

به نام حق



دانشکده فنی و مهندسی
کارشناسی ارشد مهندسی کامپیوتر نرم افزار
گروه مهندسی کامپیوتر و فناوری اطلاعات

گزارش درس سمینار

موضوع:

کاربرد تکنیک های داده کاوی برای مدیریت ریسک مالی: یک چارچوب طبقه بندی

نگارش:

سیما محسنی

استاد راهنما:

جناب آقای دکتر آرش قربان نیا دلاور

چکیده:

طی چند دهه گذشته جهان شاهد افزایش اعتماد به خدمات مالی (به عنوان مثال بانکی ، کارت های اعتباری ، بیمه) بوده است ، در حالی که ظهور اینترنت منجر به افزایش شدید تعداد معاملات آنلاین شده است. هر دوی این عوامل باعث افزایش شیوع کلاهبرداری های مالی می شود و نیاز به رویکرد جدیدی در کشف و مدیریت ریسک مالی را افزایش می دهد. یک راه حل که به دلیل در دسترس بودن فضای ذخیره سازی زیاد و توان محاسباتی که در دهه گذشته ظاهر شده ، عملی شده است ، داده کاوی است.

در این مقاله سعی شده است استفاده از داده کاوی برای شناسایی و کاهش خطرات مالی ناشی از تقلب های مالی مورد بررسی قرار گیرد. این مطالعه از یک مجموعه داده کاگل استفاده کرده و آزمایشاتی را با استفاده از چندین معیار مختلف طبقه بندی انجام داده است. بهترین عملکرد برای شناسایی مشتریان معتبر در بانک ها با طبقه بندی Random Forest حاصل شد.

کلید واژه ها:

جنگل تصادفی ، داده کاوی ، کیسه گذاری ، ماشین بردار پشتیبانی ،

ریسک مالی ، ریسک اعتباری

1. معرفی

فناوری یک جنبه اساسی در تجزیه و تحلیل ریسک مالی و همچنین ابزاری برای ارائه هشدار با توجه به روندهای آینده است. داده کاوی برای جمع آوری و انتخاب اطلاعات ارزشمند استفاده می شود [1]. این مدل در مدل‌های داده بزرگ استفاده می شود زیرا از روشهای یادگیری ماشین و هوش مصنوعی برای یافتن اطلاعات مربوط به یک سناریو در انبوهی از داده‌ها استفاده می کند [2] [3]. بنابراین، استفاده از مجموعه داده های مربوط به مدیریت ریسک مالی در زمان واقعی را تسهیل می کند. سازمان ها از مجموعه داده های بزرگ برای ایجاد الگوهای ضروری برای کاهش ریسک عملیاتی، نقدینگی، بازار، حقوقی و اعتباری استفاده می کنند [4] [5]. با این حال، برای استفاده از نتیجه گیری و اتوماسیون داده محور، تهیه مدل هایی برای پرداختن به چالش های زمان واقعی ضروری است [6]. داده کاوی شامل تلفیق تکنیک های مدل سازی با آماری، هوش مصنوعی و یادگیری ماشین است [7]. ترکیبی از این ابزارها تضمین می کنند که شرکت ها می توانند بر اساس بینش حاصل از یادگیری بدون نظارت و نظارت و همچنین تصمیم گیری، شرایط مختلفی را مرتفع کنند. در این راستا، الگوریتم های یادگیری و پردازش داده ها برای دستیابی به اهداف سازمان مالی، به ویژه در جمع آوری و جمع آوری اطلاعات ایجاد شده اند [8]. داده کاوی به دلیل توانایی پردازش اطلاعات از چندین سیستم عامل و همچنین پایگاه داده ضمن دستیابی به نتایج مطلوب، کارآمد است. پردازش چنین اطلاعاتی برای انسان و دستیابی به کارایی حاصل از یادگیری عمیق و ماشینی می تواند طولانی تر باشد [9]. از این رو، داده کاوی اکنون در مدیریت ریسک مالی برای ایجاد مشخصات، روندها و بینش های زمان واقعی مورد استفاده قرار می گیرد.

2. تکنیک های اساسی داده کاوی

داده کاوی فرایند استخراج اطلاعات یا کشف دانش پنهان و ارزشمند بر اساس برخی الگوها از داده های بزرگ است. داده کاوی برای شرکت های مختلف مطلوب است ، به عنوان مثال ، تولید ، تبلیغات ، مدیریت ریسک و غیره. بر این اساس ، یک فرایند استاندارد داده کاوی مورد نیاز است. این فرایند داده کاوی باید قابل اعتماد باشد. بعلاوه ، این رویه باید توسط تجار با دانش یا تخصص عملاً صفر درباره علم داده تکرار شود.

شکل 1-فرآیند داده کاوی

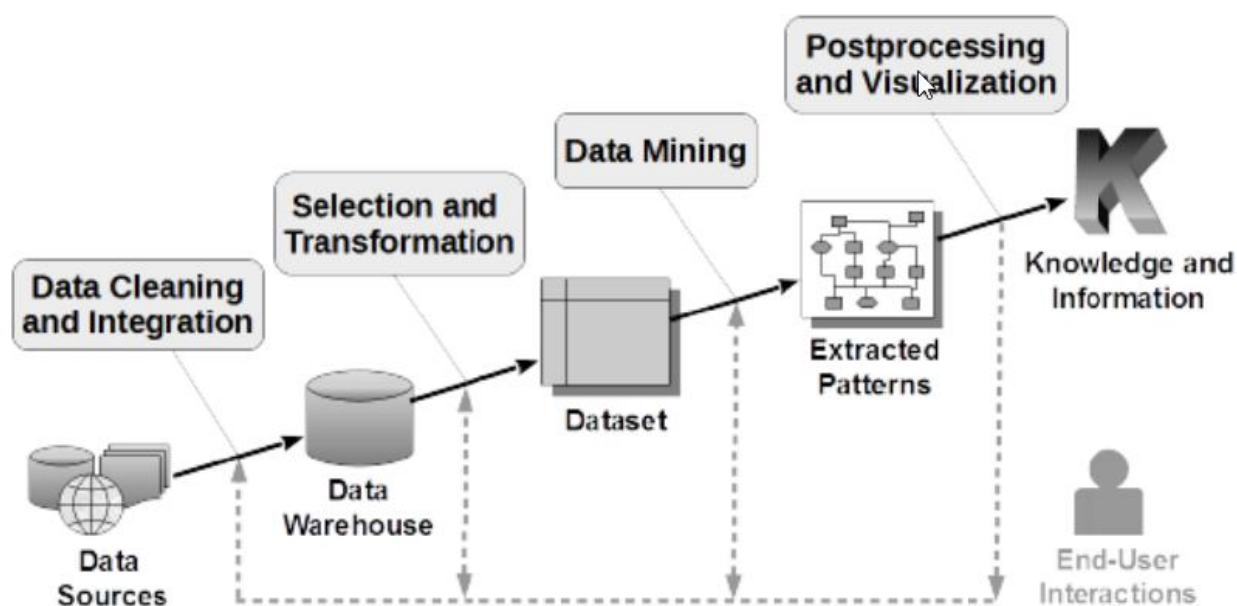


Fig. 1 Data Mining Process

فرآیند داده کاوی در دو مرحله طبقه بندی می شود: اول) آماده سازی داده ها / پیش پردازش داده ها و دوم) داده کاوی. فرآیند آماده سازی داده ها شامل تمیز کردن داده ها ، یکپارچه سازی داده ها ، انتخاب داده ها و تبدیل داده ها است. مرحله دوم شامل داده کاوی ، ارزیابی الگو و بازنمایی دانش است.

آ. تمیز کردن داده ها

مشاهده شده است که داده های دنیای واقعی که به آنها داده های خام نیز گفته می شود ، معمولاً متناقض یا غیر قابل اعتماد هستند. اولین قدم تمیز کردن داده های هدف است. پر کردن مقادیر از دست رفته و محاسبه پارامترها / مقادیر مورد نیاز.

ب یکپارچه سازی داده ها

داده های خام یعنی داده های جمع آوری شده از چندین منبع در قالب ها و مکان های مختلف قرار دارند ، به عنوان مثال در چندین پایگاه داده ، صفحه گسترده یا اسناد مختلف پخش می شوند. داده های همه این منابع برای کاهش خطاها در فرآیند ادغام داده ها ، باید با استفاده از یک فراداده رایج یا فرهنگ لغت داده ، قالب بندی شوند. ادغام داده ها همچنین سعی می کند افزونگی را بدون از دست دادن داده ها کاهش دهد.

ج انتخاب داده

در مرحله انتخاب داده ، داده های مربوط به تجزیه و تحلیل از منابع داده گرفته می شود. این روش داده های بزرگ ذخیره شده یا داده های تاریخی را بوجود می آورد و زیر مجموعه ای از داده ها را برای دستیابی به اهداف تعیین شده تجزیه و تحلیل ایجاد می کند.

د تبدیل داده ها

تبدیل داده ها فرایندی است برای جمع آوری داده ها از چندین منبع مختلف و جمع آوری آنها با هم و مناسب کردن آنها برای کاوش.

ه. داده کاوی

این فرآیند شناسایی و استخراج الگوهای یافت شده در مجموعه داده های هدف است. داده کاوی در چندین مرحله مانند طبقه بندی ، پیش بینی و خوشه بندی و سایر مراحل انجام می شود.

