

مجتبی حاجی غلامی^۱، پریسا دانشجو^۲ پیش بینی قیمت سهام با استفاده از روش داده کاوی

^۱ مجتبی حاجی غلامی، آزاد تهران غرب / دانشکده فنی، m.hajigholami@yahoo.com

^۲ پریسا دانشجو، آزاد تهران غرب / دانشکده فنی، daneshjoo.p@wtiau.ac.ir

چکیده

زمینه و هدف:

مقاله حاضر به بررسی گونه‌های متنوعی از روش‌های داده‌کاوی می‌پردازد که در حوزه تجزیه و تحلیل بازارهای مالی کاربرد دارند و به طور ویژه تمرکز خود را بر پیش‌بینی تحرکات و روندهای بازار سهام معطوف می‌دارد. مطالعه ما نشان می‌دهد، از آنجایی که بازارهای مالی پویا و متغیر هستند و همواره تحت تأثیر عوامل اقتصادی، سیاسی و اجتماعی قرار دارند، استفاده از روش‌های یادگیری ماشینی و داده‌کاوی می‌تواند در پیش‌بینی دقیق‌تر حرکات قیمتی سهام مؤثر واقع شود.

با توجه به داده‌های گسترده و پیچیده موجود در بازارهای مالی، روش‌های داده‌کاوی می‌توانند ظرفیت فراوانی در کشف الگوهای پنهان و تعیین ارتباط میان متغیرهای مختلف داشته باشند. در این راستا، الگوریتم‌های مختلف یادگیری ماشینی مانند شبکه‌های عصبی مصنوعی، ماشین‌های بردار پشتیبان و جنگل‌های تصادفی به عنوان مثال مورد استفاده قرار می‌گیرند و در کنار تجزیه و تحلیل‌های آماری، به بهبود قابلیت‌های تحلیل‌گران و سرمایه‌گذاران در تصمیم‌گیری‌های اقتصادی کمک می‌کنند.

استفاده از داده‌های بزرگ و تحلیل‌های پیچیده همچنین در توسعه راهبردهای معاملاتی هوشمندانه که می‌توانند به بهینه‌سازی بازگشت سرمایه کمک کنند، تأثیری بسزا داشته‌است. برای نمونه، تحلیل‌گران می‌توانند با ادغام داده‌های احساسی حاصل از شبکه‌های اجتماعی در مدل‌های پیش‌بینی خود، دقت تخمین‌های خود را ارتقا دهند. این مطالعه تأکید دارد که توسعه پایدار در بازارهای مالی نیازمند درک عمیق‌تر و تحلیلی دقیق‌تر از داده‌ها است که در نهایت منجر به تصمیم‌گیری‌های مبتنی بر داده و فرآیندهای معاملاتی قوی و مستحکم‌تر می‌شود.

کلید واژه‌ها:

توسعه پایدار، داده‌کاوی، داده‌های بزرگ، شبکه عصبی مصنوعی، یادگیری ماشینی

۱-۱. مقدمه

پیش‌بینی بازار سهام با استفاده از تکنیک‌های داده‌کاوی، توجه دانش‌گامیان، محققان و سرمایه‌گذاران را به خود جلب کرده است. بازار سهام حجمی عظیم از داده‌های ارزشمند را تولید می‌کند. به‌منظور ارائه اطلاعات مفید به سرمایه‌گذاران برای تصمیم‌گیری، پیگیری روش‌های مؤثر برای تجزیه و تحلیل و خلاصه کردن داده‌های بازار سهام به عنوان یک چالش اساسی در عصر حاضر مطرح شده است. بازار سهام آشفته است و داده‌های آن به دلیل آسیب‌پذیری چشم‌انداز سرمایه‌گذاران، عوامل سیاسی، ساختار اقتصادی و بلایای طبیعی، آشفته، غیرخطی و غیر ایستا است. برای مواجهه با چنین داده‌هایی، روش‌های داده‌کاوی موثرترین فناوری شناخته شده‌اند. این روش‌ها برای کمک به محققان در کشف الگوهای پنهان و ناشناخته طراحی شده‌اند. دسترسی و فراوانی داده‌های مالی مستلزم استفاده از ابزارهای داده‌کاوی در تحقیقات بازار سهام است. داده‌کاوی هنر و علم تحلیل پایگاه‌های داده برای یافتن روندها و ناهنجاری‌ها است و عنصری از کشف دانش گسترده‌تر در فرآیند پایگاه‌های داده (KDD) است. یک فرآیند تکراری است که شامل جمع‌آوری داده‌ها، تبدیل آن، کاهش داده‌ها و سپس داده‌کاوی می‌شود. این روش را می‌توان به صورت رسمی به صورت «فرآیند سازمان‌یافته شناسایی الگوهای معتبر، بدیع، مفید و قابل درک از مجموعه داده‌های بزرگ و پیچیده» تعریف کرد. داده‌کاوی شامل استخراج اطلاعات از داده‌های موجود با بررسی الگوها و روابط بین ویژگی‌ها با استفاده از روش‌هایی مانند الگوریتم‌های یادگیری ماشین² (ML) است. نوع الگوریتم داده‌کاوی که باید استفاده شود، بر اساس وضعیت موجود و طبیعت داده‌های موجود تعیین می‌شود. از آنجا که روش کشف دانش در پایگاه داده‌ها (KDD) تکراری است، برخی از فرایندها ممکن است نیاز به اعمال مجدد و چندین بار تجدیدنظر داشته باشند تا نتایج خوبی به دست آیند. روش‌های سنتی محدودیت‌های مختلفی، از جمله عدم کارایی در مدیریت مجموعه داده‌های بزرگ و متغیرهای چند بعدی، مدل‌های پیش‌بینی و برآورد ضرایب عوامل دارند.

این تکنیک‌ها، به جزء اساسی تحقیقات پیش‌بینی سهام در سال‌های اخیر تبدیل شده‌اند. به‌طور سنتی، محققان از تکنیک‌های مرسوم پیش‌بینی مانند میانگین متحرک، هموارسازی نمایی، رگرسیون خطی و مدل‌های میانگین متحرک تلفیقی اتورگرسیون (arima) برای تصمیم‌گیری استفاده می‌کردند. با این حال، پیش‌بینی مبتنی بر تکنیک داده‌کاوی، جایگزین قدرتمندی برای روش‌های سنتی شده است. تکنیک‌های سنتی نتایج معناداری را در مورد مؤلفه‌های قابل شمارش و ابعاد پیش‌بینی عددی متناهی ارائه می‌دهند، اما داده‌های

انبوه و چندبعدی فقط می‌توانند با استفاده از تکنیک‌های داده‌کاوی انجام شوند. تکنیک‌های داده‌کاوی نسبت به روش‌های سنتی در پیش‌بینی پیچیده بهتر هستند، زیرا این روش‌ها توانایی بهبود دقت پیش‌بینی‌ها را دارند و تابع زیان را به حداقل می‌رسانند که با مجموع مربعات خطا در پیش‌بینی محاسبه می‌شود.

1. Knowledge Discovery in Databases (KDD)
2. Machine Learning (ML)
3. Artificial Neural Network (ANN)
4. Support Vector Machine (SVM)
5. Principal Component Analysis (PCA)

یکی از مهم‌ترین وظایف مالی داده‌کاوی، پیش‌بینی قیمت سهام است. این امر مستلزم تحلیل روندهای بازار، تدوین راهبردهای سرمایه‌گذاری، تعیین زمان بهینه خرید سهام و تصمیم‌گیری نهایی برای خرید سهام است. تصمیم‌گیرندگان به شدت علاقه‌مند به پیش‌بینی روند شاخص بازار سهام و روندهای قیمت سهام برای ایجاد معاملات سودآور هستند. هدف از پیش‌بینی کاهش ریسک (خطر) تصمیم‌گیری مالی بد است. با این حال، پیش‌بینی بازارهای سهام یک روش چالش‌برانگیز است، زیرا دینامیک بازار، بحرانی و اساساً غیر خطی است. اگرچه تجزیه و تحلیل رفتار حرکت سهام یک کار چالش‌برانگیز است، تکنیک‌های داده‌کاوی می‌توانند سرمایه‌گذار را راهنمایی کنند تا تصمیمات سرمایه‌گذاری بهتری بگیرد.

