

تکنیک های داده کاوی در پیش بینی قیمت سهام بورس اوراق بهادار

اسدالله مهر آرا*

دانشگاه آزاد اسلامی واحد قائم شهر، گروه مدیریت دولتی، قائم شهر، ایران
Mehrara_a@yahoo.com

زهرا عطف

دانشگاه آزاد اسلامی واحد قائم شهر، دانشجوی کارشناسی ارشد مدیریت دولتی،
قائم شهر، ایران
Atf.zahra@yahoo.com

زهرا عسکری

دانشگاه آزاد اسلامی واحد قائم شهر، دانشجوی کارشناسی ارشد مدیریت دولتی،
قائم شهر، ایران
Farideh9022@yahoo.com

چکیده

دستیابی به رشد اقتصادی و ایجاد انگیزه جهت سرمایه گذاری زمانی در یک کشور تسريع می گردد که آن کشور دارای بازارهای سرمایه فعال و قابل اعتماد باشد. وجود بازارهای بورس فعال همواره سرمایه گذاران متعددی را به تکاپو واداشته و جریان سرمایه و منابع مالی را به بخش های مولده تسريع می نماید. از روش های دقیق تری که می توان در هر نوع پیش بینی از آن استفاده کرد تکنیک های داده کاوی از جمله شبکه عصبی و درخت های تصمیم و رگرسیون منطقی، تحلیل تفکیکی، مدل افزودنی کلی، روش Boosting است. پیش بینی قیمت سهام شرکتهای پذیرفته شده در بورس اوراق بهادار تهران با استفاده از تکنیک های داده کاوی و با توجه به معیارهای مالی از جمله موضوع های با اهمیتی است که متأسفانه تاکنون در ارتباط با آن تحقیقات اندکی انجام شده است. در این مقاله پس از بررسی مفاهیم بورس اوراق بهادار، انواع تکنیک های داده کاوی مورد بررسی قرار خواهد گرفت.

کلید واژه: داده کاوی، بورس اوراق بهادار، پیش بینی قیمت.

*۱-نویسنده مسؤول

مقدمه

بورس کلمه ای فرانسوی است که معنای آن در زبان فرانسه "کیف پول" می‌باشد. بورس یک نهاد سازمان یافته‌ای است که از جمله نهادهای عمدۀ و اساسی در بازار سرمایه محسوب می‌شود و در کنار سایر مؤسسات و سازمان‌ها، وظایف چندگانه ای را بر عهده دارد. کارکردهای اساسی بورس، مدیریت انتقال ریسک و توزیع آن، شفافیت اطلاعات، کشف قیمت، ایجاد بازار رقابتی و همچنین یکی از کارکردهای مهم آن جمع آوری سرمایه‌ها و پس اندازهای کوچک برای تأمین سرمایه مورد نیاز فعالیت‌های اقتصادی است. بازار بورس یک بازار ثانویه است، یعنی در بورس اوراق بهادر شرکتی تأسیس نمی‌شود و فقط سهام شرکت‌های پذیرفته شده (که قبلًا تاسیس شده‌اند) مورد معامله قرار می‌گیرند. به همین دلیل بازار بورس، بازار ثانویه است. (اسکندری، ۱۳۷۱) امروزه بازارهای بورس‌نه تنها از پارامترهای اقتصاد کلان، بلکه از بسیاری از عوامل دیگر متاثر می‌شوند. تعدد و ناشناخته بودن عوامل مؤثر بر بازار بورس، معمولاً موجب عدم اطمینان در زمینه سرمایه‌گذاری می‌شود. بنابراین بطور طبیعی تمام تلاش سرمایه‌گذار کاوش عدم اطمینان است و از این جهت پیش‌بینی بازار بورس یکی از ابزارهای کاوش عدم اطمینان و به عنوان یکی از مهمترین شاخه‌های علمی مطرح شده است و روز به روز توسعه و پیشرفت می‌نماید. مدیران بخش‌های مختلف اقتصادی و بازرگانی، به دلیل وجود انبوه متغیرهای تأثیرگذار ترجیح می‌دهند مکانیزمی را در اختیار داشته باشند که بتواند آن‌ها را در امور تصمیم‌گیری شان یاری و مشاوره دهد. به همین دلیل، سعی در روی آوردن به روش‌هایی در پیش‌بینی دارند که به واسطه آن‌ها تخمین‌هایشان به واقعیت نزدیک و خطایشان بسیار کم باشد. از آنجا که بورس اوراق بهادر تهران یکی از ارکان بازار سرمایه کشور به شمارمی‌آید و پیشرفت و گسترش این سازمان می‌تواند رشد و توسعه اقتصاد ملی را در پی داشته باشد، بررسی همه عوامل اثر گذار بر این سازمان، از جمله عواملی که منجر به گرایش هر چه بیشتر سرمایه‌گذاران به بورس اوراق بهادر می‌شود اهمیت ویژه دارد. در سالهای اخیر فن آوری‌های تولید و گرد آوری داده‌ها به سرعت در حال رشد بوده است. مسئله پیش‌روی سازمان‌ها به ویژه در حوزه پیش‌بینی‌های مالی دیگر جمع آوری صرف داده‌ها نیست بلکه دستیابی به توان استخراج دانش مفید نهفته در داده‌ها دغدغه اصلی سازمان‌ها است. در چنین شرایطی است که باید از رشد فن آوری برای استفاده مؤثرات این دانش بالقوه سود جست و داده کاوی² یک جواب مناسب برای استخراج این شروط است.

تکنیک‌ها و روش‌های متعددی مانند رگرسیون، شبکه‌های عصبی، دسته بندی، طبقه بندی، درخت تصمیم و ... برای داده کاوی وجود دارد که ریشه تمام آنها در ریاضیات، آمار، هوش محاسباتی و یادگیری ماشینی است. در تئوری، بیشتر تکنیک‌های داده کاوی تایل به جموعه‌های داده ای بزرگ دارند زیرا جموعه‌های داده ای بزرگتر مدل‌های دقیق‌تری را به همراه خواهد داشت.

تاریخچه ای درباره بورس اوراق بهادر

تاریخچه تأسیس اولین بورس به اوخر قرن چهارده و اوایل قرن پانزده در کشور هلند بر می‌گردد. اما بورس به صورت مدون از دو قرن پیش شروع به

².Data Mining

فعالیت کرده است. در حال حاضر با پیشرفت تکنولوژی و استفاده از ابزارهای الکترونیکی بیشتر معاملات در بورس از طریق رایانه صورت می‌گیرد.

واژه بورس از نام خانوادگی شخصی به نام واندربورس^۳ اخذ شده که در اوایل قرن پانزدهم در شهر بروز^۴ در بلژیک می‌زیسته و صرافان شهر در مقابل خانه او گرد می‌آمدند و به داد و ستد کالا، پول و اوراق بهادر می‌پرداختند. نام بورس بعدها به کلیه اماکنی اطلاق شد که محل داد و ستد کالا، پول و اسناد تجاری و مالی بوده است. تشکیل اولین بورس اوراق بهادر جهان، در سال ۱۴۶۰ روی داده است. در این سال در شهر آنورس^۵ بلژیک که موقعیت تجاری قابل ملاحظه ای داشت، اولین بازار مشکل سرمایه به وجود آمد. (رجب زاده، ۱۳۷۶، ۱۳۸۱) و (صرفنحواده، ۱۳۸۱، ۱۳۸۱)

اندیشه اصلی ایجاد بورس اوراق بهادر در ایران به سال ۱۳۱۵ باز می‌گردد. در این سال یک کارشناس بلژیکی به نام وان لوترفلد^۶ به همراه یک کارشناس هندی، به درخواست دولت وقت ایران، درباره تشکیل بورس اوراق بهادر مطالعاتی نمودند و طرح تأسیس و اساسنامه آن را تهیه نمودند. در همان سال ها بانک ملی نیز، به عنوان تنها متصدی امور بانکی در کشور، مطالعات مشابهی را در

در سال ۱۳۴۱ با تشکیل کمیسیونی مشکل از بانک مرکزی، بانک توسعه صنعتی و معدنی ایران و به میزانی وزارت بازرگانی و با همکری بلژیکی ها، موضوع تأسیس سازمان بورس اوراق بهادر به صورت جدی دنبال شد.

