



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده صنایع

معاملات الگوریتمی مالی ترکیبی با استفاده از

KNN و ANN ، SVC ، RF

نگارش: مسعود هادی

استاد راهنما: دکتر مسعود ماهوتچی

آبان ماه 1400

چکیده:

روند تغییرات شاخص همواره به عنوان یکی از معیارهای سرمایه‌گذاری مدنظر قرار می‌گیرد.

این پژوهش یک سیستم پشتیبان تصمیم برای معاملات الگوریتمی در بازارهای مالی ارائه می‌دهد که از رویکردی جدید در اتخاذ تصمیمات خودکار استفاده می‌کند.

این رویکرد با استفاده از الگوریتم‌های جنگل تصادفی، ماشین بردار پشتیبان، شبکه عصبی و نزدیک‌ترین همسایه و ترکیب آن‌ها با استفاده از متالگوریتم بگینگ، به پیش‌بینی سیگنال‌های خرید، نگهداری یا فروش می‌پردازد.

فهرست مطالب



01

مقدمه



02

مبانی نظری و
مروری بر ادبیات
پژوهش



03

مروری بر پیشینه
پژوهش



04

روش‌شناسی
پژوهش



05

بررسی نتایج



06

جمع‌بندی،
نتیجه‌گیری و
پیشنهادهات

مقدمه

01

یکی از مهم‌ترین دستاوردهای علم: قابلیت پیش‌بینی بخشیدن به متغیرها و پدیده‌ها

پژوهشگران علوم مالی نیز با استفاده از ابزارهای مختلف به دنبال طراحی مدل‌هایی هستند که به وسیله آن متغیرها و حوادث مدنظرشان در بازارهای مالی را پیش‌بینی کنند.

پیش‌بینی قیمت سهام از مهم‌ترین اهداف در علوم مالی و سرمایه‌گذاری است.

پیش‌بینی قیمت سهام از دو جنبه حائز اهمیت است:

2

افزایش بازدهی در سطح مشخصی
از ریسک و بالعکس با استفاده از
روش‌های پیش‌بینی مطلوب

1

تاثیرگذاری در بهینه‌سازی پرتفویهای
سرمایه‌گذاری و اجرای راهبردهای
سرمایه‌گذاری

شاخص

به عنوان یک معیار آماری، قابلیت مقایسه وضعیت کنونی را نسبت به گذشته فراهم آورده و بررسی و تحلیل آن اطلاعات مفیدی را به کارشناسان و افراد ذی ربط در آن حوزه ارائه می‌دهد. شاخص‌های قیمت سهام در تمام بازارهای مالی دنیا، به مثابه یکی از مهم‌ترین معیارهای سنجش عملکرد بازار سهام، اهمیت زیادی دارند و شاید مهم‌ترین دلیل توجه روزافزون به آن‌ها، این نکته باشد که شاخص‌ها از تجمیع حرکات‌های قیمتی سهام تمام شرکت‌ها یا طبقه‌ی خاصی از شرکت‌های موجود در بازار به دست می‌آیند.

پیچیدگی‌های شاخص کل قیمت:

- غیرخطی بودن
- عدم قطعیت
- نوسان
- پویایی

که پیش‌بینی آن را دشوار می‌کنند و نتایج پیش‌بینی را با عدم قطعیت زیادی مواجه می‌سازند که خود تاثیر شایان توجهی در بازده سرمایه‌گذاران به همراه دارد.

تحلیل‌های شناخته‌شده در بازارهای مالی

تحلیل مالی-رفتاری

حوزه‌ای از دانش مالی است که از نظریه‌های مبتنی بر روانشناسی برای توضیح رفتار بازارهای مالی بهره می‌گیرد.

هوش مصنوعی

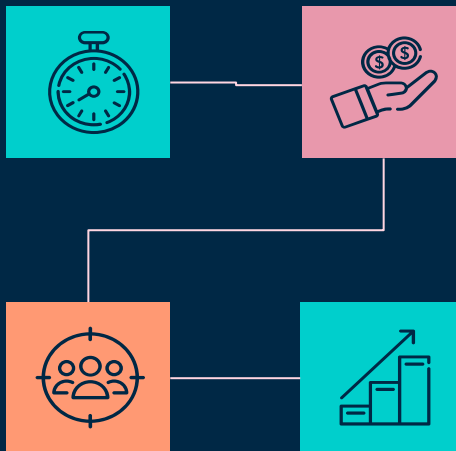
در سال‌های اخیر برای پیش‌بینی بازار از هوش مصنوعی استفاده شده است که ترکیب آن با تحلیل تکنیکال می‌تواند منجر به ایجاد سیستم‌های خودکار معاملاتی و الگوریتمی شود.

تحلیل بنیادی

بر پایه‌ی عملکرد شرکت‌ها و رشد سودآوری آن‌ها بنا شده است.

تحلیل تکنیکال

بر پایه سابقه معاملات یک دارایی مالی از طریق نمودار قیمت و فرمول‌های ریاضی که اندیکاتورهای تکنیکال نامیده می‌شوند، بنا شده است.



در این پژوهش

تلاش شده است تا با بهره‌گیری از اندیکاتورهای تحلیل تکنیکال در مرحله آغازین قابلیت پیش‌بینی‌پذیری را در اثر کاهش نویز ورودی به مدل افزایش دادو با ارائه‌ی ترکیب جدیدی از الگوریتم‌های داده‌کاوی، در یک فرآیند سه مرحله‌ای به تولید سیگنال‌های معاملاتی برای روز آتی پرداخته شود.

1. در مرحله اول سهم‌های دارای بیشترین تاثیر انتخاب می‌شوند
2. سپس از چهار الگوریتم مختلف برای پیش‌بینی استفاده می‌شود
3. در نهایت با استفاده از یک متالگوریتم سیگنال‌های تولید شده در مرحله قبل ترکیب شده و سیگنال نهایی ارائه می‌شود.

بررسی‌های صورت گرفته نشان‌دهنده‌ی صحت قابل قبول سیگنال‌های تولید شده و بازده مناسب‌تر مدل از استراتژی "خرید - نگهداری" است.

مبانی نظری و مروری بر ادبیات پژوهش

02

دسته‌بندی، شکلی از تحلیل داده‌ها تلقی می‌شود که
در آن مدلی جهت توصیف کلاس‌های مهمی از داده‌ها
استخراج می‌شود.