۲-۱. اهمیت و ضرورت انجام تحقیق

با توجه به دسترسی به طیف وسیعی از تکنیک‌های داده‌کاوی، ابزارهای پردازش و تکنیک‌های مختلف برای پیامدها، نیاز به شناخت تکنیک‌های مناسب و قابل‌اعتماد در حوزه پیش‌بینی سهام و درک نقاط قوت و ضعف داده‌کاوی وجود دارد. علاوه بر این، کشف انواع داده‌های ورودی، روش‌های پردازش و معیارهای خروجی برای کمک به سرمایه‌گذاران، محققان و دانشگاهیان نیز لازم است. مطالعات قبلی با هدف تعیین تکنیک‌های دقیق داده‌کاوی انجام شده است، اما اطلاعاتی بیشتر در مورد کاربرد این تکنیک‌ها مانند پیش‌بینی سودآوری، حرکت کوانتومی سهام و استفاده از مدل‌های ترکیبی می‌توان استخراج کرد. بنابراین، پژوهش حاضر با هدف بررسی پتانسیل (ظرفیت) تکنیک‌های داده‌کاوی انجام شده است و بدین وسیله دانشگاهیان، سرمایه‌گذاران و محققان را قادر می‌سازد تصمیمات بهتری اتخاذ کنند.

۳-۱. جنبه جدید بودن و نوآوری در تحقیق

آخرین تحقیقات و بررسی‌ها نشان می‌دهند که تکنیک‌های جدید و پیشرفته‌تری از داده‌کاوی، از جمله یادگیری عمیق و یادگیری ماشین برای تحلیل بازار سهام استفاده می‌شوند. این تکنیک‌ها قادر به کشف الگوهای پیچیده‌تر و غیرخطی در داده‌ها هستند که می‌تواند به پیش‌بینی دقیق‌تر و موثرتر بازار سهام منجر شود.

مثال‌هایی از نوآوری‌ها :

- استفاده از شبکه‌های عصبی کانولوشنی (CNN) و شبکه‌های عصبی بازگشتی (RNN) برای تحلیل توالی زمانی قیمت سهام.
 - استفاده از تکنیک‌های تقویت یادگیری (Reinforcement Learning) برای تصمیم‌گیری در مورد خرید و فروش سهام.
 - استفاده از داده‌های بزرگ و انواع مختلف داده‌ها مانند اخبار، شبکه‌های اجتماعی و داده‌های مالی برای تحلیل و پیش‌بینی بازار سهام.
- به طور کلی، نوآوری در این حوزه به معنای استفاده از تکنیک‌های پیشرفته‌تر و داده‌های متنوع‌تر برای تحلیل و پیش‌بینی دقیق و صحیح‌تر بازار سهام است. با گسترش فناوری و دسترسی به داده‌های بیشتر، این حوزه همچنان در حال توسعه و پیشرفت است.

۴-۱. اهداف مشخص تحقیق

۱. بررسی و ارزیابی کارایی تکنیک‌های مختلف داده‌کاوی در پیش‌بینی بازار سهام.
۲. توسعه و ارائه مدل‌های پیشرفته‌تر بر اساس تکنیک‌های داده‌کاوی برای پیش‌بینی بازار سهام بر اساس داده‌های موجود.
۳. تأکید بر نقش داده‌های بزرگ و متنوع در افزایش دقت پیش‌بینی‌ها.
۴. مقایسه تکنیک‌های داده‌کاوی مبتنی بر یادگیری عمیق با روش‌های سنتی تر.
۵. کشف الگوها و روابط پنهان در داده‌های بازار سهام که ممکن است برای سایر تحلیل‌گران و سرمایه‌گذاران غیرقابل تشخیص باشد.

در عصر کنونی، با پیشرفت‌های چشمگیر در فناوری اطلاعات و تجمع حجم انبوهی از داده‌ها، روش‌های نوینی برای تحلیل و درک بازارهای مالی به وجود آمده‌اند. یکی از این روش‌ها، داده‌کاوی است که به عنوان ابزاری

قدرتمند در پیش‌بینی حرکت‌های بازار سهام شناخته می‌شود. توانایی تشخیص الگوهای پیچیده و پیش‌بینی روندهای آتی در داده‌های گسترده، داده‌کاوی را تبدیل به حوزه‌ای جذاب و مورد توجه در مطالعات مالی کرده‌است.

با توجه به اهمیت ویژه پیش‌بینی رفتار بازار سهام در توسعه سیاست‌های سرمایه‌گذاری و مدیریت ریسک، این تحقیق قصد دارد تا با مروری جامع بر ادبیات موجود، پژوهش‌هایی را که به استفاده از تکنیک‌های داده‌کاوی برای پیش‌بینی قیمت‌های سهام پرداخته‌اند، بررسی نماید. این بررسی شامل تحلیل عملکرد الگوریتم‌های مختلف، از جمله شبکه‌های عصبی، ماشین‌های بردار پشتیبان و درخت تصمیم‌گیری، در بازارهای مالی خواهد بود. همچنین، رونق استفاده از داده‌های بزرگ و تکنیک‌های یادگیری ماشین در این حوزه بیانگر نیاز به تحلیل دقیق‌تر و عمیق‌تر این ابزارها و ارزیابی پتانسیل (ظرفیت) آن‌ها برای بهبود راهبردهای معاملاتی است.

علاوه بر این، این مطالعه به تجزیه و تحلیل اهمیت توسعه و کاربرد مدل‌های داده‌کاوی پایداری می‌پردازد که می‌توانند در برابر تغییرات بازار و شرایط غیرقابل پیش‌بینی مقاوم باشند. مرور ادبیات نظری در این زمینه، به ارزیابی پیشرفت‌های اخیر و دستاوردهای موجود در عرصه پیش‌بینی بازار سهام با استفاده از داده‌کاوی خواهد پرداخت و بحث در مورد دورنمای کاربرد آن‌ها در یک محیط دینامیک بازاری را تسهیل می‌کند.

در پایان، این ادبیات نظری قرار است نقشه راهی برای تحقیقات آتی ارائه دهد، به چالش‌های کلیدی بر سر راه پیش‌برد تکنیک‌های داده‌کاوی در بازار سهام اشاره کند و پیشنهادهایی برای رفع این موانع و پیشبرد دانشکده مالی پیش‌بینی‌کننده ارائه دهد.

5-1. جدول نظریه‌ها:

شرح	دسته بندی	نام نظریه
این نظریه‌ها به تعیین میزان بازدهی که سرمایه‌گذاران انتظار دارند روی دارایی‌ها به دست آورند، می‌پردازند.	نظریه‌های باز	نظریه بازده مورد انتظار (Expected Return Theories)

<p>این نظریه‌ها که شامل مدل‌های قیمت‌گذاری دارایی‌های سرمایه‌ای (CAPM) و نظریه قیمت‌گذاری اوراق بهادار چند عاملی می‌شوند، به تحقیق در مورد رابطه بین ریسک و بازده و تأثیر عوامل متعدد روی قیمت‌گذاری دارایی‌ها می‌پردازند.</p>		<p>نظریه قیمت‌گذاری اوراق بهادار (Security Pricing Theories)</p>
<p>بیان می‌کند که قیمت‌های سهام در هر زمان تمام اطلاعات موجود را منعکس می‌کنند و بنابراین «کارا» هستند.</p>		<p>نظریه بازار کارا (Efficient Market Hypothesis - EMH)</p>
<p>این نظریه‌ها به تفاوت بین بازارهای کامل-که در آنها هیچ مانعی برای ورود و خروج از بازار وجود ندارد- و بازارهای ناکامل که ممکن است شامل موانعی برای ورود یا اطلاعات نابرابر باشند، می‌پردازند.</p>		<p>نظریه‌های بازار کامل و ناکامل (Perfect and Imperfect Market Theories)</p>
<p>این نظریه بررسی می‌کند که چگونه اطلاعات جدید تأثیراتی بلافاصله و گاهی اوقات پیچیده روی قیمت‌های بازار دارند.</p>		<p>نظریه تأثیر اطلاعاتی (Information Impact Theory)</p>
<p>این نظریه که بیشتر در مدیریت سرمایه‌گذاری به کار می‌رود، بر میزان دارایی‌های قابل معامله و چگونگی تأثیر آن بر قیمت‌ها در بازار سهام تمرکز دارد.</p>		<p>نظریه موجودیت (Inventory Theory)</p>
<p>این دیدگاه به مطالعه تأثیر روان‌شناسی انسانی بر رفتار سرمایه‌گذاران و نتایج بازار می‌پردازد.</p>	<p>نظریه‌های رفتاری</p>	<p>نظریه رفتار مالی (Behavioral Finance Theory)</p>
<p>بررسی این موضوع می‌پردازد که چگونه شرکت‌ها از طریق سیگنال‌های خود مثل تقسیم سودها یا اعلانات مختلف، اطلاعات را به بازار منتقل می‌کنند.</p>		<p>نظریه سیگنالینگ (Signaling Theory)</p>