و. ارزیابی الگوها

ارزیابی الگو فرآیندی است که می تواند الگوهای کاملاً در حال افزایش را نشان دهد که بر اساس معیارهای ارائه شده است. در حالی که ، الگویی جالب است اگر بالقوه مفید و به راحتی قابل درک باشد.

ز. نمایندگی دانش

بازنمایی دانش ارائه داده ها به روشی ساخت یافته و جذاب برای مخاطبان هدف است.

فرآیند استاندارد صنعت برای داده کاوی

فرآیند استانداردهای صنعت شامل 4 مرحله است که در یک روند چرخه ای رخ می دهد:

اول. منابع داده

پس از شناسایی نیازهای کاربر ، تلاش برای شناسایی داده های موجود و منابع آن متمرکز می شود.

دوم. کاوش و تهیه داده ها

این یک مرحله چند مرحله ای است که در آن چندین فعالیت مختلف مانند بارگذاری داده ، یکپارچه سازی داده ها انجام شده است. پس از جمع آوری داده های هدف ، ویژگی های "سطح" داده های به دست آمده باید مورد بررسی و گزارش قرار گیرند. بعداً با استفاده از پرس و جو ، گزارش و داده های تجسم برای مقابله با سوالات یا اهداف داده کاوی مورد بررسی قرار می گیرد. در آخر ، کیفیت داده ها با اطمینان از کامل بودن داده های بدست آمده و عدم وجود مقادیر از دست رفته و غیره ارزیابی می شوند.

سوم. مدل سازی

این مرحله ای است که در متن هدف تجاری ، یک تکنیک مدل سازی مناسب انتخاب می شود و یک سناریو آزمون برای تأیید کیفیت و کارایی مدل انتخاب شده تولید می شود. سپس ، با استفاده از ابزارهای مدل سازی ،

یک یا چند مدل روی مجموعه داده باید تهیه شود. در آخر ، برای اطمینان از مطابقت مدل ها با ابتکارات تجاری ، مدل های انتخاب شده به شرکت کنندگان / ذینفعان پروژه ارائه می شوند.

چهارم. مدل استقرار

اطلاعات جمع آوری شده از مراحل فوق لازم بود که به ذینفعان ارائه شود تا آنها بتوانند این کار را برای تصمیم گیری در تجارت انجام دهند. این اطلاعات باید برای تأمین نیازهای تجاری ارائه شود. این مرحله می تواند به سادگی ایجاد گزارش یا پیچیده به عنوان یک فرآیند داده کاوی تکراری در سازمان باشد. یک طرح استقرار همچنین شامل یک برنامه نگهداری برای محافظت از یکپارچگی داده های پایه است. گزارش نهایی ، معمولاً شامل یک بینش پروژه ، نتایج و یک بازبینی جامع پروژه برای شناسایی پیشرفت ها و پیشرفت های آینده است.

شکل 2- فرآیند داده کاوی در صنعت مالی

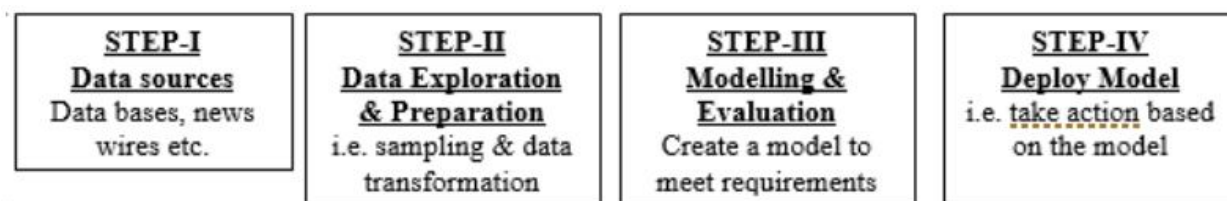


Fig. 2 Data Mining Process in Financial Industry

داده کاوی شامل تکنیک های مختلفی در جمع آوری و پیش پردازش داده ها است. یکی از تکنیک های مورد استفاده در طبقه بندی و پیش بینی داده ها درخت تصمیم است (اولسون و وو ، 2017). این روش شامل روش هایی برای الگوریتم یادگیری استقرایی است. یادگیری در این روش شامل به دست آوردن قوانین مناسب برای داده های نامنظم و غیر سازمان یافته است. این با الگوهای مقایسه ای ساده ای همراه است ، که می تواند به زبان های جستجوی پایگاه داده نیز منتقل شود. این ویژگی برای اطمینان از هماهنگی داده ها از پایگاه های مختلف با توجه به ضرورت وجود داده های بزرگ ، ضروری است [7]. یکی دیگر از ویژگی های اساسی درخت تصمیم ،

دقت بیشتر از لحاظ شباهت مجموعه داده ها است. برای پیدا کردن روابط بین مجموعه داده ها از تکنیک Association Rules استفاده می شود. عمدتاً ، این به عنوان یک فناوری استخراج برای یافتن روابط بین داده های چندین پایگاه داده مورد استفاده قرار می گیرد. این رویکرد به نتایج دقیق ، مفید و واضحی دست می یابد [10]. علاوه بر این ، می توان از آن در داده کاوی غیر مستقیم استفاده کرد. بعلاوه ، می توان از آن در پرداختن به داده های طولانی استفاده کرد. این روش از الگوریتم هایی مانند FP-رشد و الگوریتم Apriori استفاده می کند. این رویکردها با یافتن پیوندهای موجود بین مجموعه داده ها مطابقت دارد. روش مهم دیگر برای داده کاوی ، تجزیه و تحلیل خوشه بندی است که به عنوان طبقه بندی بدون نظارت نیز شناخته می شود. این یک ابزار حیاتی برای بدست آوردن دقت بالاتر و همچنین شباهت برای دسته های داده است. این مستلزم تقسیم داده ها به دسته ها است ، و پس از آن برای گروه های مختلف بحث می شود تا شکاف به حداقل برسد [11]. این روش شامل چهار مرحله است: (من) ویژگی ها و انتخاب. (ii) تعیین یا محاسبه شباهت ؛ (iii) گروه بندی ؛ (iii) و (iv) نتایج خوشه بندی [1]. الگوریتم هایی مانند BIRCH ، k-medoids ، k-means ، ROCK و CURE ، از تجزیه و تحلیل خوشه بندی استفاده می شود. این روش برای سازماندهی اطلاعات در حالت طبقه بندی و همچنین گروه بندی داده هایی که توصیف ندارند مناسب است.

3. بررسی ادبیات

کلاهبرداری مالی در حال تبدیل شدن به یک مشکل روزمره برای دنیای مالی است. شکل 3 نشان می دهد که انواع مختلفی از کلاهبرداری های مالی رایج در صنعت وجود دارد (به عنوان مثال کلاهبرداری بانکی ، کلاهبرداری بیمه ای و غیره). همه این نوع تقلب ها با استفاده از تکنیک های داده کاوی ، مانند طبقه بندی ، خوشه بندی ، تشخیص دور از انتظار ، پیش بینی ، رگرسیون و تجسم قابل شناسایی است.

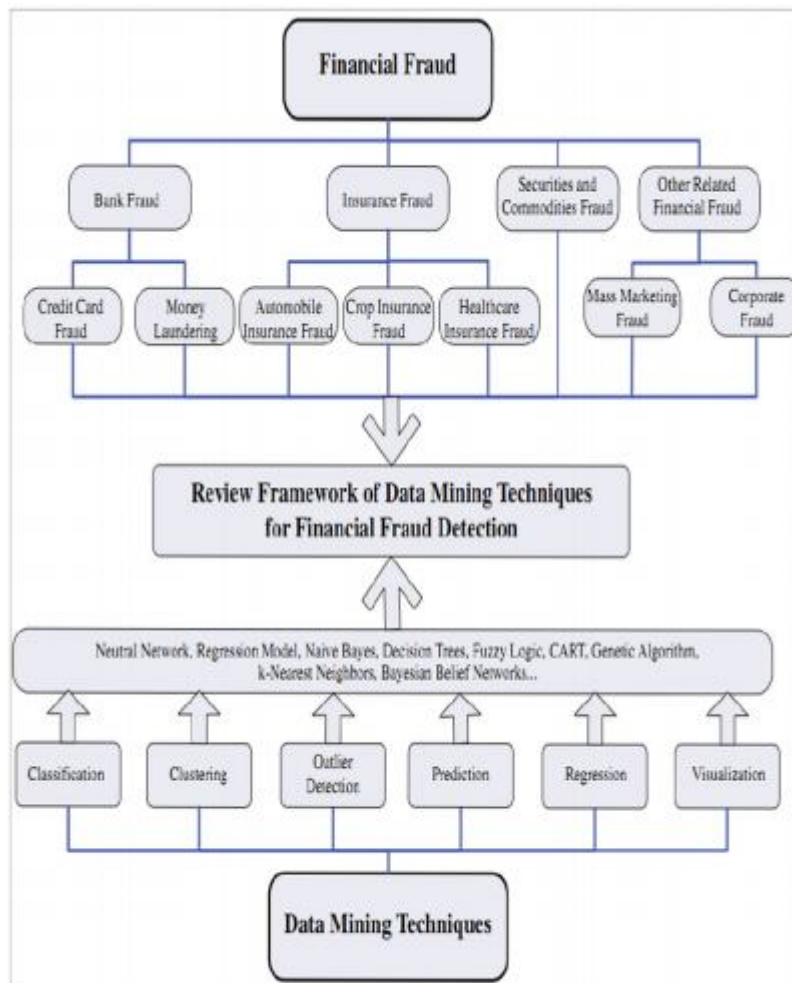


Fig. 3 Financial Fraud Detection

داده کاوی برای مدیریت ریسک اعتباری استفاده از کارت های اعتباری در حال تبدیل شدن به یک قاعده در دنیای مالی است. بانک ها و صادرکنندگان کارت اعتباری از تاریخ اعتبار مشتری برای ارزیابی شایستگی اعتبار مشتری و ایجاد رتبه بندی اعتباری استفاده می کنند. شکل زیر توضیح می دهد که چگونه یک موسسه مالی از داده های ذخیره شده در سابقه مالی مشتری برای رتبه بندی شایستگی مشتری در صدور یا رد درخواست کارت اعتباری استفاده می کند.