در سال ۱۳۴۵ قانون و مقررات تشکیل سازمان بورس اوراق بهادر تهران تهیه و به جلس ارائه شد و در اردیبهشت سال ۱۳۴۵ به تصویب رسید ولی باز هم آمادگی شروع به کار این سازمان تا بهمن سال ۱۳۴۶ به وجود نیامد. پس از تصویب قانون گسترش مالکیت و به دنبال آن، تأسیس سازمان ملی گسترش مالکیت، عرضه سهام در بورس و انتقال آن به مردم تسهیل شد. از نیمه دوم سال ۱۳۵۷، همان با اعتمادهای مبارزات مردم معاملات این بازار به دلیل بی اعتمادی به دولت و وضع مالی شرکت ها، فرار سرمایه، با سقوط قیمت سهام و کاهش معاملات روبرو شد و این روند به دلیل افزایش تورم، ملی شدن بانک ها، وقوع جنگ تحمیلی و مشخص نبودن مالکیت سهام ادامه پیدا کرد به طوری که معامله سهام و اوراق مشارکت تا سال ۱۳۶۱ تقریباً متوقف شد. در سال ۱۳۶۲، تا حدودی تقاضا برای سهام وجود داشت، ولی به دلیل پایین بودن قیمت های پیشنهادی خریداران، عرضه کنندگان چندان زیاد نبودند. در سال ۱۳۶۳، به دنبال تصمیم شورای بورس مبنی بر واگذاری تعدادی از کارخانجات صنعتی دولت به کارگزاران و سایر افراد بخش خصوصی، مبادلات سهام اندکی افزایش یافت و تا سال ۱۳۶۷ افزایش حجم مبادلات ادامه یافت. بورس اوراق بهادر تهران در سال ۱۳۶۸ در چارچوب سیاست های دولت و به منظور اجرای مبنی از برنامه پنج ساله توسعه اقتصادی و به دنبال سیاست خصوصی سازی و انتقال سهام شرکت ها به مردم از طریق بورس اوراق بهادر، فعالیت خود را به طور گسترده ای آغاز کردند.

³. Vanderbourse

⁴. Bruges

⁵. Anvers

⁶. Vanluterfeld

اهمیت بورس اوراق بهادار

رشد و توسعه هر کشور، مستلزم به کارگیری منابع در شکل بهینه و هدایت آنها در مسیری درست است. در هر جامعه ای نهادهای مختلفی می‌توانند در جهت این هدف گام برداشته و نقش مؤثری را در این رابطه ایفا کنند. بازارهای مالی و نهادهای مربوط به آن، یکی از مهم ترین عوامل مؤثر بر این فرآیند هستند. به رغم تنوع و تعدد نهادهای مربوط به بازارهای مالی، بازار بورس اوراق بهادار مهم ترین و اصلی ترین مرکز برای مبادلات دارایی‌های مالی تلقی می‌شود. بی‌شك فعالیت صحیح و درست بازار بورس اوراق بهادار، می‌تواند نقش مؤثری در تخصیص بهینه منابع در سطح کلان اقتصادی داشته و سبب رشد و توسعه اقتصادی کشور می‌گردد. نقش اصلی بورس اوراق بهادار جذب و هدایت پس اندازها و نقدینگی سرگردان و پراکنده در اقتصاد به سوی مسیرهای بهینه آن است، به گونه‌ای که منجر به تخصیص بهینه منابع کمیاب مالی شود. اما این امر مهم منوط به وجود کارآیی بازارهای مالی است.

برای موفقیت در رشد اقتصادی، فعالیت و عملکرد بورس اوراق بهادار اهمیت و حساسیت خاصی پیدا می‌کند. یکی از عملکردهای اقتصادی بازارهای مالی از جمله بازار سرمایه، انتقال وجهه از دارندگان وجهه اضافی برای سرمایه‌گذاری به کسانی که به وجودی برای سرمایه‌گذاری دارایی‌های مشهود نیاز دارند، است. علاوه بر این، کارآبودن بازار از اهمیت زیادی برخوردار است، چرا که در صورت کارآبودن بازار سرمایه، هم قیمت اوراق بهادار به درستی و عادلانه تعیین می‌شود و هم تخصیص سرمایه که مهم ترین عامل تولید و توسعه اقتصادی است به صورت مطلوب و بهینه انجام می‌شود. به طوری که هر چه میزان کارآیی بورس اوراق بهادار افزایش پیدا کند به همان اندازه اطمینان سرمایه‌گذاران نسبت به بورس بیشتر شده و رغبت آنها جهت سرمایه‌گذاری، افزایش خواهد یافت. با مطالعه تحقیقات تجربی کارآیی بازار بورس در ایران مشاهده می‌شود که به رغم این که بورس اوراق بهادار ایران در بهمن ۱۳۴۶ شروع به فعالیت کرده، این تحقیقات عدم کارآیی این بازار را به اثبات می‌رساند. با این وجود در طی سال‌های اخیر اقدامات مناسبی در جهت پیشرفت، توسعه و افزایش شفافیت بازار اوراق بهادار انجام شده است که می‌توان اثرات آن را در فرهنگ سهامداری مردم، گسترش جغرافیایی بازار، افزایش حجم معاملات بورس و افزایش تعداد سهامداران دید. اما با این حال، طبق آخرین مطالعاتی که در زمینه بررسی کارآیی بورس اوراق بهادار انجام شده، کارآیی بازار تأیید نشده است. در این راستا، عدم کارآیی بورس اوراق بهادار تهران موجب می‌شود شرکت‌های پذیرفته شده در این بازار، در روند تأمین مالی خود با مشکلات و چالش‌هایی روبرو شوند. (زارعی، ۱۳۷۰)

شاخص کل قیمت بورس (TEPIX)

کلمه شاخص^۷ در کل به معنای نمودار یا نشان دهنده یا نماینده می‌باشد. شاخص، کمیتی است که نماینده چند متغیر همگن می‌باشد و وسیله‌ای برای اندازه‌گیری و مقایسه پدیده‌هایی است که دارای ماهیت و خاصیت مشخصی هستند که بر مبنای آن می‌توان تغییرات ایجاد شده در متغیرهای معینی را در طول یک دوره بررسی نمود. در هر بازار بورس اوراق بهادار انجام شده، کارآیی بازار تأیید نشده است. در این شاخصهای زیادی را تعریف و محاسبه نمود. در تمام بورس‌های دنیا نیز شاخص‌های زیادی برای گروه‌ها و شرکت‌های مختلف محاسبه می‌شود. بعنوان مثال در بازار

⁷. Index

سهام امریکا شاخص داو جونز^۸، تغییرات ۳۰ شرکت صنعتی، ۲۰ شرکت گمل و نقل و ۱۵ شرکت خدماتی را نشان می‌دهد و یا شاخص نزدک^۹ تغییرات سهام خارج از بورس را نشان می‌دهد. شاخص کل قیمت که خفف کلمه TEHRAN PRICE INDEX می‌باشد یکی از اصلی ترین شاخصهای بورس اوراق بهادار تهران است که روش محاسبه آن بر اساس فرمول زیر (فرمول لاسپیرز) انجام می‌گیرد.

