دوازدهمین کنگره ملی سراسری
فناوریهای نوین در حوزه توسعه پایدار ایران
12th National Congress of
the New Technologies in Sustainable Development of Iran
senaconf.ir

چالش های پایگاه داده نوین
در بستر هوش مصنوعی و شبکه های عصبی

بابک یوسفی مکری
کارشناس ارشد هوش مصنوعی شبکه، دانشگاه امیر کبیر
Email:babakyousefimokri@gmail.com

چکیده

امروزه برای ارزیابی فاکتور ها برای اعطای امتیاز، بر اساس مجموعه داده های تراکنش، بسیاری از روش های تجزیه و تحلیل صلاحیت دار در شبکه های عصبی استفاده می شود. ارزیابی مجموعه داده های صلاحیت برای اعطای امتیاز منجر به انتخاب رد درخواست تراکنش می شود. این مکانیزم کاری چالش برانگیز شامل ارزیابی عمیق دیتاست یا اطلاعات تراکنش است. رویکرد داده کاوی، به عنوان متداول ترین روش برای تجزیه و تحلیل صلاحیت برای اعطای امتیاز با تمرکز بر الگوریتم های مختلف، مانند شبکه های عصبی، توصیف می شود. این فرایند مهم شامل جمع آوری، تجزیه و تحلیل و تصمیم گیری نهایی صلاحیت از فاکتورهای مختلفی است که برای ارزیابی برنامه های صلاحیت دار استفاده می شود. این منابع مانند فرم های درخواست صلاحیت، اشتراک داده ها و کلیدی ترین داده های داخلی بانک است. روش امتیازدهی در اعطای امتیاز اغلب به عنوان امتیاز بندی شناخته می شود که در حقیقت روشی برای ارزیابی درخواست دسترسی است. در عین حال آرشیو و سوابق امتیاز گذشته برای شناسایی ویژگی هایی که تأثیر مهمی در تبعیض برنامه های صلاحیت برای اعطای امتیاز به صورت خوب و بد دارند، ارزیابی می شوند.

کلمات کلیدی:صلاحیت برای اعطای امتیاز، شبکه های عصبی، تراکنش ها، شبکه های عصبی، شبکه عصبی مصنوعی، Bayesian

۱- مقدمه

همانطور که می دانید چنانچه گیرنده امتیاز خطایی انجام دهد یا از تکمیل آن ناتوان باشد، سیستم حق سلب امتیاز را خواهد داشت. دو دلیل اصلی نیاز به یک سیستم پشتیبانی متخصص عبارتند از:

۱) عدم وجود روش های دقیق اندازه گیری

۲) کمبود سیستم بررسی ریسک صلاحیت عمومی و صلاحیت برای اعطای امتیاز

در بسیاری از پایگاه داده ها است. ارزیابی ریسک در دسترسی های پایگاه داده باید به عنوان اهمیت ریسک پذیری درک شود. به منظور رتبه بندی تراکنش ها در کلاس های خوب و بد، سیستم پایگاه داده دقت مجموعه داده ها را اندازه گیری می کند. تراکنش های که در کلاس های خوب قرار دارند به احتمال پول را به پایگاه داده باز می گردانند. تراکنش ها کلاس بد احتمالاً وجه نقد را به پایگاه داده باز نمی گردانند و بنابراین درخواست های بعدی دسترسی ها آن ها اصولا بررسی نخواهد شد.

انواع مختلفی از روش های ارزیابی فاکتور ها برای اعطای امتیاز برای به حداقل رساندن نرخ تخریب داده های دسترسی استفاده می شود. از آنجایی که ارزیابی فاکتور ها برای اعطای امتیاز نقشی اساسی در بخش پایگاه داده دارد و بسیار حیاتی است، در اصل یک چالش اساسی پیش روی پایگاه داده ها است. طبقه بندی دقیق اطلاعات دسترسی به منظور جلوگیری از ضرر اقتصادی نقش بسزایی دارد.

در صورت اعطای دسترسی از پایگاه داده یا پایگاه داده نوین، امتیاز گیرنده (تراکنش) موظف به بازپرداخت مزیت اصلی و آتی آن در چارچوب قراردادی است که به رویت و امضا رسانده است. مهمترین فاکتور مبلغ دسترسی گرفته شده و نرخ مزیت پایگاه داده است. ابتدا پایگاه داده هنگام تأیید درخواست امتیاز، مشخصات و مدارک وی را بررسی می کند. هر پایگاه داده برای هر تراکنش امتیاز صلاحیت دار در نظر می گیرد که براساس پرونده فاکتورها دسترسی گیرنده به صورت اعداد بیان می شود. مقدار فعالیت بخش فاکتورها در شرایط مارکتینگ در هسته اصلی پایگاه داده نهفته است. فاکتور مهم دیگر ارزیابی ریسک در دسترسی های پایگاه داده است که باید به عنوان اهمیت ریسک درک شود.