دسته‌بندی داده‌ها فرآیندی است که کار خود را در دو گام انجام می‌دهد:

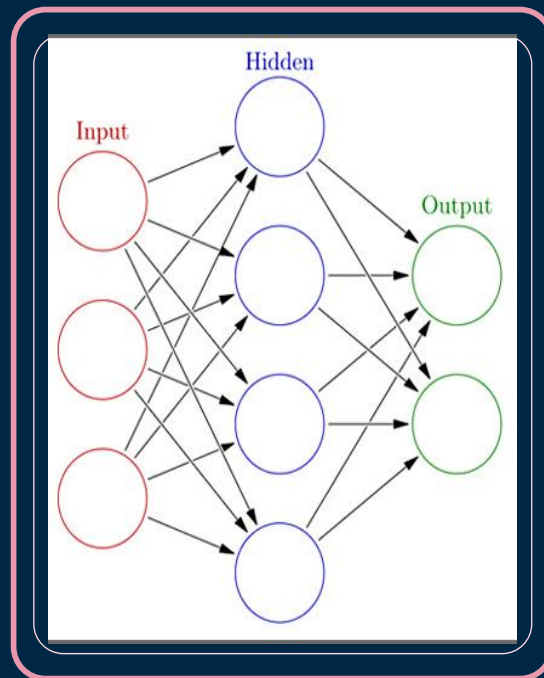
1. گام یادگیری (که در آن با استفاده از داده‌های ورودی و برچسب آن‌ها، مدل ساخته می‌شود)
 2. گام دسته‌بندی (که در آن از مدل ساخته شده در گام اول جهت پیشگویی برچسب داده‌های جدید استفاده می‌شود)
- به دلیل آن که برچسب کلاس هر یک از ردیف‌های آموزشی مشخص شده‌اند، این گام همچنین به عنوان یادگیری با ناظر شناخته می‌شوند. در مقابل یادگیری بی‌ناظر (یا خوشه‌بندی) است، که در آن برچسب کلاس تاپل‌های آموزشی شناخته شده نیست و ممکن است تعداد یا مجموعه دسته‌هایی که در نهایت به دست خواهند آمد نیز از قبل مشخص نباشند.

شبکه عصبی مصنوعی

Artificial Neural Network

الگوریتم ANN از سیستم عصبی بیولوژیکی الهام گرفته شده است و می‌توان آن را به عنوان یک ارگانیسم کامل متشکل از تعداد زیادی واحد محاسباتی در نظر گرفت که برای حل مسئله با یکدیگر تعامل دارند.

هر نورون سیگنال‌های سلول‌های عصبی همسایه را جمع‌آوری کرده و آن‌ها را به لایه بعدی منتقل می‌کند و در نتیجه سیگنال‌های "تحریک کننده" یا "بازدارنده" ایجاد می‌کند. از این رو، هر نورون می‌تواند به عنوان یک پردازنده دیده شود که محاسبه ساده‌ای انجام می‌دهد؛ مانند تصمیم‌گیری در مورد ارسال کردن یا نکردن سیگنال به سلول‌های عصبی دیگر. یادگیری زمانی اتفاق می‌افتد که اثرات سیناپس‌ها تغییر کند، به عنوان مثال تأثیر یک نورون بر روی یک نورون دیگر تغییر می‌کند.

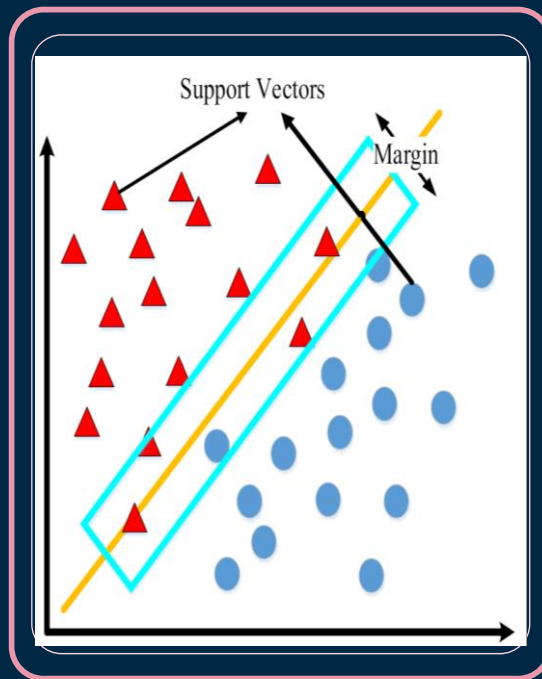


ماشین بردار پشتیبان

Support Vector Machine

الگوریتم SVM به کمک یک نگاشت غیرخطی داده‌های آموزشی اولیه را به یک بعد بالاتر تبدیل می‌کند.

در این بعد جدید به دنبال ابرصفحه‌ای بهینه می‌گردد، که تاپل‌های * یک کلاس را از دیگر کلاس به صورت خطی تفکیک می‌کند. با یک نگاشت غیرخطی مناسب به یک بعد بالای کافی، داده‌های دو کلاس را همیشه می‌توان به کمک یک ابرصفحه تفکیک نمود. الگوریتم SVM این ابرصفحه را با کمک بردارهای پشتیبان (که اساساً تاپل‌های آموزشی هستند) و حاشیه‌ها (که با کمک بردارهای پشتیبان تعریف می‌شوند) پیدا می‌کند.



* Tuple در لغت به معنای «چندتایی» بوده و در جایگاه‌های مختلف معانی متفاوتی دارد و این در حالی است در زبان‌هایی مانند پایتون، لیست و غیره، تاپل به مجموعه مقادیری گفته می‌شود که به صورت مرتب شده پشت سر هم آمده باشند.

درخت تصمیم

Decision Tree

یک درخت تصمیم همانطور که از نام آن مشخص است، یک ساختار درختی شبیه به فلوچارت دارد. هر گره داخلی (گره غیربرگ) در این درخت آزمونی را بر روی یک صفت خاصه نشان می‌دهد و هر شاخه نتیجه‌ی آزمون را نمایش می‌دهد و در هر گره برگ (یا گره پایانی) یک برچسب کلاس نگهداری می‌شود. برای ساخت درختان تصمیم به هیچ دانش خاص یا تنظیم پارامتری نیاز نیست. بنابراین برای یافتن اکتشافی دانش مناسب است. درختان تصمیم می‌توانند داده‌های چندبعدی را کنترل کنند. از نقطه نظر بصری، هضم دانش ارائه شده در درختان تصمیم برای انسان راحت است. گام‌های دوگانه‌ی یادگیری و دسته‌بندی در استقراء درختان تصمیم ساده و سریع است و به طور کلی دارای صحت مناسبی هستند.



