



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده مهندسی برق و کامپیوتر



بهینه‌سازی پرتفولیو به روش میانگین-واریانس مارکویتز با استفاده از پیش‌بینی سهام مبتنی بر یادگیری ماشین

پایاننامه برای دریافت درجه کارشناسی در رشته مهندسی برق گرایش کنترل

نام

محمد عسکری

شماره دانشجویی

۸۱۰۱۹۸۴۴۱

استاد راهنما:

دکتر تورج عباسیان

شهریورماه ۱۴۰۴

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

تعهدنامه اصالت اثر

باسمه تعالی

اینجانب محمد عسکری تأیید می کنم که مطالب مندرج در این پایان نامه حاصل تلاش اینجانب است و به دستاوردهای پژوهشی دیگران که در این نوشته از آنها استفاده شده است مطابق مقررات ارجاع گردیده است. این پایان نامه قبلاً برای احراز هیچ مدرک هم سطح یا بالاتر ارائه نشده است.

کلیه حقوق مادی و معنوی این اثر متعلق به دانشکده فنی دانشگاه تهران می باشد.

نام و نام خانوادگی دانشجو :

محمد عسکری

امضای دانشجو :

محمد عسکری - ۱۴۰۴/۰۶/۱۸

تشکر و قدردانی:

با سپاس فراوان از جناب دکتر عباسیان که در مراحل طراحی و به ثمر رساندن پروژه همواره راهنمای توانمندی برای اینجانب بودند.

چکیده

این پروژه به بررسی بهینه‌سازی سبد سهام با استفاده از پیش‌بینی سری‌های زمانی و ترکیب روش‌های یادگیری عمیق با مدل کلاسیک مارکویتز پرداخته است. داده‌های تاریخی قیمت بسته‌شدن سهام در بازه‌ی ۲۰۱۳ تا ۲۰۱۸ جمع‌آوری و با استفاده از برآوردگر مقاوم Huber (Huber's M-estimator) از نویز پاک‌سازی شدند. سپس یک مدل ترکیبی شامل شبکه عصبی کانولوشنی (CNN) برای استخراج الگوهای محلی و شبکه حافظه کوتاه‌مدت دوسویه (BiLSTM) برای مدل‌سازی وابستگی‌های بلندمدت آموزش داده شد تا روند آتی قیمت‌ها را پیش‌بینی کند. بر اساس بازده پیش‌بینی‌شده، ۱۰ سهم برتر در هر بازه انتخاب و با استفاده از مدل میانگین-واریانس مارکویتز، پرتفوی بهینه تشکیل شد.

نتایج حاصل از ۱۰ بازه‌ی متوالی ۲۱ روزه نشان داد که پرتفوی با وزن مساوی سهام منتخب، به طور میانگین دارای بازده حدود ۹,۸٪ و **Sharpe Ratio** حدود ۱,۱ بوده است. در مقابل، پرتفوی بهینه‌ی مارکویتز میانگین بازده حدود ۱۳٪ و **Sharpe Ratio** حدود ۱,۶ را به دست آورد. در برخی بازه‌ها این اختلاف بسیار محسوس بود؛ به‌طور مثال در بازه‌ی هشتم، پرتفوی مساوی **Sharpe Ratio** برابر ۱,۴۶ داشت، در حالی که پرتفوی بهینه **Sharpe Ratio** برابر ۲,۰۰ به دست آورد. این تفاوت‌ها نشان می‌دهند که استفاده از بهینه‌سازی مارکویتز توانسته است بدون افزایش محسوس در ریسک، بازده بیشتری ایجاد کند.

به طور کلی، نتایج بیانگر آن است که انتخاب هوشمندانه‌ی وزن‌ها با مدل مارکویتز کارایی پرتفوی را نسبت به روش‌های ساده‌ی مساوی‌سازی وزن‌ها به‌طور قابل توجهی ارتقا می‌دهد. بنابراین ترکیب پیش‌بینی دقیق قیمت‌ها با بهینه‌سازی وزنی مارکویتز می‌تواند در عمل سبدهایی با عملکرد پایدارتر و نسبت بازده به ریسک مطلوب‌تر ایجاد نماید.

کلمات کلیدی:

Portfolio Optimization, Time Series Forecasting

Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (BiLSTM)

Robust Statistics (Huber's M-estimator), Markowitz Mean-Variance Model

فهرست مطالب

فصل ۱: مقدمه و بیان مساله	۱
مقدمه	۲
۱-۱- تاریخچه‌ای از موضوع تحقیق	۲
۱-۲- شرح مسئله تحقیق	۳
۱-۳- تعریف موضوع تحقیق	۳
۱-۴- اهداف و آرمان‌های کلی تحقیق	۴
۱-۵- روش انجام تحقیق	۴
۱-۶- ساختار پایان‌نامه	۵
فصل ۲: مفاهیم اولیه و پیش زمینه	۶
۲-۱- مقدمه	۷
۲-۲- بازار سرمایه و شاخص S&P 500	۸
۲-۲-۱- مقدمه‌ای بر بازار سهام و شاخص S&P 500	۸
۲-۲-۲- ویژگی‌های سری زمانی مالی	۹
2-3- پیش‌پردازش داده‌های مالی و آمار مقاوم	۹
2-4- شبکه‌های عصبی عمیق برای پیش‌بینی سری زمانی	۱۰
2-5- بهینه‌سازی سبد سهام و مدل مارکویتز	۱۰
2-6- معیارهای ارزیابی عملکرد	۱۲
2-7- خلاصه و جمع‌بندی	۱۴
فصل ۳: مدل‌سازی و شبیه‌سازی	۱۵

۱-۳- مقدمه..... ۱۶

۲-۳- روش پیشنهادی..... ۱۶

۳-۳- ابزارهای مورد نیاز..... ۱۸

۴-۳- نتایج بدست آمده..... ۱۸

۵-۳- خلاصه و جمع بندی..... ۲۵

فصل ۴: پیاده سازی مدل پیش بینی کننده..... ۲۶

1-4-مقدمه ۲۶

2-4-پیاده سازی مدل ها..... ۲۶

فصل ۵: جمع بندی و نتیجه گیری..... ۳۰

1-5-جمع بندی..... ۳۰

2-5.نتیجه گیری..... ۳۱

فصل ۱:

مقدمه و بیان مساله

مقدمه

در سال‌های اخیر، با رشد چشمگیر بازارهای مالی و افزایش تنوع ابزارهای سرمایه‌گذاری، انتخاب ترکیب بهینه‌ای از دارایی‌ها به یکی از چالش‌های مهم سرمایه‌گذاران و مدیران پورتفولیو تبدیل شده است. هدف اصلی سرمایه‌گذار، دستیابی به بیشترین بازده ممکن در عین کنترل و مدیریت ریسک است. نظریه‌ی کلاسیک سبد سهام مارکویتز (Mean-Variance) یکی از شناخته‌شده‌ترین رویکردها برای بهینه‌سازی پورتفولیو است که بر اساس برآورد میانگین بازده و واریانس دارایی‌ها، وزن‌های بهینه را تعیین می‌کند. با این حال، دقت این مدل به شدت وابسته به صحت داده‌های ورودی و برآوردهای بازده آتی است. از این رو، ترکیب مدل‌های پیش‌بینی قوی با الگوریتم‌های بهینه‌سازی می‌تواند منجر به بهبود چشمگیر عملکرد پورتفولیو شود. پیشرفت‌های اخیر در حوزه یادگیری عمیق، به‌ویژه مدل‌های ترکیبی CNN و BiLSTM، امکان استخراج ویژگی‌های پیچیده سری‌های زمانی مالی و پیش‌بینی دقیق‌تر قیمت‌ها را فراهم کرده است.

۱-۱- تاریخچه‌ای از موضوع تحقیق

در دهه ۱۹۵۰، هری مارکویتز نظریه مدرن سبد سهام را معرفی کرد که نقطه عطفی در مدیریت سرمایه‌گذاری به شمار می‌آید. پس از آن، پژوهش‌های متعددی به توسعه مدل‌های بهینه‌سازی ریسک-بازده پرداخته‌اند. از دهه ۲۰۱۰ به بعد، با رشد توان محاسباتی و توسعه الگوریتم‌های یادگیری عمیق، استفاده از مدل‌های پیش‌بینی سری زمانی در کنار بهینه‌سازی کلاسیک مورد توجه قرار گرفت. مدل‌های LSTM و BiLSTM به دلیل توانایی در مدل‌سازی وابستگی‌های بلندمدت، در پیش‌بینی قیمت سهام کاربرد گسترده‌ای یافته‌اند. اخیراً، ترکیب شبکه‌های کانولوشنی (CNN) برای استخراج ویژگی‌های محلی با LSTM یا BiLSTM برای درک وابستگی‌های زمانی، نتایج بهتری نسبت به مدل‌های تک‌لایه ارائه داده است.

