دانشگاه پیام نور
دانشکده فنی و مهندسی

پایان‌نامه
برای دریافت درجه کارشناسی ارشد
رشته مهندسی کامپیوتر - گرایش نرم‌افزار
گروه مهندسی کامپیوتر و فناوری اطلاعات

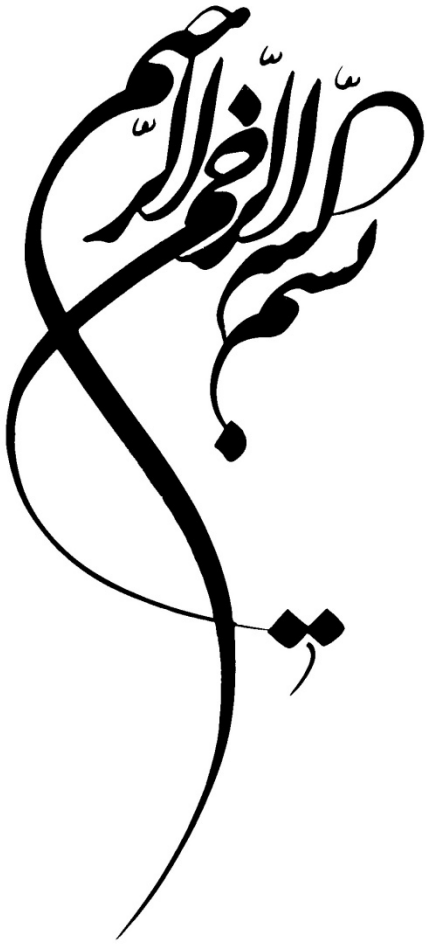
ارائه یک الگوریتم به منظور کشف قوانین وابستگی در تغییرات قیمت سهام (مطالعه موردی: بورس تهران)

نگارش:
مجتبی علاء

استاد راهنما:
دکتر محمد صنیعی آباده

استاد مشاور:
دکتر داود کریم زادگان مقدم

اردیبهشت ۱۳۹۱


بسم الله الرحمن الرحیم

دانشگاه پیام نور
دانشکده فنی و مهندسی
دانشگاه پیام نور مرکز تهران

پایان‌نامه
برای دریافت درجه کارشناسی ارشد
رشته مهندسی کامپیوتر -گرایش نرم‌افزار
گروه مهندسی کامپیوتر و فناوری اطلاعات

ارائه یک الگوریتم به منظور کشف قوانین وابستگی در تغییرات قیمت سهام (مطالعه موردی: بورس تهران)

نگارش:
مجتبی علاء

استاد راهنما:
دکتر محمد صنیعی آباده

استاد مشاور:
دکتر داود کریم زادگان مقدم

تقدیم به:

پدر و مادر عزیزم

تشکر و قدردانی:

در انجام این تحقیق، بر خود واجب می‌دانم از زحمات استاد ارجمند **جناب آقای دکتر محمد صنیعی آباده**، که مسئولیت راهنمایی پایان‌نامه را بر عهده داشتند، **جناب آقای دکتر داود کریم زادگان مقدم**، استاد محترم مشاور و نیز **جناب آقای دکتر رضا عسگری مقدم**، که با دقت و حوصله فراوان داوری این پایان‌نامه را به انجام رساندند، صمیمانه تشکر کنم.

همچنین باید سپاس فراوان خود را تقدیم دوست ارجمند جناب آقای حمید محمدی نمایم که با حمایت‌های خود، زمینه هر چه بهتر شدن این تحقیق را فراهم ساخت.

چکیده

کاوش قوانین وابستگی یک تکنیک مهم در داده‌کاوی است که در آن یافتن الگوهای تکراری، یک مرحله بسیار زمان‌بر و پرهزینه به شمار می‌رود. در این تحقیق یک روش جدید برای یافتن الگوهای تکراری بر روی داده‌هایی که مقادیر آن‌ها به شکل نوسانی (افزایش، کاهش یا بدون تغییر) هستند مانند داده‌های تغییرات قیمت سهام در بورس اوراق بهادار ارائه می‌شود. در این روش، فقط تعداد کمی از تکرار مجموعه عناصر از پایگاه داده جستجو می‌شود و تکرار بقیه مجموعه عناصر به کمک تکرار مجموعه عناصر بدست آمده و روابط پیدا شده بین تکرار آن‌ها محاسبه می‌شود. بدین ترتیب نیاز به کاوش کل الگوهای تکراری از پایگاه داده نیست. این امر باعث صرف زمان بسیار کمتری برای یافتن الگوهای تکراری و در نهایت یافتن قوانین وابستگی می‌شود. با توجه به نکات یاد شده، در این تحقیق با پیاده‌سازی الگوریتم FP-Math (مخفف Frequent Patterns-Math) کارایی آن بر روی داده‌های واقعی بورس تهران بررسی و نشان داده می‌شود. همچنین نتایج آزمایشات نشان می‌دهد الگوریتم FP-Math در مقایسه با الگوریتم‌های پیشین مانند الگوریتم Apriori، که یکی از مشهورترین و پر کاربردترین آن‌ها در این زمینه است، تقریباً سه برابر سریع‌تر اجرا می‌شود. در مجموع، بهره‌گیری از الگوریتم پیشنهادی توانست الگوهای تکراری و قوانین وابستگی جالبی را از داده‌های بورس اوراق بهادار تهران استخراج نماید.