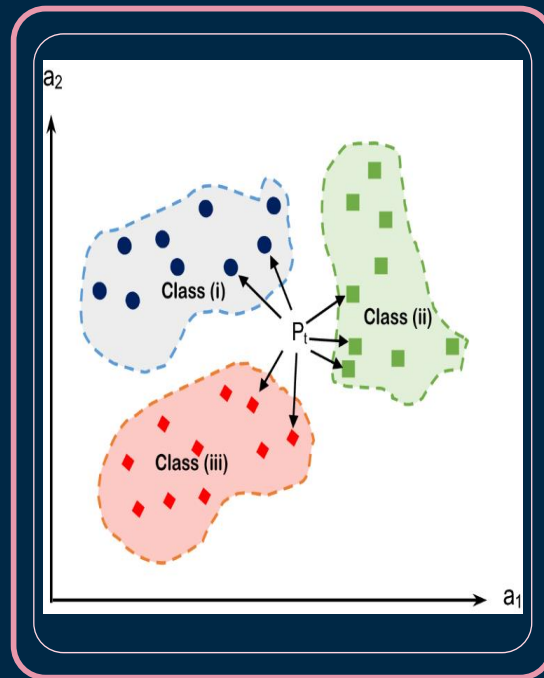
یادگیرنده‌های گُند

در رویکرد کند، یادگیرنده تا آخرین دقیقه انتظار می‌کشد و قبل از آن مدلی برای دسته‌بندی یک تاپل آزمایشی ساخته نمی‌شود. بنابراین در مواجهه با یک تاپل آموزشی یک یادگیرنده‌ی کند آن را ذخیره می‌کند (یا تنها پردازش کمی را انجام می‌دهد) و تا دریافت یک تاپل آزمایشی صبر می‌کند. تنها این یادگیرنده هنگامی که با تاپل آزمایشی روبه‌رو می‌شود، به تلاش برای دسته‌بندی تاپل مذکور بر اساس شباهت آن با تاپل‌های آموزشی ذخیره شده می‌پردازد. بر خلاف روش‌های **یادگیری مشتاق**، یادگیرنده‌های کند کار کمتری را هنگام آموزش و کار بیشتری در هنگام دسته‌بندی یا پیشگویی عددی انجام می‌دهند. این در حالی است که در یادگیرنده‌های مشتاق، قبل از دریافت یک تاپل جدید برای دسته‌بندی، با کمک مجموعه تاپل‌های آموزشی یک مدل کلی ساخته می‌شود.

دسته‌بندی‌های k همسایه‌ی نزدیک

K Nearest Neighbors

دسته‌بندی‌های نزدیک‌ترین همسایه، یادگیری خود را بر اساس تشابه انجام می‌دهند، این کار با مقایسه‌ی داده آزمایشی و داده‌های آموزشی مشابه با آن صورت می‌گیرد. داده‌های آموزشی با کمک n صفت مشخص توصیف می‌شوند و هر داده در واقع نمایش نقطه‌ای در فضای n بعدی است. بدین ترتیب تمام داده‌های آموزشی در یک فضای n بعدی ذخیره می‌شوند. هرگاه با یک داده ناشناخته روبه‌رو می‌شوید، دسته‌بند k نزدیک‌ترین همسایه به دنبال k تاپل آموزشی است که شبیه‌ترین داده‌ها به داده ناشناخته هستند. این k داده آموزشی k همسایه‌ی نزدیک داده ناشناخته هستند.



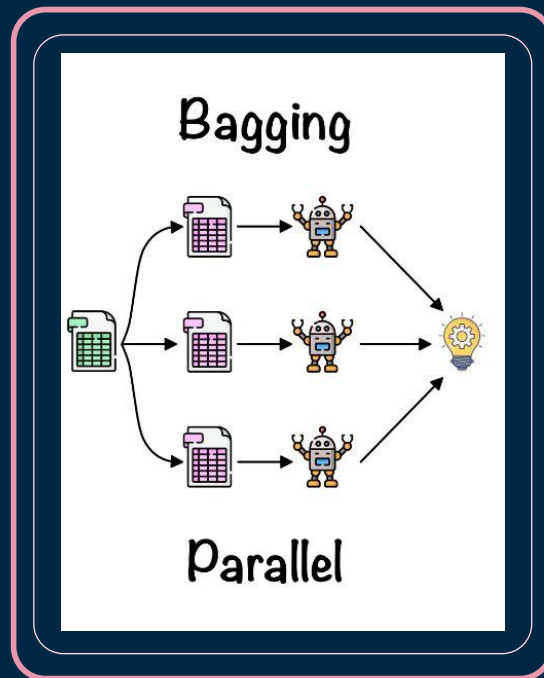
The background is a dark blue field decorated with a pattern of thin white vertical lines and small squares in teal, pink, and orange. The text is centered in the middle of the slide.

تکنیک‌هایی جهت بهبود صحت دسته‌بندی

Bagging

در روش های موازی، مدل های متفاوت مستقل از یکدیگر آموزش می بینند و می توان آنها را به صورت همزمان آموزش داد.

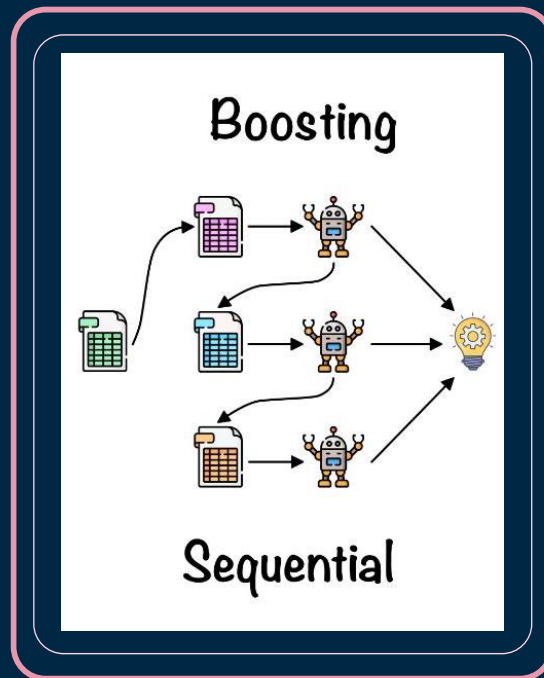
شناخته شده ترین این روش bagging است که کوتاه شده کلمات Bootstrapping (خودراه انداز) و Aggregation (تجمیعی) است و هدف آن ایجاد مدلی ترکیبی است که پایدارتر (more robust) مدل های پایه سازنده خود می باشد.



Boosting

در روش های ترتیبی (Sequential methods) مدل های پایه ترکیبی، به صورت مستقل از یکدیگر آموزش نمی بینند. ایده اصلی این است که مدل ها به صورت تکراری آموزش ببینند به طوری که آموزش مدل در یک مرحله، وابسته به مدل های آموزش دیده در مراحل قبلی است. boosting از مشهورترین این مدل ها است و مدل های ترکیبی ایجاد می کند که بایاس آن کمتر از مدل های پایه سازنده است.

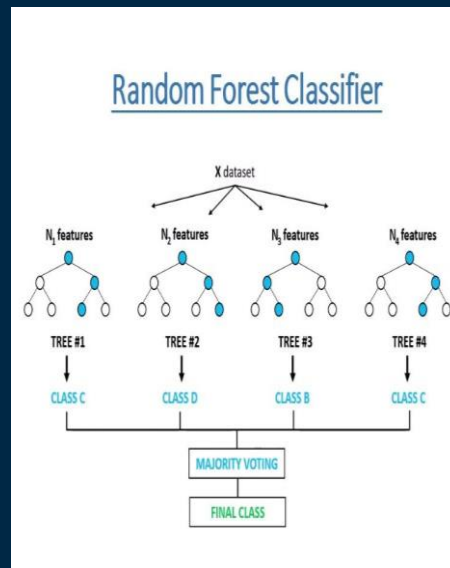
روش boosting با همان ایده روش bagging کار می کند: تعدادی مدل ایجاد می کنیم که برای به دست آوردن مدلی قویتر با کارایی بهتر.