۱-۱ هدف

ریسک جز عملکردهای سیستم شبکه های عصبی است که با استفاده از روش های مناسب و نوین به طور کامل از بین نرفته اما کاهش می یابد. یکی از اهداف اصلی در سیستم شبکه های عصبی حفظ یک سیستم صلاحیت دار پایدار و سالم است که از برنامه ریزی صلاحیت آغاز شده و با بسته شدن آن پایان می یابد. صلاحیت برای اعطای امتیاز، که ضروری ترین شکل ریسک برای پایگاه داده ها است، به شدت با اندازه گیری و مدیریت برتری این روش ارتباط دارد.

۲- چالش های شبکه های عصبی و شبکه های عصبی بر امتیاز

۱-۲ پیشینه تحقیق

برنام ها و الگوریتم های امتیازدهی احتمال بازپرداخت صلاحیت کلاینت از قبل را پیش بینی می کنند و به این مسئله می پردازند که آیا امتیاز در هر لحظه پیش فرض است یا خیر. ویژگی امتیازی برنامه براساس طبقه بندی خوب یا بد، از نظر صلاحیت برای اعطای امتیاز

ایجاد می شود. برنامه های دسترسی گذشته برای شناسایی ویژگی هایی که تأثیر مهمی در تبیین تفاوت ها بین برنامه های صلاحیت برای اعطای امتیاز خوب و بد دارند، ارزیابی می شوند [۴].

در صورت اعطای دسترسی از پایگاه داده تراکنش موظف به بازپرداخت اصلی و مزیت آتی امتیاز است. مهمترین ویژگی مبلغ دسترسی گرفته شده و نرخ مزیت تعریف شده است. به طور کلی در ابتدا پایگاه داده هنگام تأیید دسترسی مشخصات و مدارک تراکنش را بررسی می کند [۱۲].

از این فرایند به عنوان ارزیابی امتیاز یاد می شود که نیاز به پارامتر زمان دارد، لیکن معمولاً یک تصمیم باینری است که نتیجه آن تایید یا رد است. دو دلیل اصلی نیاز به یک سیستم مدیریتی و نظارتی، عدم وجود روش های دقیق اندازه گیری و عدم وجود سیستم ریسک صلاحیت عمومی و صلاحیت برای اعطای امتیاز در بسیاری از پایگاه داده ها است [۵۷].

بسیاری از خطرات مرتبط با دسترسی پایگاه داده، برای شبکه های عصبی و تراکنش های که دسترسی می گیرند. ارزیابی ریسک در دسترسی های پایگاه داده باید به عنوان اهمیت ریسک درک شود. صلاحیت برای اعطای امتیاز ریسکی است که دسترسی به موقع یا به طور کامل پس داده نمی شود. ریسک نقدینگی در اصل حالتی است که می تواند بیش از حد سریع از سپرده ها برداشت شده و پایگاه داده پرداخت وجه نقد فوری را منع کند. همچنین ریسک نرخ بهره، این خطر را دارد که نرخ های بهره قیمت گذاری شده در دسترسی های پایگاه داده برای کسب وجه کافی در پایگاه داده بسیار پایین باشد [۲۶].

به منظور رتبه بندی تراکنش ها در کلاس های خوب و بد، سیستم پایگاه داده دقت مجموعه داده ها را اندازه گیری می کند. تراکنش های که در کلاس خوب قرار دارند به احتمال زیاد بازپرداخت وجه را به پایگاه داده انجام خواهند داد. تراکنش ها ثبت شده درکلاس بد احتمالاً وجه نقد را به پایگاه داده باز نمی گردانند، بنابراین ملزم به پردخت جریمه و خسارات احتمالی خواهند شد. انواع مختلفی از روش های ارزیابی فاکتور ها برای اعطای امتیاز برای به حداقل رساندن داده های دسترسی استفاده می شود [۷].

مزایای مجموعه داده های قابل اعتماد صلاحیت برای اعطای امتیاز این است که هزینه های امتیاز دهی کاهش می یابد، تصمیم گیری های سطح بالا زمان کمتری را می‌گیرد و از خطرات ناشی از جلوگیری از عدم اعطای امتیاز جلوگیری می شود. از آنجا که ارزیابی فاکتور ها برای اعطای امتیاز نقشی اساسی در بخش پایگاه داده دارد طبقه بندی دقیق اطلاعات برای جلوگیری از ضررهای بنگاه های اقتصادی نقش بسزایی دارد [۴۸].