شکل 4-مدلی برای رتبه بندی مشتری برای کارت اعتباری

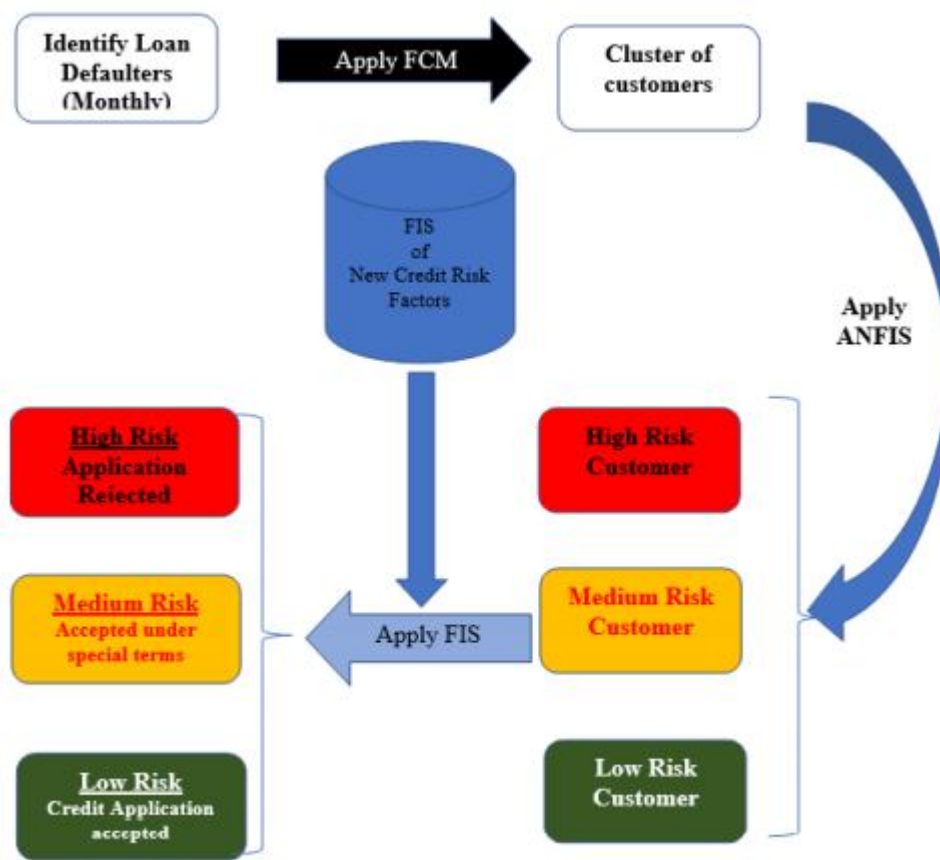


Fig. 4 Model for Customer Classification for Issuing Credit Cards

داده کاوی در مدیریت ریسک مالی مورد استفاده قرار می گیرد ، به موجب آن مدل های مرتبط به طور گسترده ای مورد مطالعه و استفاده قرار می گیرند. مدل های رتبه بندی اعتبار و ارزیابی ریسک وام از داده کاوی استفاده می کنند. از فناوری داده کاوی مربوط به درجه بندی کمی بصری برای ارزیابی اعتبار حساب بانکی استفاده می شود [4]. امتیازات اعتباری از مدل های داده کاوی برای بدست آوردن شاخص هایی برای رتبه بندی مشتری استفاده می کنند. در نتیجه ، موسسه مالی تصمیم می گیرد که آیا محدودیت مورد نظر را به متقاضی ارائه دهد یا درخواست را رد کند. علاوه بر این ، داده کاوی از اطلاعات مهم دیگری مانند استفاده غیر معمول از کارت اعتباری ، ضرر و زیان غیرقانونی پول ، کلاهبرداری یا هزینه های شدید برای مشتری استفاده می کند [4]. بینش و آموخته های حاصل از فرآیندهای گسترده داده کاوی ، شرکت مالی را با دیدی مناسب از مشتری ، تعیین می کند که این امر تعیین کننده تصویب یا عدم تأیید وام است. اعتبار اعتبار مربوط به خسارات وارده توسط مشتری

یا طرفی است که نتواند پرداخت قراردادی یا بدهی خود را انجام دهد. سایر معیارهای ریسک با توجه به افزایش ریسک نسبت به دوره معامله در نظر گرفته می شوند. قبل از توسعه ابزارهای یادگیری ماشین و هوش مصنوعی ، موسسات مالی برای مدل سازی ریسک اعتباری به رگرسیون های خطی logit، probit و کلاسیک متکی بودند. امروزه ، سازمان ها به هوش مصنوعی روی آورده اند ، که شامل چندین روش برای جمع آوری و توسعه ارتباطات برای مجموعه داده ها است (Davenport، 2018). یادآوری این نکته ضروری است که سازوکارهای سنتی در پرداختن به ماهیت پویا و جنبه های ناپایدار سیستم های اقتصادی ، که بر ریسک اعتبار تأثیر می گذارد و همچنین توانایی مشتریان در تحقق شرایط قراردادی ، کامل نبوده اند. قابلیت های مدیریت ریسک اعتبار از طریق استراتژی های مدرن داده کاوی ، به ویژه شبکه های عصبی و الگوریتم های یادگیری ماشین ، به طور قابل توجهی افزایش می یابد. در حالی که هر دو یادگیری ماشین و هوش مصنوعی همچنان در مدل های ارزیابی ریسک اعتباری محبوبیت کسب می کنند ، این مکانیسم ها ابزارهای مدرن و روش های آماری را برای دستیابی به مکانیزم های دقیق مدل سازی ترکیب می کنند. مشاهده شده است که رشد مبادله پیش فرض اعتبار (CDS) به عناصر غیرقابل پیش بینی مقیاس بزرگ مورد استفاده در تعیین هزینه نکول کمک کرده است [12]. ثابت شده است که داده کاوی برای یادگیری ماشین از نظر دقت و تاسیسات اقدامات حفاظتی بهتر از مکانیسم های سنتی و همچنین مدل های معیار عمل می کند. داده کاوی در توسعه بنگاههای کوچک و متوسط (SME) و وامهای مصرف کننده با توجه به در دسترس بودن داده های بالقوه اعمال می شود. استفاده از سازوکارهای مدرن ، مانند یادگیری ماشین ، برای داده کاوی ، اثبات شده است که به موجب آن در مقایسه با مدل های سنتی به پس انداز بیشتری کمک می کنند. استفاده از یادگیری ماشینی با تشخیص چند متغیره در تخمین ریسک اعتبار مفید است [13]. نتایج مربوطه بر تصمیمات وام تأثیر می گذارد [13]. مدل های مختلفی برای ارزیابی قابلیت اطمینان برای کاهش خطرات عدم پرداخت ساخته شده است. م institutions مالی موظفند سیستم های ارزیابی خود را براساس موقعیت های مبتنی بر زمینه توسعه دهند [14]. برخی از رویکردهای گذشته نتوانسته اند رفتار مشتری را به درستی پیش بینی کنند ، از این رو منجر به ضرر می شود. بیشتر این مدل هایی که سازمان های مالی از آنها استفاده کرده اند ثابت