{ ارزش جاری سهم در بازار*تعداد سهام منتشره) تقسیم بر (ارزش پایه سهام*تعداد سهام در سال پایه) } ضرب در عدد ۱۰۰

که در این محاسبه سال پایه سال ۱۳۶۹ می‌باشد. این شاخص گویای آن است که ارزش کل بازار نسبت به سال پایه (یا همان سال ۱۳۶۹) چند برابر شده است.

شاخص کل قیمت در بورس تهران برای سه گروه محاسبه می‌شود که عبارتند از:

-۱- شاخص قیمت کل بازار : که در محاسبه آن قیمت سهام تمام شرکتهای معامله شده تأثیر داده می‌شود.

-۲- شاخص قیمت تالار اصلی : که در محاسبه آن فقط قیمت سهام شرکتهای معامله شده در تابلو اصلی تأثیر داده می‌شود.

-۳- شاخص قیمت تالار فرعی : که در محاسبه آن فقط قیمت سهام شرکتهای معامله شده در تابلوی فرعی تأثیر داده می‌شود.

عوامل مؤثر بر نوسان قیمت سهام

معمولاً در بورس‌های اوراق بهادار حساسیت زیادی نسبت به روند قیمت سهام نشان داده می‌شود و چگونگی تحولات مرتبط با آن مورد بررسی قرار می‌گیرد. از آن جایی که بازار باید در معرض ترافیک دوطرفه از اطلاعات باشد، ما نباید در بازار شاهد این باشیم که در یک مقطعی از زمان همه خریدار و در مقطعی دیگر همه فروشنده باشند. وقوع چنین رویدادی نشان از عدم استقلال رأی و ارزیابی بازار دارد و این ها هم به دلیل آن است که عوامل واقعی که باید تغییرات آن ها بر روی قیمت‌ها تأثیر بگذارند، برای همه شناخته شده نیست. عوامل مؤثر بر قیمت سهام را می‌توان به دو دسته داخلی و خارجی تقسیم بندی کرد. از گمله عوامل داخلی می‌توان به موارد ذیل اشاره کرد: «سیاست تقسیم سود، ساختار سرمایه، کیفیت اطلاعات مالی، مدیریت شرکت، صنعت، وابستگی ارزی، وضعیت کالا، تغییرات عده قیمت محصولات نهایی، مصوبات جمع عمومی سهام داران. عوامل خارجی موثر بر قیمت سهام عبارتند از: عوامل اقتصادی (تورم، سیاست پولی و مالی دولت، بدھی خارجی دولت، تغییرات نرخ ارز، نرخ بهره بانک ها...)، عوامل فرهنگی و اجتماعی، عوامل سیاسی، عوامل فنی، تعزیرات حکومتی و محدودیت شرکت‌ها، عملکرد بورس اوراق بهادار خصوصاً در زمینه قانون گذاری و خواه اجرای آن و استفاده از ابزارهای حمایتی و...» (Miao,2007)

کارآیی بازار سرمایه

کارآیی بازار سرمایه یک مفهوم با اهمیت است و بیانگر عملکرد بازار سرمایه است. کارآ بودن بازار سرمایه به رشد و رونق اقتصاد کشور کمک شایانی می-

⁸. Dow Jones

⁹. NASDAQ

کند. در صورت کارآ بودن بازار سرمایه قیمت سهام به ارزش ذاتی سهام نزدیک بوده و پس اندازها به صورت بهینه به سرمایه گذاری های سودآور اختصاص پیدا می کند. به طوری که این امر هم به نفع سرمایه گذاران و هم به اقتصاد کشور کمک شایانی می کند. فرضیه بازار کارآ نقش مهمی در ادبیات مالی و اقتصاد داشته و بر استفاده کارآ از اطلاعات توسط عاملان بازار تأکید دارد. واژه «کارآیی» جهت توصیف بازاری مورد استفاده قرار می گیرد که اطلاعات در قیمت دارایی های مالی منعکس شده باشد.

منظور از کارآیی بازار سرمایه، اشاره به این مسئله است که تا چه میزان بازار در تعیین قیمت اوراق بهادار موفق عمل کرده است. موفق بودن بازار به این معنی است که قیمت ها به طور پیوسته منعکس کننده اطلاعات جدید باشند و اطلاعات نیز جمجمه ای از داده ها است که مربوط به شرکت ها و اوراق بهادار آنها است و قیمت ها در بازار کارآ باید تأثیرپذیر از این اطلاعات باشند. بنابراین، بازاری را می توان کارآ نامید که کارآیی لازم را برای پردازش اطلاعات داشته باشند. به عبارت دیگر، قیمت ها در هر زمان نشان دهنده ارزیابی صحیحی از اطلاعات موجود است، در نتیجه قیمت ها منعکس کننده کامل اطلاعات موجود خواهند بود. به طوری که کارآیی بازار سرمایه از همان آغاز یکی از موضوعات بحث برانگیز است.

کارآیی بازار سرمایه به دو جنبه مهم در تعیین قیمت ها توجه ویژه ای دارد و این دو جنبه عبارتند از: سرعت و کیفیت تعیین قیمت ها. اگر در بازاری عدم کارآیی در رابطه با سرعت و کیفیت تعیین قیمت ها وجود داشته باشد، سرمایه گذاران مطلع و کسانی که گوش به زنگ چنین موقعیت هایی هستند، می توانند سود بالایی را برای خود ایجاد کنند. اگر قیمت های جاری منعکس کننده اطلاعات با ارزش باشند، بسیار مشکل خواهد بود که اوراق بهادار ارزانی را یافت که بازده بالایی ایجاد کند یا در اوراق بهاداری سرمایه گذاری شود که قیمت آن بالا و بازده آن پایین باشد. تنها در صورتی می توان سرمایه گذاری خوبی داشت که بتوان آینده را به خوبی پیش بینی کرد. فرضیه بازار کارآ فرض می کند که قیمت سهام به سرعت با اطلاعات جدید تعدیل می شود و قیمت جاری هم اطلاعات را به طور کامل منعکس می کند. (جهانخانی، ۱۳۷۶)

در ادبیات مالی واژه «بازار کارآ» توسط فاما^{۱۰} (۱۹۶۵) ارائه شد. به طوری که کارآیی بازار سرمایه به صورت های مختلفی مورد بحث قرار گرفته است. ساده ترین تعریف را می توان به تعریف فاما از بازار کارای سرمایه اختصاص داد: در صورتی که در یک بازار، قیمت اوراق بهادار در هر لحظه از زمان انعکاس کاملی از همه اطلاعات در دسترس باشد، می توان آن بازار را کارآ نامید. (Fama, 1965).

مبانی تحلیل تکنیکی در بازار سرمایه

سرمایه‌گذار می‌باید ابزاری در اختیار داشته باشد تا به کمک آنها بتواند مسیر حرکت سهام را پیش‌بینی کند. در بازار سرمایه این ابزار به طور کلی بر دو نوع است:

(۱) تحلیل بنیادی: از نظر جهانخانی (۱۳۷۴)، تحلیل بنیادی عبارت است از مطالعه و بررسی شرایط اقتصادی، صنعت و وضعیت شرکت. هدف اصلی از انجام این گونه مطالعات، تعیین ارزش واقعی سهام و ارزش ذاتی شرکتهاست. تحلیل بنیادی معمولاً بر اطلاعات و آمار کلیدی مندرج در صورتهای مالی شرکت مرکز مینماید تا مشخص کند که آیا قیمت سهام به درستی ارزشیابی شده است یا نه؟

¹⁰. Fama

بخش قابل ملاحظه‌ای از اطلاعات بنیادی، بر اطلاعات و آمار اقتصاد ملی، صنعت و شرکت تمرکز دارد. رویکرد معمول در تحلیل شرکت، در برگیرنده چهار مرحله اساسی است :

(۱) تعیین وضعیت کلی اقتصاد کشور

(۲) تعیین وضعیت صنعت

(۳) تعیین وضعیت شرکت

(۴) تعیین ارزش سهام شرکت (جهانگانی، ۱۳۷۶)

۲) تحلیل تکنیکی : روش تحلیل تکنیکی برپایه سه اصل اساسی بنا شده است که در تمام بازارهای جهان صدق می‌کنند و عبارتند از :

(۱) همه چیز یک سهم، در قیمت آن سهم منعکس می‌گردد.