بر این اساس، یک شبکه عصبی یک مقدار داده پیچیده است و می تواند به عنوان یک شی معمولی ذخیره شود. رویکرد شی گرا راحت ترین و طبیعی ترین مدل طراحی را برای شبکه های عصبی ایجاد می کند (و به نظر ما ثابت کرده است) [۶۸]. از نظر سیستم های پایگاه داده شی گرا، شبکه های عصبی به طور کلی به عنوان اشیاء پیچیده تلقی می شوند. این سیستم ها در مدیریت و مدیریت چنین اشیایی در مناطق مختلف مانند طراحی به کمک رایانه، پایگاه های داده جغرافیایی، مدیریت ساختارهای مؤلفه و غیره بسیار ارزشمند بودند.

دومین جنبه بسیار مهم، استفاده از شبکه های عصبی به عنوان بخشی از مولفه قانون یک سیستم پایگاه داده مبتنی بر دانش است [۷۱]. شبکه های عصبی ذاتا دانش را با پردازش در گره ها نشان می دهند [۷۰]. شبکه های عصبی آموزش دیده شبیه به قوانین در مفهوم نمادین متعارف هستند. بنابراین، یک رویکرد بسیار امیدوارکننده، تعبیه شبکه های عصبی به طور مستقیم در چارچوب دانش تعمیم یافته یک سیستم پایگاه داده مبتنی بر دانش است.

برای اینکه کاربر بتواند پارادایم های شبکه عصبی را به صورت پویا تعریف کند، یک عملکرد مناسب باید توسط چارچوب سیستم پایگاه داده ارائه شود. این را می توان توسط عملگرهای قابل تعریف، که در تعدادی از سیستم های پایگاه داده موجود، مانند [۶۶و۷۲] در دسترس هستند، به دست آورد.

بنابراین، تجسم سیستم های پایگاه داده برای انجام وظیفه مدیریت داده (در مورد دانش ما) یک توسعه ساده بود. انتقال یک سیستم پایگاه داده به یک سیستم مدیریت مبتنی بر دانش منجر به چندین مزیت مهم می شود، مانند پایداری، رابط پرس و جو تعاملی، کنترل یکپارچگی، مدیریت همزمان، بازیابی و مدیریت داده های ثانویه [۶۷].

به طور کلی ویژگی های داده شده شامل قابلیت یک سیستم پایگاه داده برای مدیریت داده های مشترک مستقل از برنامه های کاربردی است [۶۹].

شبکه های عصبی یک رویکرد متفاوت به معنای دانش را امکان پذیر می کنند. نمایش دانش را می توان به عنوان پردازش شبکه عصبی [۷۰] مشاهده کرد. قوانین جهانی مورد نیاز نیستند، اما ذاتاً بخشی از پردازش در گره های پردازش فردی هستند (مشابه رفتار نورون ها).
تعداد زیادی مقاله در چند سال گذشته ظاهر شد که از شبکه عصبی به عنوان نمایش دانش عمیق استفاده می کرد. یک مثال معمولی منطقه در حال تکامل بازیابی اطلاعات است، جایی که شبکه های عصبی بسیار ارزشمند بودند [۶۵]. در عمل شبکه های عصبی در تعداد زیادی از کاربردها مفید هستند. آنها خود را به عنوان ابزاری برای رسیدگی به مسائل، که برای راه حل های الگوریتمی مرسوم پیچیده یا پرهزینه هستند، معرفی کرده اند.
شبکه‌های عصبی عمیق با الگوریتم‌های هوش مصنوعی در یادگیری ماشین (ML) بهترین طراحی را به‌طور خاص برای مقابله با حجم وسیعی از داده‌ها برای تجارت خرده‌فروشی تشکیل می‌دهند. رویکرد تحقیق محدود به سمت کاهش مصرف حافظه در ادغام الگوریتم‌های ML بر روی سیستم مدیریت داده است. علاوه بر این، مزیت اصلی هوش مصنوعی و ML در پیاده سازی آسان و کم هزینه بودن معرفی در مدیریت پایگاه داده موجود است [۷۳].
یک نقشه راه برای استفاده از تکنیک های هوش مصنوعی و داده های بزرگ (BD) برای اهداف مختلف مدلسازی، طراحی، نظارت، ساخت و بهره برداری از کاربردهای مختلف ابررسانا. تولید این پایگاه‌های اطلاعاتی ممکن است قبل از پیاده‌سازی سیستم دشوار یا حتی غیرممکن باشد. مدل‌سازی مبتنی بر فیزیک اجازه می‌دهد که داده‌های آموزشی نماینده تولید شود اما یکپارچه‌سازی شود .تعاملات پیچیده مورد نیاز برای پیش بینی های دقیق چالش برانگیز است[۷۳].