هستند ، اما الگوهای ریسک اعتباری تحت تأثیر عوامل مختلفی از جمله نوسانات سیاسی قرار دارند [14]. در نتیجه ، عواملی مانند تحریم ها در شکست پرداخت اعتبار نقش داشته اند. اساساً ، عوامل سیاسی و اجتماعی بر الگوی ریسک مالی تأثیر می گذارد. مدل سازی پویا می تواند روندهای اقتصادی ، اجتماعی و سیاسی را در خود جای دهد. تجزیه و تحلیل داده ها در ارزیابی ریسک مالی ، قضاوت انسان را تا حدودی مغرضانه از بین می برد [15]. برای دستیابی م effectively به کارایی داده کاوی در مدیریت ریسک مالی ، توصیف پیش بینی کننده ها ضروری است. یک سیستم استنتاج فازی مبتنی بر شبکه (ANFIS) می تواند در ارائه مشخصات مشتری موثر باشد [15]. مدل های مالی مربوطه باید انعطاف پذیر و سازگار با شرایط زندگی واقعی ، به ویژه عوامل سیاسی و اقتصادی باشد. روش پیشنهادی برای داده کاوی و پردازش توسط [15] مشخصات مشتریان بد را برای یک دوره خاص بررسی می کند. نتایج حاصل در ارزیابی مشتری استفاده می شود ، و تغییرات در مشخصات عمومی مشتری در سیستم تکرار می شود. مهمترین ویژگی این مدل این است که عوامل ارزیابی مشتری به صورت پویا ایجاد شده است. طراحی تکنیک های مناسب برای داده کاوی مستلزم در نظر گرفتن توانایی های مورد نیاز است. الگوریتم های مورد استفاده ، از جمله سیستم های طبقه بندی یادگیری ، الگوریتم های مبتنی بر نمونه ، شبکه های Bayesian (BN) - نزدیکترین همسایه ، ماشین های بردار پشتیبانی ، مدل سازی فازی ، شبکه های عصبی ، القا rule قاعده و درخت های تصمیم ، در نظر گرفته می شوند قابلیت ها [15] [16]. این تکنیک ها به عنوان یادگیری ماشین ، هوش مصنوعی یا آمار کلاسیک دسته بندی می شوند. مدل ترکیبی در ایجاد الگوریتم هایی برای بررسی ریسک اعتبار استفاده شده است. این الگوریتم انتخاب و طبقه بندی ویژگی ها را بر اساس آموزش گروه به کار می برد [1]. مدل ترکیبی شامل سه مرحله است: من) جمع آوری و پیش پردازش داده ها. (ii) انتخاب ویژگی ؛ (iii) طبقه بندی. الگوریتم های مختلفی در انتخاب ویژگی مفید است ، از جمله ارزیابی ویژگی تسکین ، نسبت کسب اطلاعات ، الگوریتم های ژنتیک و تجزیه و تحلیل م componentلفه ها [15]. پس از آن ، مدل مناسب در الگوریتم طبقه بندی گروه استفاده می شود. الگوریتم Feature Selection (FS) نیز می تواند در این مرحله ادغام شود [15]. مرحله سوم ، که شامل طبقه بندی مجموعه داده است ، نشان می دهد که افزایش

انطباقی می تواند به دقت بالاتری دست یابد. بهینه سازی معیارها عامل اساسی دیگری است که در بهبود الگوریتم داده کاوی مورد توجه قرار گرفته است. از تابع هسته برای نگاشت نقاط ورودی و تابع فازی برای بهینه سازی چند معیاره استفاده می کند [15]. این توابع از فضای ویژگی های بعدی استفاده می کنند که مورد دوم در هر نقطه داده استفاده می شود ، در حالی که مورد اول برای نقشه برداری از ویژگی های بعدی بالا استفاده می شود. از ویژگی های مجازات نابرابر در کاهش طبقات نامتعادل استفاده می شود ، از این رو بر چالش های مرتبط با ناسازگاری فاکتورها غلبه می کند. در نظر گرفتن یکپارچه سازی مدل حسابداری برای افزایش ظرفیت پیش بینی ضروری است [15]. توجه به این نکته ضروری است که تکنیک های به کار رفته در مدل ترکیبی برتری را نشان نمی دهند زیرا آنها ویژگی های مختلف مدیریت ریسک اعتباری را به دست می آورند. علاوه بر این ، سیاست های ساختاری همچنین می توانند در طراحی سیستم برای افزایش اعتماد به نفس بانکداران و کاهش ریسک های اعتباری در تأمین مالی مبتنی بر دارایی ادغام شوند. رویکرد ترکیبی داده کاوی شامل طبقه بندی و خوشه بندی برای امتیازدهی اعتبار از طریق شناسایی همگنی و ناهماهنگی است که سپس برای محرومیت یا انزوا استفاده می شود [15]. یکی از مزایای این مدل این است که می تواند چندین طبقه بندی داده را به جای دسته های مشخص یا بد استفاده کند. مدل های متنوع طبقه بندی برای اطمینان از اینکه مشتریان خوب و بد براساس یک نقطه برش تفکر دقیق طبقه بندی می شوند ، مفید هستند.

از آنجا که متغیرهای اقتصاد کلان تأثیر قابل توجهی بر ریسک های اعتباری دارند ، تهیه معیارهایی برای پیش بینی روند محیط خارجی ضروری است. در نتیجه ، از مدل ترکیبی ماشین بردار پشتیبانی (SVM) برای ارزیابی نمرات اعتبار استفاده می شود با استفاده از: مجموعه داده های همسایگی برای انتخاب ورودی. شبکه جستجوی پارامتر هسته ورودی ترکیبی بهینه ؛ و تعیین انتخاب دقت بین روشها [15]. مقایسه رگرسیون لجستیک ، تحلیل تفکیکی خطی و طبقه بندی مبتنی بر SVM ، SVM را دقیق تر نشان داد [4]. یکی از نگرانی ها ، توانایی ترکیب طبقه بندی کننده ها در یادگیری ماشین است. پیش بینی اعتبار در گروه های طبقه بندی دقیق تر است ، از این

رو گزینه ارجح است. بنابراین ، الگوریتم ها و مدل ها براساس صحت کلی آنها انتخاب می شوند [15]. برخی از مدل ها بر وثیقه های کلان برای تأمین وام و همچنین افزایش دقت در اندازه گیری بازپرداخت مشتری تأکید دارند. با این وجود ، این مدل ها در شرایط غیرقابل پیش بینی و غیرمعمول ، مانند تحریم ها ، نادرست هستند [4]. رفتار مشتری نیز با گذشت زمان تغییر می کند. حسابداری برای ریسک و رفتار اعتباری مشتری مستلزم تدوین الگوی پیچیده ای است که شرایط غیرقابل پیش بینی یا بحران ها را برطرف می کند [17]. در این راستا ، [18] ، یک مدل ارزیابی ریسک را ارائه داد که برخی از عوامل نگرانی را پوشش می دهد. این مدل به دسته ای از داده های تولید شده توسط یک مدل افزودنی تعمیم یافته نیاز دارد که به موجب آن واحدهای یادگیری و نظارت در مورد آموزش محیط زیست به عنوان واحدهای ساکن در نظر گرفته شوند. برای اندازه گیری داده های ماه قبل از ضریب جینی استفاده شده است [15]. در حالی که می توان از یک پنجره زمان حافظه کامل برای الحاق تغییرات جدید با توجه به مشخصات مشتری استفاده کرد ، ثابت شده است که سازگاری کمتری با تغییرات عمده روند دارد. برای مقابله با این محدودیت ، می توان یک حافظه کوتاه ثابت را اجرا کرد. علیرغم این مدل پیچیده که از تکنیک های ایستا خارج می شود ، با توجه به جلوگیری از ضررهای آینده ، به دلیل وجود هنجارهای عملیاتی در صنعت بانکداری معایبی دارد [19]. توجه به این نکته ضروری است که صنعت بانکداری معمولاً از مدل های ساکن استفاده می کند که از سوابق بلند مدت برای مدل های امتیازدهی اعتبار استفاده می کنند.

داده کاوی برای مدیریت ریسک عملیاتی و سازمانی

سازمان ها ، چه کوچک ، چه متوسط و چه بزرگ ، می توانند از ابزارهای داده کاوی ، داده های بزرگ و ابزارهای هوش تجاری برای به دست آوردن بینش و همچنین بهبود عملکردهای تجاری استفاده کنند [7]. داده کاوی را می توان در مدیریت به کار برد تا الگوهایی ایجاد کند که بتواند در تحلیل رفتار و روندها مورد استفاده قرار گیرد [15]. در نتیجه ، از بینش حاصل برای تدوین استراتژی پایدار استفاده می شود. برای کاهش بیشتر ریسک شرکت

، از داده کاوی در گرفتن روندهای عملیاتی و مبتنی بر صنعت و همچنین نگرانی ها استفاده می شود ، از این رو رقابت تجاری و سهم بازار شرکت بهبود می یابد.

ضررهای مالی مستقیم و غیرمستقیم می توانند منجر به خرابی عملیات شوند. حوادث داخلی و خارجی می تواند یک موسسه مالی را در معرض خطر قرار دهد [20]. بلایای طبیعی ، غفلت از رویه ، خطاهای عملیاتی ، کلاهبرداری و سیستم های ناموفق می تواند سازمان را در معرض خطرات بی شماری قرار دهد [19]. طبق [21] ، راه حل های یادگیری ماشین برای پرداختن به برخی از این مسائل از ابزارهای داده کاوی برای جمع آوری اطلاعاتی که به شما کمک می کند تا در معرض خطرهای عملیاتی و پیچیدگی در توضیح سناریوهای متنوع کمک کند ، استفاده می کنند [22]. در همان یادداشت ، هوش مصنوعی و یادگیری ماشین یک روش ضروری برای شناسایی و همچنین اندازه گیری قرار گرفتن در معرض خطر برای تعیین تأثیرات آن بر عملیات است [22]. تغییر ریسک معاملات مستلزم ایجاد یک استراتژی مناسب برای کاهش ریسک است. علاوه بر این ، داده کاوی می تواند در جمع آوری داده ها برای فرآیندهای تکراری و اطلاعات گسترده سند مورد استفاده قرار گیرد. یکی از مزایای فرآیند تکراری با توجه به ارائه داده های عملیاتی این است که می توان آنها را در مجموعه داده های بزرگ جمع کرد ، از این رو اثربخشی یادگیری ماشین.