۲-۱- شرح مسئله تحقیق

مسئله اصلی این تحقیق، بهینه‌سازی سبد سهام شاخص S&P500 در بازه زمانی ۲۱ روزه است، به‌گونه‌ای که با پیش‌بینی دقیق‌تر بازده آتی، نسبت شارپ افزایش یافته و ریسک کاهش یابد. برای این منظور، ابتدا باید قیمت‌های آتی سهام با استفاده از یک مدل هوش مصنوعی پیش‌بینی شوند. سپس، بر اساس بازده پیش‌بینی‌شده، تعدادی از سهام با بالاترین میانگین بازده انتخاب شده و وزن‌های بهینه آن‌ها با استفاده از مدل مارکوویتز تعیین می‌گردد. چالش اصلی، انتخاب مدلی است که علاوه بر دقت پیش‌بینی بالا، در برابر نویزهای موجود در داده‌های مالی مقاوم باشد.

۳-۱- تعریف موضوع تحقیق

این تحقیق بر توسعه و ارزیابی یک سیستم هوشمند مدیریت پورتفولیو مبتنی بر پیش‌بینی سری‌های زمانی مالی متمرکز است. در این سیستم چندین مدل یادگیری عمیق و شبکه‌های عصبی بازگشتی برای پیش‌بینی قیمت‌های آتی سهام شاخص S&P500 به کار گرفته و با یکدیگر مقایسه می‌شوند. چارچوب پیشنهادی شامل مراحل زیر است:

۱. پیش‌پردازش داده‌ها: استفاده از آمار مقاوم (Huber's M-estimator) برای کاهش نویز و داده‌های پرت، محاسبه بازده‌های ۲۱ روزه (returns) و نرمال‌سازی آن‌ها در دستورکار قرار گرفت.
۲. آموزش مدل‌های پیش‌بینی: پیاده‌سازی و ارزیابی چند معماری از جمله:
 - LSTM (مدل پایه‌ی بازگشتی)
 - BiLSTM (با وابستگی‌های دوطرفه زمانی)
 - CNN+BiLSTM (ترکیب استخراج ویژگی محلی با مدل بازگشتی)
۳. انتخاب سهام برتر: بر اساس دقت پیش‌بینی بازده‌ها و معیارهای خطای مختلف (MSE, MAE, SMAPE) سهام برتر شناسایی شده و برای تشکیل پورتفولیو انتخاب می‌شوند.
۴. بهینه‌سازی پورتفولیو: با استفاده از مدل Mean-Variance مارکوویتز، ترکیب بهینه‌ی سهام منتخب به دست می‌آید و عملکرد پورتفولیوهای ساخته‌شده بر اساس هر مدل پیش‌بینی با یکدیگر و همچنین با روش مرجع (مانند وزن‌دهی مساوی) مقایسه می‌شود.

۴-۱- اهداف و آرمان‌های کلی تحقیق

هدف اصلی:

افزایش بازده و کاهش ریسک سبد سهام از طریق بهبود فرآیند انتخاب و وزن‌دهی سهام و انتخاب سبدي که بیشترین نسبت بازده به ریسک را دارد.

اهداف فرعی:

ارزیابی و مقایسه دقت مدل‌های مختلف یادگیری عمیق

مقایسه عملکرد پورتفولیوی حاصل از این مدل‌ها با استراتژی‌های مرجع ساده مانند وزن مساوی

۵-۱- روش انجام تحقیق

در این تحقیق ابتدا داده‌های تاریخی قیمت بسته‌شدن سهام شاخص S&P500 گردآوری شد. پس از پاک‌سازی داده‌ها و حذف نمادهای ناقص، با استفاده از آمار مقاوم Huber's M-estimator سری قیمت‌ها هموارسازی و اثر نویز و داده‌های پرت کاهش یافت. سپس قیمت‌های اصلاح‌شده نرمال‌سازی شده و برای ایجاد داده‌های ورودی مدل، به توالی‌هایی با طول مشخص (پنجره زمانی ۸ روزه) تقسیم گردید. چندین مدل پیش‌بینی شامل LSTM، BiLSTM و CNN-BiLSTM آموزش داده شدند تا مسیر آتی قیمت سهام را پیش‌بینی کنند. بر اساس پیش‌بینی‌ها، بازده تحقق‌یافته در افق‌های ۲۱ روزه محاسبه شد و ۱۰ سهم برتر با بالاترین بازده پیش‌بینی‌شده انتخاب گردید. در ادامه، به‌منظور تشکیل پرتفوی بهینه، وزن‌های سهام منتخب با استفاده از مدل Mean-Variance مارکویتز تعیین شد. در نهایت، عملکرد مدل‌ها و پرتفوی‌های حاصل بر اساس شاخص‌هایی چون میانگین بازده، انحراف معیار و نسبت شارپ ارزیابی و با روش‌های مرجع مانند پرتفوی با وزن مساوی مقایسه گردید.

۶-۱- ساختار پایان نامه

در فصل دوم، تعاریف اساسی مربوط به حوزه‌ی مدیریت پورتفولیو، مفاهیم اولیه شبکه‌های عصبی عمیق، و مروری بر پژوهش‌های پیشین در زمینه پیش‌بینی سری‌های زمانی مالی و بهینه‌سازی سبد سهام ارائه خواهد شد.

فصل سوم شامل شرح کامل مدل پیشنهادی، معماری شبکه ها، روش پیش‌پردازش داده‌ها و الگوریتم بهینه‌سازی مارکویتز است.

در فصل چهارم، محیط پیاده‌سازی، جزئیات داده‌ها، تنظیمات آموزش، و نتایج آزمایش‌ها و مقایسه عملکرد مدل‌ها ارائه می‌شود.

در فصل پنجم، جمع‌بندی کلی، بیان دستاوردها، محدودیت‌ها و پیشنهادهایی برای ادامه کار در آینده مطرح خواهد شد. در نهایت، در فصل پنجم، نتیجه‌گیری‌های کلی حاصل شده در این تحقیق، پیاده‌سازی‌ها/نوآوری‌های/... انجام شده و محدودیت‌ها مورد بحث قرار می‌گیرد و پیشنهادهایی برای ادامه‌ی مسیر به علاقمندان این حوزه‌ی ارائه خواهد شد.

فصل ۲: مفاهیم اولیه و پیش زمینه

۱-۲- مقدمه

در سال‌های اخیر، موضوع پیش‌بینی قیمت سهام و بهینه‌سازی سبد سرمایه‌گذاری توجه بسیاری از پژوهشگران حوزه‌ی مالی و هوش مصنوعی را به خود جلب کرده است. ترکیب مدل‌های پیش‌بینی سری‌های زمانی با مدل‌های کلاسیک مدیریت ریسک و بازده مانند مارکویتز، یکی از رویکردهای پرکاربرد در این زمینه است.

به عنوان نمونه، [این مقاله](#) با استفاده از رویکردی ترکیبی شامل شبکه‌ی عصبی کانولوشنی (CNN) و حافظه‌ی کوتاه‌مدت دوسویه (BiLSTM)، قیمت‌های آینده‌ی سهام را پیش‌بینی کرده و سپس با مدل **Mean-Variance** مارکویتز وزن‌های بهینه‌ی پرتفوی را تعیین کردند. آن‌ها نشان دادند که پرتفوی‌های بهینه‌شده با این روش عملکرد بهتری نسبت به پرتفوی‌های وزن مساوی دارند.

در [مطالعه‌ای دیگر](#)، بازده سهام را با مدل‌های سری زمانی مانند ARIMA-GARCH و مدل‌های یادگیری ماشین نظیر XGBoost پیش‌بینی کرده و آن‌ها را در چارچوب مارکویتز به کار گرفتند. نتایج پژوهش آن‌ها تأیید کرد که پرتفوی مبتنی بر بازده‌های پیش‌بینی‌شده نسبت به پرتفوی سنتی مبتنی بر داده‌های تاریخی Sharpe Ratio بالاتری دارد.

از سوی دیگر، [این پژوهش](#) با معرفی مفهوم **Decision-Focused Learning** نشان داد که می‌توان فرآیند پیش‌بینی و بهینه‌سازی پرتفوی را به صورت یکپارچه طراحی کرد. در این رویکرد، مدل پیش‌بینی به گونه‌ای آموزش می‌بیند که مستقیماً معیار عملکرد پرتفوی (مانند نسبت شارپ) بهینه شود، نه صرفاً دقت پیش‌بینی بازده‌ها.

در همین راستا، [مقاله‌ی منتشرشده](#) در **SCIRP (2023)** ترکیب مدل‌های مختلف یادگیری عمیق مانند BiLSTM و BiGRU را برای پیش‌بینی به کار گرفت و نشان داد که استفاده از مدل‌های ترکیبی موجب بهبود پایداری پیش‌بینی‌ها و در نتیجه بهینه‌تر شدن پرتفوی مارکویتز می‌شود.

همچنین در [این پژوهش](#)، **Lu و Yang** با توسعه‌ی مدل **Black-Litterman** و ترکیب آن با روش‌های کاهش نویز و مدل‌های پیش‌بینی پیشرفته نظیر TCN نشان دادند که می‌توان کارایی سبد سهام را بیش از پیش ارتقاء داد. این پژوهش چشم‌انداز جدیدی را برای ترکیب مدل‌های کاهش نویز، پیش‌بینی سری‌های زمانی و بهینه‌سازی پرتفوی در اختیار قرار می‌دهد.