۳- برخی تعاریف در شبکه های عصبی و امتیاز

نزدیکی قابلیت رقابت و مزیت پایگاه داده ها و مدیر پایگاه داده ، در اصل کیفیت سبد فاکتورهای پایگاه داده است. با افزایش میزان تراکنش های که دارای صلاحیت بالایی هستند، کیفیت دسترسی پایگاه داده نیز افزایش می یابد. امتیاز دهی سیستم اصلی تصمیم گیری (در شبکه های عصبی) است که برای ارزیابی فاکتور ها کلاینت استفاده می شود. بنابراین، امتیاز بندی فاکتورها و امتیاز می تواند به عنوان روشی برای مدل سازی صلاحیت کلاینت تعریف شود.
در ادبیات برای امتیاز بندی صلاحیت تعاریف مختلفی ارائه شده است. منشأ سیستم های امتیازدهی را می توان در دهه ۱۹۳۰ جستجو کرد که برخی از مشاغل سفارش نامه پستی از سیستم امتیاز بندی برای کاهش اختلاف بین تحلیل گران دسترسی استفاده می کردند.

مدیریت صلاحیت برای اعطای امتیاز مسئله ای مهم برای شرکت های فعال در حوزه مالی بوده است. زیرا وظیفه اجباری تحلیلگران فاکتورها برای انجام خدمات نظامی در طول جنگ جهانی دوم و در نتیجه عدم وجود متخصصان صلاحیت برای اعطای امتیاز بوده است. بنابراین، شرکت ها می خواستند تحلیلگران، قوانینی را که برای اعطای صلاحیت استفاده می کردند به نگارش در آورند.

ارزیابی خوب از برنامه های ویژگی صلاحیت دار از طریق روش های رتبه بندی صلاحیت دار، باعث شد که پایگاه داده ها از رتبه بندی صلاحیت دار برای دسترسی های مسکن و امتیاز استفاده کنند. مقدار این قبیل برنامه ها آنقدر زیاد شده است که استفاده از استراتژی های سنتی به کار رفته در آن، از نظر اقتصادی دشوار شده و کارشناس درخواست ها را یک به یک ارزیابی کرده است. تمایل تراکنش ها صلاحیت و پایگاه دادهها برای ارزیابی درخواستها در یک لحظه عادلانه باعث ترغیب پایگاه دادهها به استفاده از سیستم های امتیازدهی فاکتورها در ارزیابی درخواستهای رسیده می شود.

به منظور ارزیابی امتیاز به تفکیک پذیرش یا رد، روش نمره صلاحیت از سوابق صلاحیت قبلی برای به دست آوردن یک مدل کمی استفاده می کند. معیارهای ارزیابی فاکتور ها به طور مداوم در کلیه درخواست های دسترسی در سیستم امتیاز دهی اعمال می شود. تصمیمات صلاحیت دار را می توان به سرعت با سیستم امتیاز دهی اتخاذ کرد. علاوه بر این، به دلیل زمان کوتاه مورد نیاز برای تکمیل درخواست دسترسی، امتیاز دهی رضایت تراکنش را افزایش می دهد. مدل های امتیازدهی فاکتورها یکی از موثرترین کاربردهای آماری در حوزه تجار و شبکه های عصبی محسوب می شود. بسیاری از روش های کمی برای اهداف امتیاز دهی در منابع پایگاه داده و کامپیوتری استفاده شده است. مطالعات مقایسه روش های آماری و روش های یادگیری ماشین برای ارزیابی فاکتور ها برای اعطای امتیاز نشان می دهد که تکنیک های یادگیری ماشین از کارآیی بیشتری نسبت به تکنیک های آماری برخوردار هستند.

۳-۱ تعریف مسئله و راهبرد پیشنهادی

مدل های امتیاز دهی یکی از موثرترین کاربردهای آماری تجاری محسوب می شود. بسیاری از روشهای کمی برای اهداف امتیازدهی در ادبیات استفاده شد. چندین روش آماری برای ارزیابی فاکتور ها برای اعطای امتیاز وجود دارد، اگرچه این روش ها در مدل سازی سیستم های پیچیده اقتصادی مشکل دارند، زیرا براساس ویژگی های ثابت و فرضیات آماری استوار هستند. مطالعات مقایسه روش های آماری و روش های یادگیری ماشین برای ارزیابی فاکتور ها برای اعطای امتیاز نشان می دهد که تکنیک های یادگیری ماشین از کارآیی بیشتری نسبت به تکنیک های آماری برخوردار هستند. یک تکنیک طبقه بندی باینری و با استفاده از Support Vector Machines[۱] برای صدور ارزیابی فاکتور ها برای اعطای امتیاز کمی ارائه شده است. روش SVM با استفاده از داده های گسترده و در عین حال کم حجم استفاده می شود که در آن محدوده زمانی معمولا کمتر از ۹۰ روز گذشته و معنای محدود بیش از ۹۰ روز گذشته است.. مقایسه عملکرد مدل ها نشان می دهد که استفاده از داده های گسترده بهتر از مدل های تولید شده با استفاده از داده های جزیی است.