سازمانها با خطرات بی شماری روبرو هستند که برای مدیریت ریسک ارجاع شده به سیستمهای خودکار یاب نیاز دارند. موسسات فناوری داده کاوی در شناسایی علائم هشدار دهنده اولیه در بخش بانکی مفید بوده است. در این بعد ، داده کاوی در تعیین ریسک مالی مقیاس بزرگ در میان ظهور بحران مالی مورد استفاده قرار می گیرد. در درجه اول ، از آن در تجزیه و تحلیل استفاده می شود که به انعکاس ورشکستگی ، بحران پرداخت و وخامت مالی کمک می کند [17]. از مدلهای هشدار دهنده با توجه به این عوامل ، برای کنترل و هشدار زودهنگام برای شرکتهای استفاده می شود. سپس مدیریت با مسئولیت مدیریت مالی محتاطانه ، و همچنین طلبکاران ، سرمایه گذاران و بینش سیاستگذاران موظف است. از روشهای مختلف آماری ، یادگیری ماشینی و هوش مصنوعی ، از جمله

رگرسیون لجستیک ، شبکه عصبی و رگرسیون خطی ، برای ایجاد یک مدل هشدار دهنده برای بحران مالی استفاده می شود [19] [23]. این علائم اولیه برای موسسات مالی مفید است ، موسساتی که باید از خطرات بیجا جلوگیری کنند.

اجرای تجزیه و تحلیل ریسک مالی مستلزم روشن سازی اهداف و اهداف تعیین ابزار مناسب و همچنین مجموعه داده های مربوطه است. برنامه های کاربردی داده کاوی در شرکت های مالی شش مرحله را دنبال می کند: (i) توضیحات محتوا و اهداف. (ii) جمع آوری داده ها ؛ (iii) پیش پردازش ؛ (iv) داده کاوی ؛ (v) ارزیابی و تفسیر ؛ (vi) و جذب دانش [24]. اولین گام شامل روشن شدن ارتباط اطلاعات یا محتوایی است که باید جمع آوری شود تا کل فرایند در پرداختن به خطرات مالی قریب الوقوع دقیق باشد. مرحله دوم شامل تهیه اکتساب داده ها از منابع مختلف ، از جمله انبار داده و سیستم های حسابداری است. فرآیند سوم شامل اصلاح مشکلاتی است که ممکن است در مجموعه داده های جمع آوری شده وجود داشته باشد [24]. هدف این مرحله رسیدگی به مشکلات مربوط به ساختارهای دست و پاگیر داده ، محتوای غیراستاندارد و مجموعه داده های ناقص است [24]. مرحله چهارم شامل داده کاوی است ، به موجب آن الگوریتم مناسب انتخاب و در جمع آوری اطلاعات مورد نظر استفاده می شود. در این مرحله از الگوریتم های مناسب و همچنین هوش تجاری برای ارزیابی و تفسیر نتایج استفاده می شود [24]. مدل های مختلفی برای تصمیم گیری در یک شرکت مالی ارائه شده است. این مدل ها باید مربوط به بدنه امور مالی باشد. مرحله نهایی و ششم برای کاربرد داده کاوی در فرآیندهای تجاری ، جذب و کاربرد دانش در سیستم اطلاعات کسب و کار است. بینش مربوطه در تجزیه و تحلیل ریسک مالی و همچنین تصمیم گیری استفاده می شود.

پیش بینی بحران های مالی با استفاده از داده کاوی یک جنبه اساسی برای مشاغل است. این در ارائه هشدارهای ریسک مالی از طریق تجزیه و تحلیل داده های مربوط به یک شرکت مفید است. نتایج در ارائه پشتیبانی فنی و تصمیم گیری مفید است (عباس و همکاران ، 2019). بنابراین ، در انتخاب فرآیندهای داده کاوی ، شاخص های

مربوطه با توجه به معیارهای خاص شرکت در نظر گرفته می شوند. در این راستا ، ویژگی های سری زمانی در داده کاوی اعمال می شود. ایده اول شناسایی و کاربرد ویژگی های سری زمانی است که می تواند بر اساس چرخه حیات باشد [25]. مرحله دوم مربوط به یافتن قوانین مربوطه از طریق داده های مربوطه است. سرانجام ، از قوانین برای ارتباط داده ها با فرایندها و ویژگی های مالی ، پیش بینی های مربوطه استفاده می شود. مالی از الگوریتم های داده کاوی و همچنین یادگیری ماشین و هوش مصنوعی برای کاهش خطرات بی شماری که می تواند بدون شناسایی اتفاق بیفتد استفاده می کنند [15]. در این راستا ، داده کاوی برای طراحی مکانیزم های کنترل مالی برای محافظت از داده ها ، سیستم ها و سرویس گیرندگان استفاده می شود. توانایی یادگیری ماشین و هوش مصنوعی برای افزایش مدل های داده کاوی ، سرعت انجام کارهای معمول را افزایش می دهد و داده های فرآیند بدون ساختار برای آشکار کردن خطر [21]. این روش ها از بسیاری جنبه ها اتوماسیون را ترویج می دهند ، از این رو خطرات مرتبط با خطاهای انسانی را کاهش می دهند. در این راستا ، ارزیابی شبکه ها و سرویس گیرنده ها آسان تر می شود [15]. با استفاده از این تجزیه و تحلیل داده ها ، می توان کارمندان و بازرگانان را نیز کنترل کرد. در مورد رفتار ، خوشه بندی و طبقه بندی از ابزارهای اساسی برای ایجاد پروفایلی است که در آن از داده های ارتباطات الکترونیکی ، تجاری و صوتی برای ایجاد الگوها و همچنین خطرات پنهان استفاده می شود [19]. هشدارها همچنین با استفاده از داده های اولویت دار اطلاعات برای تعیین فعالیت های مشکوک ایجاد می شوند. ابزارهای اصلی هوش مصنوعی مورد استفاده برای شناسایی تقلب مالی ، از داده کاوی در انتخاب مجموعه داده های مناسب استفاده می کنند.

داده کاوی برای مدیریت ریسک بازار

مجموعه داده های زیادی برای تجارت و سرمایه گذاری در بازارهای مالی موجود است. استراتژی های مدیریت ریسک بازار سعی در کاهش مواجهه با پیامدهای منفی تغییر مربوطه در اوراق بهادار دارند. داده کاوی و پردازش با استفاده از یادگیری ماشین در مدیریت ریسک بازار مفید است [6]. توجه به این نکته ضروری است که تکنیک

های مناسب و روش های سیستماتیک برای اعتبار سنجی و مدل سازی اساسی هستند و تعیین نقش مدل های بازار در معاملات مهم است. هر مکانیزم برای یک هدف خاص مناسب است. یکی از جنبه های اساسی که به داده کاوی و مکانیسم زبان ماشین نیاز دارد ، تست تنش برای مدل های بازار است [14]. در این مورد ، نقش داده کاوی جمع آوری و همچنین پردازش داده ها برای تعیین خطر ظهور با توجه به رفتار تجارت است.

هم یادگیری ماشین و هم شبکه های عصبی برای تجزیه و تحلیل بازار مالی و همچنین رفتار ذینفعان ، مطابق با بهبود استراتژی های فعلی مدل سازی که ناقص ، نامعتبر یا نادرست هستند. از این رو ، داده کاوی از جنبه های مهم مدیریت ریسک مدل و همچنین بهبود مدل های فعلی است [19]. یادگیری ماشین برای داده کاوی به دلیل کارایی بدست آمده پس از انجام شبیه سازی ماشین ، برای اعتبارسنجی مدل در شرکت های مختلف مفید بوده است. در این حالت ، فرآیندهای داده کاوی در اجرای روشهای بدون نظارت یا نظارت متفاوت هستند [8]. این مدل ها می توانند الگوهای جدیدی ایجاد کنند. در اصل ، مدل های مدیریت ریسک بازار برای نظارت بر زمان واقعی ، آزمایش انحراف و اعتبار سنجی استفاده می شوند [6]. از آنجا که تکنیک های یادگیری ماشین و هوش مصنوعی روشهای مربوطه را به کار می اندازند ، روش های داده کاوی برای جمع آوری و پیش پردازش آزمون های داده ضروری است.

داده کاوی همچنین در مدیریت ریسک برای شرکت های بزرگی که دارایی آنها تأثیر قابل توجهی در قیمت گذاری بازار دارد ، کمک می کند. داده کاوی ناشی از یادگیری ماشین و هوش مصنوعی و در نتیجه استفاده از تکنیک های خوشه بندی برای اجرای اقدامات مقابله ای در معاملات بازارهای غیر نقد مفید است [5]. توجه به این نکته ضروری است که شبیه سازی ها ، و همچنین اقدامات واقعی ، مستلزم تخصیص روش های داده کاوی است که در صورت ورود یک شرکت بزرگ به بازار ، سود و زیان را به طور دقیق اندازه گیری می کند [26]. به همین ترتیب ، هر زمان که دارایی یا دارایی قابل توجهی به بازار عرضه شود ، می توان اقداماتی را تدوین کرد. داده کاوی یک فرایند اساسی در تکنیک های یادگیری ماشین است که برای شناسایی رابطه بین انتشار دارایی ها به بازار و

تأثیر کمی قابل مشاهده آنها استفاده می شود. داده کاوی همچنین در الگوریتم های معاملاتی مورد استفاده قرار می گیرد که به موجب آن یادگیری تقویت می شود ، بنابراین واکنش های بازار تغییراتی را ایجاد می کند که در فعالیت های آینده نهفته است. تحلیلگران معاملات ارزی از درختان تصمیم گیری و شبکه های عصبی برای هشدار به بازرگانان استفاده می کنند [6] [26].