جنگل تصادفی

Random Forest

تصور کنید دسته‌بندی‌های استفاده شده در روش تلفیقی همگی از نوع درخت تصمیم هستند؛ بدین ترتیب این مجموعه تشکیل یک جنگل را خواهند داد. هر یک از درختان تصمیم با استفاده از یک انتخاب تصادفی صفات خاصی موجود در هر گره جهت تعیین انشعاب ساخته می‌شوند. به عبارت دیگر هر درخت بر اساس مقادیر یک بردار تصادفی ساخته می‌شود. این مقادیر دارای توزیع یکسانی برای تمام درختان موجود در جنگل هستند و به صورت مستقلی نمونه‌گیری می‌شوند. برای دسته‌بندی نیز هر درخت رای خود را صادر و نتیجه نهایی با رای اکثریت تعیین می‌شود.



مروری بر پیشینه پژوهش

03

پنج گروه از پژوهشگران معتقد هستند که پیش‌بینی قیمت سهام امکان‌پذیر نیست



کسانی که به فرضیه بازار
کارا اعتقاد دارند.



کسانی که به علت موثر بودن
عوامل متعدد بر تغییرات
قیمت سهام و رفتار آشوب‌گونه
و غیر خطی تغییرات قیمت
سهام، پیش‌بینی قیمت سهام
را امری غیرممکن می‌دانند.



کسانی که عامل مهمی به نام
اقتصاد اطلاعات را مطرح
می‌سازند.



کسانی که ماهیت بازی به
جمع صفر بودن در بازارهای
مالی را دلیل اصلی دشواری
پیش‌بینی در این حوزه
می‌دانند.



کسانی که نوع رویدادها در
بازارهای مالی را دلیل اصلی
عدم پیش‌بینی پذیری قیمت
می‌دانند.

- ماهیت پیچیده، تکاملی، داشتن خاصیت دینامیکی و غیرخطی که به دلیل تعامل حوادث و شرایط اقتصادی بوجود می‌آید و انتظارات غیرعقلانی سرمایه‌گذاران، پیش‌بینی قیمت سهام را به امری دشوار مبدل می‌سازد نه امری محال.

ناکاموری، 2005

پیشینه پژوهش

از شبکه عصبی با دو لایه پنهان برای پیش‌بینی معاملات آتی استفاده کردند.

برگس و همکاران
2000

فونته و همکاران
2006

با به کار بردن الگوریتم ژنتیک سعی در بهینه کردن پارامترهای اندیکاتورهای مختلف تکنیکال داشته است.

با بهره‌گیری از الگوریتم شبکه‌های عصبی به پیش‌بینی سیگنال معاملاتی روز آتی شاخص سهام معمولی استرالیا پرداخته اند.

تیلاکاراتن و همکاران
2007

تاکور و کومار
2017

روشی جدید از ترکیب الگوریتم‌های ماشین بردار پشتیبان وزنی و جنگل تصادفی برای ایجاد سیگنال‌های معاملاتی ارائه کرده اند.

پیشینه پژوهش

پیش‌بینی شاخص کل سهام
بورس اوراق بهادار تهران با
استفاده از شبکه های عصبی
موجکی



فلاح پور و علي پور
1393

دستپاک و رستگار
1394



طراحی یک سیستم معاملات
الگوریتمی با استفاده از
اندیکاتورهای تکنیکال

طراحی یک سیستم معاملات
الگوریتمی از نوع معاملات
زوجی در بورس اوراق بهادار
تهران



فلاح پور و حکیمیان
1395

مشاری و همکاران
1396



بهینه‌سازی متغیرهای پژوهش
به کمک الگوریتم ژنتیک و
ارائه‌ی مدلی جهت پیش‌بینی
نقاط طلایی

روش شناسی پژوهش

04

این پژوهش از نظر رویکرد کاربردی بوده و مبتنی بر
پژوهش‌های میدانی است.
هدف این پژوهش ارائه‌ی مدل پیش‌بینی جدیدی
برای شاخص S&P500 می‌باشد.

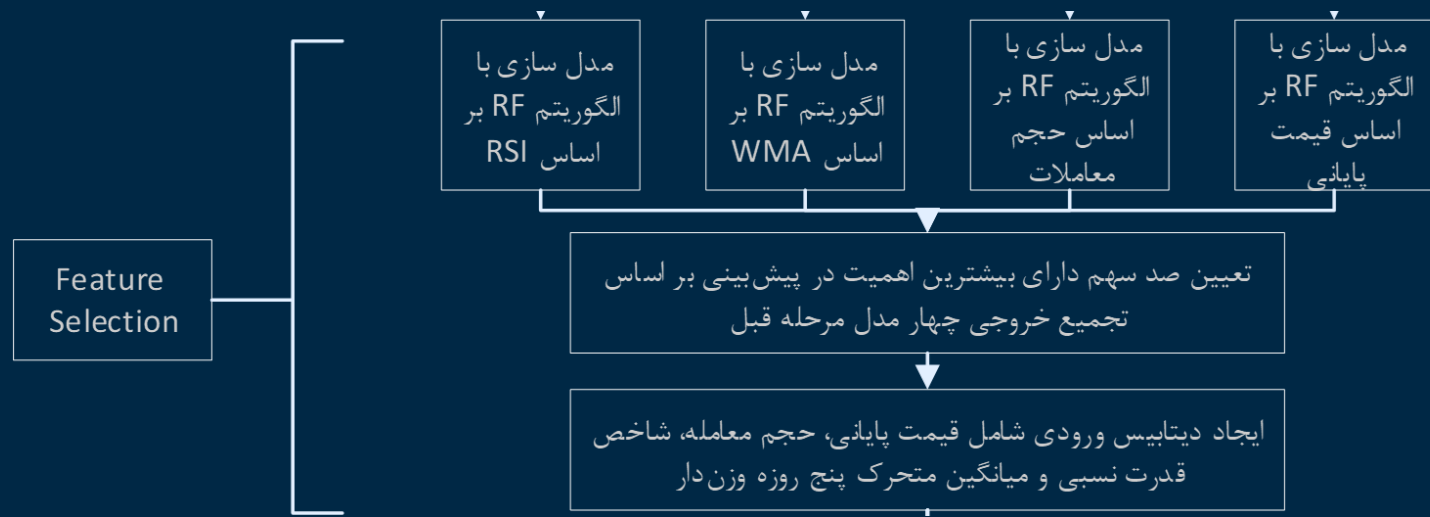
مرحله اول: شناسایی و جمع‌آوری متغیرهای مسئله

اطلاعات قیمت پایانی و حجم معاملات روزانه مربوط به 500 سهام موجود در شاخص S&P500 و قیمت نماد SPY در بازه زمانی ابتدای سال 2015 تا پایان ماه اکتبر سال 2021 از طریق کتابخانه مالی یاهو در زبان برنامه نویسی پایتون استخراج گردید و در راستای کاهش نویز ورودی به مدل، با استفاده از این اطلاعات اندیکاتورهای شاخص قدرت نسبی (14روزه) و میانگین متحرک وزنی (پنج روزه) نیز، همانطور که در پژوهش آدریان (2011) توصیه شده است، برای هر یک از نمادها در بازه زمانی ذکر شده محاسبه شد. همچنین به دلیل تفاوت در مقیاس داده‌ها، در انتها تمامی داده‌ها نرمال می‌شوند.