[۱] SVM

سلسله تحقیق دیگری با استفاده از SVM برای ارزیابی فاکتور ها برای اعطای امتیاز همراه با انتخاب پارامترهای توسعه نیز به انجام رسیده است.

در مرحله بعدی، از محاسبه و ارزیابی این نتایج عملکرد تراکنش پیش بینی شده به عنوان ورودی به الگوریتم درختان طبقه بندی شده و رگرسیون استفاده می شود. در همین حال گزارش های صلاحیت دار تراکنش و اطلاعات دموگرافیک نیز به کار می آید. نتایج حاصل، برای تهیه دستورالعمل های تصمیم گیری به جهت تعیین اینکه آیا صلاحیت به کلاینت اعطا می شود، برای تعیین محدوده دسترسی، نرخ درصد سالانه و سایر سطوح محصولات پایگاه داده استفاده می شود. بر اساس این نتایج، محققان نتیجه گرفتند که گزارش صلاحیت دار یک دارنده رکورد، بیشترین توضیح را در مورد امتیاز دهی دارد و عدسترسیل جمعیتی از اهمیت کمتری برخوردار هستند زیرا از اثربخشی امتیاز دهی کمتری برخوردار می باشند.

۲-۳ روش های مقایسه

مطالعات اخیر نشان داده است که تکنیک های سنتی تجزیه و تحلیل آماری در کنار هوش مصنوعی (AI[۲]) معمولاً در انتخاب ویژگی هایی اعمال می شوند که می توانند دقت شناسایی صلاحیت برای اعطای امتیاز را افزایش دهند. با توجه به تکنیک های آماری سنتی، تحقیقات موجود عدسترسیلی را بررسی می کند که صلاحیت برای اعطای امتیاز تراکنش ها را تحت تأثیر قرار می دهند که عمدتا با روش های آماری مانند تجزیه و تحلیل چندگانه متمایز، رگرسیون منطق چندگانه و زنجیره مارکوف صورت می گیرند.

با استفاده از مدل های پیشنهادی مانند NLM، ایجاد یک مدل رگرسیون منطقی می توان شاخص های ارزیابی اعطای امتیاز را بررسی کرد. برای بررسی ویژگی موثر در صلاحیت برای اعطای امتیاز، از مدل های سفارشی منظم رگرسیون نیز بارها استفاده شده است. در تحقیق دیگری با تکنیک تعریف Fuzzy و روش تحت آزمون شرایط صلاحیت آنها ارزیابی شد [۵]. همچنین با بکار گیری الگوریتم مبتنی بر ژنتیک تجزیه و تحلیل رتبه صلاحیت دار تراکنش ها به انجام رساندند [۶۱]. یک مدل پنهان مارکوف نیز برای پیش بینی رتبه بندی فاکتورها و ارزیابی عملکرد قابل پیش بینی قابل اطمینان تری ارائه داد [۴۹].

از آنجایی که اغلب ویژگی های ریسک بر شاخص های مالی یا داده های خصوصی تراکنش ها متمرکز است، ممکن است متغیرهای اقتصاد کلان نادیده گرفته شود. ثانیاً، بیشتر سیستم های رتبه بندی فقط بر روی دقت طبقه بندی متمرکز هستند اما قادر به انتخاب متغیرهای اصلی که بر تمایل تراکنش ها به بازپرداخت ها تأثیر می گذارند نیستند.

۳-۳ انواع طبقه بندی برای تجزیه و تحلیل ریسک صلاحیت دار

۱-۳-۳ طبقه بندی کننده Bayesian

شبکه Bayesian یک مدل آماری – نموداری مستقیم یا دایره ای شکل است. هر گره در نمودار یک متغیر تصادفی را نشان می دهد که در آن لبه ها نماد عملکردهای متغیر مربوطه هستند. شبکه Bayesian همچنین یک نمودار حلقوی است که نمایانگر توزیع احتمال مشترک در یک متغیر تصادفی است. به منظور محاسبه پارامترهایی مانند میانگین و واریانس، یک متغیر مهم دیگر برای طبقه بندی Naïve Bayes، مقدار کمی از داده، ضروری است.

۲-۳-۳ شبکه باورهای Bayesian

[۲] Artificial Intelligence

یک شبکه باورهای Bayesian یک نمودار حلقوی مستقیم است که شامل چندین گره است که متغیرهایی را با یک گروه محدود از حالت ها و لبه ها نشان می دهد. این نماد نشانگر وابستگی علیت احتمالی به متغیرها است. نمودار حلقوی مستقیم نشان دهنده ساختار وابستگی ما بین گره ها است. بنابراین رابطه بین متغیرها و حالات مربوطه بخشی کمی را شامل می شود که در برگیرنده یک جدول احتمالی مشروط است. قانون زنجیره ای بیان می کند که یک شبکه باورهای Bayesian توزیع عمومی همه متغیرها را نشان می دهد که مبتنی بر نمودار حلقوی مستقیم است. برای هر گره از شبکه، اندازه گیری حاشیه و شرایط احتمال وجود دارد.