واژه‌های کلیدی: *داده کاوی، کاوش قوانین وابستگی، الگوهای تکراری، بورس.*

فهرست مطالب

عنوان **صفحه**

فهرست شکل‌ها

عنوان **صفحه**

فصل ۶ جمع بندی و پیشنهادها

فهرست جداول

فهرست علائم اختصاری

AR	Association Rules	قوانین وابستگی
DM	Data Mining	داده کاوی
DB	Database	پایگاه داده
FP	Frequent Patterns	الگوهای تکراری
KDD	Knowledge Discovery from Database	استخراج دانش از پایگاه داده
TID	Transaction Identify	شناسه تراکنش

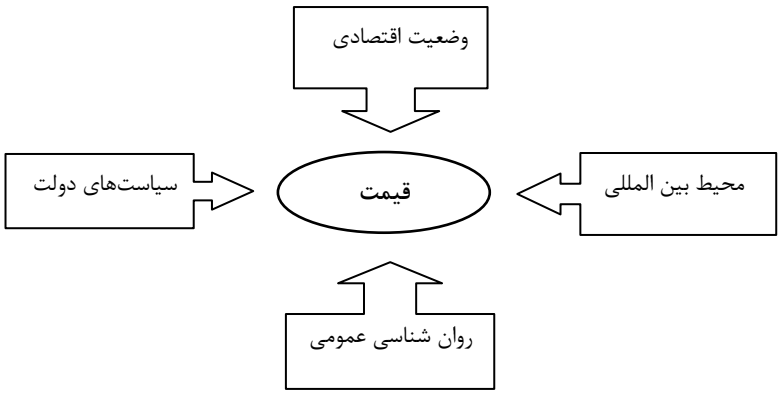
فصل۱

مقدمه

۱-۱. مقدمه

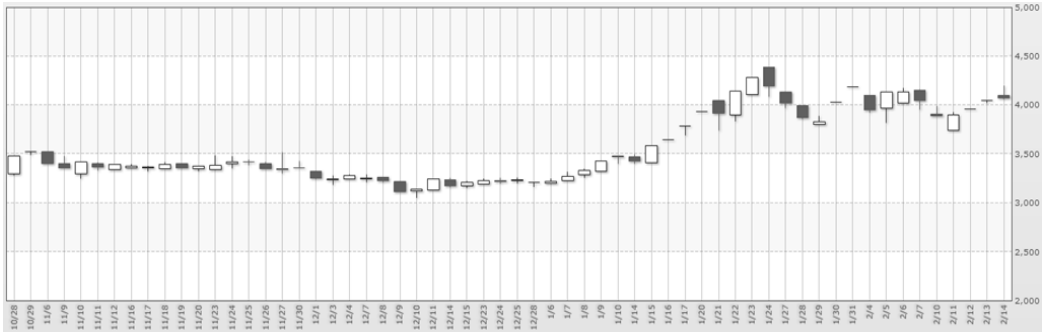
بورس اوراق بهادار تهران، بازار رسمی و سازمان یافته سرمایه است که در آن اوراق بهادار پذیرفته شده از قبیل سهام شرکت‌ها و اوراق مشارکت، تحت قوانین و مقررات خاصی داد و ستد می‌شوند. رسالت این بازار، جذب سرمایه‌های اندک و سرگردان عموم و هدایت آن به سمت فعالیت‌هایی است که متضمن منفعت عامه هستند.

قیمت سهام تحت تأثیر عوامل جالب زیادی مانند محیط بین‌المللی، سیاست‌های دولت، وضعیت آب و هوا، وضعیت اقتصادی، روان‌شناسی عمومی و حتی شایعات دچار تغییر و نوسان (مثبت و منفی) می‌شود. شکل ۱-۱ به خوبی بیانگر تأثیرگذاری عوامل مختلف بر روی قیمت سهام است.