(۲) قیمتها به صورت روندهایی حرکت می‌کنند که این روندها در برابر تغییرات مقاومت می‌کنند.

(۳) روندهای بازار تکرار می‌شوند

در بازار سرمایه، اشخاص استفاده کننده از روش تحلیل بنیادی را بنیادگرا و اشخاص استفاده کننده از روش تحلیل تکنیکی را تکنسین یا چارتیست می‌نامند. اگر جواهیم منصفانه و با دید علمی به قضیه نگاه کنیم. باید بگوییم که این دو روش در واقع مکمل یکدیگرند و در کنار هم تصویر شفاف از تحولات بازار سرمایه ارائه می‌دهند. (هادی پور، ۱۳۸۲)

فرآیند داده کاوی

حتی در بعضی محیط‌های حرفه‌ای این نظر وجود دارد. داده کاوی شامل انتخاب و بکارگیری ابزار مبتنی بر کامپیوتر برای حل مسائل فعلی و بدست آوردن راه حلی به طور خود کار می‌باشد. این برداشت در حقیقت یک تصور اشتباه بر پایه آرمان گرایی تصنیعی جهان است. البته دلایل زیادی برای این برداشت اشتباه وجود دارد. یکی از این دلایل این است که داده کاوی فقط جمجمه‌ای از ابزارهای ایزووله شده نیست بلکه هر کدام کاملاً از یکدیگر متفاوت بوده و به دنبال حل مسائل هستند . دلیل دوم در انطباق یک مسئله با یک تکنیک نهفته است. خیلی به ندرت این حالت وجود دارد که یک پرسش تحقیقی با دقت کافی، کاربرد ساده و منحصر بفرد، یک روش و تکنیک را بیان کند. در حقیقت در عمل چه روی می‌دهد که داده کاوی به یک فرآیند تکرار پذیر تبدیل می‌گردد. به طور دقیق تر ما داده‌ها را مطالعه می‌کنیم، آنها را با استفاده از بعضی تکنیک‌های تحلیلی بررسی و بالاخره به آنها از دیدگاه‌ها و زوایای دیگری توجه می‌کنیم. شاید آنها را اصلاح کنیم و در این راستا لازم باشد کار را از ابتدا شروع و ابزار تحلیل داده دیگری را اعمال کنیم که حاصل همه اینها، شاید رسیدن به یک نتیجه جدید یا بهتر باشد. این فرآیند می‌تواند بارها و بارها تکرار شود. هر تکنیک برای بررسی جنبه‌های تا حدودی متفاوت داده‌ها استفاده می‌شود. (کانتارادزیک، ۲۰۰۵)

روش‌های تجربی معمول با مسائل داده کاوی منطبق شد و مراحل زیر را در بر می‌گیرد:

۱- بیان مسئله و فرموله کردن فرضیه^{۱۱}

^{۱۱}. State the problem and formulate the hypothesis

بیشتر مطالعات بر مبنای مدلسازی در زمینه کاربردی خاصی اجرا می‌شوند. بنابراین، دانش و تجربه در حوزه ای خاص معمولاً به منظور ذکر یا بیان معنی دار مسئله ضروری است. متأسفانه بسیاری از مطالعات کاربردی به جای مرکز بر مسئله تمایل دارند بر تکنیک‌های داده کاوی تکیه کنند. (Hand, et al, 2005)

۲- جمع آوری داده ها^{۱۲}

این مرحله در ارتباط با چگونگی تولید و جمع آوری داده‌ها می‌باشد. به طور کلی، دو امکان متمایز وجود دارد. اولین امکان زمانی است که فرآیند تولید داده‌ها تحت کنترل یک متخصص (مدلساز) می‌باشد. این روش به نام آزمون طراحی شده^{۱۳} شناخته می‌شود. امکان دوم زمانی مطرح می‌شود که متخصص نمی‌تواند فرآیند تولید را تحت تأثیر قرار دهد. این روش به نام روش یا شیوه دیداری (مشاهده ای)^{۱۴} مطرح می‌شود. (Wiley,2003)

۳- پیش‌پردازش داده ها^{۱۵}

در روش و تنظیم دیداری، داده‌ها معمولاً از پایگاه داده‌های موجود در انبار داده‌ها و مرکز داده‌ها "جمع آوری" می‌شود. فرآیند پیش‌پردازش داده‌ها معمولاً شامل حداقل دو مرحله متدال می‌باشد:

۱- آشکار سازی و (حذف) داده‌های غیر عادی . داده‌های غیر معمول یا غیر عادی در حقیقت مقادیر داده نامنطبق و غیرمرسوم هستند که با بیشترین مشاهدات سازگار نیستند. معمولاً داده‌های غیرعادی نتیجه سنجش خطاهای، کدنویسی و ثبت خطاهای بوده و گاهی اوقات نیز مقادیر طبیعی و غیر عادی هستند. چنین نمونه‌های غیرقابل ارائه، می‌توانند بعداً مدل‌های ایجاد شده را کاملاً تحت تأثیر قرار دهند. در اینجا دو خط مشی در خصوص داده‌های غیر عادی وجود دارد:

الف. تشخیص و حذف علمی داده‌های غیر عادی به عنوان بخشی از مرحله پیش‌پردازش، یا

ب. توسعه روش‌های قوی مدلسازی که به داده‌های غیر عادی و ناخواسته غیر حساس باشند.

۲- جنبه‌ها و ویژگی‌های مقیاس‌بندی، رمزگذاری و انتخاب . پیش‌پردازش داده‌ها شامل چندین مرحله است از جمله مقیاس‌بندی متغیر و انواع مختلف رمزگذاری. برای مثال یک مشخصه با دامنه [۰، ۱]^{۱۶} و دیگری با دامنه [۱۰۰۰، ۱۰۰] دارای ارزش مشابهی در تکنیک اعمال شده نیستند. آنها همچنین بر روی نتایج داده کاوی به طور متفاوتی تأثیر خواهند گذاشت. بنابراین، توصیه می‌شود که آنها را مقیاس‌بندی نموده و برای تحلیل و بررسی بیشتر هر دو جنبه را به وزن یکسانی تبدیل کنیم. (Wiley,2003)

۴- برآورد و ارزیابی مدل^{۱۷}

انتخاب و پیاده سازی تکنیک‌های مناسب داده کاوی کار اصلی این مرحله است. این فرآیند معمولاً چندان روش و واضح نیست. در عمل، چندین مدل بطور همزمان پیاده سازی شده و سپس بهترین آنها انتخاب می‌شود که یک وظیفه و کار اضافی