به طور کلی، این مطالعات نشان می‌دهند که ترکیب مدل‌های پیش‌بینی مبتنی بر یادگیری ماشین و یادگیری عمیق با مدل‌های کلاسیک مالی مانند مارکویتز و Black-Litterman می‌تواند رویکردی قدرتمند برای افزایش بازده و کاهش ریسک پرتفوی باشد. بر همین اساس، در این پژوهش نیز تلاش شده است با بهره‌گیری از ترکیب شبکه‌های عصبی عمیق و مدل مارکویتز، چارچوبی برای بهینه‌سازی پرتفولیو ارائه گردد.

برای درک کامل مراحل و روش‌های به‌کاررفته در این تحقیق، آشنایی با مفاهیم پایه در دو حوزه‌ی اصلی ضروری است: پیش‌بینی سری‌های زمانی مالی با استفاده از شبکه‌های عصبی عمیق و روش‌های کاهش نویز، و بهینه‌سازی سبد سهام با مدل میانگین-واریانس مارکویتز. در این فصل، ابتدا به معرفی مفاهیم مرتبط با بازار سرمایه و شاخص S&P500 پرداخته می‌شود. سپس تکنیک‌های آماری و مدل‌های یادگیری عمیق مورد استفاده (Huber's M-estimator، CNN، BiLSTM) معرفی می‌گردند. در ادامه، چارچوب بهینه‌سازی سبد سهام و معیارهای ارزیابی عملکرد پورتفولیو شرح داده می‌شود.

۲-۲- بازار سرمایه و شاخص S&P 500

۲-۲-۱- مقدمه‌ای بر بازار سهام و شاخص S&P 500

بازار سهام محلی است که در آن سرمایه‌گذاران به خرید و فروش اوراق بهادار می‌پردازند. شاخص S&P500 شامل ۵۰۰ شرکت بزرگ و فعال در بورس آمریکا است و به‌عنوان یکی از مهم‌ترین شاخص‌های عملکرد اقتصادی در سطح جهان شناخته می‌شود. تغییرات این شاخص نمایانگر وضعیت کلی بازار سرمایه آمریکا است.

۲-۲-۲- ویژگی‌های سری زمانی مالی

قیمت سهام یک سری زمانی است که در آن مقادیر متوالی در فواصل زمانی منظم ثبت شده‌اند. سری‌های زمانی مالی معمولاً دارای ویژگی‌هایی مانند نویز بالا، ناپایداری و وابستگی‌های زمانی پیچیده هستند که مدل‌سازی آن‌ها را دشوار می‌کند.

2-3 پیش‌پردازش داده‌های مالی و آمار مقاوم

2-3-1 نویز در داده‌های مالی

نویز ناشی از عوامل مختلفی همچون نوسانات کوتاه‌مدت، خطاهای ثبت داده یا معاملات غیرعادی می‌تواند باعث کاهش دقت مدل‌های پیش‌بینی شود.

2-3-2 آمار مقاوم و برآوردگر هابر (Huber's M-estimator)

آمار مقاوم شاخه‌ای از آمار است که هدف آن ارائه روش‌هایی برای برآورد پارامترها در حضور داده‌های پرت و نویز بالاست، به‌گونه‌ای که برآورد تحت تأثیر شدید مقادیر غیرعادی قرار نگیرد. برآوردگر هابر (Huber's M-estimator) یکی از مهم‌ترین ابزارهای این حوزه است که رویکردی بینابینی میان میانگین و میانه ارائه می‌دهد: در نواحی نزدیک به مرکز توزیع، مشابه میانگین عمل می‌کند (برای بهره‌مندی از کارایی بالا در داده‌های نرمال) و در نواحی دورتر، وزن مشاهدات پرت را کاهش می‌دهد تا مانند میانه نسبت به این نقاط مقاوم باشد.

تابع هزینه‌ی هابر به‌صورت زیر تعریف می‌شود:

$$L_{\delta}(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2 & \text{for } |y - f(x)| \leq \delta \\ \delta \cdot (|y - f(x)| - \frac{1}{2}\delta) & \text{otherwise} \end{cases}$$

که در آن k پارامتر آستانه است. هرچه کوچک‌تر باشد، برآوردگر به میانه نزدیک‌تر می‌شود و هرچه بزرگ‌تر باشد، به میانگین نزدیک‌تر خواهد شد.

مزیت اصلی این روش در کاربردهایی نظیر پیش‌بینی سری‌های زمانی مالی، کاهش اثر نوسانات ناگهانی و داده‌های پرت بر فرآیند آموزش مدل‌های یادگیری ماشین است، بدون آنکه اطلاعات مفید موجود در داده‌ها

از بین برود. در این تحقیق، با استفاده از Huber's M-estimator سری قیمت بسته‌شدن هموار شده و مقادیر robust حاصل با پنجره زمانی ۴ روز به‌عنوان ورودی مدل‌های پیش‌بینی مورد استفاده قرار گرفته‌اند.

2-4 شبکه‌های عصبی عمیق برای پیش‌بینی سری زمانی

2-4-1 شبکه عصبی کانولوشنی (CNN)

CNN با اعمال فیلترهای کانولوشنی روی داده‌ها، ویژگی‌های محلی و الگوهای کوتاه‌مدت را استخراج می‌کند. در حوزه مالی، CNN می‌تواند نوسانات کوتاه‌مدت قیمت را شناسایی کرده و به‌عنوان ورودی برای لایه‌های بعدی استفاده شود.

2-4-2 شبکه حافظه کوتاه‌مدت (LSTM) و نسخه‌ی دوسویه (BiLSTM)

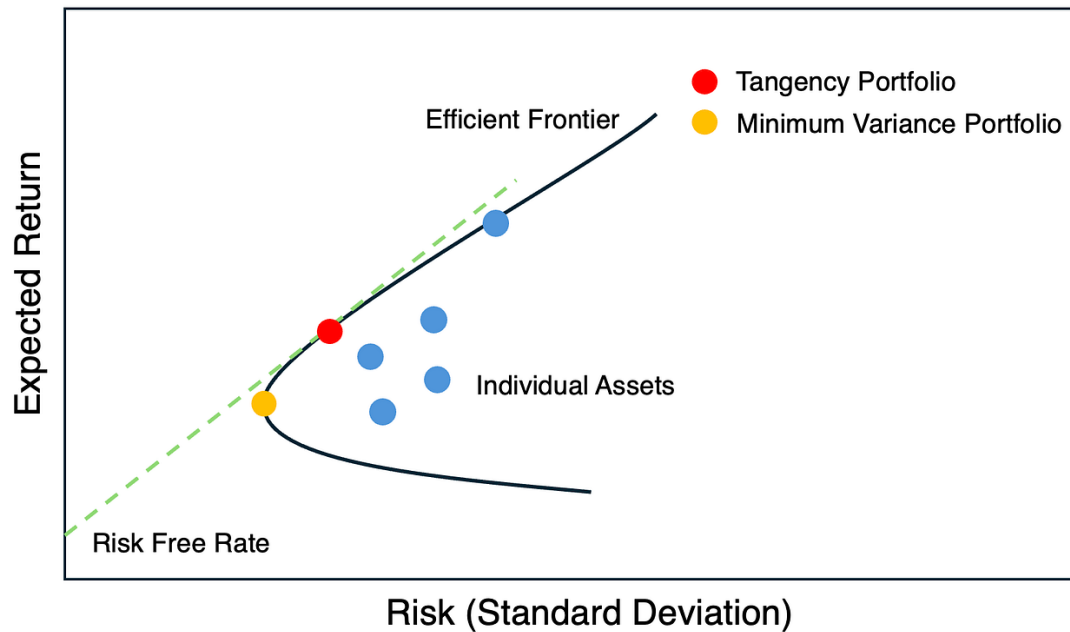
LSTM برای مدل‌سازی وابستگی‌های بلندمدت در داده‌های ترتیبی طراحی شده است و در پیش‌بینی سری‌های زمانی مالی کاربرد فراوان دارد. BiLSTM توسعه‌یافته‌ی LSTM است که داده‌ها را هم در جهت گذشته به آینده و هم در جهت آینده به گذشته پردازش می‌کند و به همین دلیل توانایی بیشتری در شناسایی الگوهای پیچیده دارد.