شکل۱-۱. عوامل تأثیرگذار بر روی قیمت سهام شرکت‌های بورس اوراق بهادار تهران

در زمینه سرمایه‌گذاری در بازار بورس، سرمایه‌گذاران در یک تحلیل تکنیکی، معتقد هستند که همه اطلاعات یک سهام در سابقه قیمت آن نهفته است و به عبارتی روند قیمت سهام تکرار شونده است به طوری که بررسی نوسانات قیمت در گذشته، آینده بازار را نشان می‌دهد؛ بنابراین اطلاعات ارزشمندی در داده‌های تاریخی بازارهای بورس اوراق بهادار از جمله بازار بورس تهران وجود دارد که تلاش برای استخراج این اطلاعات و استفاده از آن در جهت کمک به تصمیم‌گیری سرمایه‌گذاران حائز اهمیت فراوان می‌باشد. شکل ۱-۲ نمونه‌ای از روند قیمت سهام را در بازار بورس تهران در یک دوره زمانی سه ماهه نشان می‌دهد.



شکل۱-۲. تغییرات قیمت سهام بانک پارسیان در یک دوره زمانی سه ماهه بر اساس نمودار کندل استیک[1]

یکی از مهم‌ترین تکنیک‌ها در داده‌کاوی که تحقیقات گسترده‌ای را در بازار سرمایه به خصوص بورس اوراق بهادار در سال‌های اخیر به خود معطوف نموده، کاوش قوانین وابستگی است. اما با وجود پیشرفت‌های صورت گرفته در این تکنیک، کاوش الگوهای تکراری در آن به دلیل جستجوی مجموعه عناصر تکراری، بسیار زمان‌بر و پرهزینه می‌باشد.

با توجه به نقطه ضعف یاد شده، هدف از انجام این تحقیق ارائه یک الگوریتم سریع برای یافتن قوانین وابستگی در داده‌های بورس اوراق بهادار می‌باشد. جهت رسیدن به این هدف، در

[1] Candlestick

مرحله کاوش الگوهای تکراری، در گام نخست با استفاده از ویژگی‌های داده‌های بورس، روابط جالبی بین تکرار مجموعه عناصر پیدا می‌شود که در گاه بعدی با به کار گیری آن کشف الگوهای تکراری سریع‌تر میسر خواهد گردید.

به منظور دست‌یابی به ساختاری منسجم و مناسب برای انجام تحقیق، در ادامه این فصل به بیان مهم‌ترین اصول و پاسخ‌گویی به سوالات اصلی پژوهش، پرداخته خواهد شد.

۱-۲. تعریف مسأله و سوالات اصلی تحقیق

داده‌کاوی[1] فرآیندی است که با نگرشی نو، به مسئله استخراج اطلاعات ناشناخته و بالقوه مفید از پایگاه داده‌ها می‌پردازد (هان[2] و کمبر[3]، ۲۰۰۶). طی دهه‌های گذشته الگوریتم‌ها و تکنیک‌های مختلفی برای بدست آوردن دانش مبتنی بر قاعده در داده‌کاوی توسعه یافته‌اند. یکی از مطرح‌ترین این روش‌ها، **کاوش قوانین وابستگی**[4] است که در آن روابط تکراری بین اشیاء مختلف با یکدیگر (الگوهای تکراری)، کشف می‌شود.

استفاده از قوانین وابستگی در تحلیل داده‌های بازار سرمایه می‌تواند برای یافتن وابستگی بین تغییرات قیمت سهام شرکت‌های بورسی و نشان دادن قواعد و الگوهای موجود در آن‌ها به منظور پیشنهاد سبد سهام و گروه‌بندی مجموعه‌های سرمایه‌گذاری سهام مورد استفاده قرار گیرد.

هر چند الگوریتم‌های متنوع زیادی وجود دارند که به کشف الگوهای تکراری و قوانین

[1] Data Mining
[2] Han, J.
[3] Kamber, M.
[4] Association Rule Mining

وابستگی در پایگاه داده‌های مختلف می‌پردازند، ولی با توجه به تعداد زیاد شرکت‌های پذیرفته شده در بورس و متراکم بودن پایگاه داده‌های بورس، کشف قوانین وابستگی در داده‌های بورس در مرحله یافتن الگوهای تکراری زمان بسیار زیادی را صرف می‌کند. به این دلیل و با در نظر گرفتن اهمیت تحلیل داده‌های بازار سرمایه، نیاز به یک الگوریتم سریع مختص این نوع داده‌ها احساس می‌شود.