¹². Collect the data

¹³. Designed Experiment

¹⁴. Observational Approach

¹⁵. Preprocessing the data

¹⁶. Estimate the model

محسوب می‌گردد. در این گام است که از تکنیک‌های مختلف داده کاوی برای استخراج مدل‌ها استفاده می‌گردد و به تناسب حجم، طبیعت و رفتار هر داده یکی از این تکنیک‌ها می‌تواند بکار گرفته شود. شاید بتوان بطور خلاصه گفت که مأموریت اصلی کاوش داده‌ها به عهده این گام بوده و خروجی این گام الگوهای کاوش شده داده‌ای است. (Wiley, 2003)

۵- تفسیر مدل و استنتاج^{۱۷}

در اکثر موارد، مدل‌های داده کاوی باید به فرآیند تصمیم‌گیری کمک کنند. پس مدل زمانی مفید است که تفسیر پذیر باشد زیرا انسان‌ها به احتمال زیاد بر پایه مدل‌های پیچیده جعبه سیاه^{۱۸} تصمیم‌گیری نمی‌کنند. (Wiley, 2003) معمولاً مدل‌های ساده بیشتر قابل تفسیر و قابل فهم هستند. در عین اینکه دقت کمتری دارند. از روش‌های جدید داده کاوی انتظار می‌رود که در استفاده از مدل‌های دارای ابعاد بزرگ نتایجی با دقت بالا ارائه دهند و مسئله تفسیر این مدل‌ها در عین مبهم بودن یک وظیفه جزا و جدأگانه است که با تکنیک‌های ویژه برای معتبسازی نتایج در نظر گرفته می‌شود. مسلماً صدھا صفحه اطلاعات حاوی ارقام و اعداد مختلف برای یک کاربر چندان مفید فایده نبوده و مورد قبول او قرار نخواهد گرفت. همچنین این کاربر نمی‌تواند این نتایج را خلاصه و تفسیر نموده و از آنها برای یک تصمیم‌گیری موفق استفاده کند. اگر چه تأکید بر روی مراحل ۳ و ۴ فرآیند داده کاوی است اما باید به این نکته اساسی توجه داشت که اینها فقط دو مرحله از یک فرآیند بسیار پیچیده هستند. همه مراحل بطور جزا و تمام فرآیند داده کاوی به عنوان یک کل، بسیار تکراری‌ذیر هستند. همانطور که در شکل ۲-۲ نشان داده شده است درک مناسب از کل فرآیند برای هر کاربرد موفق، مهم است. بدون توجه به اینکه چه میزان روش داده کاوی به کار رفته در مرحله ۴ توأم‌نند است، اگر داده‌ها به درستی جمع آوری و پیش‌پردازش نشوند یا اگر فرمول مسئله معنی دار نباشد، مدل حاصل آنچنان معتبر نخواهد بود. (Wiley, 2003)

مراحل داده کاوی

داده کاوی شامل مراحل مختلفی می‌باشد که عبارتند از:

۱. تعیین اطلاعات گذشته.
۲. تمیز کردن داده‌ها^{۱۹} و پردازش اولیه. در این مرحله خطاهای داده‌ها تصحیح می‌شوند و داده‌های اشتباه جایگزین می‌شوند. این مرحله ممکن است تا ۶۰ درصد از زمان داده کاوی را در بر گیرد
۳. یکپارچه سازی داده‌ها. معمولاً داده‌ها از منابع متفاوتی جمع آوری می‌شوند و باید به صورتی درآیند که یک مخزن داده‌ای^{۲۰} مناسب ایجاد شود تا بتوان عملیات داده کاوی را بهتر انجام داد.
۴. انتخاب جموعه داده‌های هدف.
۵. یافتن ویژگی‌های مورد استفاده و تعیین ویژگی‌های جدید.
۶. نمایش داده‌ها به صورتیکه بتوان برای داده کاوی استفاده نمود.

^{۱۷}. Interpret the model and draw conclusions

^{۱۸}. Black Box

^{۱۹}. Data Cleaning

^{۲۰}. Data Warehouse

۷. انتخاب عملیات داده کاوی (دسته بنده، خوشه بنده، پیش‌بینی) .
۸. انتخاب روش داده کاوی (شبکه‌های عصبی، درخت تصمیم، خوشه بنده و نظایر آن) .
۹. داده کاوی و جستجو برای یافتن الگوی مناسب.
۱۰. ارزیابی و تحلیل الگوی به دست آمده و حذف الگوهای نامناسب.
۱۱. تفسیر نتایج داده‌ها و استنتاج از اطلاعات با ارزش.

باید توجه داشت که جمع آوری و محافظت از داده‌ها نکته بسیار مهمی می‌باشد. اصولاً چون قالب و نوع داده‌ها در طول زمان تغییر می‌کند ممکن است بسیاری از داده‌های موجود در قالب‌های متفاوت باشند و همچنین بسیاری از داده‌های قدیمی از بین رفته و دور ریخته شوند. در حالیکه ممکن است اهمیت این داده‌ها از داده‌های جدید به هیچ وجه کمتر نباشد. همچنین به علت اینکه داده‌ها از منابع مختلف داخلی و خارجی مانند کارکنان شرکت، مدیران، مشتریان، کارفرمایان، پیمانکاران باشند باز هم ممکن است قالب داده‌ها با هم یکسان نباشد. به همین دلیل انتخاب داده کاوی درست و یکپارچه سازی قالب آن‌ها به منظور استفاده در داده کاوی از اهمیت بسیار بالایی برخوردار می‌باشد.

عناصر داده کاوی

توصیف و کمک به پیش‌بینی دو کارکرد اصلی داده کاوی هستند. تحلیل داده مربوط به مشخصه‌های انتخابی متغیرها، از گذشته و حال، و درک الگو مثالی از تحلیل توصیفی است. برآورده ارزش آینده یک متغیر و طرح ریزی کردن روند، مثالی از توانایی پیشگویانه داده کاوی است. (غضنفری، ۱۳۸۷) برای عملی شدن هریک از دو کارکرد فوق الذکر داده کاوی، چند گام ابتدایی اما مهم باید اجرا شوند که از این قرارند:

۱. انتخاب داده‌ها
۲. پاک‌سازی داده‌ها
۳. غنی‌سازی داده‌ها
۴. کدگذاری داده‌ها

با دارا بودن هدف کلی در مطالعه، انتخاب جموعه داده‌های اصلی برای تحلیل، اولین ضرورت است. رکوردهای لازم می‌تواند از انبار داده‌ها و یا بانک اطلاعاتی عملیاتی استخراج شود. این رکوردهای داده جمع آوری شده، اغلب از آنچه آلودگی داده‌ها نامگذاری شده است رنج می‌برند و بنابراین لازم است پاکسازی شوند تا از یکدستی فرمت (شکلی) آنها اطمینان حاصل شود، موارد تکراری حذف شده و کنترل سازگاری دامنه به عمل آید. ممکن است داده‌های گردآوری شده از جنبه‌های خاصی ناقص یا ناکافی باشند. در این صورت داده‌های مشخصی باید گردآوری شوند تا بانک اطلاعات اصلی را تکمیل کنند. منابع مناسب برای این منظور باید شناسایی شوند. این فرآیند مرحله غنی‌سازی داده‌ها را تکمیل می‌کند. یک سیستم کد گذاری مناسب معمولاً جهت انتقال داده‌ها به فرم ساختاربندی شده جدید، مناسب برای عملیات داده کاوی تعییه می‌شود.