<p>این نظریه‌ها به تحلیل روش‌هایی که سرمایه‌گذاران برای ترکیب دارایی‌ها و مدیریت ریسک اتخاذ می‌کنند، می‌پردازد.</p>	<p>نظریه‌های مدیریت ریسک و سرمایه‌گذاری</p>	<p>۶. نظریه پرتفوی و انتخاب سرمایه (Portfolio and Capital Choice Theories)</p>
<p>این نظریه‌ها به بررسی رویکردها و تمایلات سرمایه‌گذاران نسبت به ریسک و چگونگی تأثیر آن بر انتخاب‌های معاملاتی و قیمت‌های بازار می‌پردازند.</p>		<p>نظریه پذیرش ریسک (Risk Appetite Theories)</p>
<p>این نظریه‌ها به بررسی تکنیک‌های مختلف سرمایه‌گذاری و تعیین میزان بهینه ریسک و بازده در انتخاب دارایی‌ها می‌پردازند.</p>		<p>نظریه‌های نوین پرتفولیو (Modern Portfolio Theories)</p>
<p>این دیدگاه به مطالعهٔ تعاملات متقابل و رفتارهای پیچیدهٔ بازارهای مالی به شکلی غیرخطی می‌پردازد که در آن اجزای مختلف بازار تأثیرات غیرمنتظره‌ای بر یکدیگر می‌گذارند.</p>	<p>نظریه‌های اقتصادی</p>	<p>نظریه‌های مربوط به بازار مالی به عنوان سیستم پیچیده (Financial Markets as Complex Systems)</p>
<p>در این نظریه به مطالعه رفتارهای رقابتی سرمایه‌گذاران در بازارهایی که تصمیمات یک شخص می‌تواند بر روی نتایج برای دیگران تأثیر بگذارد، پرداخته می‌شود.</p>		<p>نظریه بازی‌ها (Game Theory)</p>
<p>این نظریه‌ها به بررسی رویکردها و تمایلات سرمایه‌گذاران نسبت به ریسک و چگونگی تأثیر آن بر انتخاب‌های معاملاتی و قیمت‌های بازار می‌پردازند.</p>	<p>نظریه‌های پردازش و تحلیل داده‌ها</p>	<p>نظریه پذیرش ریسک (Risk Appetite Theories)</p>
<p>در مدل‌های پیش‌بینی داده‌کاوی، از الگوریتم‌های یادگیری ماشین مانند شبکه‌های عصبی، ماشین‌های بردار پشتیبان (SVM)، درخت تصمیم و یادگیری عمیق (Deep Learning) برای پیش‌بینی روندهای قیمت سهام استفاده می‌شود.</p>		<p>الگوریتم‌های ماشین برای یادگیری (Machine Learning Algorithms)</p>
<p>این نظریه به هزینه‌هایی که به هنگام خرید و فروش دارایی‌های مالی متحمل می‌شویم و چگونگی تأثیر این هزینه‌ها بر تصمیم‌های سرمایه‌گذاری می‌پردازد.</p>	<p>نظریه‌های مبتنی بر هزینه‌ها</p>	<p>نظریه هزینه معاملات (Transaction Cost Theory)</p>

6-1. مرور ادبیات و سوابق مربوطه

در مرور ادبیات و سوابق مربوطه به «کاربرد و عملکرد تکنیک‌های داده‌کاوی در بازار سهام» باید به چند بخش مهم توجه شود. این مرور به معرفی مباحث اصلی، بررسی کاربردها و ارزیابی نتایج تحقیقات قبلی، به چالش کشیدن فرضیات موجود و طرح مسائل و جهات تحقیقی آینده می‌پردازد. در زیر، به تفصیل، این بخش‌ها را بررسی می‌کنیم. تحلیل و پیش‌بینی دقیق رفتار بازارهای مالی، همواره یکی از مهم‌ترین چالش‌های اقتصاددانان، محققان علوم داده و سرمایه‌گذاران بوده است. پیچیدگی‌های ذاتی در حرکات بازار سهام و تأثیر عوامل گوناگون بر آن، موجب شده است تا استفاده از رویکردهای مبتنی بر داده برای دستیابی به پیش‌بینی‌های موثر افزایش یابد. در اینجا نگاهی اجمالی بر تحقیقات گذشته و دستاوردهای کنونی در این زمینه خواهیم داشت.

۱. مقدمه و زمینه تحقیق:

در این بخش، به بیان اهمیت موضوع و دلایل انتخاب آن می‌پردازیم. باید بررسی کنیم که چرا مطالعه تکنیک‌های داده‌کاوی در بازار سهام مهم است و چگونه این تکنیک‌ها می‌توانند به درک بهتر بازارهای مالی و تصمیم‌گیری‌های سرمایه‌گذاری کمک کنند. همچنین، باید به تشریح دیدگاه‌ها و نظریات مربوط به کارایی بازار بپردازیم و تأثیر آن‌ها بر تکامل ابزارهای تحلیلی و تکنیک‌های معاملاتی را مورد بررسی قرار دهیم.

- Malkiel, B.G. (2003). *The Efficient Market Hypothesis and Its Critics*. Journal of Economic Perspectives, 17(1), 59-82.

۲. تاریخچه و تکامل تکنیک‌های داده‌کاوی در بازار مالی:

استفاده از تکنیک‌های داده‌کاوی در حال افزایش پیش‌بینی بازار سهام است. با دسترس به نرم افزارهای پیچیده و افزایش علاقه محققان و متخصصان، انتظار می‌رود دامنه به سرعت رشد کند. مطالعه حاضر برتری تکنیک‌های داده‌کاوی را در پیش‌بینی بازار سهام در مقایسه با تکنیک‌های سنتی مانند رگرسیون خطی و مدل‌های گردش تصادفی نشان داده است. علاوه بر این، یک تغییر واضح از استفاده از یک تکنیک به تکنیک‌های ترکیبی با هدف بهبود دقت مدل‌های پیش‌بینی وجود دارد. در حال حاضر، حوزه تحقیق به‌طور عمده بر بهبود دقت ریاضی مدل‌های پیش‌بینی سهام مبتنی بر داده‌کاوی تمرکز دارد. نیاز به حرکت بیش از دقت مدل برای سودآوری مدل وجود دارد.

Application and performance of data mining techniques in stock market: A review 2022

۳. رویکردهای مختلف داده‌کاوی و معرفی الگوریتم‌ها:

در این مقاله به بررسی جامع الگوریتم‌های مختلف داده‌کاوی از جمله شبکه‌های عصبی، ماشین‌های بردار پشتیبان، روش‌های درخت تصمیم و یادگیری تقویتی که در تحلیل داده‌های مالی به کار رفته‌اند، می‌پردازیم. همچنین، به نحوه اعمال این روش‌ها بر داده‌های قیمتی و مقایسه عملکرد آن‌ها در پیش‌بینی بازار اشاره می‌کنیم.

- Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques, Third Edition*. Morgan Kaufmann Publishers.

۴. بررسی مطالعات موردی: کاربردهای واقعی و نتایج:

در این مقاله مفهوم یک بستر داده‌کاوی بازار مصنوعی ارائه شده است که ترکیبی از شبیه‌سازی بازار مصنوعی و داده‌کاوی است. مزیت این سکو این است که همه چیز در بازار مصنوعی می‌تواند کنترل و مشاهده شود. بنابراین با تغییر عوامل موجود در بازار مصنوعی و یا مؤلفه‌های حاکم بر آن بازار می‌توان شرایط بسیاری را آزمایش کرد. از این رو، در ترکیب با روش‌های داده‌کاوی، با توجه به داده‌های واقعی و بررسی اینکه چگونه عوامل مختلفی در پیش‌بینی بازارهای مالی نقش دارند، فرض بر این بود که این سکو به ما امکان ارزیابی نسبی روش‌های داده‌کاوی برای بازار مالی را می‌دهد، همچنین یک مدل عملی از مفهوم پیشنهادی خود را به عنوان مثال نشان دادیم سپس با استفاده از مدل عملی به بررسی نتایج پرداختیم. به عنوان یک مثال عملی، ما یک مدل شبیه‌سازی بازار مصنوعی سطح زمانی tick و داده‌کاوی برای پیش‌بینی تغییرات کوتاه مدت قیمت و بررسی تاثیر چهار عامل بازار مالی بر عملکرد مدل‌های داده‌کاوی ساختیم. از میان این عوامل، مشاهده شد که مقیاس نویز بازارهای مالی تاثیر معناداری بر عملکرد پیش‌بینی روش‌های داده‌کاوی دارد که اثر پیچیده‌ای خواهد داشت. در نهایت به بررسی پتانسیل‌ها و کاربردهای آینده مفهوم خود می‌پردازیم. در این مقاله، اعتبار و مزایای مدل مفهومی و عملی خود در آینده را مورد بررسی قرار می‌دهیم و امیدواریم که محققان دیگر از این مفهوم استفاده کنند و بتوانند پاسخگوی پرسش‌های مختلف در بازارهای مالی باشند. همچنین قصد داریم تا موارد عملی‌تری از مفهوم خود بسازیم و سودمندی این مفهوم را حتی بیشتر هم نشان دهیم.