مرحله دوم: انتخاب صد متغیر دارای بیشترین اهمیت

در این مرحله با استفاده از قابلیت تعیین اهمیت متغیرها در الگوریتم جنگل تصادفی، ابتدا با استفاده از چهار ورودی مختلف که عبارت‌اند از قیمت پایانی، حجم معاملات، شاخص قدرت نسبی (14روزه) و میانگین متحرک وزنی (پنج روزه)، مدلی جهت تعیین اهمیت هر یک از 500 سهام در پیش‌بینی شاخص ایجاد شده و بر اساس این مدل صد سهام دارای بیشترین اهمیت در پیش‌بینی، انتخاب می‌شوند. سپس پایگاه داده ورودی مدل اصلی که شامل قیمت پایانی، حجم معاملات، شاخص قدرت نسبی و میانگین متحرک وزنی پنج روزه برای صد سهم انتخاب شده، است، تشکیل می‌شود.



مرحله سوم: پیش‌بینی و ارزیابی اولیه

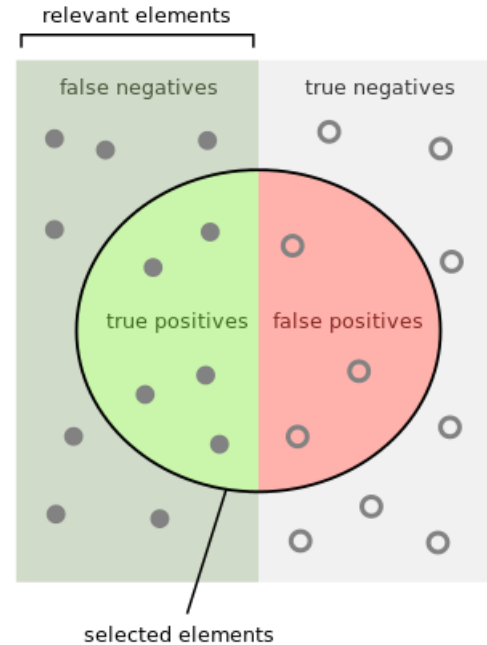
در این گام الگوریتمی برای انتخاب بازه‌های زمانی تصادفی پانصد روزه ایجاد می‌شود، به گونه‌ای که چهارصد روز ابتدایی هر بازه زمانی به عنوان داده آموزش و صد روز پایانی بازه به عنوان داده آزمایش در نظر گرفته می‌شود. در ادامه هر یک از الگوریتم‌های KNN، SVM، RF، MLP توسط داده‌های آموزشی برای پیش‌بینی سیگنال روز آتی آموزش داده می‌شوند و نتایج عملکرد هریک از الگوریتم‌ها برای پیش‌بینی توسط داده‌های آموزشی، با استفاده از ماتریس اختلال و دو معیار دقت و اطمینان سنجیده خواهد شد.

Precision (positive)= $TP/(TP+FP)$

Precision(negative)= $TN/(TN+FN)$

Recall(positive)= $TP/(TP+FN)$

Recall(negative)= $TN/(TN+FP)$



How many selected
items are relevant?

$$\text{Precision} = \frac{\text{Green Circle}}{\text{Green Circle} + \text{Red Circle}}$$

How many relevant
items are selected?

$$\text{Recall} = \frac{\text{Green Circle}}{\text{Green Circle} + \text{Green Square}}$$

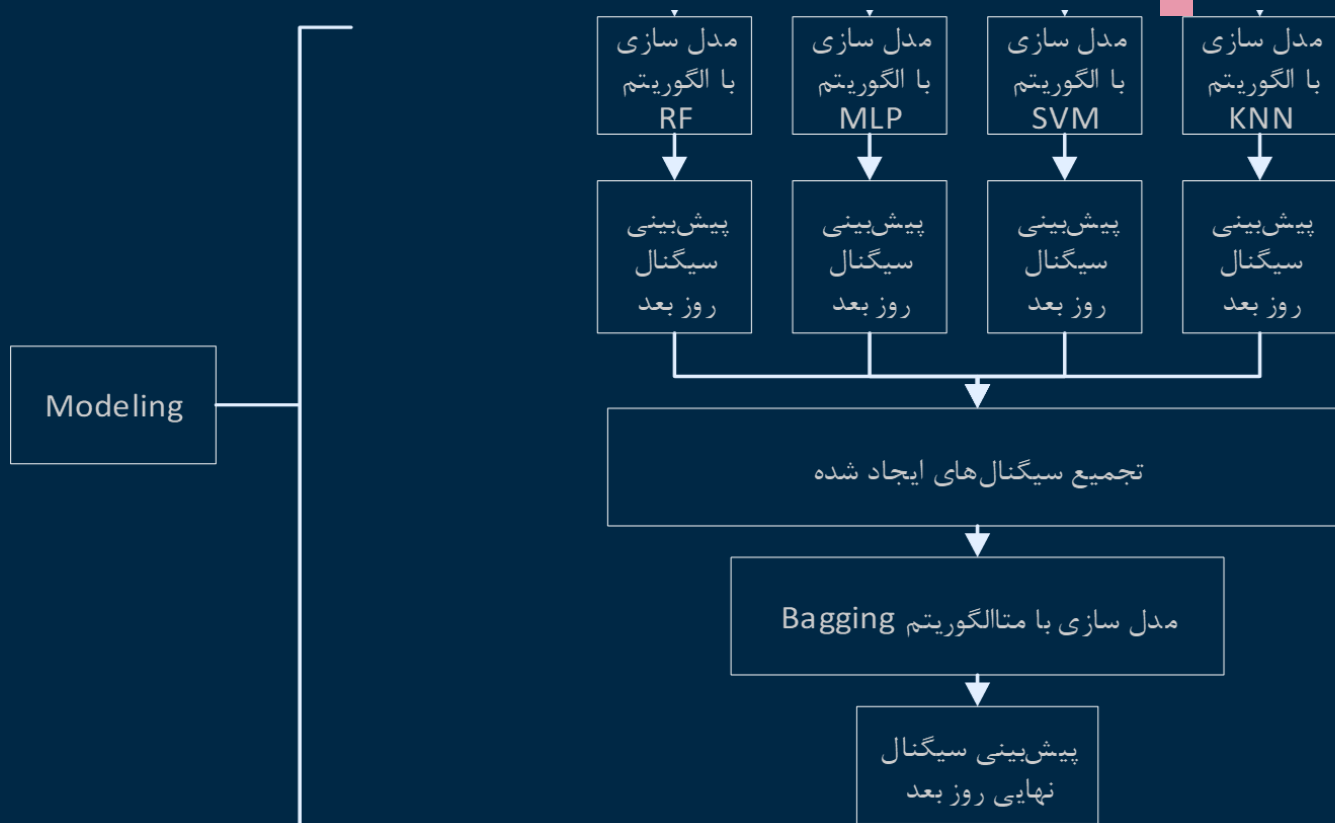
مرحله چهارم: پیش‌بینی بر اساس داده‌های آزمایشی و ارزیابی مدل‌ها