داده کاوی برای مدیریت ریسک قانونی

انطباق با توجه به کنترل های اعمال شده ، به ویژه پس از بحران مالی ، یک م critical فلسفه اساسی برای موسسات مالی است. فرایندهای مدیریت ریسک با توابع بوروکراتیک برای انطباق نظارتی ناسازگار تلقی می شوند ، اما از نظر ارتباط با سیستم های مدیریت ریسک مرتبط هستند [27]. بنابراین ، انطباق با پیامدهای متنوع مدیریت ریسک در شرکت ارتباط دارد. به همین ترتیب ، خطرات عملیاتی ، بازاری و اعتباری در مدیریت ریسک انطباق در نظر گرفته می شوند [27]. داده کاوی ، یادگیری ماشینی و همچنین هوش مصنوعی برای اطمینان از ارزیابی مجموعه داده های حجیم برای تضمین انطباق آنها با مقررات قانونی استفاده می شود [2]. علاوه بر این ، داده کاوی در تعیین داده های غیر متعارف ، که چهارچوب ها و مقررات نظارتی را نقض می کند ، حیاتی است [27]. در بیشتر موارد ، داده کاوی به طور مستمر برای نظارت بر جنبه های عملکرد و اعتبار یک شرکت انجام می شود. با این حال ، پیگیری به صورت زمان واقعی در جلوگیری از نقض مربوط به انطباق بسیار مهم است. شرکت های مالی همچنین می توانند از داده کاوی برای به حداقل رساندن سرمایه نظارتی که سالانه هزینه می شود استفاده کنند.

افراد از هر قشر اجتماعی در خدمات بانکی و مالی تحت تأثیر تصمیمات خودکار حساس اجتماعی قرار می گیرند ، که بعضی اوقات برخی گروه ها را بر اساس مدل سازی و الگوریتم های مورد استفاده برای مرتب سازی پروفایل ها از مطالعه خارج می کنند [28]. امروزه شرکت ها می توانند از داده های شخصی و جمعی برای ارزیابی تصمیمات حساس اجتماعی استفاده کنند. ابزارهای داده کاوی می توانند استفاده ناعادلانه از تصمیمات خودکار را ایجاد

کنند ، همچنین به عنوان وب پرده شناخته می شود [28]. نتایج شبکه سازی می تواند نسبت به بازاریابی و قیمت گذاری کالاها ناعادلانه باشد. داده کاوی با توجه به مفاد قانونی ، الگوریتم های کاربردی را مورد تجزیه و تحلیل قرار می دهد تا تعیین کند که آیا استفاده از آنها ، به عمد یا غیر عمد ، به یک گروه خاص آسیب می رساند یا تبعیض آمیز است [28]. تمرکز الگوریتم های مربوطه برای داده کاوی حذف تعصبات با توجه به نژاد ، جنسیت یا سایر سوگیری های اجتماعی است. بنابراین ، کارشناسان در تلاشند تا پیامدهای الگوریتم های مورد استفاده موسسات مالی را برای تعیین پیامدهای اجتماعی آنها در سطح تجارت و جلوگیری از اثرات ناشی از آن در مورد چارچوب های نظارتی و قانونی درک کنند.

موسسات می توانند از داده کاوی آگاه از تبعیض (DADM) برای رسیدگی به اقدامات و نتایج تبعیض آمیز استفاده کنند ، برخی از آنها همچنین می توانند آنها را تحت پیگرد قانونی قرار دهند [28]. این روش بر استخراج و تحلیل بر اساس قوانین سوگیری اجتماعی متمرکز است. برخورد با جنبه حقوقی مدل های مورد استفاده برای پردازش داده ها در م institutions مالی ضروری است. اساساً ، برای مدل های فعلی داده کاوی دشوار است که از پس چالش های مربوط به تبعیض خودکار برآییم ، زیرا مشاهدات آماری کافی نیست [28]. در عوض ، این فرایندها دارای چند حوزه قضایی هستند و شامل چندین حوزه حقوقی می شوند ، از جمله شامل قوانین محافظت از داده و برابری نیستند. به عنوان مثال ، در ایالات متحده ادعا می شود در چارچوب قانونی به تبعیض خودکار پرداخته شده است [28]. در حالی که ممکن است در سطح نهادی با تبعیض قانونی برخورد شود ، الگوریتم ها و مدل های داده کاوی ممکن است تبعیض اجتماعی را ارتقا دهند. بنابراین ، از مدل های یادگیری ماشینی و هوش مصنوعی مشابه باید برای تعیین عدالت پیش پردازش در رابطه با برابری و تبعیت از مقررات قانونی استفاده شود.

4. بررسی تجربی

این بخش عملکرد و دقت الگوریتم پیشنهادی را ارائه می دهد. بر اساس پارامترهای ماتریس سردرگمی (جدول 1) F-size ، صحت ، MCC و ROC محاسبه و برای ارزیابی صحت الگوریتم پیشنهادی استفاده شد.

		Actual Values	
		Defected	Non-defective
Predicted Values	Defective	True Positive (TP)	False Positive (FP)
	Non-defective	False Negative (FN)	True Negative (TN)

Table 1: Confusion Matrix

در ماتریس سردرگمی فوق:

TP = یک نمونه مثبت شناسایی شده مثبت است.

FP = نمونه منفی مثبت تشخیص داده شد.

FN = نمونه مثبت منفی توسط الگوریتم.

TN = نمونه منفی توسط الگوریتم.

در زیر شرح مختصری از اقدامات عملکردی برای ارزیابی صحت الگوریتم ها استفاده شده است:

F-Measure با ارزیابی دقیق (به عنوان مثال سهمیه واقعی مثبت (TP) با توجه به تعداد کل مواردی که به

عنوان مثبت طبقه بندی شده اند) و اندازه گیری فراخوان (به عنوان مثال سهمیه واقعی مثبت (TP) با محاسبه

محاسبه می شود) به کل موارد مثبت).

$$Precision = \frac{TP}{(TP + FP)}$$

$$Recall = \frac{TP}{(TP + FN)}$$

حال

$$F1 - Measure = \frac{Precision * Recall * 2}{Precision + Recall}$$

دقت عبارت است از نسبت موارد طبقه بندی شده صحیح به تعداد کل موارد.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

AUC تخمین می زند که یک الگوریتم چقدر می تواند بین کلاس های معیوب و غیر معیوب تمایز قائل شود.

$$AUC = \frac{1 + TP_Y - FP_Y}{2}$$

5. آزمایش کنید

برای آزمایش ، مجموعه داده کاگل (<https://www.kaggle.com/uciml/german-credit>) به تصویب رسید [29]. مجموعه داده اصلی شامل 1000 مدخل با 20 ویژگی طبقه ای / نمادین است که توسط پروفیسور هافمن تهیه شده است. در این مجموعه داده ، هر ورودی شخصی را نشان می دهد که توسط یک بانک اعتبار به وی اعطا می شود. هر شخص با توجه به مجموعه ویژگی ها در ردیف های اعتباری خوب یا بد طبقه بندی می شود. برای طبقه بندی ، به ترتیب از Gaussian Naïve Bayes ، Random Forest ، J48 و MLP به عنوان زبان آموز پایه برای مجموعه داده های روان و کیسه بندی استفاده شد. این آزمایشات با استفاده از Pycharm در دامنه Python انجام شد. ماهیت طبقه بندی کننده ها در این آزمون با استفاده از دقت ، دقت ، فراخوان ، ROC و F-نمره طبقه بندی برآورد شد. لازم به ذکر است که برای تعیین این اندازه گیری ها از نرمال وزن دار استفاده شده است. نکته ای که در مورد تصمیم گیری برای نرمال وزن دار وجود داشت ، ارزیابی اندازه گیری برای هر برچسب گروه و روشن شدن منحصر به فرد بودن برچسب بود. اجرای طبقه بندی با مجموعه داده های آموزشی