2-5 بهینه‌سازی سبد سهام و مدل مارکویتز

2-5-1 تئوری مدرن پورتفولیو

تئوری مدرن پورتفولیو که توسط هری مارکویتز در سال ۱۹۵۲ ارائه شد، چارچوبی ریاضی برای انتخاب ترکیب بهینه دارایی‌ها با هدف ایجاد تعادل میان بازده و ریسک است. در این رویکرد، بازده مورد انتظار پرتفوی به‌عنوان میانگین وزنی بازده دارایی‌ها و ریسک به‌عنوان واریانس یا انحراف معیار بازده‌ها سنجیده می‌شود. مدل میانگین-واریانس مارکویتز با استفاده از ماتریس کوواریانس بازده‌ها، اثر همبستگی بین دارایی‌ها را در نظر می‌گیرد و نشان می‌دهد که از طریق تنوع‌بخشی (Diversification) می‌توان ریسک

غیرسیستماتیک را کاهش داد. هدف اصلی این تئوری، یافتن مرز کارا (Efficient Frontier) است که مجموعه‌ای از پرتفوی‌ها با بیشترین بازده ممکن برای هر سطح ریسک یا کمترین ریسک برای هر سطح بازده را نشان می‌دهد



$$\min \sigma^2 = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \sigma_{ij}$$

$$\text{s.t.} \quad \sum_{i=1}^n w_i E(r_i) = E(r), \quad \sum_{i=1}^n w_i = 1$$

2-5-2. قیدها و سناریوهای بهینه‌سازی

در این تحقیق، قیدهای عدم منفی بودن وزن‌ها و مجموع یک بودن آن‌ها اعمال شده است. دو سناریوی اصلی مقایسه شده‌اند: (۱) مدل بهینه مارکویتز، (۲) وزن‌دهی مساوی

2-6 معیارهای ارزیابی عملکرد

میانگین بازده (Mean Return): میانگین بازده، یکی از ابتدایی‌ترین و مهم‌ترین معیارهای ارزیابی عملکرد پورتفولیو است که میانگین بازده‌های دوره‌ای (در پروژه ما ۲۱ روزه) را در یک بازه زمانی مشخص نشان می‌دهد. این شاخص، تصویر کلی از سودآوری سرمایه‌گذاری ارائه می‌کند و مبنایی برای مقایسه عملکرد پرتفوی‌ها یا دارایی‌های مختلف است. محاسبه آن به‌صورت مجموع بازده‌ها تقسیم بر تعداد دوره‌ها انجام می‌شود و معمولاً به‌صورت درصد بیان می‌گردد. هرچند میانگین بازده نمایانگر سطح کلی سود است، اما به‌تنهایی نمی‌تواند میزان نوسان یا ریسک سرمایه‌گذاری را نشان دهد.

$$Return = \frac{price_{close} - price_{open}}{price_{open}}$$

انحراف معیار (Risk): انحراف معیار در مدیریت سرمایه‌گذاری، به‌عنوان معیاری کمی برای سنجش میزان نوسان یا پراکندگی بازده‌های پورتفولیو نسبت به میانگین بازده به کار می‌رود. هرچه انحراف معیار بالاتر باشد، به معنای تغییرات شدیدتر بازده و در نتیجه ریسک بیشتر است. این شاخص به‌ویژه در تئوری مدرن پورتفولیو اهمیت زیادی دارد، زیرا بخشی از ارزیابی مرز کارا و انتخاب پرتفوی بهینه به تحلیل نوسانات وابسته است. انحراف معیار هم شامل نوسانات مثبت و هم منفی می‌شود و بنابراین به‌تنهایی توان تفکیک ریسک نامطلوب از مطلوب را ندارد. در عمل، برای محاسبه‌ی آن، ابتدا بازده‌های روزانه یا دوره‌ای را در طول بازه‌ی مورد نظر (مثلاً ۲۱ روز) محاسبه کرده و یک آرایه از این بازده‌ها تشکیل می‌دهیم؛ سپس انحراف معیار این آرایه به‌عنوان ریسک سبد در آن بازه در نظر گرفته می‌شود.

$$\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n (w_i * w_j * Cov(R_i, R_j))$$

نسبت شارپ (Sharpe Ratio): نسبت شارپ که توسط ویلیام شارپ معرفی شد، معیاری برای سنجش کارایی پورتفولیو با در نظر گرفتن ریسک است. این نسبت با تقسیم بازده مازاد (بازده پورتفولیو منهای نرخ بازده بدون ریسک) بر انحراف معیار بازده‌ها محاسبه می‌شود. مقدار بالاتر نسبت

شارپ نشان‌دهنده این است که سرمایه‌گذار به‌ازای هر واحد ریسک، بازده بیشتری کسب کرده است. از آنجا که این معیار ریسک را به‌طور مستقیم در نظر می‌گیرد، ابزاری کلیدی برای مقایسه کارایی سرمایه‌گذاری‌های مختلف با سطوح ریسک متفاوت محسوب می‌شود.

$$Shape\ ratio = \frac{Return_p}{\sigma_p}$$

میانگین مربعات خطا (Mean Squared Error – MSE): شاخصی متداول برای سنجش دقت مدل‌های رگرسیون که میانگین مربعات اختلاف مقادیر پیش‌بینی‌شده و واقعی را نشان می‌دهد. مقدار کمتر آن نشان‌دهنده خطای پیش‌بینی کمتر است.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

میانگین قدرمطلق خطا (Mean Absolute Error – MAE): میانگین قدرمطلق اختلاف بین مقادیر واقعی و پیش‌بینی‌شده را محاسبه می‌کند و در مقایسه با MSE حساسیت کمتری به داده‌های پرت دارد.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

درصد خطای متقارن (Symmetric Mean Absolute Percentage Error – SMAPE): SMAPE یکی از معیارهای پرکاربرد در ارزیابی مدل‌های پیش‌بینی سری زمانی است که خطای نسبی بین مقادیر واقعی و پیش‌بینی‌شده را به‌صورت درصدی بیان می‌کند. تفاوت اصلی SMAPE با MAPE در این است که برای نرمال‌سازی خطا، مجموع قدرمطلق مقدار واقعی و مقدار پیش‌بینی‌شده در مخرج قرار می‌گیرد، نه فقط مقدار واقعی؛ به همین دلیل، این معیار «متقارن» نامیده می‌شود و مشکل نامتقارن بودن MAPE را ندارد.

$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$

2-7 خلاصه و جمع‌بندی

در این فصل مفاهیم کلیدی شامل ساختار بازار سرمایه، ویژگی‌های سری زمانی مالی، آمار مقاوم، مدل‌های LSTM و CNN، ترکیب این مدل‌ها در قالب، تئوری مدرن پورتفولیو مارکوویتز و معیارهای ارزیابی عملکرد سبد سهام مرور شد. این مفاهیم، پیش‌نیاز درک روش پیشنهادی و نتایج ارائه‌شده در فصل‌های بعدی هستند.

فصل ۳: مدل سازی و شبیه سازی

۱-۳- مقدمه

در این فصل، ابتدا ساختار کلی سیستم پیشنهادی برای پیش‌بینی قیمت سهام و بهینه‌سازی سبد معرفی می‌شود. سپس معماری مدل ترکیبی **CNN+BiLSTM** و مراحل پیش‌پردازش داده‌ها تشریح خواهد شد. در ادامه، فرآیند انتخاب سهام برتر و بهینه‌سازی پورتفولیو با مدل **Mean-Variance** مارکویتز توضیح داده شده و ابزارها، معیارهای ارزیابی و نتایج آزمایش‌ها ارائه می‌شود.

۲-۳- روش پیشنهادی

در این تحقیق چارچوبی ترکیبی برای پیش‌بینی قیمت سهام و بهینه‌سازی پرتفوی ارائه می‌شود. این چارچوب شامل چهار مرحله اصلی است:

۱. پیش‌پردازش داده‌ها

- داده‌های قیمت بسته‌شدن سهام شاخص S&P500 از دیتاست جمع‌آوری شده است.
- به‌منظور کاهش اثر نویز و داده‌های پرت، از آمار مقاوم *Huber's M-estimator* استفاده شده و سری قیمتی «robust» تولید شده است.
- سپس داده‌ها نرمال‌سازی شده و به توالی‌های ثابت با طول پنجره ۸ روزه تقسیم گردیده‌اند تا به‌عنوان ورودی مدل‌های یادگیری عمیق مورد استفاده قرار گیرند.

۲. مدل‌سازی پیش‌بینی قیمت

- سه معماری مختلف یادگیری عمیق طراحی و آموزش داده شده‌اند:
- **LSTM** به‌عنوان مدل پایه برای مدل‌سازی وابستگی‌های زمانی بلندمدت.
- **BiLSTM** برای در نظر گرفتن وابستگی‌های دوطرفه در سری زمانی.
- **CNN-BiLSTM** ترکیبی از شبکه‌های کانولوشنی برای استخراج الگوهای محلی و **BiLSTM** برای وابستگی‌های زمانی

مدل **CNN-BiLSTM** با ترکیب قدرت استخراج الگوهای محلی شبکه‌های کانولوشنی و توانایی درک وابستگی‌های زمانی دوطرفه در **BiLSTM**، چارچوبی قدرتمند برای پیش‌بینی سری‌های زمانی مالی فراهم می‌کند. در حالی که مدل **BiLSTM** به تنهایی نیز عملکرد قابل‌قبولی ارائه داد، نتایج نشان داد که **CNN-BiLSTM** از نظر پایداری در برابر

نویز و توانایی تعمیم در شرایط متغیر بازار، برتری نسبی دارد. با توجه به ماهیت پرنوسان داده‌های مالی، استفاده از CNN-BiLSTM گزینه‌ای مناسب‌تر برای کاربردهای عملی در مدیریت پرتفویو محسوب می‌شود.