در این مرحله هر یک از الگوریتم‌ها بر اساس داده‌های مربوط به ابتدای سال 2015 تا یک روز قبل از روز مورد پیش‌بینی، به تولید سیگنال خرید، نگهداری یا فروش می‌پردازند (همانند معامله‌گری که در پایان روز فعلی برای پیش‌بینی سیگنال روز بعد، اطلاعات مربوط به روز جاری را نیز در اختیار خواهد داشت). در پایان نتایج پیش‌بینی هر یک از الگوریتم‌ها با معیار دقت مدل سنجیده می‌شود.

$$\text{Accuracy(ACC)} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{TN} + \text{FN})$$

مرحله پنجم: ترکیب نتایج پیش‌بینی الگوریتم‌ها و تولید سیگنال نهایی

در این مرحله نتایج الگوریتم‌های مرحله قبل با استفاده از متالگوریتم بگینگ، که یک الگوریتم ترکیبی است، سیگنال نهایی برای روز آتی را بر اساس رای اکثریت تولید می‌کند. به این صورت که متالگوریتم بگینگ ابتدا توسط خروجی الگوریتم‌های RF، MLP، SVM و KNN در بازه زمانی آموزش، آموزش می‌بیند و سپس در بازه زمانی آزمون با دریافت خروجی مدل‌های مذکور در هر روز، تصمیم‌گیری نهایی را انجام دهد.



بررسی نتایج

05

جزئیات پیاده‌سازی مدل

جامعه آماری این پژوهش، اطلاعات جمع آوری شده از قیمت و حجم معاملات 500 سهام موجود در شاخص S&P500 در بازه زمانی ابتدای سال 2015 تا پایان ماه اکتبر سال 2021 می‌باشد. این پژوهش درصدد است تا مدل پیش‌بینی جدیدی برای پیش‌بینی قیمت سهام صندوق SPY به عنوان معیار شاخص S&P500 ارائه نماید.

پارامترهای اصلی الگوریتم‌های RF و MLP توسط قابلیت جستجوی شبکه‌ای کتابخانه سایکیت‌لرن و با اعتبارسنجی متقابل پنجگانه (که بیش از هزار و پانصد حالت مختلف برای پارامترها را بررسی کرده و بهترین پارامترها جهت پیش‌بینی را تعیین می‌کند) انجام شده‌است.



در این مطالعه با توجه به بررسی‌های موريس در رابطه با توزیع داده‌ها و مرزبندی تعیین سیگنال‌ها، فرض می‌شود که 0.5% افزایش (یا کاهش) در قیمت پایانی به اندازی کافی معقول است که حرکت مربوطه به عنوان سیگنال خرید (یا فروش) در نظر گرفته شود همچنین در صورتی که تغییر قیمت کمتر از 0.5% افزایش (یا کاهش) باشد، سیگنال نگهداری تولید خواهد شد.

سنجش عملکرد مدل

- برای سنجش مدل، تابعی طراحی شد تا سری های زمانی تصادفی به طول 500 روز کاری انتخاب شوند. سپس چهارصد روز ابتدای هر بازه زمانی به عنوان داده آموزش و صد روز پایانی به عنوان داده آزمایش در نظر گرفته شده اند.
- ابتدا داده های آموزش به هر یک از چهار مدل RF، MLP، SVC و KNN داده شده و دقت هریک از مدل ها در بازه آموزش سنجیده شد.
- سپس هریک از مدل ها توسط داده های آزمایش بررسی شده و دقت آن ها سنجیده شده است.
- در ادامه خروجی هر یک از چهار مدل مذکور به عنوان ورودی به متالگوریتم بگینگ داده شده و خروجی آن به عنوان سیگنال نهایی در نظر گرفته شده است که دقت آن نیز مورد بررسی قرار گرفته است.

به جهت سنجش بهتر مدل پیشبینی، ده مرتبه و در بازه‌های زمانی متفاوت مدل را اجرا کرده و میانگین، انحراف از معیار، کمترین مقدار و بیشترین مقدار دقت مدل محاسبه شده‌است. همچنین در انتها بازده مدل نیز با استراتژی خرید و نگهداری مقایسه شده است.



	Train				Test					Return	
	MLP	RF	SVC	KNN	MLP	RF	SVC	KNN	BAG (Final Model)	Buy&Hold	Model
Avg	70.00%	100.00%	59.10%	67.58%	59.30%	56.30%	62.50%	56.70%	56.97%	4.76%	9.58%
Std	8.18%	0.00%	8.50%	4.70%	8.97%	9.14%	8.50%	9.90%	7.87%	7.80%	5.17%
Min	58.75%	100.00%	49.75%	61.75%	50.00%	41.00%	53.00%	47.00%	50.51%	-15.50%	-1.75%
Max	82.75%	100.00%	69.50%	73.25%	76.00%	74.00%	78.00%	78.00%	73.74%	11.48%	17.71%

مقایسه بازده

- در بازه‌های زمانی متفاوت بازده معاملات بر مبنای مدل پیشبینی با استراتژی "خرید - نگهداری" سهام مورد مقایسه واقع شده است و میانگین، انحراف از معیار، کمترین مقدار و بیشترین مقدار بازده هریک از استراتژی‌ها نیز محاسبه است.

نتایج نشان دهنده این موضوع است که مدل پیشنهادی نه تنها بصورت میانگین بازده بهتری از خود نشان می‌دهد، بلکه انحراف از معیار کمتری نیز در مقایسه با استراتژی خرید و نگهداری از خود به‌جا گذاشته است.