در این پایان‌نامه، ضمن بررسی ویژگی‌های داده‌های بورس، با تمرکز بر روی روابط بین تکرار الگوها در این نوع داده‌ها، روش سریعی برای کشف الگوهای تکراری پیدا خواهد شد که با استفاده از آن، یک الگوریتم بسیار سریع مختص داده‌های با نوع مقدار نوسانی، مانند داده‌های بورس، ارائه می‌شود. همچنین با توجه به قابلیت قیاس بین سبد سهام[1] و سبد بازار، از الگوریتم پیشنهادی برای پیدا کردن وابستگی‌ها در بین داده‌های بورس تهران استفاده خواهد شد. الگوهای تکراری در بین چند شرکت فعال‌تر[2] در صنایع مختلف در بورس تهران کشف خواهد شد که ارتباط بین تغییرات قیمت سهام آن‌ها را نشان می‌دهد. در نتیجه، برای سرمایه‌گذاران اطلاعات مفیدی درباره رفتار بازار سهام فراهم خواهد شد که به کمک این اطلاعات می‌توانند سیاست سرمایه‌گذاری‌هایشان را به صورت دقیق‌تری مشخص کنند. در نهایت از داده‌های واقعی بورس تهران جهت بررسی عملکرد این الگوریتم استفاده خواهد شد.

بر این اساس، پرسش پژوهش پیش رو این است که «آیا تکنیک قوانین وابستگی با استفاده از الگوریتم پیشنهادی می‌تواند به نحو مؤثرتری الگوهای تکراری را از تغییرات قیمت سهام در داده‌های بورس تهران استخراج کند؟».

[1] Portfolio

[2] ۵۰ شرکت به عنوان شرکت‌های فعال‌تر در هر دوره ۳ ماه توسط بورس اوراق بهادار تهران معرفی می‌شود.

۱-۳. فرضیه‌ها

- پایگاه داده بورس، متراکم[1] است و چون نمادها به ازای هر روز معاملاتی دارای یک تراکنش هستند، شامل حجم بزرگی از اطلاعات می‌باشد.
- برای کاوش قوانین وابستگی، از نوع مقدار داده به صورت تغییرات (افزایش، کاهش یا بدون تغییر) قیمت سهام استفاده می‌شود.
- محدودیت حافظه وجود ندارد.
- در دوره مورد بررسی پارامترهای اقتصادی بورس اوراق بهادار تهران نباید تحت تأثیر رخدادهای کاملاً ویژه و بحرانی قرار گیرد.

۱-۴. اهداف تحقیق

هدف اصلی پژوهش حاضر آن است که برای کشف قوانین وابستگی، یک الگوریتم جدید با سرعت بسیار بالاتر، مختص داده‌های بورس، ارائه دهد. با استفاده از این الگوریتم، کشف وابستگی‌ها در داده‌های بورس، به طور مؤثرتری انجام خواهد شد.

اهداف فرعی از این پژوهش نیز عبارتند از:

- کشف الگوهای تکراری و قوانین وابستگی در تغییرات قیمت سهام چند شرکت فعال‌تر در بورس اوراق بهادار تهران؛

[1] متراکم بودن پایگاه داده بورس به این معنی می‌باشد که تمام نمادهای در تمام تراکنش‌ها حتماً دارای یکی از سه مقدار افزایش، کاهش و یا بدون تغییر می‌باشند.

- تسهیل سرمایه‌گذاری در بورس اوراق بهادار؛
- ایجاد اطمینان خاطر به سرمایه‌گذاران در پیش‌بینی قیمت سهام؛

۱-۵. روش تحقیق

این پژوهش به روش تحلیلی- تطبیقی صورت خواهد گرفت و در آن سعی خواهد شد به دو صورت مطالعه تاریخی (مطالعه طرح‌ها و روش‌های مختلف موجود در راستای موضوع تحقیق) و تطبیقی (مقایسه با روش‌های مرتبط موجود) سوالات و فرضیات تحقیق مورد ارزیابی و آزمون قرار گیرد.

۱-۶. مراحل انجام تحقیق

- شناسایی و بیان کامل مسئله
- ارزیابی روش‌های موجود
- تشریح مسائل مربوط به کاوش الگوهای تکراری بر روی داده‌های با نوع مقدار نوسانی
- ارائه یک روش جدید برای کشف قوانین وابستگی در داده‌های بورس و پیاده‌سازی الگوریتم آن
- ارزیابی عملکرد الگوریتم جدید از طریق مقایسه با الگوریتم‌های پیشین
- استخراج الگوها و قوانین وابستگی از داده‌های تاریخی بازار بورس توسط الگوریتم جدید