مدل‌ها و الگوریتم‌های داده کاوی

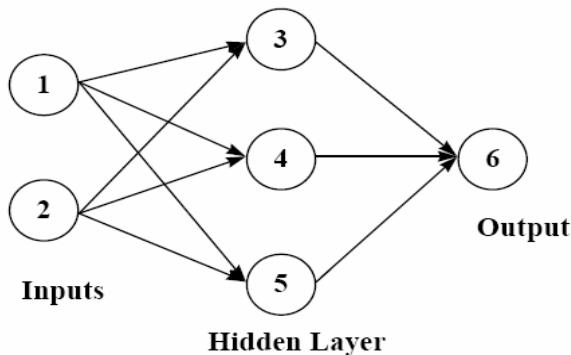
بسیاری از محصولات تجاری داده کاوی از جموعه‌ای از این الگوریتم‌ها استفاده می‌کنند و معمولاً هر کدام آنها در یک بخش خاص قدرت دارند و برای استفاده از

یکی از آنها باید بررسی های لازم در جهت انتخاب متناسب ترین مصوب توسط گروه متخصص در نظر گرفته شود. نکته مهم این است که در بین این مدل ها و الگوریتم ها، بهترین وجود ندارد و با توجه به داده ها و کارایی مورد نظر باید مدل انتخاب گردد. در اینجا به توضیح برخی از آنها خواهیم پرداخت:

^{۲۱}- شبکه های عصبی

شبکه های عصبی از پر کاربردترین و عملی ترین روش های مدل سازی مسائل پیچیده و بزرگ که شامل صدها متغیر هستند می باشد. شبکه های عصبی می توانند برای مسائل کلاس بندی (که خروجی یک کلاس است) یا مسائل رگرسیون (که خروجی یک مقدار عددی است) استفاده شوند. هر شبکه عصبی شامل یک لایه ورودی^{۲۲} می باشد که هر گره در این لایه معادل یکی از متغیرهای پیش بینی می باشد. گره های موجود در لایه میانی به تعدادی گره در لایه نهان^{۲۳} وصل می شوند. هر گره ورودی به همه گره های لایه نهان وصل می شود. گره های موجود در لایه نهان می توانند به گره های یک لایه نهان دیگر وصل شوند یا می توانند به لایه خروجی^{۲۴} وصل شوند. لایه خروجی شامل یک یا چند متغیر خروجی می باشد (Hand, et.al, 2005).

در شکل زیر یک شبکه عصبی با یک لایه پنهان نشان داده شده است.



شکل ۴-۲ - شبکه عصبی با یک لایه پهان

هریال که بین نودهای X, Y می باشد دارای یک وزن است که با $W_{x,y}$ نمایش داده می شود. این وزن ها در محاسبات لایه های میانی استفاده می شوند و طرز استفاده آنها به این صورت است که هر نود در لایه های میانی (لایه های غیر از لایه اول) دارای چند ورودی از یال مختلف می باشد که همانطور که گفته شد هر کدام یک وزن خاص دارد. هر نود لایه میانی میزان هر ورودی را در وزن یال مربوطه آن ضرب می کند و حاصل این ضرب ها را با هم مجموع می کند و سپس یکتابع از پیش تعیین شده (تابع فعال سازی) روی این حاصل اعمال می کند و نتیجه را به عنوان خروجی به نودهای لایه بعد می دهد. وزن یال ها پارامترهای ناشناخته ای هستند که توسط تابع آموزش^{۲۵} و داده های آموزشی که به سیستم داده می شود تعیین می گردند. تعداد گره ها و تعداد لایه های نهان و نحوه وصل

²¹. Neural Networks

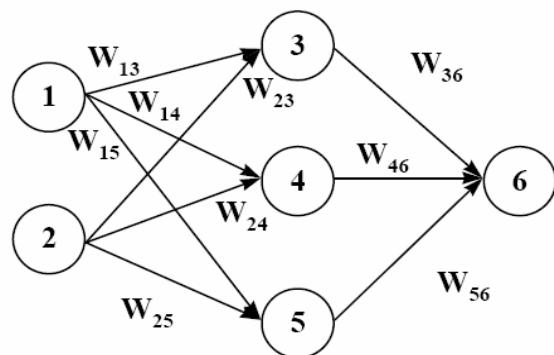
²². Input Layer

²³. Hidden Layer

²⁴. Output Layer

²⁵. Training method

شنونده‌ها به یکدیگر معماری (توبولوژی) شبکه عصبی را مشخص می‌کند. کاربر یا نرم افزاری که شبکه عصبی را طراحی می‌کند باید نودها، تعداد لایه های نهان، تابع فعال سازی و محدودیت‌های مربوط به وزن یال‌ها را مشخص کند.



شکل ۲-۵ - وزن یال بین Y و X است.

۲- درخت های تصمیم^{۲۶}

درخت های تصمیم روشی برای نمایش یک سری از قوانین هستند که منتهی به یک ردیف یا مقدار می‌شوند. برای مثال، می‌خواهیم متقاضیان وام را به دارندگان ریسک اعتبار خوب و بد تقسیم کنیم. شکل ۲-۶، یک درخت تصمیم را که این مسئله را حل می‌کند نشان می‌دهد و همه مؤلفه‌های اساسی یک درخت تصمیم در آن نشان داده شده است: نود تصمیم، شاخه‌ها و برگ‌ها.



شکل ۲-۶- درخت تصمیم گیری

²⁶. Decision Trees

بر اساس الگوریتم، ممکن است دو یا تعداد بیشتری شاخه داشته باشد. برای مثال، CART درختانی با تنها دو شاخه در هر شاخه دو نود ایجاد می‌کند. هر شاخه منجر به نود تصمیم دیگر یا یک نود برگ می‌شود. با پیمایش یک درخت تصمیم از ریشه به پایین به یک مورد یک رده یا مقدار نسبت می‌دهیم. هر نود از داده‌های یک مورد برای تصمیم گیری درباره آن انشعاب استفاده می‌کند. درخت های تصمیم از طریق جداسازی متوالی داده‌ها به گروه‌های مجزا ساخته می‌شوند و هدف در این فرآیند افزایش فاصله بین گروه‌ها در هر جداسازی است. یکی از تفاوت‌ها بین متدهای ساخت درخت تصمیم اینست که این فاصله چگونه اندازه گیری می‌شود. درخت‌های تصمیمی که برای پیش‌بینی متغیرهای دسته‌ای استفاده می‌شوند، درختهای Classification نامیده می‌شود زیرا نونه‌ها را در دسته‌ها یا رده‌ها قرار می‌دهند. درختهای تصمیمی که برای پیش‌بینی متغیرهای پیوسته استفاده می‌شوند، درخت‌های Regression نامیده می‌شوند. هر مسیر در درخت تصمیم تا یک برگ معمولاً قابل فهم است. از این لحاظ یک درخت تصمیم می‌تواند پیش‌بینی‌های خود را توضیح دهد، که یک مزیت مهم است. با این حال این وضوح ممکن است گمراه کننده باشد. برای مثال، جداسازی‌های سخت در درخت تصمیم دقتی را نشان می‌دهند که کمتر در واقعیت نمود دارندیک اشکال معمول درخت‌های تصمیم اینستکه آنها تقسیم کردن را بر اساس یک الگوریتم حریصانه انجام می‌دهند که در آن تصمیم گیری اینکه بر اساس کدام متغیر تقسیم انجام شود، اثرات این تقسیم در تقسیم‌های آینده را در نظر نمی‌گیرد. بعلاوه الگوریتم‌هایی که برای تقسیم استفاده می‌شوند، معمولاً تک متغیری هستند: یعنی تنها یک متغیر را در هر زمان در نظر می‌گیرند. در حالیکه این یکی از دلایل ساخت سری مدل است، تشخیص رابطه بین متغیرهای پیش‌بینی کننده را سخت‌تر می‌کند. (Hand et.al,2001)