۳-۳-۳ درخت تصمیم

درخت تصمیم مدلی است که نظرات مربوط به عناصر هر شاخه را به منظور نتیجه گیری یک مقدار هدف در برگ ها ترسیم می کند. این روش یکی از بهترین تکنیک های نظارتی است. هر گره داخلی یا گره غیر برگ با عملکرد ورودی در این روش یادگیری مشخص شده است. هر گره برگ نیز در درخت دارای توزیع طبقاتی یا طبقه ای احتمالی است.

۳-۳-۴ K- نزدیکترین همسایه[۳]

روش KNN تکنیک طبقه بندی و مبتنی بر رگرسیون است که روشی غیرپارامتری است. این شیوه شامل یک بسته آموزش مطلوب و نامطلوب است.گاهی به آن الگوریتم بی رمق نیز گفته می شود. برای تعمیم الگوریتم از هیچ نقطه داده ای برای تست استفاده نمی شود. این عملکرد بدان معنی است که مرحله آموزش بسیار سریع است و تمام اطلاعات آموزش حفظ می شود. در طول مرحله آزمون،به تمامی داده های آموزش احتیاج است. در این میان از فاصله اندازه گیری شده برای تعیین موارد k مجموعه داده های آموزشی استفاده می شود. در این روش از متداولترین فاصله اقلیدسی برای متغیر ورودی با ارزش واقعی استفاده می شود. در رگرسیونKNN، پیش بینی براساس میانگین یا بیشتر موارد مشابه K استوار است. مکانیزم KNN را می توان به عنوان یک کلاس با بیشترین فرکانس در بیشتر موارد قابل مقایسه وقتی KNN برای طبقه بندی استفاده می شود، محاسبه کرد.

۳-۳-۵ K-Means

هدف این الگوریتم کشف گروه هایی در اطلاعات با تعداد گروه های K است. این الگوریتم برای اختصاص دادن یک گروه K به هر نقطه داده با توجه به ویژگی های ارائه شده عمل می کند. به طور تصادفی مقدار گره را وسیله k برای مرکز خوشه انتخاب کرده و نزدیک ترین نقطه داده را به خوشه اختصاص می دهد. روش
K-Means از رایج ترین الگوریتم های خوشه بندی هستند.

۳-۳-۶ شبکه های عصبی مصنوعی[۴]

گروهی از شبکه های عصبی متصل به گره وزنی، یک شبکه عصبی مصنوعی را تشکیل می دهند. هر گره می تواند نورون موجودی را تولید کرده و اتصالات سیناپسی بین نورون همان ارتباط بین گره ها است. شبکه عصبی شامل سه لایه است، به نام پرسپترون چند لایه، که لایه ورودی، لایه پنهان و لایه خروجی است. در MLP شبکه لایه ها به عنوان یک لایه واحد ورودی به لایه واحدهای پنهان متصل می شوند که سپس به یک لایه واحد خروجی متصل می شوند.

---

[۳] K Near Neighbor
[۴] Multi-Layer Perceptron

۴- مدل های شبکه عصبی

از میان مدل های متنوع، شبکه های عصبی مصنوعی (ANN) به دلیل ویژگی های یادگیری تطبیقی غیر خطی و غیر پارامتری یکی از قدرتمندترین ابزارها برای طبقه بندی الگوها بوده و با موفقیت در محاسبه امتیازات صلاحیت استفاده شده است. بسیاری از مطالعات انجام شده است، که ANN ها را با سایر تکنیک های طبقه بندی سنتی در زمینه مدل های امتیاز دهی مقایسه کرده است، زیرا دقت پیش بینی از پیش تعیین شده ANN ها از LDA و LR کلاسیک بهتر است.

۴-۱ هوش مصنوعی-اسناد-مجموعه-مهم

فناوری هوش مصنوعی بسیار مهم است زیرا به نرم افزارها اجازه می دهد تا عملکردهای حوزه انسانی (درک، استدلال، برنامه ریزی، ارتباط و ادراک) را به طور فزاینده ای موثر، کارآمد و مقرون به صرفه انجام دهند.

