما آن را به صفحه ی کاربری خود در کارگزاری ها وصل می کنید و پس از مجوزات و دستورات برنامه به صورت خودکار به پردازش و یادگیری بپردازد و زمانی که نتایج عالی را به دست آورد به خرید و فروش سهم ها به طور اتوماتیک اما با میزان مشخصی سرمایه چرا که هنوز به طور کامل قابل اعتماد نیست.

یکی از کار هایی که باعث می شود برنامه ی ما حتی با درصد بالا دقت نتواند کامل مورد اعتماد قرار گیرد این است که روش ما بر اساس و پایه ی تحلیل تکنیکال است و تحلیل های بنیادی و... در نظر گرفته نشده است.



سید محمد جواد امیرحسینی مدرک
کاردانی خود در رشته ی نرم افزار
کامپیوتر در دانشگاه فنی پسران قم در
سال ۱۳۹۶ دریافت کرده و در
سال ۱۳۹۸ مدرک کارشناسی خود را
در دانشگاه شهاب دانش و در رشته ی

نرم افزار کامپیوتر دریافت کرده است و در حال حاضر مشغول
تحصیل در مقطع کارشناسی ارشد رشته ی نرم افزار کامپیوتر
در دانشگاه شهاب دانش می باشد و زمینه پژوهش و علایق
؛ یادگیری ماشین و ترکیب اندیکاتور ها را دارا می باشد.
نشانه رایانامه ایشان عبارتند از:

smja.ic@gmail.com



آرش خسروی مدرک کارشناسی
خود را در رشته مهندسی نرم‌افزار
در سال ۱۳۸۲ از دانشگاه صنعتی
اصفهان، مدرک کارشناسی ارشد
خود را در رشته مهندسی فناوری
اطلاعات در سال ۹۲ و مدرک

دکتری خود را در رشته مهندسی فناوری اطلاعات، گرایش
دستگاه‌های اطلاعاتی در سال ۹۶ از دانشگاه صنعتی مالزی اخذ
کرده است. ایشان در حال حاضر به‌عنوان هیات علمی مرکز
آموزش عالی محلات مشغول به کار هستند. زمینه‌های پژوهشی
موردعلاقه ایشان عبارت‌اند از: هوش تجاری، سیستم های
پیشنهاددهنده، مدیریت دانش مشتری، داده‌کاوی، متن‌کاوی و
فناوری اطلاعات در پزشکی.

نشانه رایانامه ایشان عبارتند از:

Khosravi.280@gmail.com

۶-منابع:

۱. Chen, Y. and Y. Hao, A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. *Expert Systems with Applications*, 2017. **80**: p. 340-355.
۲. Laboissiere, L.A., R.A. Fernandes, and G.G. Lage, Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks. *Applied Soft Computing*, 2015. **35**: p. 66-74.
۳. Zahedi, J. and M.M. Rounaghi, Application of artificial neural network models and principal component analysis method in predicting stock prices on Tehran Stock Exchange. *Physica A: Statistical Mechanics and its Applications*, 2015. **438**: p. 178-187.
۴. Zhong, X. and D. Enke, Forecasting daily stock market return using dimensionality reduction. *Expert Systems with Applications*, 2017. **67**: p. 126-139.
۵. Rezaei, H., H. Faaljou, and G. Mansourfar, Stock price prediction using deep learning and frequency decomposition. *Expert Systems with Applications*, 2021. **169**: p. 114332.
۶. Chen, W., et al., Mean-variance portfolio optimization using machine learning-based stock price prediction. *Applied Soft Computing*, 2021. **100**: p. 106943.
۷. Kou, G., et al., Machine Learning Methods for Systemic Risk Analysis in Financial Sectors. *Technological and Economic Development of Economy*, 2019. **25**(5): p. 716-742.
۸. Nabipour, M., et al., Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis. *IEEE Access*, 2020. **8**: p. 150199-150212.
۹. Rundo, et al., Machine Learning for Quantitative Finance Applications: A Survey. *Applied Sciences*, 2019. **9**(24): p. 5574.
۱۰. Huang, B., et al., Automated trading systems statistical and machine learning methods and hardware implementation: a survey. *Enterprise Information Systems*, 2018. **13**(1): p. 132-144.
۱۱. Umer, M., M. Awais, and M. Muzammul, Stock market prediction using machine learning (ML) algorithms. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 2019. **8**(4): p. 97-116.
۱۲. Zhong, X. and D. Enke, Predicting the daily return direction of the stock market using hybrid machine learning algorithms. *Financial Innovation*, 2019. **5**(1): p. 1-20.
۱۳. Sebastião, H. and P. Godinho, Forecasting and trading cryptocurrencies with machine learning under changing market conditions. *Financial Innovation*, 2021. **7**(1): p. 1-30.
۱۴. Xiong, W. and L. Xiong, Smart contract based data trading mode using blockchain and machine learning. *IEEE Access*, 2019. **7**: p. 102331-102344.

Presenting a stock price prediction method by means of improved combined machine learning algorithms

A. Khosravi¹ S. Amirhussaini²

^{1,2} Faculty of Engineering, Mahallat Institute of Higher Education, Mahallat, Iran

Abstract

Now a day, investing in online financial markets, analyzing and predicting stock prices for buying and selling with more profit is very important. If we want to be able to prepare and adjust a program that can predict the price of stocks and give us information about the best time in which to buy or sell; The analyzing of the past prices of the stock by indicators are very helpful. The process of machine learning to obtain a formula that predicts the stock price in the future and informs us whether we should buy, sell is usefull. In this research, some usefull indicators are combibe and the result of them was used by machine learning to predect the price of market. According to our results, we predicted the price with high accuracy and got high profit from our transactions in online financial markets.

روش ارجاع: آ. خسروی، س. امیرحسینی، ارائه ی یک روش پیش بینی قیمت سهام بورس به وسیله ی الگوریتم های بهبود یافته ی ترکیبی یادگیری ماشین، دوفصلنامه محاسبات و سامانه های توزیع شده، سال ششم، شماره ۱، شماره پیاپی ۱۱، صفحه ۷۴ تا ۸۸، سال ۱۴۰۲

How to cite: A. Khosravi, S. Amirhussaini Presenting a stock price prediction method by means of improved combined machine learning algorithms, Journal of Distributed Computing and Systems (JDCS), Vol 6, Issue 1, Page 74-88, 2024.