مدل	لایه‌های اصلی	تعداد نرون/فیلتر	تابع فعال‌سازی	توضیح ساختار
LSTM	LSTM → Dense(1)	LSTM(64)	tanh	یک لایه LSTM با ۶۴ واحد حافظه و خروجی یک‌بعدی؛ ساده‌ترین مدل بازگشتی
BiLSTM	Bidirectional(LSTM) → Dense(1)	LSTM(64) دوطرفه	tanh	وابستگی‌های زمانی هم از گذشته و هم آینده را یاد می‌گیرد
CNN-BiLSTM	Conv1D → BiLSTM → Dense(1)	فیلتر 32 Conv1D (kernel=3) + BiLSTM(64)	ReLU (Conv1D), tanh (LSTM)	ابتدا ویژگی‌های محلی را با CNN استخراج می‌کند و سپس وابستگی‌های زمانی را با BiLSTM مدل‌سازی می‌کند

آموزش مدل‌ها با استفاده از داده‌های ۸۰٪ آموزش و ۲۰٪ آزمون، در ۳۰ اپاک به همراه مکانیزم توقف زودهنگام (Early Stopping) انجام شده است.

۳. انتخاب سهام برتر

پس از پیش‌بینی سری قیمت‌ها، بازده‌های تحقق‌یافته در افق ۲۱ روزه محاسبه شده‌اند. در هر بازه، ۱۰ سهم با بالاترین بازده پیش‌بینی‌شده انتخاب شده و به‌عنوان گزینه‌های ورودی به مرحله بهینه‌سازی پرتفویو در نظر گرفته می‌شوند.

۴. بهینه‌سازی پرتفویو

از مدل کلاسیک **Mean-Variance** مارکویتز برای تعیین وزن‌های بهینه سهام منتخب استفاده شده است. قیدهای اعمال‌شده شامل: جمع وزن‌ها برابر با یک و عدم فروش استقرایی (وزن‌ها غیرمنفی).

عملکرد پرتفویهای بهینه شده بر اساس معیارهای میانگین بازده، انحراف معیار و نسبت شارپ ارزیابی و با استراتژیهای مرجع (پرتفوی با وزن مساوی و انتخاب تصادفی) مقایسه شده است.

۳-۳- ابزارهای مورد نیاز

برای پیاده سازی این مراحل، زبان برنامه نویسی **Python 3.10** به کار گرفته شد و از کتابخانه هایی نظیر **NumPy**، **TensorFlow/Keras**، **Pandas**، **scikit-learn** و **Matplotlib** جهت پردازش داده ها و طراحی مدل ها استفاده گردید. منبع داده ها دیتاست آماده ی شاخص S&P500 بوده و در بخش بهینه سازی پرتفولیو نیز کتابخانه های **NumPy** و **cvxopt** برای پیاده سازی مدل مارکویتز به کار گرفته شدند.

۳-۴- نتایج بدست آمده

برای ارزیابی دقیق عملکرد مدل، داده های بخش آزمون به صورت ۱۰ بازه ی متوالی، هر کدام به طول ۲۱ روز معاملاتی تقسیم بندی شدند. به این ترتیب، در هر بازه یک دوره ی مستقل از پیش بینی و تشکیل پرتفولیو بررسی گردید.

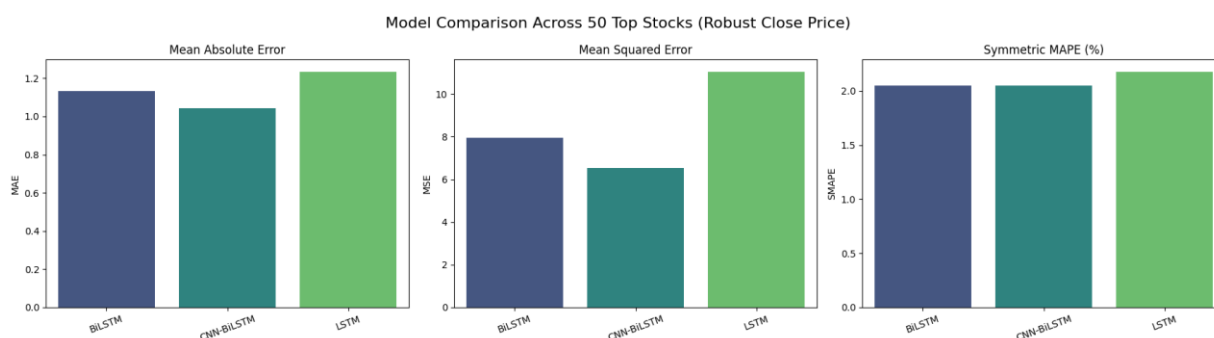
این تقسیم بندی دو مزیت اساسی داشت:

۱. پوشش بازه های زمانی متنوع: با تقسیم داده های آزمون به ۱۰ بخش، نتایج مدل در شرایط مختلف بازار (دوره های صعودی، نزولی و خنثی) سنجیده شد.
۲. امکان مقایسه ی بین بازه ای: عملکرد مدل ها و پرتفوی ها در هر بازه به صورت جداگانه ارزیابی و سپس با هم مقایسه شدند. این موضوع نشان داد که کارایی روش پیشنهادی تنها محدود به یک دوره خاص نبوده و در اکثر بازه ها پایدار باقی مانده است.

در فصل چهارم، نتایج حاصل از این ۱۰ بازه ی ۲۱ روزه به تفصیل گزارش شده و مقایسه ی پرتفولیو های مساوی، تصادفی و بهینه شده مارکویتز ارائه می شود.

۱. مقایسه دقت مدل های پیش بینی

مدل	MAE	MSE	SMAPE
BiLSTM	1.132393	7.9955113	2.049177
CNN-BiLSTM	1.042707	6.540083	2.047712
LSTM	1.235456	11.057618	2.180617