مجموعه کتابخانه های علوم داده مهمی که در دسترس همه محققین هستند عبارتند از:

۱) پانداها:

پردازش و تجزیه و تحلیل داده ها ممکن است با کمک زبان برنامه نویسی پایتون با استفاده از بسته نرم افزاری Pandas انجام شود. به طور خاص، ساختارهای داده و رویه های لازم برای دستکاری جداول عددی و سری های زمانی را فراهم می کند. این نرم افزار منبع باز است که با مجوز BSD سه بند توزیع شده است. عبارت "پانل" که در اقتصاد سنجی استفاده می شود، منشأ کلمه "پانل" است که به مجموعه داده هایی اشاره دارد که شامل مشاهدات در چندین دوره زمانی برای افراد مشابه است. نام آن یک جناس برای عبارت "تجزیه و تحلیل داده های پایتون" است که در نام آن نیز گنجانده شده است.

۲) NumPy:

Numpy کتابخانه ای برای زبان برنامه نویسی پایتون است که پشتیبانی از آرایه ها و ماتریس های چند بعدی و بزرگ را اضافه می کند، به علاوه تعداد زیادی توابع ریاضی سطح بالا که می توانند برای کار بر روی این آرایه ها استفاده شوند. Numpy توسط بنیاد نرم افزار پایتون توسعه یافته است. سلف NumPy، معروف به Numeric، اولین بار توسط Jim Hugunin با کمک تعدادی از توسعه دهندگان نرم افزار دیگر توسعه یافت. در سال ۲۰۰۵، Travis Oliphant NumPy را با ادغام جنبه‌های یک محصول رقیب به نام Numarray در Numeric و ایجاد تعدادی تغییرات دیگر توسعه داد. NumPy نرم افزاری است که به صورت رایگان در دسترس عموم است و چندین مشارکت کننده دارد.

۳) SciPy:

Scipy یک کتابخانه پایتون است که برای محاسبات فنی و علمی استفاده می شود. این منبع باز و رایگان برای استفاده است. SciPy دارای ماژول هایی برای مواردی مانند بهینه سازی، جبر خطی، یکپارچه سازی، درون یابی، توابع ویژه، FFT، پردازش سیگنال و تصویر، حل کننده های ODE و انواع کارهای دیگر است که در مهندسی و علم رایج است.

۴) Matplotlib:

یک کتابخانه نمودار برای زبان برنامه نویسی پایتون و پسوند NumPy برای ریاضیات عددی، Matplotlib بخشی از کتابخانه استاندارد پایتون است. این یک رابط برنامه نویسی برنامه شی گرا (API) برای جاسازی نمودارها در برنامه ها با استفاده از جعبه ابزار رابط کاربری

گرافیکی همه منظوره مانند Tkinter، wxPython، Qt یا GTK ارائه می دهد. همچنین یک رابط رویه ای موجود است که به آن "Pylab" گفته می شود. این رابط بر روی یک ماشین حالت (مثل OpenGL) ساخته شده است و به گونه ای طراحی شده است که بسیار شبیه به MATLAB است. با این حال، استفاده از این رابط توصیه نمی شود.

۵) SeeBorn

کاربران پایتون که به دنبال ساخت تصاویر آماری هستند، Seaborn را یک ماژول مفید خواهند یافت. این برنامه در بالای Matplotlib توسعه یافته است و به شدت با پشته PyData مرتبط است و از ساختارهای داده Numpy و Pandas و همچنین روش‌های آماری Scipy و StatsModels پشتیبانی می‌کند. این یک رابط سطح بالا برای ایجاد تصاویر آماری ارائه می دهد که هم جذاب و هم مفید هستند.

۶) Scikit- Learning:

بسته یادگیری ماشینی معروف به scikit-learn که قبلاً به عنوان scikit-learn شناخته می شد و همچنین به عنوان sklearn شناخته می شد، به صورت رایگان به عنوان بخشی از زبان برنامه نویسی پایتون در دسترس است. این برای تعامل با کتابخانه‌های عددی و علمی پایتون NumPy و SciPy طراحی شده است و به انواع الگوریتم‌های طبقه‌بندی، رگرسیون و خوشه‌بندی مجهز است. برخی از این الگوریتم‌ها شامل ماشین‌های بردار پشتیبانی، جنگل‌های تصادفی، تقویت گرادیان، k-means و DBSCAN هستند.

۷) TensorFlow:

TensorFlow یک کتابخانه نرم افزاری است که منبع باز و رایگان برای استفاده است. برای یادگیری ماشین و هوش مصنوعی استفاده می شود. ممکن است برای مشاغل مختلف مورد استفاده قرار گیرد، اما آموزش و استنتاج شبکه‌های عصبی عمیق جایی است که تاکید اصلی آن است. تیم Google Brain TensorFlow را برای استفاده داخلی در گوگل، به‌ویژه برای استفاده در تحقیق و تولید ایجاد کرد. سال ۲۰۱۵ شاهد انتشار نسخه افتتاحیه بود که تحت مجوز آپاچی ۲,۰ انجام شد. TensorFlow 2.0 نامی بود که گوگل به نسخه اصلاح شده TensorFlow که در سپتامبر ۲۰۱۹ منتشر شد داده بود.