^{۲۷}- رگرسیون منطقی

رگرسیون منطقی یک حالت عمومی‌تر از رگرسیون خطی می‌باشد. قبل این روش برای پیش‌بینی مقادیر باینری یا متغیرهای دارای چند مقدار گستره (کلاس) استفاده می‌شد. از آنجایی که مقادیر مورد نظر برای پیش‌بینی مقادیر گستره می‌باشند نمی‌توان آنرا به روش رگرسیون خطی مدل‌سازی کرد برای این منظور این متغیرهای گستره را به رویی به متغیر عددی و پیوسته تبدیل می‌کنیم و همچنین مقدار لگاریتم احتمال متغیر مربوطه را در نظر می‌گیریم و برای این منظور احتمال پیشامد را بدین صورت در نظر می‌گیریم:

احتمال اتفاق نیفتادن پیشامد / احتمال اتفاق افتادن پیشامد

و تفسیر این نسبت مانند تفسیری است که در بسیاری از مکالمات روزمره در مورد مسابقات یا شرط‌بندهای موارد مشابه بکار می‌رود. مثلاً وقتی می‌گوییم شانس بردن یک تیم در مسابقه ۳ به ۱ است در واقع از همین نسبت استفاده کرده و معنی آن این است که احتمال بردن آن تیم ۷۵٪ است. وقتی که ما موفق شدیم لگاریتم احتمال مورد نظر را بدست آوریم با اعمال لگاریتم معکوس می‌توان نسبت مورد نظر و از روی آن کلاس مورد نظر را مشخص نمود. (Han, and Kamber,2001

^{۲۸}- تحلیل تفکیکی

²⁷. Logistic regression

این روش از قدیمی ترین روش‌های ریاضی^{۲۹} وار گروه بندی داده‌ها می‌باشد که برای اولین بار در سال ۱۹۳۶ توسط فیشر^{۳۰} استفاده گردید. روش کار بدین صورت است که داده‌ها را مانند داده‌های جند بعدی بررسی کرده و برای داده‌ها مرزهایی ایجاد می‌کنند (برای داده‌های دو بعدی، خط جداکننده و برای داده‌های سه بعدی سطح جداکننده و...) که این مرزها مشخص کننده کلاس‌های مختلف می‌باشند و بعد برای مشخص کردن کلاس مربوط به داده‌های جدید فقط باید محل قرار گیری آن را مشخص کنیم. این روش از ساده ترین و قابل رشد ترین روش‌های کلاس‌بندی می‌باشد که در گذشته بسیار استفاده می‌شد. این روش به سه دلیل خوبیت خود را از دست داد: اول اینکه این روش فرض می‌کند همه متغیرهای پیش‌بینی به صورت نرمال توزیع شده اند که در بسیاری از موارد صحت ندارد. دوم اینکه داده‌هایی که به صورت عددی نمی‌باشند مثل رنگها در این روش قابل استفاده نمی‌باشند. سوم اینکه در این روش فرض می‌شود که مرزهای جداکننده داده‌ها به صورت اشکال هندسی خطی مثل خط یا سطح می‌باشند حال اینکه این فرض همیشه صحت ندارد. نسخه‌های اخیر تحلیل تفکیکی بعضی از این مشکلات را رفع کرده‌اند. به این طریق اجازه می‌دهند مرزهای جداکننده بیشتر از درجه ۲ نیز باشند که باعث بهبود کارایی و حساسیت در بسیاری از موارد می‌گردد.

۵-مدل افزودنی کلی (GAM)

این روش‌ها در واقع بسطی بر روش‌های رگرسیون خطی و رگرسیون منطقی می‌باشند. به این دلیل به این روش افزودنی می‌گویند، که فرض می‌کنیم می‌توانیم مدل را به صورت مجموع چندتابع غیرخطی (هر تابع برای یک متغیر پیش‌بینی کننده) بنویسیم. GAM می‌تواند هم به منظور رگرسیون و هم به منظور کلاس‌بندی داده‌ها استفاده گردد. این ویژگی غیرخطی بودن توابع باعث می‌شود که این روش نسبت به روش‌های رگرسیون خطی بهتر باشد.

۶-روش Boosting

در این روش‌ها مبنی کار این است که الگوریتم پیش‌بینی را چندین بار و هر بار با داده‌های آموزشی متفاوت (که با توجه به اجرای قبلی انتخاب می‌شوند) اجرا کنیم و در نهایت آن جوابی که بیشتر تکرار شده را انتخاب کنیم. این روش اگر چه وقت گیر است ولی جواب‌های آن مطمئن تر خواهد بود. این روش اولین بار در سال ۱۹۹۶ استفاده شد و در این روزها با توجه به افزایش قدرت محاسباتی کامپیوترها بر مقبولیت آن افزوده گشته است. (Fayyad, 1996) تلاش اصلی تحقیقات روی تحلیل الگوریتم‌های داده کاوی جهت استخراج الگوها از داده‌ها بوده است.

نتیجه گیری

توسعه صنعتی و اقتصادی کشور در سالهای آینده منوط به توجه و سرمایه‌گذاری هر چه بیشتر به شرکتهای پذیرفته شده در بورس می‌باشد. دستیابی به رشد اقتصادی و ایجاد انگیزه جهت سرمایه‌گذاری زمانی در یک کشور تسريع می‌گردد که آن کشور دارای بازارهای سرمایه فعال و قابل اعتماد باشد. وجود بازارهای بورس فعال همواره سرمایه‌گذاران متعددی را به

²⁸. Discriminat analysis

²⁹. Fisher

³⁰. Generalized Additive Models

تکاپو ودادشته و جریان سرمایه و منابع مالی را به بخش‌های مولده تسریع می‌نماید. یکی از فرضیاتی که در خصوص این بازار وجود دارد، فرضیه بازار کارا (EMH)³¹ است. براساس فرضیه بازار کارا قیمت اوراق بهادار به ارزش ذاتی آنها نزدیک است. به عبارت دیگر، ویژگی مهم بازار کارا این است که قیمت تعیین شده در بازار، شاخص مناسبی از ارزش واقعی اوراق بهادار است. (جهانخانی، ۱۳۸۰) از مواردی که فرضیه بازار کارا را ضعیف‌تر می‌کند قابلیت پیش‌بینی در این بازارهاست. برای پیش‌بینی قیمت سهام در بازارهای بورس دنیا از روش‌های مختلفی از تحلیل‌ها نظریه تکنیک‌های داده کاوی، رگرسیون‌ها، سری‌های زمانی و... استفاده شده است. (جهانخانی، ۱۳۸۰، صفرنواده، ۱۳۸۰)