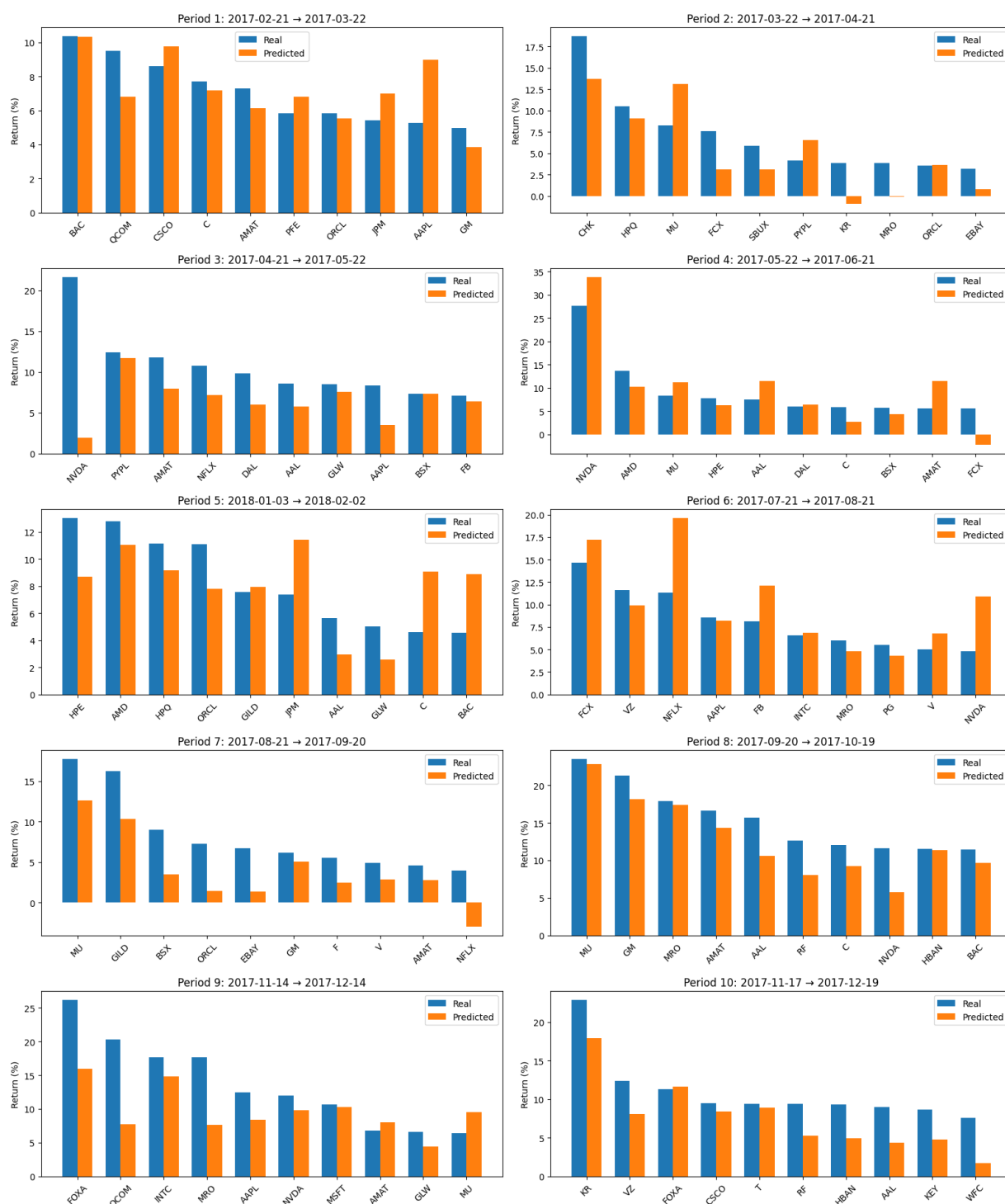


نتایج به دست آمده نشان می دهد که مدل های مختلف شبکه های بازگشتی عملکردهای متفاوتی در پیش بینی قیمت سهام داشته اند. مدل LSTM به عنوان ساده ترین معماری، توانسته روندهای کلی را یاد بگیرد اما دقت آن در مقایسه با سایر مدل ها پایین تر بوده است. مدل BiLSTM با استفاده از پردازش دوطرفه، روابط زمانی را دقیق تر استخراج کرده و بهبود کمی در معیارهای MSE و SMAPE داشته است. در ادامه، مدل CNN-BiLSTM با ترکیب توانایی استخراج ویژگی های محلی توسط CNN و قابلیت مدل سازی وابستگی های زمانی توسط BiLSTM، توانسته نتایجی پایدارتر و دقیق تر به ویژه در شاخص MAE ارائه دهد. این ویژگی نشان می دهد که CNN-BiLSTM نه تنها از نظر دقت پیش بینی برتری نسبی دارد، بلکه برای داده های مالی پرنوسان نیز گزینه ای کارآمد محسوب می شود. بنابراین، در مقایسه کلی میان مدل های بررسی شده، CNN-BiLSTM متوازن ترین و قابل اعتمادترین عملکرد را از خود نشان داده است.

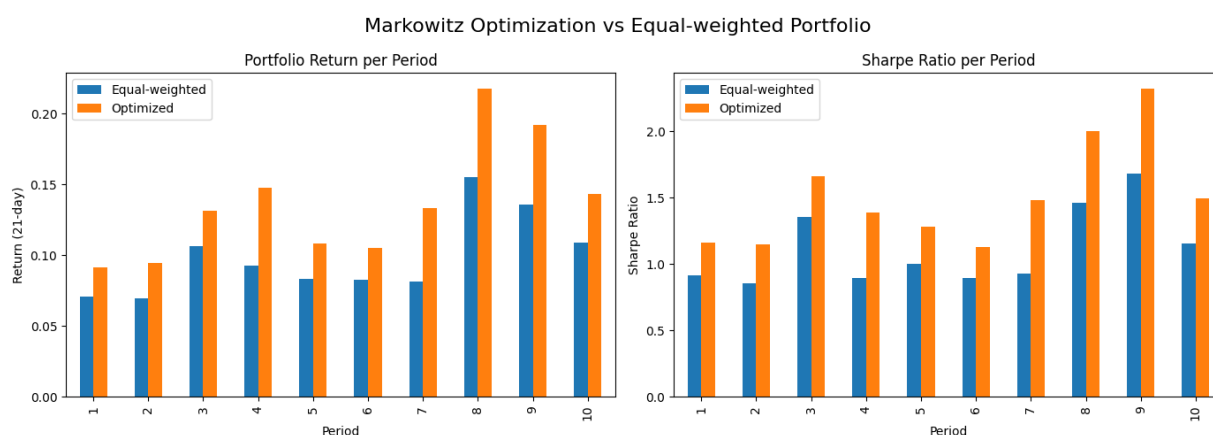
۲. نتایج پیش بینی مدل ها

در هر یک از ۱۰ بازه زمانی تست، نتایج پیش‌بینی بازده ۲۱ روزه توسط مدل‌های پیش‌بینی در کنار بازده واقعی این سهام ارائه شده است:

Top 10 Stocks by 21-Day Return (Real vs Predicted, per Period)



در ادامه به مقایسه ی بازده واقعی ۲۱ روزه سبد با اوزان مساوی و سبد بهینه سازی شده میپردازیم:



۳. مقایسه پورتفولیو ها

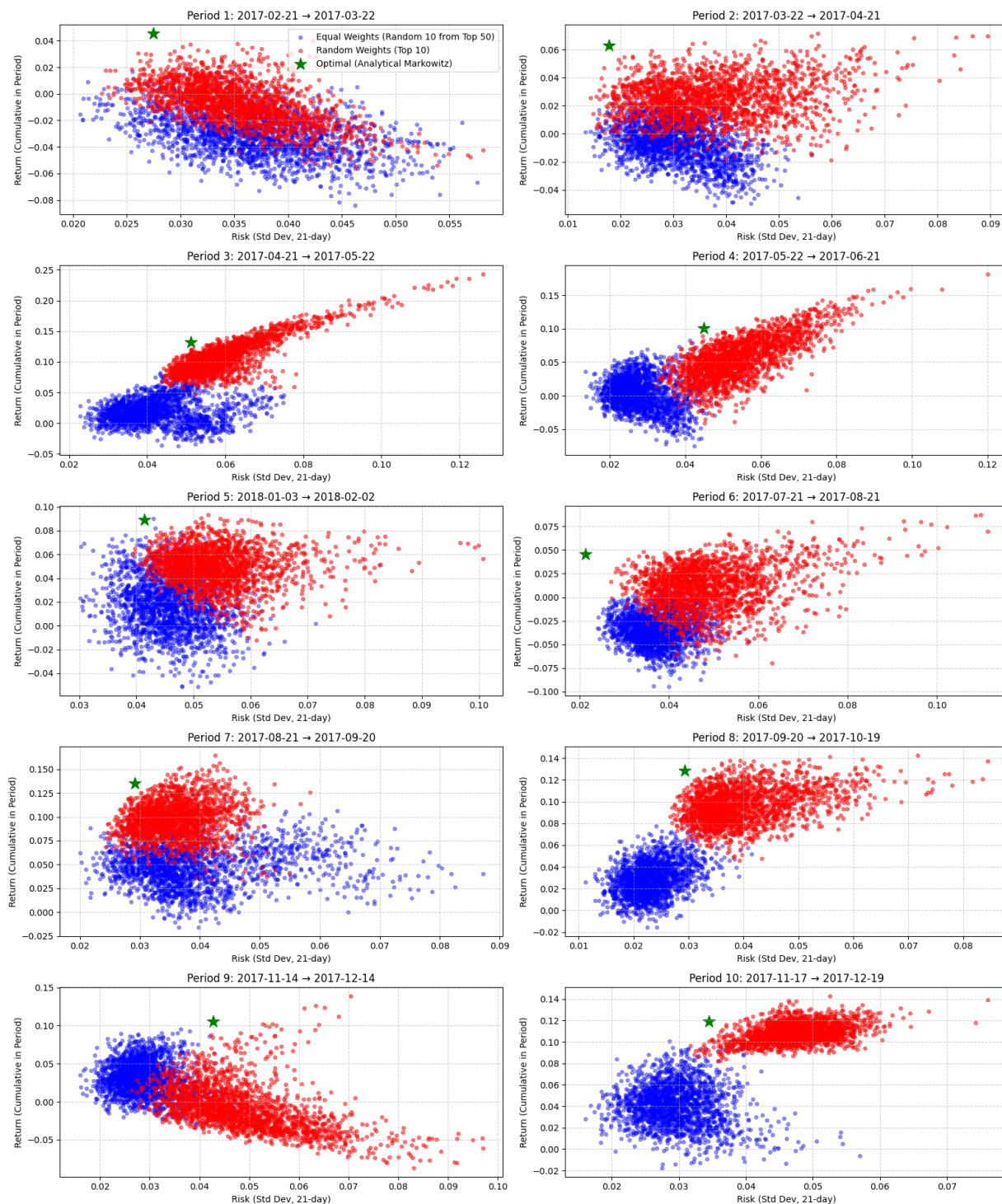
در این پژوهش در بازه های زمانی مختلف، دو دسته پرتفوی شبیه سازی و مقایسه شدند:

۱. پرتفوی های مساوی (آبی): شامل ۱۰ سهم تصادفی انتخاب شده از میان ۵۰ سهم با بیشترین حجم معاملات، با وزن های مساوی.
 ۲. پرتفوی های برتر (قرمز): شامل ۱۰ سهم برتر هر بازه که توسط مدل پیش بینی انتخاب شدند، با وزن های تصادفی.
- نتایج نشان می دهد که در اکثر بازه ها، سبدهای دسته دوم (قرمز) که بر اساس سهم های برتر تشکیل شده اند، بازده بالاتری نسبت به پرتفوی های دسته اول دارند. این موضوع بیانگر آن است که انتخاب سهام برتر می تواند نقش تعیین کننده ای در بهبود عملکرد پرتفوی داشته باشد.

علاوه بر این، برای هر بازه بهینه سازی پرتفوی بر اساس مدل مارکویتز نیز انجام شد. در این روش با استفاده از ماتریس کوواریانس بازده ها و بردار میانگین بازده مورد انتظار، پرتفویی بهینه با بیشترین نسبت شارپ (Max Sharpe Ratio) استخراج گردید. نتایج نشان داد که:

- پرتفوی بهینه مارکویتز (ستاره سبز) عموماً در ناحیه‌ای با ریسک کنترل شده و بازده بالاتر نسبت به پرتفوی‌های تصادفی قرار دارد.
- موقعیت این پرتفوی نسبت به سایر نقاط نشان می‌دهد که ترکیب بهینه وزن‌ها می‌تواند بازدهی را بدون افزایش قابل توجه ریسک بهبود دهد.
- مقایسه سه دسته پرتفوی (آبی، قرمز و سبز) حاکی از آن است که ترکیب استراتژی انتخاب سهام برتر و بهینه‌سازی وزنی (مدل مارکویتز) می‌تواند بهترین عملکرد را در اکثر بازه‌ها به همراه داشته باشد.