Concept and Practice of Artificial Market Data Mining Platform 2022.

۵. مسائل و چالش‌ها در داده‌کاوی بازار سهام:

بزرگترین چالش استفاده از پایگاه‌های داده برای تصمیم‌گیری، استخراج دقیق محتوای اطلاعات پنهان در داده‌های خام است.

به طور معمول، غربالگری دستی برای چنین اهدافی به کار گرفته می‌شود که با بزرگ شدن حجم داده‌ها غیر عملی و ناکارآمد می‌شوند. همچنین پیش‌بینی بازار سهام شامل پردازش مقادیر زیادی از قیمت‌های سهام و داده‌های حجم است. به منظور رفع این مشکل، در این پژوهش سعی شده است از روش طبقه‌بندی داده‌کاوی استفاده شود. علاوه بر این، این مطالعه با استفاده از روش طبقه‌بندی داده‌های قیمت سهام به منظور هدایت بهتر سرمایه‌گذاران در تدوین راهبردهای تجاری سودآور انجام شده است. طبقه‌بندی داده‌ها را می‌توان به روش‌های مختلفی انجام داد: از جمله این روش‌ها طبقه‌بندی با درخت تصمیم است. این روش نمایش گرافیکی تمام نتایج ممکن و مسیرهای زیربنایی برای رسیدن به آنها را ارائه می‌دهد. (با توجه به این تحقیق، از مدل درخت تصمیم، فرآیند استاندارد صنعت متقاطع برای داده‌کاوی برای تحلیل و پیش‌بینی قیمت سهام نمونه شرکت از پنج بخش مختلف پذیرفته شده در بورس اوراق بهادار پاکستان استفاده شده است.)

(Ampomah, Qin, & Nyame, 2020; Basak, Kar, Saha, Khaidem, & Dey, 2019; Enke & Thawornwong, 2005; Guresen, Kayakutlu, & Daim, 2011; Niaki & Hoseinzade, 2013; Shah, Isah, & Zulkernine, 2019; Wang, Wang, Zhang, & Guo, 2011)

۶. کاربرد تکنیک‌های داده‌کاوی در مدیریت ریسک و بهینه‌سازی:

یافته‌های تحقیق نشان می‌دهد که سرمایه‌گذاران و مدیران پرتفوی (سود سهام) می‌توانند از روش‌های داده‌کاوی برای پیش‌بینی قیمت سهام با دقت قابل قبول در PSX استفاده کنند. علاوه بر این، یافته‌های تحقیق می‌تواند به شرکت‌کنندگان بازار در تدوین راهبردهای تجاری سودآور در مقایسه با روش‌های مرسوم پیش‌بینی سهام کمک کند. به ویژه در بازارهای در حال ظهور، این رویکرد برای سرمایه‌گذاران به منظور بررسی تصمیمات سرمایه‌گذاری آنها مفید است، زیرا تحقیقات گذشته نشان داده‌است که بازارهای نو ظهور، در حال افزایش هستند. با این حال، میزان نسبتاً پایین دقت نیز، با توجه به اینکه قیمت سهام در برابر افراد مختلف آسیب پذیر است، دشواری پیش‌بینی قیمت سهام را فقط با استفاده از ویژگی‌های مطالعاتی مربوط به تاریخچه قیمت و حجم سهام برجسته می‌کند.

- Forecasting on Stock Market Time Series Data Using Data Mining Techniques 2022.

۷. چشم‌انداز آینده و فناوری‌های نوظهور:

ارتباط بلاکچین با داده‌کاوی و پیش‌بینی در بازارهای مالی به این صورت است که بلاکچین به عنوان یک فناوری ثبت و انتقال داده‌ها، داده‌های مختلفی از جمله تراکنش‌های مالی، قراردادهای هوشمند و داده‌های

مربوط به زنجیره تأمین را در خود ذخیره می‌کند. این داده‌ها معمولاً حاوی اطلاعاتی مفید برای تحلیل الگوها، روندها و پیش‌بینی رویدادهای آینده هستند.

در این زمینه، داده‌کاوی به عنوان یک فرآیند استخراج الگوهای مفید و اطلاعات قابل استفاده از داده‌ها، می‌تواند به بررسی داده‌های موجود در بلاکچین کمک کند. این فرآیند می‌تواند الگوها، روندها و اطلاعات مفیدی را کشف کند که به تصمیم‌گیری‌های مالی، پیش‌بینی روندهای آینده و بهبود فرآیندهای مالی کمک کند.

به عنوان مثال، داده‌کاوی می‌تواند به تحلیل الگوهای تراکنش‌های مالی در بلاکچین کمک و الگوهای تقاضا و عرضه را شناسایی کند. همچنین، با استفاده از داده‌کاوی می‌توان الگوهای تقلب و فعالیت‌های غیرقانونی را نیز شناسایی کرد.

بنابراین، این ارتباط می‌تواند به بهبود تصمیم‌گیری‌های مالی و پیش‌بینی روندهای بازار مالی کمک کند.

- The Sustainable Development of Financial Topic Detection and Trend Prediction by Data Mining 2021.

۸. نتیجه‌گیری و جهت‌گیری‌های آتی تحقیق:

به طور خلاصه، مدل‌های مختلفی از پیش‌بینی سهام وجود دارد که نتایج خوبی به دست آورده‌اند. با این حال، مشکلات زیادی وجود دارد که هنوز به طور کامل حل نشده‌اند. چالش‌ها و دیدگاه‌های احتمالی تحقیقات آینده عمدتاً در چهار حوزه زیر هستند. اغلب مطالعات از یک الگوریتم واحد برای ساخت مدل‌های پیش‌بینی استفاده کرده‌اند، در حالی که مطالعات ترکیب‌کننده مدل‌های یادگیری ماشینی آماری با روش‌های سنتی تحلیل سهام (تحلیل بنیادی و تحلیل شاخص فنی) فاقد آن هستند. فرآیند پیش‌بینی، اغلب ارزش ذاتی سهام و تاثیر خارجی عوامل بازار را نادیده می‌گیرد. در آینده می‌توان از مزایای الگوریتم‌های بهینه، تنظیم ویژگی‌های مدل، بهینه‌سازی میانگین وزنی مدل برای ساخت یک مدل ترکیبی برای پیش‌بینی استفاده کرد. ما باید از تحلیل عاملی و روش‌های دیگر برای انتخاب ویژگی‌های شاخص مربوط به سهام استفاده کنیم. همچنین می‌توان از تصاویر بازار سهام استفاده کرد که می‌تواند به طور کامل قوانین بازارهای سهام مختلف را ثبت و بازار سهام را عمیق‌تر کشف کند. اغلب روش‌های پیش‌بینی قیمت سهام فعلی، بدون در نظر گرفتن مطالعات متنی مانند اخبار مالی، گزارش‌های سود شرکت و نظرات نوار سهام که ممکن است بر سرمایه‌گذاران تاثیر داشته باشد، از داده‌های تجاری هیستوریکیال (تاریخی) برای تحلیل و پیش‌بینی استفاده می‌کنند، بنابراین، تاثیر اطلاعات متنی بر بازار سهام باید به طور کامل در نظر گرفته شود. یادگیری عمیق باید برای پیش‌بینی قیمت سهام به کار رود. ما داده‌های متنی مانند تصمیمات

گفتاری سرمایه‌گذاران و اخبار مالی را کمی و توجه سرمایه‌گذاران را با هم ترکیب می‌کنیم. همچنین مدل‌هایی برای ترکیب آنها با داده‌های تاریخی تجارت سهام و تجزیه و تحلیل و پیش‌بینی با استفاده از منابع مطالعاتی چندگانه می‌سازیم.