وابسته به اعتبار سنجی 10 برابر محاسبه شد. الگوریتم حاصل برای انجام آزمایشات استفاده شد. ابتدا یک مجموعه داده کاگل و لیستی از طبقه بندی ها داده شد و سپس در مجموعه داده ها تکرار شد ، همانطور که در خط 7 نشان داده شده است. همانطور که در خط 8 نشان داده شده است ، برنامه مجموعه های آموزشی و آزمایشی را بر اساس اعتبار سنجی 10 برابر با تنظیم مجدد داده ها قبل از تقسیم حفظ می کند. . برای هر برابر ، حلقه در خطوط 9-20 بر آموزش طبقه بندی ها ، به دست آوردن پیش بینی ها و ارزیابی معیارهای ارزیابی متمرکز بود. میانگین امتیاز محاسبه شد. روش تعریف شده در خطوط 7-28 در کل مجموعه داده تکرار شد.

```

Input: Data-set, Classifiers
Results: AvgAccuracy, AvgPrecision, AvgF-Score, and AvgAUC
1: Data set: Kaggle Dataset for Fake News
2: Classifiers  $\leftarrow$  {RF,NB,MLP,J48,
    Bagging(RF,NB,MLP,J48), Smooteddataset(RF,NB,MLP,J48)}
3: All Accuracy Scores  $\leftarrow$  {}.
4: All Recall Scores  $\leftarrow$  {}.
5: All Precision Scores  $\leftarrow$  {}.
6: All F- Scores  $\leftarrow$  {}.
7: All AUC Scores  $\leftarrow$  {}.
8 for DS  $\in$  Datasets do
9   for Xtrain, Xtest  $\in$  KFold (nsplits = 10, shuffle = True).split(DS) do
10    (Xtrain, Xtest)  $\leftarrow$  PerformStandardScaler(Xtrain, Xtest);
11  For DS ResampledXtrain, ResampledYtrain  $\leftarrow$  SMOTE(Xtrain, Ytrain);
12    for clf  $\in$  Classifiers do
13      clf  $\leftarrow$  TrainClassifier(clf, ResampledXtrain, XtrainLabels);
14      predictions  $\leftarrow$  predict(clf, Xtest);
15      Accuracy  $\leftarrow$  ComputeAccuracy(predictions, XtestLabels);
16      Recall  $\leftarrow$  ComputeRecall(predictions, XtestLabels);
17      Precision  $\leftarrow$  ComputePrecision(predictions, XtestLabels);
18      F-score  $\leftarrow$  ComputeFmeasure(predictions, XtestLabels);
19      AUC  $\leftarrow$  ComputeAUC(predictions, XtestLabels);
20      AllAccuracyScores  $\leftarrow$  AllAccuracyScores  $\cup$  Accuracy;
21      AllRecallScores  $\leftarrow$  AllRecallScores  $\cup$  Recall;
22      AllPrecisionScores  $\leftarrow$  AllPrecisionScores  $\cup$  Precision;
23      AllFScores  $\leftarrow$  AllFScores  $\cup$  F-score;
24      AllAUCScores  $\leftarrow$  AllAUCScores  $\cup$  AUC;
25    end
26  end
27  AvgAccuracy  $\leftarrow$  ComputeAvgAccuracy(AllAccuracyScores);
28  AvgRecall  $\leftarrow$  ComputeAvgRecall(AllRecallScores);
29  AvgPrecision  $\leftarrow$  ComputeAvgPrecision(AllPrecisionScores);
30  AvgF-score  $\leftarrow$  ComputeAvgFmeasure(AllFScores);
31  AvgAUC  $\leftarrow$  ComputeAvgAUC(AllAUCScores);
32 end
33 return AvgAccuracy, AvgRecall, AvgPrecision, AvgF-score, AvgAUC

```

Fig. 5 Experiment for classification

6. نتایج:

نتایج تجربی در جداول 2-5 و شکل‌های 6-8 نشان داده شده است. جدول‌ها معیارهای مختلف طبقه بندی را نشان می دهد. بهترین عملکرد از نظر شناسایی اعتبار خوب در بانکها با طبقه بندی طبقه بندی شده Random Forest. Bagging Random Forest با 73,5٪ و 75,3٪ دقت و بهترین عملکرد در مجموعه داده صاف توسط Random Forest با 74,1٪ دقت به دست آمد. برای یادآوری و F-Measure همان عملکرد

Random Forest با مقادیر 75,4٪ ، 74,1٪ ، 72,2٪ برای فراخوان و 75٪ ، 74,1٪ ، 72,6٪ برای F-size بهتر است. به طور کلی ، AUC 0.5 نشانگر عدم تبعیض است ، 0,7 تا 0,8 عالی در نظر گرفته می شود و بیش از 0,9 برجسته است Again Random Forest یک عملکرد عالی با مقادیر 74,0٪ ، 74,8 و 75,8٪ است.

Table 2: Recall

Base Learner				Smoothed Dataset				Bagging Smoothed dataset			
RF	NB	MLP	J48	RF	NB	MLP	J48	RF	NB	MLP	J48
72.2	69.8	66.1	70.4	74.1	69.7	69.1	72.9	75.4	71.3	70.8	73.5

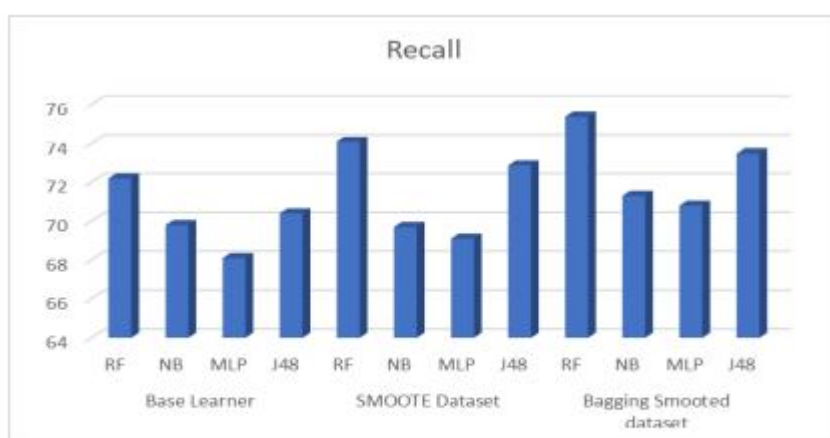


Fig. 6 Performance of classifiers (Recall) for bank credit

Table 3: F-measure

Base Learner				SMOOTE Dataset				Bagging Smoothed dataset			
RF	NB	MLP	J48	RF	NB	MLP	J48	RF	NB	MLP	J48
72.6	70	68.4	71.3	74.1	70.3	69.3	72.4	75.3	70.3	71	73.2

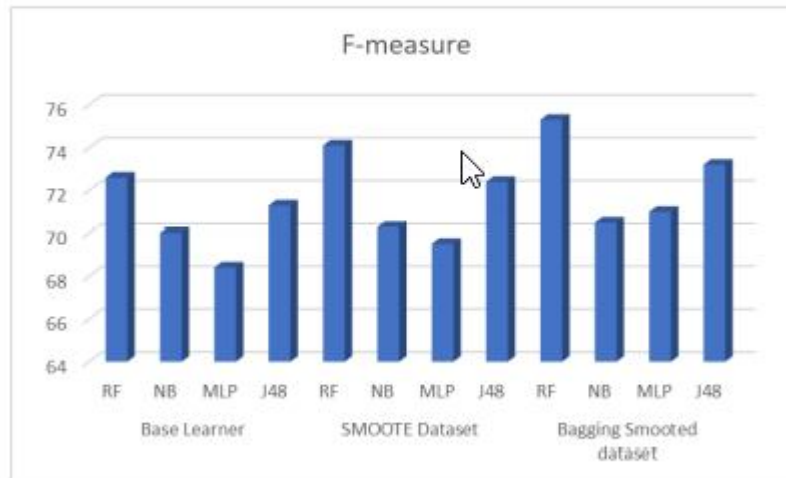


Fig. 7 Performance of classifiers (F-measure) for bank credit

Table 4: AUC

Base Learner				Smoothed Dataset				Bagging Smoothed dataset			
RF	NB	MLP	J48	RF	NB	MLP	J48	RF	NB	MLP	J48
74	74.6	66.1	69	74.8	73.4	70.3	69.2	75.8	73.6	72.1	74.6

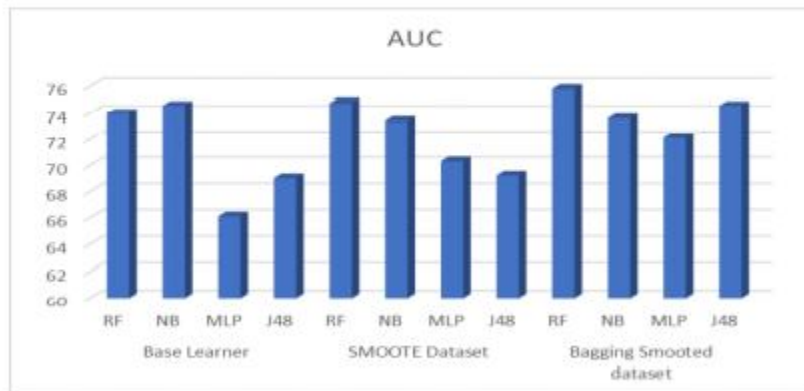


Fig. 8 Performance of classifiers (AUC) for bank credit

Table 5: ACCURACY

Base Learner				Smoothed Dataset				Bagging Smoothed dataset			
Random forest	NB	MLP	J48	RF	NB	MLP	J48	RF	NB	MLP	J48
73.5	72	68.8	70.4	74.1	69.7	69.1	72	75.3	70	70.8	73