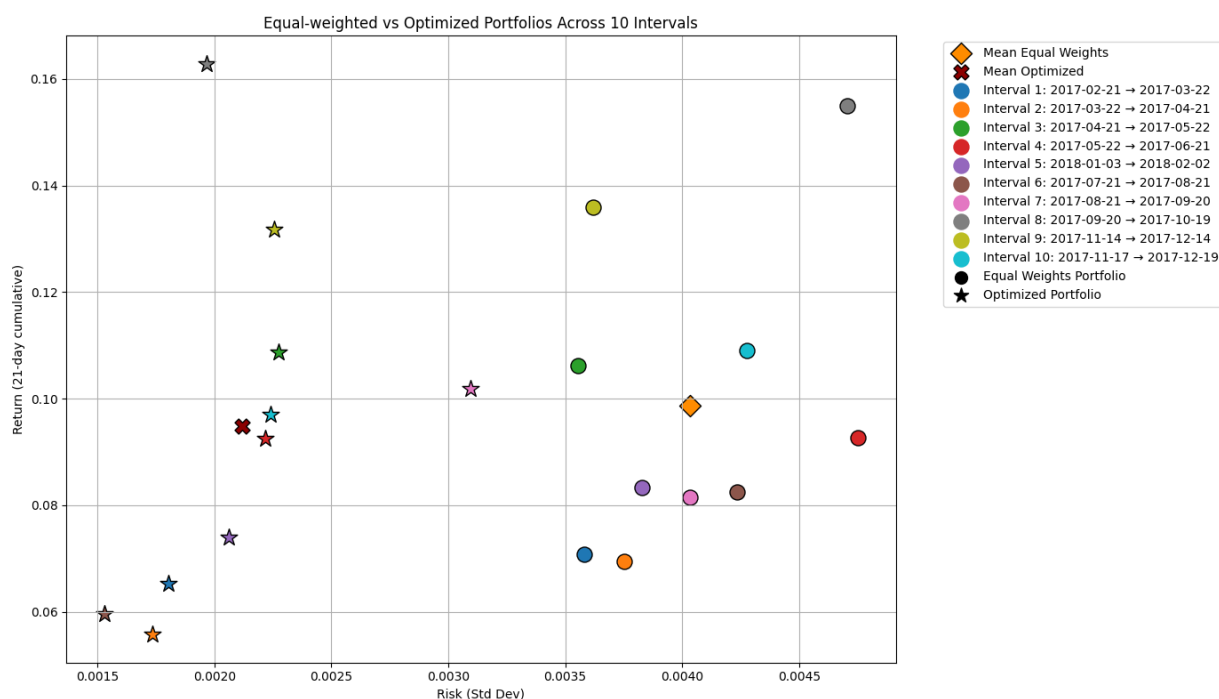
Portfolio Comparison with Analytical Markowitz Optimization (Cumulative Returns per Period)



۴. عملکرد پرتفولیو ها در ۱۰ بازه تاریخی

در شکل زیر مقایسه‌ی پرتفوی‌های مساوی (دایره‌های رنگی) و پرتفوی‌های بهینه‌شده با مدل مارکوویتز (ستاره‌های رنگی) در ۱۰ بازه‌ی زمانی نشان داده شده است. هر رنگ معرف یک بازه‌ی زمانی مشخص است و بازه‌ها در راهنمای نمودار (Legend) همراه با تاریخ شروع و پایان مشخص شده‌اند. همان‌طور که مشاهده می‌شود، تقریباً در تمام بازه‌ها موقعیت ستاره‌ها نسبت به دایره‌های هم‌رنگ بالاتر قرار دارد. این به آن معناست که پرتفوی بهینه‌شده با مدل مارکوویتز توانسته در همان سطح ریسک (انحراف معیار بازده‌ها) بازده بیشتری نسبت به پرتفوی مساوی کسب کند. در برخی بازه‌ها این اختلاف به‌ویژه محسوس است، مانند بازه‌های هشتم و نهم که بهبود نسبت شارپ به‌طور چشمگیری مشاهده می‌شود.

علاوه بر مقایسه در سطح هر بازه، میانگین کلی عملکرد دو رویکرد نیز روی نمودار نشان داده شده است. لوزی نارنجی میانگین پرتفوی‌های مساوی و ضربدر قرمز میانگین پرتفوی‌های بهینه‌شده است. فاصله‌ی این دو نقطه نشان می‌دهد که در افق کلی، پرتفوی‌های بهینه‌شده نه تنها بازده بالاتری داشته‌اند بلکه نسبت بازده به ریسک (Sharpe) نیز در آن‌ها مطلوب‌تر بوده است. به‌طور خلاصه، این نمودار تصویری روشن از برتری روش بهینه‌سازی مارکوویتز نسبت به تخصیص وزن مساوی ارائه می‌دهد و نشان می‌دهد که بهینه‌سازی وزنی می‌تواند در کنار انتخاب صحیح سهام، کارایی پرتفوی را به شکل معناداری افزایش دهد.



۵-۳- خلاصه و جمع بندی

در این فصل، روش پیشنهادی و معماری مدل **CNN + BiLSTM** برای پیش بینی قیمت سهام و بهینه سازی سبد معرفی شد. مراحل جمع آوری و پردازش داده ها، آموزش مدل، انتخاب سهام و بهینه سازی پورتفولیو تشریح گردید. نتایج اولیه نشان دهنده برتری روش پیشنهادی نسبت به مدل های پایه و روش های ساده تخصیص سرمایه است

فصل ۴: پیاده سازی مدل پیش بینی کننده

4-1 مقدمه

در این فصل، روند پیاده سازی مدل های پیش بینی قیمت سهام تشریح می شود. ابتدا مراحل آماده سازی داده ها بیان شده و سپس مدل های یادگیری عمیق شامل LSTM، BiLSTM و CNN-BiLSTM معرفی و پیاده سازی می شوند. در ادامه معیارهای ارزیابی عملکرد مدل ها تعریف شده و نتایج اجرای مدل ها روی داده های ۵۰ سهم منتخب گزارش می گردد.

4-2. پیاده سازی مدل ها

مدل های LSTM، BiLSTM و CNN-BiLSTM مطابق با مشخصات ذکر شده در فصل ۳، در محیط Python و با استفاده از کتابخانه های Keras و TensorFlow پیاده سازی شدند. برای آموزش مدل ها از تنظیمات زیر استفاده شد:

- بهینه ساز Adam: با نرخ یادگیری ۰,۰۰۱
- تعداد دوره های آموزشی: حداکثر ۳۰ اپک
- معیار توقف زودهنگام (Early Stopping) با آستانه ۸ اپک بدون بهبود
- تقسیم داده ها: ۸۰٪ آموزش، ۱۰٪ اعتبارسنجی، ۱۰٪ آزمون

برای جلوگیری از طولانی شدن آموزش، از **TqdmCallback** برای نظارت بر روند آموزش استفاده شد.

مدل‌ها روی داده‌های ۵۰ سهم پرجم اجرا شدند. برای هر سهم، قیمت‌های واقعی و پیش‌بینی‌شده در بازه‌ی آزمون استخراج شد و سه معیار MAE ، MSE و $SMAPE$ محاسبه گردید.

- مدل LSTM عملکرد پایه‌ی مناسبی داشت اما در برخی سهم‌ها دچار بیش‌برازش شد.
- مدل BiLSTM به دلیل بهره‌گیری از اطلاعات دوطرفه، بهبود نسبی در دقت پیش‌بینی ارائه کرد.
- مدل ترکیبی CNN-BiLSTM بهترین عملکرد را داشت.

به طور کلی، نتایج نشان دادند که ترکیب استخراج ویژگی‌های محلی با کانولوشن و یادگیری وابستگی‌های زمانی توسط BiLSTM باعث افزایش توانایی مدل در پیش‌بینی شد.

پس از انتخاب ۱۰ سهم برتر در هر بازه‌ی ۲۱ روزه بر اساس نتایج مدل پیش‌بینی، مرحله‌ی دوم فرآیند تشکیل پرتفوی، تعیین وزن‌های بهینه با استفاده از مدل مارکویتز بود.

در این پژوهش از نسخه‌ی کلاسیک مدل میانگین-واریانس مارکویتز استفاده شد. به طور مشخص، مسئله‌ی بهینه‌سازی به صورت زیر تعریف گردید:

$$\begin{aligned} & \underset{w}{\text{maximize}} && \frac{w^T \mu - r_f}{\sqrt{w^T \Sigma w}} \\ & \text{subject to} && \mathbf{1}^T w = 1, \quad w \geq 0, \end{aligned}$$

که در آن:

- w بردار وزن‌ها،
- μ بردار بازده مورد انتظار سهام،
- Σ ماتریس کوواریانس بازده‌ها،
- و قیدها بیانگر این هستند که مجموع وزن‌ها باید برابر با یک باشد و وزن‌ها منفی نشوند.