Progress and prospects of data-driven stock price forecasting research 2023

این توضیحات می‌توانند به تعیین گام‌های بعدی در تحقیق کمک کنند و برای فهمیدن زمینه‌هایی که نیاز به مطالعات بیشتری دارند، مهم هستند.

7-1. نتیجه‌گیری

استفاده از تکنیک‌های داده‌کاوی در حال افزایش پیش‌بینی بازار سهام است. با دسترسی به نرم‌افزارهای پیچیده و افزایش علاقه محققان و متخصصان، انتظار می‌رود دامنه به سرعت رشد کند. مطالعه حاضر برتری تکنیک‌های داده‌کاوی را در پیش‌بینی بازار سهام در مقایسه با تکنیک‌های سنتی مانند رگرسیون خطی و مدل‌های گردش تصادفی نشان داده‌است. علاوه بر این، یک تغییر واضح از استفاده از یک تکنیک به تکنیک‌های ترکیبی با هدف بهبود دقت مدل‌های پیش‌بینی وجود دارد. در حال حاضر، حوزه تحقیق عمدتاً بر بهبود دقت ریاضی مدل‌های پیش‌بینی سهام مبتنی بر داده‌کاوی تمرکز دارد. تکنیک‌ها و الگوریتم‌های متنوع مورد استفاده در این زمینه، از شبکه‌های عصبی و ماشین‌های بردار پشتیبان گرفته تا درختان تصمیم، به دانشمندان داده و تحلیلگران مالی این امکان را می‌دهند که با دقت و سرعت بیشتری به تحلیل بازار اقدام کنند.

با وجود پیشرفت‌های قابل توجه در این عرصه، همچنان چالش‌هایی نیز وجود دارد، مانند مدیریت داده‌های نامتعرف و اعتبارسنجی مدل‌ها. علاوه بر این، تاکید زیادی بر روی نیاز به توسعه مدل‌هایی است که بتوانند در برابر نوسانات بازار مقاومت نشان دهند و نتایج قابل اتکایی را فراهم آورند. با وجود پیشرفت‌های اخیر در تکنیک‌های داده‌کاوی و یادگیری ماشین، مطالعات نشان می‌دهند که هنوز پتانسیل (ظرفیت) زیادی برای بهبود و توسعه بیشتر وجود دارد. پیوستگی بین تحقیقات علمی و کاربردهای عملی نیازمند تلاش مضاعف در زمینه‌های دقت و پایداری مدل‌های پیش‌بینی است. برای نتیجه‌گیری، به نظر می‌رسد که مدل‌های ترکیبی که داده‌کاوی و یادگیری ماشینی را یکپارچه می‌کنند، پتانسیل داشته‌اند تا عملکرد قابل توجهی در بازارهای سهام مختلف از خود نشان دهند. منطق نهفته در این عمل، این است که با استفاده از نقاط قوت متعدد روش‌ها، می‌توان خطاهای مربوط به هر روش فردی را متعادل کرد.

در ادامه به یک نتیجه‌گیری فرضی اشاره می‌کنیم و جدولی ارائه می‌دهیم که روش‌های رایج پیش‌بینی بازار سهام را با یکدیگر مقایسه می‌کند. برای دقت بیشتر، باید این اطلاعات را با تحلیل داده‌های واقعی و مطالعات تجربی مورد بررسی قرار داد. بر اساس مرور ادبیات انجام شده، سه روش برجسته در پیش‌بینی بازار سهام شناسایی شده‌اند: شبکه‌های عصبی، ماشین‌های بردار پشتیبان (SVM) و تحلیل زمانی سری‌های زمانی. هر یک از این تکنیک‌ها نقاط قوت و ضعف خاص خود را دارند. شبکه‌های عصبی به دلیل قابلیت تقریب توابع غیرخطی و توانایی در یادگیری پیچیده‌ترین الگوهای بازار مورد توجه قرار گرفته‌اند. ماشین‌های بردار پشتیبان در دسته‌بندی و پیش‌بینی‌های دقیق، عملکرد خوبی دارند، به ویژه زمانی که مقدار داده محدود است. تحلیل زمانی سری‌های زمانی که شامل روش‌های ARIMA می‌شود، برای داده‌های مالی که نمایش منظم ترند و فصلیت دارند مفید است.

جدول مقایسه روش‌های پیش‌بینی بازار سهام:

نقاط ضعف	نقاط قوت	نام روش
۱. حساس به تنظیمات پارامتر و ممکن است دچار بیش‌برازش شود. ۲. نیازمند منابع قدرتمند و محاسبات سنگین. ۳. تفسیرپذیری کمتر مدل‌ها	۱. توانایی مدل‌سازی روابط غیرخطی پیچیده میان ورودی‌ها و خروجی‌ها. ۲. با دیتاست‌های بزرگ و پیچیده خوب عمل می‌کند. ۳. قابلیت خودیادگیری از داده‌ها و بهبود عملکرد با افزایش داده.	شبکه‌های عصبی (ANN)
۱. انتخاب پارامترها مانند C و هسته می‌تواند دشوار باشد. ۲. نیازمند تکنیک‌های مقیاس‌بندی موثر داده‌ها پیش از آموزش. ۳. کارایی کم بر روی دیتاست‌های بسیار بزرگ به علت افزایش هزینه محاسباتی	۱. عملکرد خوب به دلیل بهینه‌یابی مرز تصمیم (Decision Boundary). ۲. کارآمد در فضاهای با ابعاد بالا. ۳. کمتر دچار بیش‌برازش می‌شود	ماشین بردار پشتیبان (SVM)

<p>۱. فرض بر ثابت بودن سری زمانی و تنظیم نامناسب در مواجهه با داده‌های غیر ثابت.</p> <p>۲. عدم قابلیت در نظر گرفتن روابط پیچیده غیر خطی.</p> <p>۳. محدودیت در مدل‌سازی فقط بر اساس داده‌های گذشته و استفاده نکردن از متغیرهای توضیحی خارجی.</p>	<p>۱. مدل‌سازی پایدار و پیش‌بینی بلندمدت در صورت ثابت بودن سری زمانی.</p> <p>۲. مناسب برای داده‌هایی با الگوهای فصلی و روند مشخص.</p> <p>۳. تکنیک‌های ارزیابی و تشخیص استاندارد و قابل درک.</p>	<p>تحلیل زمانی سری‌های زمانی (مانند ARIMA)</p>
<p>۱. نیاز به تنظیمات گسترده و دقیق مولفه‌های الگوریتم.</p> <p>۲. ممکن است در بهینه‌یابی محلی گیر کند و به جواب‌های جهانی نرسد.</p> <p>۳. زمان محاسباتی به نسبت بالا.</p>	<p>۱. توانایی در پیدا کردن راه‌حل‌های نزدیک به بهینه حتی در فضاهاى جست‌وجوی بزرگ.</p> <p>۲. قابلیت هم‌زمان‌سازی و تطبیق‌پذیری با تغییرات محیطی</p>	<p>الگوریتم‌های ژنتیکی (Genetic algorithm)</p>
<p>۱. ممکن است دچار بیش‌برازش شوند و نیاز به هرس کردن دارند.</p> <p>۲. کارایی کمتر در مقابل داده‌هایی که دارای روابط پیچیده و غیر خطی هستند.</p>	<p>۱. مدل‌های ساده و قابل فهم که می‌توانند تفسیرپذیر باشند.</p> <p>۲. قادر به کار با انواع داده‌ها، چه عددی و چه دسته‌بندی شده.</p>	<p>درخت تصمیم (Decision Trees)</p>
<p>۱. نیاز به مقادیر زیادی از داده‌های برچسب‌دار برای آموزش دارد.</p> <p>۲. تفسیر نتایج مشکل‌تر است.</p>	<p>توانایی عالی در کشف پیچیدگی‌های غیر خطی و الگوهای داده‌های بسیار بزرگ و پیچیده.</p> <p>پیشرفت‌های بزرگ در تشخیص الگو به واسطه معماری‌های پیشرفته مانند Convolutional Neural Networks (CNNs) و Recurrent Neural Networks (RNNs).</p>	<p>یادگیری عمیق (Deep Learning)</p>
<p>۱. فرضیات محدودیت‌های خاصی را درباره توزیع داده‌ها دارد و در مواقعی که این فرضیات برقرار نباشد، ممکن است دقت پایینی داشته باشد.</p> <p>۲. قادر به مدل‌سازی روابط پیچیده نیست.</p>	<p>۱. مدل ساده‌ای است که می‌توان به سرعت آموزش داد و برای تفسیر، قابلیت بالایی دارد.</p> <p>۲. احتمالات خروجی و ارتباط مستقیم با حالت‌های ورودی را نشان می‌دهد.</p>	<p>مدل رگرسیون لجستیک (Logit Model)</p>