از روش‌های دقیق‌تری که می‌توان در هر نوع پیش‌بینی از آن استفاده کرد تکنیک‌های داده کاوی از جمله شبکه عصبی³² و درخت‌های تصمیم و... است. پیش‌بینی قیمت سهام شرکتهای پذیرفته شده در بورس اوراق بهادار تهران با استفاده از تکنیک‌های داده کاوی و با توجه به معیارهای مالی از جمله موضوع‌های با اهمیتی است که متأسفانه تاکنون در ارتباط با آن تحقیقات اندکی انجام شده است. تمرکز پیش‌بینی کنندگان بازار بورس روی شیوه‌های توسعه برای موفقیت پیش‌بینی ارزش‌های شاخص یا قیمت سهم، دید سود‌های زیاد برای استفاده خوب برای تعیین استراتژی‌های تجاری است. ایده اصلی برای پیش‌بینی بازار بورس موفق، دستیابی به نتایج بهتر با استفاده از حداقل داده‌های اولیه مورد نیاز و کمترین مدل بازار بورس پیچیده است. طبقه‌بندی‌ها بر حسب داده‌های اولیه، متداول‌تری پیش‌بینی، ارزیابی عملکرد و استفاده از معیارهای عملکرد ساخته شده است. که به واسطه این تحقیق نشان می‌دهد که تکنیک‌های محاسبه در بسیاری موارد برای مطالعه و ارزیابی رفتار بازار بورس مورد پذیرش می‌باشد. بدون شک پیش‌بینی قیمت‌ها و بازده‌های سهام مشکل است، زیرا که تغییرات بازار به تصرف و تسخیر در استفاده و اجرا و تکمیل مدل‌ها نیاز دارد. پیشرفت‌های اخیر در تکنیک‌های داده کاوی ایجاد می‌کند که ابزارهای مفیدی را در پیش‌بینی محیط‌های شلوغ مانند بازارهای بورس در تسخیر رفتارهای غیر خطی مورد استفاده قرار داد. امروزه سازمان‌ها قادرند با هزینه کم اطلاعات وسیعی از وضعیت کسب و کار خود جمع و نگهداری کنند و این موجب شده است که استفاده از روش‌های داده کاوی، ارزش‌های قابل توجهی را برای سازمان به دست آورده. رویکردهای موجود به مسئله داده کاوی متنوع است. در این تحقیق نگاهی به آخرین دستاوردهای این زمینه علمی انداخته خواهد شد. امروزه افزایش سریع حجم پایکاوه داده‌ها به شکلی است که توانایی انسان برای درک این داده‌ها بدون ابزارهای پر قدرت میسرنگی باشد. در این وضعیت تصمیم‌گیری‌ها به جای تکیه بر اطلاعات، بر درک مدیران و کاربران تکیه دارند، چرا که تصمیم‌گیرندگان ابزار قوی برای استخراج اطلاعات با ارزش را در دست ندارند. در واقع شرایط فعلی توصیف کننده حالتی است که ما از لحاظ داده غنی، اما از لحاظ اطلاعات ضعیف هستیم.

داده کاوی کمک می‌کند تا سازمان‌ها با کاوش بر روی داده‌های یک سیستم، الگوها، روندها و رفتارهای آینده را کشف و پیش‌بینی کرده و

³¹. Efficient Market Hypothesis

³². Neural Network

بهتر تصمیم بگیرند. همچنین با استفاده از تحلیل وقایع گذشته یک تحلیل اatomاتیک و پیش‌بینانه ارائه می‌نماید و به سؤالاتی جواب می‌دهد که پاسخ آنها در گذشته ممکن نبوده و یا به زمان زیادی نیاز داشت. ابزارهای داده کاوی الگوهای پنهانی را کشف و پیش‌بینی می‌کند که متخصصان ممکن است به دلیل اینکه این اطلاعات و الگوها خارج از انتظار آنها باشد، آنها را مدنظر قرار ندهند و به آنها دست نیابند. در داده کاوی مواردی که در آن‌ها با تعداد بسیار زیادی متغیر، مدل و یا فرضیه مواجه هستیم، فراوان است. از طرفی داده کاوی یک فرآیند اکتشافی و تکراری است به این معنی که در خلال تحلیل داده‌ها اطلاعات جدید کشف می‌شوند، فرضیه‌های قبلی اصلاح و فرضیه‌های جدید ارائه می‌شوند و این کار ممکن است با داده‌های زیاد، بارها تکرار شود. لذا از دیدگاه آمار روش‌های با کارایی حساباتی بالا، تحلیل‌های حساباتی و تحلیل‌های تقریبی، مورد توجه خاص داده کاوی است. تأکید بیشتر داده کاوی بر بعضی روش‌های آماری، به معنی عدم استفاده از سایر روش‌های آماری نیست و در عمل از طیف گسترده‌ای از روش‌های آماری برای تحلیل داده‌ها استفاده می‌شود. (Berson, 2004) با توجه به مطالعات انجام شده در این مقاله به حقیقین توصیه می‌شود که تحقیقاتی درخصوص پیش‌بینی بازده سهام شرکتهای موجود در بورس اوراق بهادار با استفاده از روش‌های داده کاوی انجام دهند.

منابع

- ۱-اسکندری، مجید.، (۱۳۷۱)، "بورس و سرمایه‌گذاری از حرف تا عمل"، مجله بورس، شماره ۱۴.
- ۲-صفر نواده، محمد، (۱۳۸۰)، "پیش‌بینی قیمت سهام در بازار بورس اوراق بهادار تهران"، پایان نامه کارشناسی ارشد، دانشگاه امام صادق.
- ۳-زارعی، حمدرضا.، (۱۳۷۰)، "بورس اوراق بهادار و چشم انداز آن"، ماهنامه تازه‌های اقتصاد، شماره ۲۱، مهرماه ، صفحه ۲۹-۲۳.
- ۴-جهانخانی، علی، (۱۳۷۶)، "بررسی عوامل مؤثر بر قیمت سهام و مروری بر روند تغییر قیمت سهام در بورس اوراق بهادار تهران"، زمینه، شماره ۶۵، ۱۳-۱۶ .
- ۵-حق دوست، شادی، (۱۳۸۵)، "مدلسازی پیش‌بینی قیمت سهام با استفاده از شبکه عصبی مصنوعی و مقایسه آن با روش‌های پیش‌بینی ریاضی"، پژوهشنامه اقتصادی ، شماره ۱۸ .
- ۶-هادی پور، حسین.، (۱۳۸۲)، "تعیین بهترین مدل جهت پیش‌بینی قیمت سهام در گروه صنایع غذایی و آشامیدنی بورس اوراق بهادار"، پایان نامه کارشناسی ارشد مدیریت بازرگانی موسسه عالی پژوهش‌های بازرگانی
- ۷-کانتاردزیک، محمد، (۱۳۸۵)، داده کاوی ، بابل، علوم رایانه.
- ۸-غضنفری، مهدی.، (۱۳۸۷)، "داده کاوی و کشف دانش"، دانشگاه علم و صنعت تهران .

9-Azarbaiejani K., Fahimifard S.M., and Bayari L.,(2007), Energy Consumption Forecasting Based on Sustainable Development: Case Study of Iran. Iran-Italy Conference on Challenge and Prospects of Sustainable Development, December 17-18, Isfahan-Iran.

10-Fama E. F.(1970), "Efficient capital market: A review of theory and Empirical work", the Journal of finance No. 2. pp. 383-417.

11- Hand, D. Mannila, H. and Smith , P,(2005), " Principles of Data Mining", Prentice Hall.

12-Wei-Sen Chen and Yin-Kuan Dua,(2008), Using neural networks and data mining techniques for the financial distress prediction model, The Journal of Finance.

13-Fayyad , U. Uthwrsusamy, R, (1996), "Data Mining and Knowledge Discovery in Databases" ,Communication of ACM.

14-Kanas, A. and Yannopoulos, A., (2001), Comparing Linear and Nonlinear Forecasts for Stock Returns, Vol.10.

کنفرانس ملی حسابداری، مدیریت مالی و سرمایه‌گذاری ۲۶ بهمن ماه ۱۳۹۱، دانشگاه
جامع علمی-کاربردی استان گلستان

The 2013 Conference on Accounting, Financial Management and Investment (CAFMI 2013)
ISSN: 2322-3510

15-Sun, J. and Li, H., (2006)," Data mining method for listed companies financial distress prediction ", Knowledge Based System, Vol. 21, Issue 1, pp: 1-5.

16-Yi-Hsien Wang,(2007), Nonlinear neural network forecasting model for stock index option price: Hybrid GJR-GARCH approach, Expert Systems with Applications, Volume 36, Issue 1, Pages 564-570

17-Berson Alex, Smith Stephen, and Thearling Kurt,(2004), " Building Data Mining Application for CRM", Tata McGraw-Hill, New York.