Sequential Least Squares Programming (SLSQP) برای حل این مسئله از الگوریتم

در کتابخانه‌ی SciPy استفاده شد. این الگوریتم به دلیل پشتیبانی از قیود خطی و کران‌ها گزینه‌ی مناسبی برای چنین مسائل محدب محسوب می‌شود.

پس از حل بهینه‌سازی، وزن‌های بهینه برای هر یک از ۱۰ سهم محاسبه گردید. سپس بازده و ریسک متناظر پرتفوی بهینه به دست آمد و به همراه پرتفوی مساوی (Equal Weights) روی نمودار مرز کارا (Efficient Frontier) ترسیم شد.

در نمودارها، پرتفوی‌های مساوی با دایره و پرتفوی‌های بهینه مارکوویتز با ستاره مشخص شدند. مقایسه‌ی این دو نشان داد که در اکثر بازه‌ها پرتفوی بهینه در موقعیتی بالاتر از پرتفوی مساوی قرار گرفت؛ به این معنا که با همان سطح ریسک توانست بازده بیشتری ایجاد کند.

در این پژوهش، بردار بازده مورد انتظار μ از بازده تجمعی در بازه‌ی ۲۱ روزه محاسبه شد و ماتریس کوواریانس Σ بر اساس بازده‌های روزانه و با مقیاس‌دهی به افق ۲۱ روزه استخراج گردید. حل این مسئله به کمک الگوریتم SLSQP در کتابخانه‌ی Scipy انجام شد.

همان‌طور که در جدول ۴-۲ مشاهده می‌شود، در تمامی ۱۰ بازه‌ی زمانی، پرتفوی بهینه‌شده با مدل مارکوویتز بازده و نسبت شارپ بالاتری نسبت به پرتفوی مساوی دارد، در حالی که سطح ریسک تقریباً یکسان باقی مانده است. این موضوع نشان می‌دهد استفاده از مدل مارکوویتز در تخصیص وزن‌ها منجر به افزایش کارایی پرتفوی می‌شود.

Period	Start_Date	End_Date	EqualW_Return	EqualW_Risk	EqualW_Sharpe	Opt_Return	Opt_Risk	Opt_Sharpe
1	2017-02-21	2017-03-22	0.070823	0.003582	0.912816	0.091155	0.003583	1.163618
2	2017-03-22	2017-04-21	0.069424	0.003753	0.854853	0.094185	0.003753	1.146148
3	2017-04-21	2017-05-22	0.106214	0.003556	1.356605	0.131327	0.003557	1.658447
4	2017-05-22	2017-06-21	0.092622	0.004752	0.891687	0.147451	0.004753	1.384856
5	2018-01-03	2018-02-02	0.083306	0.003829	0.998756	0.107882	0.003829	1.278926
6	2017-07-21	2017-08-21	0.082406	0.004236	0.893914	0.104952	0.004236	1.126658
7	2017-08-21	2017-09-20	0.081439	0.004035	0.927709	0.132996	0.004035	1.479982
8	2017-09-20	2017-10-19	0.154967	0.004708	1.464557	0.217675	0.004708	2.003716
9	2017-11-14	2017-12-14	0.135955	0.003618	1.684644	0.192065	0.003618	2.323596
10	2017-11-17	2017-12-19	0.109016	0.004277	1.156889	0.142884	0.004277	1.493619

فصل ۵

فصل ۵: جمع بندی و نتیجه گیری

5-1 جمع بندی

در این پژوهش ابتدا داده‌های پنج‌ساله‌ی ۵۰ سهم با بیشترین حجم معاملات مورد استفاده قرار گرفت. با استفاده از شبکه‌های عصبی بازگشتی LSTM، BiLSTM و (CNN-BiLSTM) تلاش شد تا روند قیمت‌های پایانی سهم‌ها پیش‌بینی شود. نتایج نشان داد که مدل‌های عمیق به‌ویژه ساختار CNN-BiLSTM توانستند عملکرد مناسبی در پیش‌بینی مقادیر داشته باشند و به طور میانگین خطاهای پیش‌بینی MAE، MSE و SMAPE در سطح قابل قبولی کاهش یافتند.

پس از مرحله‌ی پیش‌بینی، از خروجی مدل‌ها برای شناسایی ۱۰ سهم برتر در هر بازه‌ی زمانی استفاده شد. معیار انتخاب، بالاترین بازده پیش‌بینی‌شده در یک افق ۲۱ روزه بود. این مرحله

اهمیت ویژه‌ای داشت زیرا ترکیب صحیح سهام ورودی، پایه‌ی اصلی ساخت پرتفوی کارا محسوب می‌شود.

در گام بعد، سه دسته پرتفوی تشکیل شد:

۱. پرتفولیو با ۱۰ سهم تصادفی از میان ۵۰ سهم پر حجم با وزن مساوی.
۲. پرتفولیو با ۱۰ سهم برتر انتخاب‌شده توسط مدل پیش‌بینی با وزن تصادفی.
۳. پرتفولیو بهینه بر اساس مدل مارکویتز که وزن‌های آن با حل یک مسئله بهینه‌سازی تعیین شد.

نتایج شبیه‌سازی‌ها نشان داد که پرتفوی‌های دسته دوم (۱۰ سهم برتر با وزن تصادفی) عموماً عملکرد بهتری نسبت به پرتفوی‌های دسته اول داشتند. این موضوع بیانگر نقش تعیین‌کننده‌ی انتخاب درست سهام در عملکرد کلی پرتفوی است. همچنین پرتفوی‌های بهینه‌شده با مدل مارکویتز (دسته سوم) نسبت به دو دسته‌ی دیگر از نظر نسبت شارپ موقعیت بهتری داشتند.

5-2. نتیجه‌گیری

مقایسه‌ی نتایج سه دسته پرتفوی نشان داد که:

- پرتفولیو‌های مساوی با انتخاب تصادفی (آبی) معمولاً در نواحی پایین‌تر مرز کارا قرار می‌گیرند؛ بازده پایین‌تر و ریسک بالاتر دارند.
- پرتفولیو‌های تصادفی بر مبنای ۱۰ سهم برتر (قرمز) به دلیل کیفیت انتخاب سهام، عملکرد بهتری داشتند و در برخی بازه‌ها توانستند به بازدهی بین ۱۰ تا ۲۰ درصد برسند.
- پرتفولیو‌های بهینه‌شده با مارکویتز (سبز) بهترین عملکرد را داشتند. این پرتفولی‌ها در بیشتر بازه‌ها نسبت شارپ بالاتری را نسبت به سایر پرتفوی‌ها نشان دادند، به این معنا که سرمایه‌گذار می‌توانست با سطح ریسک مشابه بازده بیشتری کسب کند.

به طور خلاصه، می‌توان نتیجه گرفت که ترکیب دو رویکرد انتخاب سهام بر اساس یادگیری ماشین و تخصیص وزن بهینه با مدل مارکویتز منجر به ایجاد پرتفویی کارا تر نسبت به روش‌های ساده‌ی تصادفی می‌شود. اگر تنها از یکی از این دو رویکرد استفاده می‌شد (فقط انتخاب سهام یا فقط بهینه‌سازی وزن‌ها)، نتایج به این میزان بهبود نمی‌یافت.

1. Chaweewanchon, A., & Chaysiri, R. (2022). Markowitz Mean-Variance Portfolio Optimization with Predictive Stock Selection Using Machine Learning. *Journal of Risk and Financial Management*, 15(3), 64. MDPI.
<https://www.mdpi.com/2227-7072/10/3/64>
2. Ślusarczyk, D., & Ślepaczuk, R. (2025). Optimal Markowitz portfolio using returns forecasted with time series and machine learning models. *Journal of Big Data*. SpringerOpen.
<https://journalofbigdata.springeropen.com/articles/10.1186/s40537-025-01164-z>
3. Wang, Z., et al. (2025). Return Prediction for Mean-Variance Portfolio Selection: How Decision-Focused Learning Shapes Forecasting Models. *arXiv preprint*.
<https://arxiv.org/html/2409.09684v3>
4. (2023). Hybrid Data-Driven and Deep Learning Based Portfolio Optimization. *Journal of Data Analysis and Information Processing*. SCIRP.
<https://www.scirp.org/journal/paperinformation?paperid=133998>
5. Yang, Z., & Lu, K. (2025). Enhancing Black-Litterman Portfolio via Hybrid Forecasting Model Combining Multivariate Decomposition and Noise Reduction. *arXiv preprint*. <https://arxiv.org/abs/2505.01781>



University of Tehran



College of Engineering

School of Electrical and Computer Engineering

**Markowitz Mean-Variance Portfolio Optimization Based on Machine
Learning-Driven Stock Price Prediction**

A thesis submitted to the Undergraduate Studies Office

In partial fulfillment of the requirements for

The degree of bachelor in

Electrical Engineering

By:

Mohammad Askari

Supervisor:

Dr Touraj Abbasian