<p>۱. برای پیش‌بینی، نیاز به محاسبات بالا و زمان‌بر است، به ویژه برای مجموعه داده‌های بزرگ.</p> <p>۲. حساسیت به داده‌های نویزی و پرت است و نیاز به تعیین مناسب پارامتر k دارد.</p>	<p>۱. این مدل آموزشی نیاز به فرضیات خاصی ندارد و درک مفهومی ساده‌ای دارد.</p> <p>۲. قادر به مدل‌سازی الگوهای پیچیده و غیرخطی است.</p>	<p>k-NN (k-Nearest Neighbors)</p>
<p>۱. مدل تقریبی است و از فرضیات محدودیت‌های خاصی درباره توزیع داده‌ها استفاده می‌کند.</p> <p>۲. نمی‌تواند الگوهای پیچیده‌تر داده‌ها را مدل کند.</p>	<p>۱. یک مدل ساده و مؤثر برای دسته‌بندی است.</p> <p>۲. در مواردی که فرضیات LDA برقرار باشند، دقت بالایی دارد و نیازمند پارامترهای کمتری است.</p>	<p>LDA (Linear Discriminant Analysis)</p>
<p>۱. نیاز به تعداد نمونه داده‌ای بالا و انعطاف بیش‌اندازه می‌تواند منجر به بیش‌برازش شود.</p> <p>۲. به دلیل پیچیدگی مدل، ممکن است در مقابل داده‌های کمتر مؤثر نباشد.</p>	<p>۱. این مدل انعطاف‌پذیری دارد و می‌تواند توزیع‌های احتمالاتی مختلف برای هر کلاس را مدل کند.</p> <p>۲. در صورتی که فرضیات QDA صدق کنند، دقت بالایی دارد.</p>	<p>QDA (Quadratic Discriminant Analysis)</p>

دقت پیش‌بینی برای هر یک از روش‌ها بستگی به مجموعه داده‌های خاص، تنظیمات مدل و حوزه کاربرد دارد. برخی تحقیقات نشان داده‌اند که مدل‌های ترکیبی که فنون مختلف را ادغام می‌کنند، ممکن است به دقت بالاتری در پیش‌بینی دست یابند. با توجه به بروز و تکامل تکنیک‌های یادگیری ماشین و داده‌کاوی، توصیه می‌شود که برای به دست آوردن دقت مطلوب، از الگوریتم‌های چندگانه برای سنجش و تطبیق عملکرد استفاده شود.

TABLE 16 Prediction performance of various techniques

Technique	ANN	SVM	Combined model	Logit model	QDA	LDA	k-NN
Kara et al. (2011)	75.74	71.52	-	-	-	-	-
Ou and Wang (2009)	-	86.00	86.40 (LS-SVM)	85.60	84.80	84.40	79.60
Yu et al. (2009)	67.56	72.61	82.66 (LS-SVM)	-	-	61.43	-
Choudhry and Garg (2008) (TCS)	-	58.09	61.73 (GA-SVM)	-	-	-	-
Kumar and Thenmozhi (2006)	62.93	68.44	-	59.60	-	56.34	-
Huang et al. (2005)	69.00	73.00	75.00	-	69.00	55.00	-
Kim (2003)	54.73	57.83	-	-	-	-	-

جدول فوق مقایسه عملکرد پیش‌بینی تکنیک‌های مختلف مورد استفاده محققان در مطالعات آنها را نشان می‌دهد. ترجیح NNS برای پیش‌بینی ممکن است ناشی از دقت و نتایج امیدوار کننده آنها باشد.

(Hammad et al., 2009; Majumder & Hussian, 2007; Mizuno et al., 1998; Tjung et al., 2010; Vojinovic et al., 2001).

برتری مدل‌های شبکه عصبی مصنوعی بر مدل‌های رگرسیونی در مطالعه Altay and Satman (2005) نشان داده شده است که درصد بالاتری از نشانه‌های پیش‌بینی شده برای ann نسبت به مدل‌های رگرسیونی را تبیین می‌کند. در مطالعه دیگر، میانگین عملکرد مدل ann (75.74 درصد) در مقایسه با مدل svm (71.52 درصد) از نظر آماری بهتر بود (Kara et al., 2011).

با این حال، دیدگاه‌های متضادی نیز وجود دارند که محققان svm ها را بر شبکه‌های عصبی ترجیح می‌دهند.

8-1. بررسی تحقیقات پیشین:

بازارهای سهام به طور گسترده‌ای به عنوان مؤلفه‌های کلیدی اقتصاد جهانی شناخته می‌شوند. در سال‌های اخیر، با پیشرفت‌های عمیق در فناوری‌های داده‌کاوی، پژوهش‌های متعددی به بررسی کارایی و تاثیر این تکنولوژی‌ها بر قابلیت پیش‌بینی بازارهای سهام پرداخته‌اند. از بین این پژوهش‌ها می‌توان به مطالعات منتشر شده در سال‌های ۲۰۲۱ تا ۲۰۲۳ اشاره کرد که جنبه‌های مختلف فرآیند پیش‌بینی و تجارت را در بازار سهام مورد کنکاش قرار داده‌اند.

طیف گسترده‌ای از مدل‌ها و چهارچوب‌های داده‌کاوی، از جمله "A data mining framework for financial prediction" در سال ۲۰۲۱، بیانگر تلاش‌های مستمر برای بهبود دقت پیش‌بینی‌های مالی است. در راستای توسعه پایدار، مطالعه "The Sustainable Development of Financial Topic Detection and Trend Prediction by Data Mining" نه تنها به کشف موضوعات مالی می‌پردازد بلکه بر روند پیش‌بینی تأکید می‌کند. در مقاله "Forecasting stock prices using a data mining method: Evidence from emerging market" منتشر شده در سال ۲۰۲۱، از تکنیک‌های استخراج داده بر اساس الگوهای سابقه‌دار در بازارهای نوظهور استفاده

شده و نتایج نشان می‌دهند که روش‌های پیشرفته داده‌کاوی می‌توانند عملکرد خوبی در پیش‌بینی قیمت‌ها حتی در بازارهای نسبتاً ناپایدار داشته باشند.

"Stock Price Prediction Based on Data Mining Combination Model" منتشر شده در سال ۲۰۲۳، نوید بخش ترکیبی از مدل‌های داده‌کاوی است که با به‌کارگیری چندین روش مختلف و مکمل، دقت پیش‌بینی‌ها را بهبود می‌بخشد. این تحقیق نمایان‌گر پیشرفت‌های حاصل از ادغام تکنیک‌های یادگیری ماشین و سایر روش‌های تحلیلی داده‌هاست.

تحلیل فرضیه‌ها و مدل‌ها:

مطالعاتی نظیر "Progress and prospects of data-driven stock price forecasting research" پیشرفت‌ها و دورنمای روش‌های پیش‌بینی قیمت سهام را محور قرار داده‌اند. در این تحقیق، روش‌های نوین داده‌کاوی و تأثیر آن‌ها بر پیش‌بینی قیمت‌ها بررسی شده و این سوال مطرح می‌شود که چگونه فناوری‌های جدید می‌توانند بهینه‌سازی مدل‌های پیش‌بینی موجود و طراحی مدل‌های نوآورانه‌تر را تسهیل کنند.

ارزیابی روش‌ها و عملکرد:

تحقیق "Application and performance of data mining techniques in stock market: A review" به ارزیابی جامعی از تکنیک‌های داده‌کاوی در عمل پرداخته و شاخص‌های عملکرد مختلف را مورد تحلیل قرار داده است. این بررسی نشان‌دهنده قابلیت اتکای روش‌های داده‌کاوی و کاربرد آن‌ها در فرآیندهای تصمیم‌گیری مالی است.