مقایسه بازده

بر اساس ده بازه زمانی متفاوت صد روزه

"خرید - نگهداری"



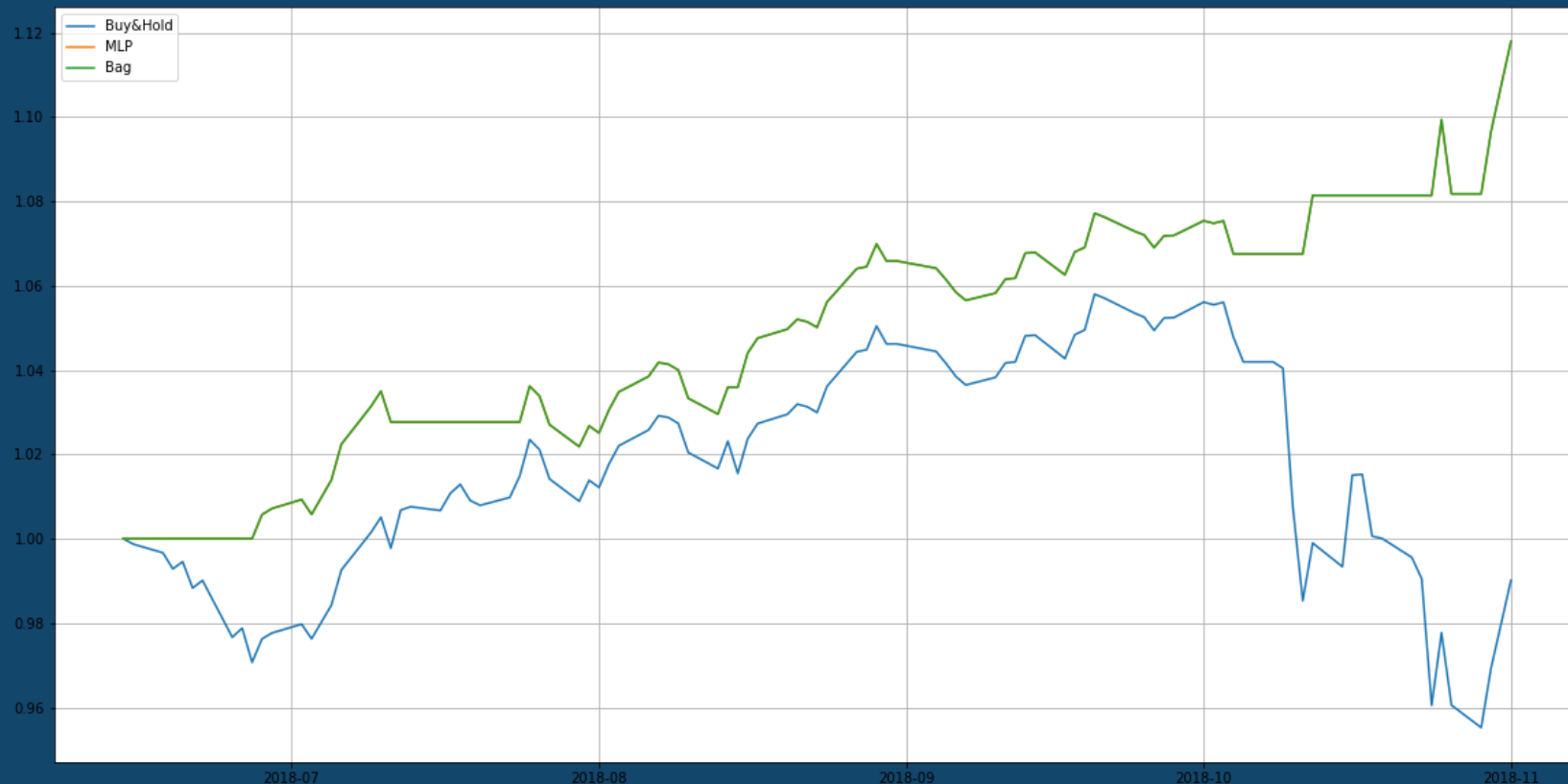
4.76% (Avg)
7.80% (Std)

مدل طراحی شده

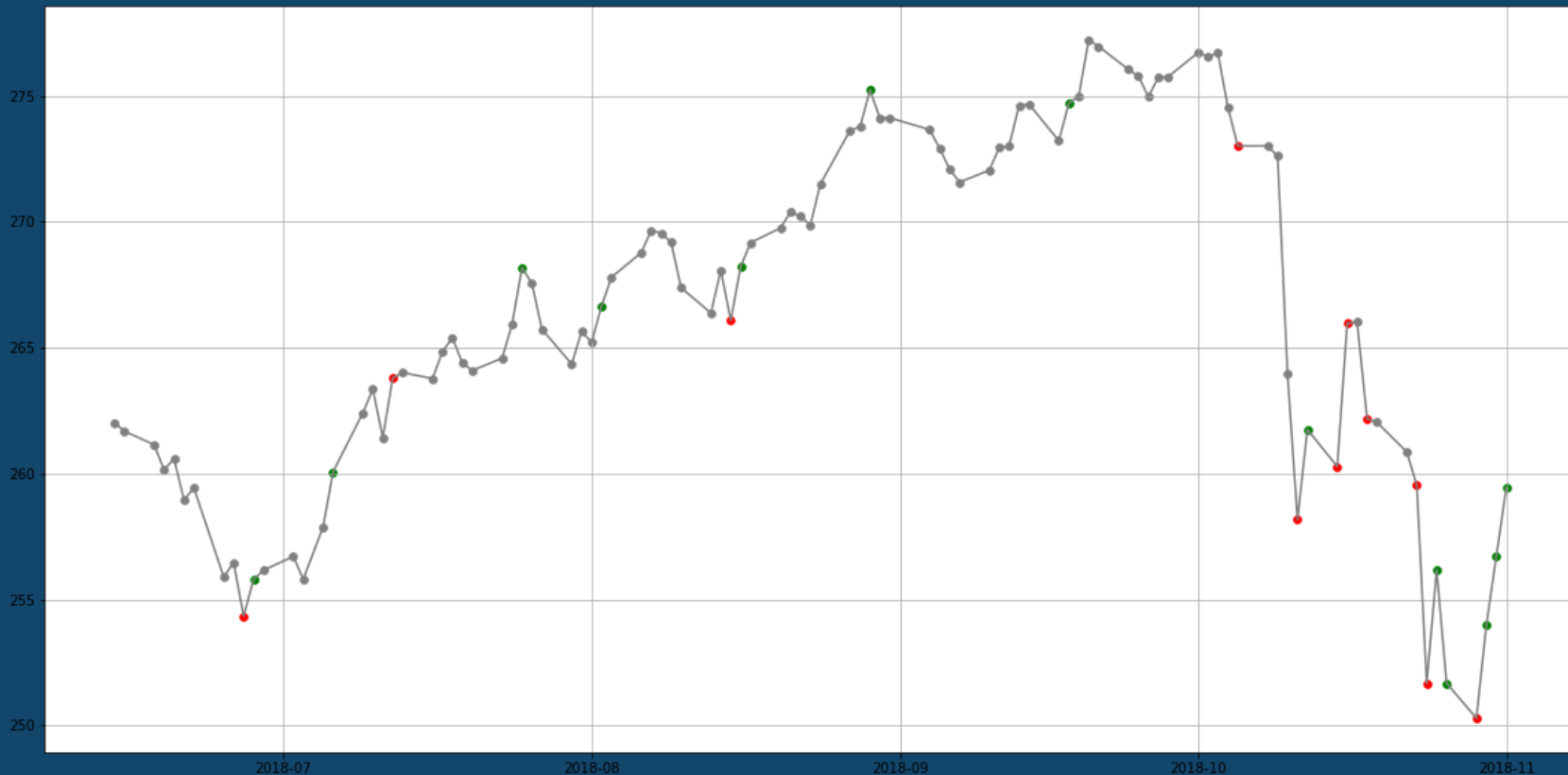


9.58% (Avg)
5.17% (Std)