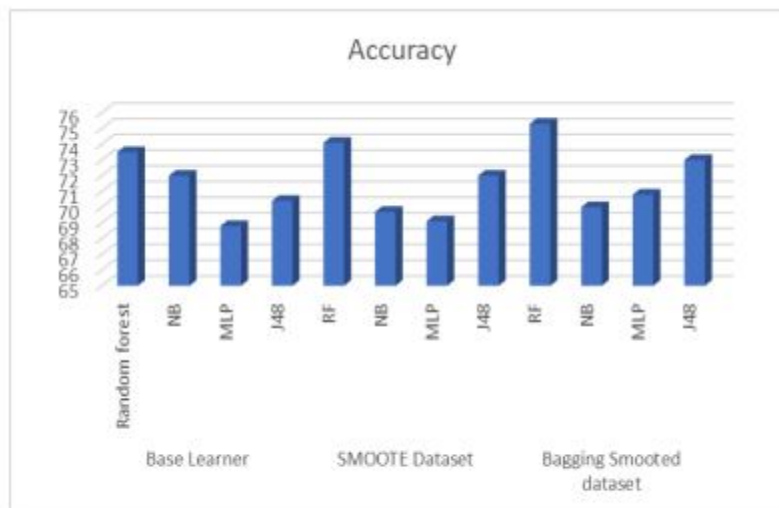


Fig. 9 Performance of classifiers (Accuracy) for bank credit

7. نتیجه گیری

داده کاوی یک فناوری اساسی است که به عنوان ابزاری برای توسعه مدل های مدرن برای مدیریت ریسک مالی مورد استفاده قرار می گیرد. این فناوری همراه با رگرسیون آماری ، یادگیری ماشین و هوش مصنوعی در مناطق وسیعی برای ارزیابی و مدیریت ریسک مالی با تجزیه و تحلیل خوشه ای ، درخت تصمیم گیری و قاعده ارتباط به عنوان رویکردهای اساسی استفاده می شود. داده کاوی عمدتاً در مدل سازی ارزیابی ریسک اعتبار و مشخصات مشتری با تعیین رتبه های اعتباری مورد استفاده قرار گرفته است. تکنیک های مختلف و همچنین الگوریتم ها ، از جمله مدل ترکیبی و SVM ، در درجه اول برای بهبود دقت طبقه بندی استفاده شده است. در مدیریت ریسک عملیاتی ، از داده کاوی برای توسعه کنترل های مالی برای کلاهبرداری و سایر موارد مهم مانند ورشکستگی و بحران های مالی استفاده شده است. مدیریت ریسک بازار در حال توسعه مدلهایی است که الگوهای زمان واقعی در بازارهای مالی را برطرف می کنند. به دنبال پیامدهای کمی تزریق دارایی به بازار ، از این مدل ها برای هشدار دادن به بازرگانان و همچنین شرکت های بزرگ استفاده می شود. در این تحقیق ، با مجموعه داده انتخاب شده ، بهترین عملکرد با طبقه بندی جنگل تصادفی در انواع معیارها ارائه شد: دقت ، AUC ، فراخوان و F-Measure

(در جداول 2-5 آورده شده و در شکل‌های 6-9 نشان داده شده است). از مدل سازی مشابه برای ایجاد هشدار برای مدیریت ریسک عملیاتی استفاده می شود. در حالی که سازمان ها بر مدل های داده کاوی عملیاتی ، بازار و اعتبار تمرکز دارند ، مدل سازی مبتنی بر قانون برای کاهش خطرات و هزینه های نقض چارچوب های قانونی ضروری است. بنابراین ، داده کاوی می تواند برای از بین بردن الگوریتم ها و مدل هایی برای اعتبار ، بازار و فعالیت های تجاری که بی عدالتی های اجتماعی را تداوم می بخشد با توسعه ابزارهایی برای رمزگشایی پیامدهای منفی مدل های موجود و همچنین از بین بردن تهدیدهای احتمالی برای تخلفات قانونی.

- [1] Hassani, H., Huang, X., & Silva, E. (2018). Digitalisation and big data mining in banking. *Big Data and Cognitive Computing*, 2(3), 18. doi: 10.3390/bdcc2030018
- [2] Dicuonzo, G., Galeone, G., Zappimbulso, E., & Dell'Atti, V. (2019). Risk management 4.0: The role of big data analytics in the bank sector. *International Journal of Economics and Financial Issues*, 9(6), 40-47. doi: 10.32479/ijefi.8556.
- [3] Wang, L., & Alexander, C. (2016). Machine learning in big data. *International Journal of Mathematical, Engineering and Management Sciences*, 1(2), 52-61. doi: 10.33889/ijmems.2016.1.2-006.
- [4] Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, 7(1), 29. doi: 10.3390/risks7010029.
- [5] Incekaraa , Ahmet and Çetinkayaa , Harun (2019). Liquidity risk management: A comparative analysis of panel data between Islamic and conventional banking in Turkey. *Procedia Computer Science*, 158, 955-963. doi: 10.1016/j.procs.2019.09.136.
- [6] Zhong, X., & Enke, D. (2017). Forecasting daily stock market return using dimensionality reduction. *Expert Systems with Applications*, 67, 126-139. doi: 10.1016/j.eswa.2016.09.027.
- [7] Davenport, T. (2018). From analytics to artificial intelligence. *Journal of Business Analytics*, 1(2), 73-80. doi: 10.1080/2573234x.2018.1543535.

- [8] Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. (2019). A systematic review on supervised and unsupervised machine learning algorithms for data science. *Unsupervised and Semi-Supervised Learning*, 3-21. doi: 10.1007/978-3-030-22475-2_1.
- [9] Heaton, J., & Polson, N. (2016). Deep learning for finance: Deep portfolios. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2838013.
- [10] Zhan, F., Zhu, X., Zhang, L., Wang, X., Wang, L., & Liu, C. (2019). Summary of association rules. *IOP Conference Series: Earth and Environmental Science*, 252, 032219. doi: 10.1088/1755-1315/252/3/032219.
- [11] Cavalcante, R., Brasileiro, R., Souza, V., Nobrega, J., & Oliveira, A. (2016). Computational intelligence and financial markets: A Survey and Future Directions. *Expert Systems with Applications*, 55, 194-211. doi: 10.1016/j.eswa.2016.02.006.
- [12] Son, Y., Byun, H., & Lee, J. (2016). Nonparametric machine learning models for predicting the credit default swaps: An empirical study. *Expert Systems with Applications*, 58, 210-220. doi: 10.1016/j.eswa.2016.03.049.
- [13] Figini, S., Bonelli, F., & Giovannini, E. (2017). Solvency prediction for small and medium enterprises in banking. *Decision Support Systems*, 102, 91-97. doi: 10.1016/j.dss.2017.08.001.
- [14] Aziz, S., & Dowling, M. (2018). Machine learning and AI for risk management. *Disrupting Finance*, 33-50. doi: 10.1007/978-3-030-02330-0_3.
- [15] Moradi, S., & Mokhatab Rafiei, F. (2019). A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks. *Financial Innovation*, 5(1). doi: 10.1186/s40854-019-0121-9.

- [16] Cerchiello, P., & Giudici, P. (2016). Big data analysis for financial risk management. *Journal of Big Data*, 3(1). doi: 10.1186/s40537-016-0053-4.
- [17] Abbas, F., Iqbal, S., & Aziz, B. (2019). The impact of bank capital, bank liquidity, and credit risk on profitability in post-crisis period: A comparative study of US and Asia. *Cogent Economics & Finance*, 7(1). doi: 10.1080/23322039.2019.1605683.
- [18] Sousa, M., Gama, J., & Brandão, E. (2016). A new dynamic modeling framework for credit risk assessment. *Expert Systems with Applications*, 45, 341-351. doi: 10.1016/j.eswa.2015.09.055.
- [19] Olson, D., & Wu, D. (2017). Data mining models and enterprise risk management. *Springer Texts in Business and Economics*, 119-132. doi: 10.1007/978-3-662-53785-5_9.
- [20] Weeserik, B., & Spruit, M. (2018). Improving operational risk management using business performance management technologies. *Sustainability*, 10(3), 640. doi: 10.3390/su10030640.
- [21] Choi, T., Chan, H., & Yue, X. (2017). Recent development in big data analytics for business operations and risk management. *IEEE Transactions on Cybernetics*, 47(1), 81-92. doi: 10.1109/tcyb.2015.2507599.
- [22] Kou, G., Chao, X., Peng, Y., Alsaadi, F., & Herrera-Viedma, E. (2019). Machine learning methods for systemic risk analysis in financial sectors. *Technological and Economic Development of Economy*, 25(5), 716-742. doi: 10.3846/tede.2019.8740.
- [23] Alzeaideen, K. (2019). Credit risk management and business intelligence approach of the banking sector in Jordan. *Cogent Business & Management*, 6(1). doi: 10.1080/23311975.2019.1675455.

- [24] Hou, Y., & Yuan, Z. (2019). Financial risk analysis and early warning research based on data mining technology. *Journal of Physics: Conference Series*, 1187(5), 052106. doi: 10.1088/1742-6596/1187/5/052106.
- [25] Jin, M., Wang, Y., & Zeng, Y. (2018). Application of data mining technology in financial risk analysis. *Wireless Personal Communications*, 102(4), 3699-3713. doi: 10.1007/s11277-018-5402-5.
- [26] Chandrinos, S., Sakkas, G., & Lagaros, N. (2018). AIRMS: A risk management tool using machine learning. *Expert Systems with Applications*, 105, 34-48. doi: 10.1016/j.eswa.2018.03.044.
- [27] Arner, D., Barberis, J., & Buckley, R. (2016). The emergence of Regtech 2.0: from know your customer to know your data. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3044280.
- [28] Carmichael, L., Stalla-Bourdillon, S., & Staab, S. (2016). Data mining and automated discrimination: A mixed legal/technical perspective. *IEEE Intelligent Systems*, 31(6), 51-55. doi: 10.1109/mis.2016.96.
- [29] Kaggle, <https://www.kaggle.com/uciml/german-credit>