شکاف‌های تحقیق:

در پایان، تحقیق "Netflix Stock Price Movements Insights from Data Mining" نمونه‌ای از چگونگی استخراج داده‌های مربوط به یک شرکت خاص و تحلیل روند قیمتی آن است. با استفاده از مطالعه موردی، این کار به ما نشان می‌دهد که چگونه داده‌کاوی در فهم بهتر رفتار بازار و کمک به سازمان‌ها برای تصمیم‌گیری‌های مالی استراتژیک به کار می‌رود.

در حالی که ادبیات تجربی تحقیق موجود پیشرفت‌های قابل توجهی را در زمینه پیش‌بینی قیمت سهام نشان می‌دهد، شکاف‌هایی نیز وجود دارند که فرصت‌هایی برای تحقیقات آینده ارائه می‌دهند. نیازمندی بیشتر به شفاف‌سازی مدل‌های نظری، دقت‌های بهبود یافته در انتخاب متغیرهای ورودی و بررسی آثار سیستماتیک عوامل ماکرواقتصادی بر بازارهای سهام از جمله این شکاف‌ها هستند.

در پایان، این بخش به طور خلاصه شکاف‌های موجود در تحقیقات و چگونگی ساخت یک زمینه محکم برای تحقیق کنونی را آشکار می‌کند و بستری را برای طرح پرسش‌های تحقیقاتی جدید فراهم می‌آورد.

روش‌شناسی تحقیق

روش‌شناسی یک پژوهش در برخورداری از تعریف چگونگی جمع‌آوری، تحلیل و تفسیر داده‌هاست. با توجه به موضوعات مقاله‌های فهرست شده، می‌توان یک روش‌شناسی عمومی برای تحقیقات مربوط به داده‌کاوی در پیش‌بینی قیمت‌های سهام مشخص کرد که عبارتند از:

۱. تعریف مشکل و فرضیات تحقیق:

- ابتدا به مشکل تحقیق از جمله پیش‌بینی قیمت سهام، ردیابی ترند(معروف)های مالی، یا توسعه راهبردهای تجاری اشاره شده و فرضیات مرتبط با تحقیق تعریف می‌شوند.

۲. جمع‌آوری داده‌ها:

- این فرآیند شامل جمع‌آوری داده‌های تاریخی سهام، داده‌های مالی، شاخص‌های اقتصادی، داده‌های خبری و سایر منابع مرتبط است.
- برای داده‌های زمانی و فرکانسی(پرتکرار و متوالی) مانند قیمت‌های سهام، ممکن است داده‌های روزانه، هفتگی یا ماهانه به کار رود.

۳. پیش‌پردازش و تهیه داده‌ها:

- پاک‌سازی داده‌ها برای حذف نویز و داده‌های پرت.
- کار با داده‌های گمشده و رسیدگی به داده‌های بی‌کیفیت.
- نرمال‌سازی و تبدیل داده‌ها برای آماده‌سازی برای تحلیل‌های داده‌کاوی.

۴. انتخاب ویژگی‌ها و کاهش بُعد:

- شناسایی ویژگی‌هایی که بیشترین تأثیر را در پیش‌بینی قیمت سهام دارند.

- استفاده از تکنیک‌هایی مانند PCA یا تحلیل مولفه‌های اصلی برای کاهش بعد داده‌ها و تمرکز بر داده‌های معنادار.

۵. استفاده از مدل‌های داده‌کاوی:

- توسعه و استفاده از مدل‌های داده‌کاوی مختلف مانند شبکه‌های عصبی، دسته‌بندی‌های درختی، روش‌های انسمبل و مدل‌های خطی.

- پیاده‌سازی چندین مدل برای مقایسه و انتخاب بهترین عملکرد با استفاده از معیارهای دقت.

۶. آزمایش و اعتبارسنجی مدل و تحلیل نتایج:

- استفاده از تکنیک‌های اعتبارسنجی مانند اعتبارسنجی متقاطع (cross-validation) برای آزمایش دقت پیش‌بینی‌های مدل بر روی داده‌های ناشناخته.

- تجزیه و تحلیل عملکرد مدل‌ها از طریق معیارهای ارزیابی سنجش خطای پیش‌بینی مانند MAE, RMSE, و ضریب تعیین (R^2).

- تفسیر نتایج و استخراج دانش قابل انتقال به تصمیم‌گیری‌های مالی و تجاری.

مرور گسترده‌ای که بر ادبیات تجربی تحقیقات مرتبط با پیش‌بینی قیمت سهام انجام گرفته، یک دید کلی از پیشرفت‌ها، چالش‌ها و فرصت‌های موجود در این حوزه تحقیقاتی پدیدینامیک را فراهم می‌آورد. در این بخش، جنبه‌های کلیدی بررسی‌های صورت گرفته، ابعاد مهم پیشرفت‌های اخیر و جهاتی که به شناسایی نیازهای تحقیقاتی بعدی منجر شده‌اند، مورد تأمل قرار گرفته‌اند.

یکی از جنبه‌های برجسته که از مرور ادبیات، مشخص می‌شود، توجه فزاینده به تکنیک‌های داده‌کاوی پیشرفته است. این رویکردهای داده‌بنیاد در مطالعاتی مانند "Stock Price Prediction Based on Data Mining Combination Model" نشان‌دهنده یک گرایش به سمت مدل‌های ترکیبی و برداربنیان است که قادر به ادغام پویایی‌های بازار و سیگنال‌های غیرخطی مربوطه هستند. این مدل‌های ترکیبی، با تثبیت دقت و کاهش وابستگی به فرضیات خاص مثل توزیع‌های آماری شناخته شده، قدمی رو به جلو در حوزه پیش‌بینی قیمت سهام را نشان می‌دهند.

با این حال، بازخوردهای عملی این تحقیقات نشان دهنده نیاز به توجه بیشتر به شفافیت مدلسازی و تفسیرپذیری نتایج است. علاوه بر این، بر لزوم توجه به نقش پیچیدگی مدل در قابلیت کاربردی تصمیم‌گیری مالی، به عنوان یک مانع بالقوه، تأکید می‌شود. ادبیات تجربی تحقیق نشان می‌دهد؛ با اینکه مدل‌های پیچیده به

گرایش‌های جدیدی در داده‌کاوی کمک می‌کنند، اما ممکن است در مواقعی، کاربردی بودن مورد نیاز کارشناسان مالی را تحت تاثیر قرار دهند.

در نهایت، مطالعات حاضر به این نتیجه رسیده‌اند که درک بهتر از ارتباطات میان اطلاعات بازار - به عنوان ورودی‌های مدل و پیش‌بینی‌های حاصل از آن‌ها - اساسی است. تحقیقات مانند "Netflix Stock Price Movements Insights from Data Mining" نمونه‌ای از کاربرد داده‌کاوی در جهت فراهم آوردن بینش‌هایی کاربردی و عملی است که می‌تواند به معامله‌گران کمک کند تا استراتژی‌هایی متناسب با داده‌های واقعی بازار طراحی کنند.

شناسایی این شکاف‌ها و نیازمندی‌های تحقیقاتی نه تنها برای پژوهش‌های آتی بلکه برای توسعه عملی تکنیک‌های پیش‌بینی در دنیای استراتژیک تجارت سهام ضروری است. تقویت ادبیات تجربی تحقیق موجود با تحقیقات دقیق‌تر و جامع‌تر می‌تواند به روشن‌تر شدن مسیر پژوهش‌های آتی کمک شایانی نماید.

نتایج تحلیل داده‌ها و مطالعات مختلف نشان‌دهنده پتانسیل چشمگیر داده‌کاوی در ارتقاء دقت پیش‌بینی قیمت سهام است. تحقیقات نمایانگر آن است که اعمال روش‌های داده‌کاوی می‌تواند به اندازه‌گیری و استخراج ویژگی‌های پیچیده بازار کمک کند و در نتیجه، امکان تدوین راهبردهای معاملاتی دقیق‌تری را فراهم آورد.

ارزیابی دقیق یافته‌ها:

در مطالعه "Forecasting stock prices using a data mining method: Evidence from emerging market 2021"، مشاهده شد که رویکردهای داده‌کاوی قادر هستند تفاوت‌های قابل توجهی را در بازارهای نوظهور ایجاد کنند. از طرف دیگر، مطالعه "Stock Price Prediction Based on Data Mining Combination Model 2023" راهبردهای ترکیبی داده‌کاوی را شناسایی می‌کند که به افزایش قدرت پیش‌بینی می‌انجامد.