بازده پیشبینی مدل (سبز رنگ) با بازه استراتژی خرید و نگهداری (آبی رنگ) در صد روز پایانی بازه زمانی 11/7/2016 تا 11/1/2018 (داده آزمایش) مورد مقایسه قرار گرفته است.



سیگنال‌های تولید شده توسط مدل را بر روی نمودار قیمت سهام SPY در صد روز پایانی بازه زمانی 11/7/2016 تا 11/1/2018 (داده آزمایشی). قابل ذکر است که سیگنال خرید با رنگ سبز، سیگنال فروش با رنگ قرمز و سیگنال نگهداری با رنگ خاکستری نمایش داده شده است.



بررسی سیگنال‌های تولید شده توسط مدل، بیانگر عملکرد مناسب مدل در مواجهه با تلاطمات قیمت می‌باشد.



جمع‌بندی، نتیجه‌گیری و پیشنهادات

06

جمع بندی و نتیجه گیری

هدف اصلی این پژوهش دستیابی به نرخ سود بالاتر در بازار سرمایه است و همانطور که اشاره شد، تلاش شده است بر اساس تغییرات قیمت سهامهای زیر مجموعه شاخص S&P500 تغییرات شاخص در روز معاملاتی بعد پیش بینی شود.

مدل ارائه شده در این پژوهش در سه مرحله و با استفاده از الگوریتم‌های جنگل تصادفی، شبکه عصبی، ماشین بردار پشتیبان و نزدیکترین همسایه به پیش‌بینی سیگنال خرید، نگهداری و یا فروش می‌پردازد. به این صورت که:

1. در ابتدا با استفاده از الگوریتم جنگل تصادفی صد سهم دارای بیشترین اهمیت در پیش‌بینی انتخاب می‌شوند،
2. سپس هر یک از الگوریتم‌های ذکر شده به پیش‌بینی سیگنال معاملاتی روز آتی می‌پردازند
3. در نهایت با استفاده از متا الگوریتم بگینگ سیگنال نهایی بر پایه سیگنال‌های تولید شده در مرحله قبل ایجاد می‌شود.

نتایج نشان دهنده این موضوع است که مدل پیش‌بینی از دقت مناسبی در تعیین سیگنال‌های روزانه برخوردار است که موجب شده است میزان بازده ناشی از معامله بر اساس سیگنال‌های روزانه تولید شده توسط مدل برای داده‌های آزمون، بیشتر از سود از ناشی استراتژی "خرید - نگهداری" باشد که بیانگر عملکرد مناسب مدل پیشنهادی است.



پیشنهادهات

1. پیشنهاد می‌شود برای توسعه‌ی این روش از داده‌های تکمیلی همچون دیگر شاخص‌های بازار سهام ایالات متحده، شاخص بازار سهام دیگر کشورها و اندیکاتورهای پیشرفته‌تر استفاده شود.
2. همچنین می‌توان با استفاده از سایر الگوریتم‌های داده کاوی و استفاده از روش‌های دیگر در جهت تعیین پارامترهای الگوریتم‌ها، عملکرد مدل را بهبود بخشید.
3. همچنین به‌کارگیری روش‌های دیگر متالگوریتم مثل Boosting ممکن است باعث بهبود نتایج پیش بینی شود.

با سپاس از توجه شما

منابع و مراجع

- | | |
|-----|--|
| [1] | مشاری، محمد؛ دیده خانی، حسین؛ خلیلی دامغانی، کاوه؛ عباسی، ابراهیم؛ "طراحی مدل هوشمند ترکیبی جهت پیش‌بینی نقاط طلایی قیمت سهام"، فصلنامه علمی پژوهشی دانش سرمایه‌گذاری، سال هشتم، شماره 29، صفحات 45-66، بهار 1398. |
| [2] | درودی، دیاکو؛ ابراهیمی، سید بابک؛ "ارائه‌ی روش هیبریدی نوین برای پیش‌بینی شاخص کل قیمت بورس اوراق بهادار"، تحقیقات مالی، دوره 18، شماره 4، صفحات 613-632، زمستان 1395. |
| [3] | رستگار، محمدعلی؛ صداقتی پور، امین؛ "ارائه سیستم معاملات الگوریتمی برای قرارداد آتی سکه طلا مبتنی بر داده‌های درون-روزی"، فصلنامه علمی پژوهشی دانش سرمایه‌گذاری، سال هفتم، شماره 28، صفحات 49-67، زمستان 1397. |
| [4] | هان، ژیاوی؛ کمبر، میشلین؛ پی، ژان؛ اسماعیلی، مهدی؛ داده‌کاوی (مفاهیم و تکنیک‌ها)، انتشارات نیازدانش، تهران، 1393. |

[5]	ERKARTAL, Bugra & Ozdamar, Linet. (2018). Generating Buy/Sell Signals for an Equity Share Using Machine Learning. Eurasian Journal of Business and Economics. 11. 85-105. 10.17015/ejbe.2018.022.04.
[6]	Mantri, J. K., Gahan, P., & Nayak, B. B. "Artificial neural networks--an application to stock market volatility. Soft-Computing in Capital Market: Research and Methods of Computational Finance for Measuring Risk of Financial Instruments", 179, 2014.
[7]	Burgess, A. N., & others. "A computational methodology for modelling the dynamics of statistical arbitrage", PHD Thesis: University of London, 2000.
[8]	Tilakaratne, C. D., Morris, S. A., Mammadov, M. A., & Hurst, C. P. "Predicting stock market index trading signals using neural networks", In Proceedings of the 14th Annual Global Finance Conference (GFC'07), Pages 171–179, 2007.
[9]	Manoj Thakur, Deepak Kumar. "A hybrid financial trading support system using multi-category classifiers and random forest", Applied Soft Computing, Volume 67, Pages 337-349, 2017.
[10]	Omer Berat Sezer, Ahmet Murat Ozbayoglu. "Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach", Applied Soft Computing, Volume 70, Pages 525-538, 2018.
[11]	Moroşan, Adrian. (2011). The relative strength index revisited. African journal of business management. 5. 5855.