تحلیل انتقادی:

حال که پژوهش‌های صورت‌گرفته در مورد "The Sustainable Development of Financial Topic Detection and Trend Prediction by Data Mining 2021" و "Netflix Stock Price Movements Insights from Data Mining 2022" نشان‌دهنده پیشرفت‌های قابل توجهی در تشخیص و پیش‌بینی روند مالی هستند، باید توجه داشت که عدم شفافیت و تفسیرپذیری ممکن است کاربرد واقعی آن‌ها را محدود کند.

با وجود پیشرفت‌های حاصله، داده‌کاوی در پیش‌بینی قیمت سهام با محدودیت‌هایی نظیر تعصب در تجزیه و تحلیل داده‌ها، اشتباهات بالقوه در داده‌های پایه و تغییرات فوری بازار که به سادگی قابل پیش‌بینی نیستند، مواجه است.

9-1. پیشنهادها

پیشنهادهای پژوهشی و تصمیم‌گیری‌های مالی:

پیشنهاد می‌شود که تحقیقات آتی بر روی بهینه‌سازی الگوریتم‌های داده‌کاوی، ترکیب داده‌های بازار متنوع‌تر و به‌کارگیری شبکه‌های عصبی پیچیده‌تر با هدف بهبود دقت و کارایی متغیرهای پیش‌بینی‌کننده تمرکز کنند. این بهبود می‌تواند از طریق انتخاب ویژگی‌های بهتر، استفاده از الگوریتم‌های پیشرفته یادگیری ماشین، تعدیل پارامترهای مدل و یا ترکیب این روش‌ها با هم انجام شود. بهبود دقت و کارایی متغیرهای پیش‌بینی‌کننده می‌تواند باعث بهبود دقت و کارایی مدل در پیش‌بینی نتایج باشد.

یک پیشنهاد عملی این است که موسسات مالی و تاجران باید در استفاده از مدل‌های داده‌کاوی به‌عنوان یک ابزار تکمیلی و نه به‌عنوان جایگزین برای تجزیه و تحلیل‌های بنیادین سهام، به دقت عمل کنند.

تحقیقات مروری نشان‌دهنده پیشرفت‌های مهمی در زمینه داده‌کاوی و پیش‌بینی قیمت سهام است، با وجود این، اتکای بیش از حد به الگوریتم‌ها بدون درک کامل و شفافیت می‌تواند به اشتباهات تصمیم‌گیری منجر شود. تحقیقات آتی باید نه تنها به بهینه‌سازی مدل‌ها بلکه به بهبود فرآیندهای تفسیرپذیری و شفاف‌سازی بینجامد.

این ساختار پیشنهادی باید بر اساس داده‌ها و یافته‌های مشخص تحقیق، تعدیل و توسعه یابد.

در نوشتن پیشنهادها برای یک بخش کامل از تحقیق، باید به صورت جامع به بررسی چگونگی بهره‌گیری از یافته‌ها در عمل، توصیه‌هایی برای تحقیقات بعدی و رویکردهایی که می‌توانند به بهینه‌سازی اقدامات فعلی کمک کنند، پرداخت. پیشنهادها باید بر اساس یافته‌های کلیدی تحقیق و همچنین چالش‌ها و محدودیت‌های کشف شده شکل گیرند.

- پیشنهادها برای تحقیقات آینده:

بررسی و توسعه الگوریتم‌های پیشرفته داده‌کاوی:

**** تحقیقات آینده باید به سمت توسعه الگوریتم‌هایی باشند که قادر به مدیریت حجم زیاد داده‌ها، پردازش سریع و کاهش پیچیدگی محاسباتی هستند. (توجه ویژه به یادگیری ماشینی و اعمال مدل‌های یادگیری عمیق برای شناسایی الگوهای پنهان در داده‌های بزرگ مالی)**

پیشنهاد می‌شود که متغیرهای جدید و بیشتری از جمله متغیرهای اقتصادی کلان، رویدادهای جهانی و داده‌های رسانه‌های اجتماعی در مدل‌سازی‌ها ادغام شوند تا به افزایش دقت پیش‌بینی سهام کمک کنند.

- پیشنهادها برای سازگاری با تغییرات بازار:

۱- پایش و پاسخگویی به تغییرات فوری بازار: لزوم ایجاد سیستم‌هایی که به صورت دینامیک به تغییرات ناگهانی و فوری بازار پاسخ دهند، به ویژه در شرایط نااطمینانی و بحران.

۲- افزایش قابلیت تفسیر الگوریتم‌ها: اگرچه الگوریتم‌های داده‌کاوی می‌توانند پیچیدگی بالایی داشته باشند، اما توسعه رویکردهایی که قابلیت تفسیر نتایج آن‌ها را بهبود بخشند، امری ضروری است.

- پیشنهادها برای بهبود تصمیم‌گیری‌های مالی:

۱- ترکیب تحلیل مالی کلاسیک و داده‌کاوی: مهندسی مالی و تحلیل‌گران باید به مدل‌هایی که هم تکنیک‌های سنتی تحلیل مالی و هم الگوریتم‌های داده‌کاوی را تلفیق می‌کنند، بپردازند.

۲- تعیین راهبردهای معاملاتی بر اساس الگوریتم‌ها: توسعه برنامه‌های تحلیلی که استراتژی‌های معاملاتی را بر اساس الگوریتم‌های داده‌کاوی بهینه می‌سازند.

- پیشنهادها برای رعایت مقررات و توسعه آموزش و برنامه‌ریزی:

۱- توجه به مسائل اخلاقی و حریم خصوصی: با افزایش استفاده از داده‌کاوی در بازارها، لزوم توجه به حریم خصوصی و اخلاق داده‌ها از اهمیت بالایی برخوردار است.

۲- رعایت مقررات و ایجاد ضوابط: تحلیل‌گران و موسسات مالی باید دائماً نسبت به مقررات جدید داده‌کاوی و تحلیل سهام به‌روز شوند و از تمامی ضوابط و مقررات مربوطه اطمینان حاصل کنند.

۳- توسعه برنامه‌های آموزشی: دوره‌ها و ورک‌شاپ‌هایی که تحلیل‌گران و مدیران مالی را با ابزارهای داده‌کاوی و تکنیک‌های مربوطه آشنا می‌کند.

۴- آموزش مداوم برای تحلیلگران: تأکید بر اهمیت آموزش مداوم برای تحلیلگران به منظور نگهداری و تقویت مهارت‌هایشان در استفاده از ابزارهای داده‌کاوی جدید.

این پیشنهادها که به صورت گسترده و جامع طراحی شده‌اند و نیازمند تلاش‌های بین‌رشته‌ای برای پیگیری و تحقق آن‌ها در آینده هستند. امید است که این پیشنهادها به عنوان نقطه عطفی برای محققان، تصمیم‌گیران و عملکردگران صنعت مالی عمل کنند تا به بهبود ادامه‌دار در فرایندها و نتایج مربوط به پیش‌بینی قیمت سهام منجر شوند.

		لیست مرجع	به بدنه مقاله مراجعه کنید
مقاله مروری		"Application and performance of data mining techniques in stock market" https://onlinelibrary.wiley.com/doi/10.1002/isaf.1518	
مقاله مجله با اطلاعات از دست رفته	شماره جلد موجود نیست	"Forecasting stock prices using a data mining method: Evidence from" 28 Jan. 2021, https://onlinelibrary.wiley.com/doi/10.1002/ijfe.2516 .	
	شماره نسخه موجود نیست	"Netflix Stock Price Movements Insights from Data Mining IEEE" https://ieeexplore.ieee.org/document/9972547/ .	
	شماره صفحه یا مقاله موجود نیست	"Concept and Practice of Artificial Market Data Mining Platform." https://ieeexplore.ieee.org/document/9776095 .	
		"Stock Price Prediction Based on Data Mining Combination Model." https://www.igi-global.com/viewtitle.aspx?titleid=296707 .	
		"A data mining framework for financial prediction Expert Systems with" 01 Jul. 2021, https://dl.acm.org/doi/abs/10.1016/j.eswa.2021.114651 .	

	"Progress and prospects of data-driven stock price forecasting research." 01 Jun. 2023, https://www.sciencedirect.com/science/article/pii/S2666307423000116 .	
	"A Study on Developing Effective Option Trading Strategy On Nifty Index" https://ieeexplore.ieee.org/document/9377179 .	
	"The Sustainable Development of Financial Topic Detection and Trend" 07 Jul. 2021, https://www.mdpi.com/2071-1050/13/14/7585 .	
	"Forecasting on Stock Market Time Series Data Using Data Mining Techniques." http://www.ijesi.org/papers/NCIOT-2018/Volume-4/2.%2006-13.pdf .	