۸) Keras

Keras یک رابط شبکه عصبی مصنوعی مبتنی بر پایتون است که توسط یک بسته نرم‌افزار منبع باز معروف به Keras ارائه می‌شود. عملکرد Keras ارائه یک رابط برای کتابخانه TensorFlow است. Keras از تعدادی بک‌اند مختلف از جمله TensorFlow، Microsoft Cognitive Toolkit، Theano و PlaidML تا نسخه ۲,۳ پشتیبانی می‌کرد. TensorFlow تنها چارچوبی است که در نسخه ۲,۴ پشتیبانی می شود. تاکید زیادی بر کاربرپسند بودن، ماژولار بودن و توسعه پذیر بودن با هدف تسهیل آزمایش سریع با شبکه های عصبی عمیق دارد.

۹) PyTorch

PyTorch یک کتابخانه یادگیری ماشین منبع باز است که بر اساس کتابخانه Torch است. این برای برنامه هایی مانند بینایی کامپیوتر و پردازش زبان طبیعی استفاده می شود و عمدتا توسط آزمایشگاه تحقیقاتی هوش مصنوعی فیس بوک (FAIR) ایجاد شده است. این نرم افزار منبع باز است که تحت مجوز مبتنی بر BSD اصلاح شده ارائه می شود. PyTorch دارای هر دو رابط Python و C++ است که اولی اصلاح‌تر شده و دومی تمرکز اصلی توسعه است.

۴-۲ انواع نوین پایگاه داده

از لحاظ تاریخی، سیستم‌های مدیریت پایگاه داده (DBMS) برنامه‌های نرم‌افزاری ساده و سخت‌افزار مرتبطی بودند که به کاربران امکان دسترسی به داده‌ها از مکان‌های جغرافیایی مختلف را می‌دادند. این سیستم به کاربران خود امکان ذخیره داده ها را بدون نگرانی در مورد تغییرات ساختاری یا مکان فیزیکی داده ها ارائه می دهد. علاوه بر این، یک سیستم مدیریت پایگاه داده (DBMS) می تواند محدودیت هایی را برای داده های مورد استفاده و خدمات در دسترس برای هر کاربر تعیین کند.

سازمان ها به طور فزاینده ای انبارهای داده و دریاچه های داده خود را در سیستم های ذخیره سازی ابری ادغام می کنند. انتقال به فضای ابری به یک سیستم مدیریت پایگاه داده (DBMS) برای کار با طیف وسیعی از قالب‌های داده جدید نیاز دارد.

گرایش های مدیریت پایگاه داده عبارتند از:

۱) DBMS مبتنی بر ابر

گزارش گارتنر The Future of the DBMS Market Is Cloud پیش بینی می کند که استفاده از DBMS های مبتنی بر ابر افزایش خواهد یافت. بازار سیستم های مدیریت پایگاه داده به طور فزاینده ای توسط سرویس های ابری هدایت می شود، و نه دیگر توسط سیستم های داخلی.

۲) مدیریت داده های افزوده (ADM)

مدیریت داده‌های افزوده شده از یادگیری ماشین و هوش مصنوعی برای خودکارسازی وظایف مدیریت داده‌ها، مانند تشخیص ناهنجاری‌ها در مقادیر زیاد داده و حل مشکلات کیفیت داده، استفاده می‌کند.

۳) امنیت داده ها (و جلوگیری از نقض داده ها)

در سال گذشته چندین مورد نقض اطلاعات در سطح بالا رخ داده است. به عنوان مثال، لینکدین در ژوئن ۲۰۲۱ نقض شد و در نتیجه ۷۰۰ میلیون کاربر اطلاعات خود را به صورت آنلاین فروختند. در ماه سپتامبر، خرده فروش Neiman Marcus با ۴,۸ میلیون مشتری تحت تأثیر قرار گرفت.

۴) پایگاه های داده در حافظه

پایگاه داده های درون حافظه به دلیل اینکه سریعتر از سیستم های سنتی پاسخ می دهند محبوبیت پیدا می کنند. یک پایگاه داده در حافظه (IMDB) درایو دیسک را حذف می کند و در عوض داده ها را در حافظه اصلی رایانه ذخیره می کند - حافظه دسترسی تصادفی یا RAM آن.

۵) پایگاه داده گراف

پایگاه داده های گراف یک راه عالی برای ایجاد و تحقیق روابط به روشی سریع و آسان است. آنها از گره ها و یال ها برای ایجاد روابط داده ها استفاده می کنند (گره ها نشان دهنده موجودیت ها و یال ها نشان دهنده روابط آنها هستند).

۵-بررسی چالش های شبکه های عصبی و شبکه های عصبی

معمولا اولین کاربرد امتیازدهی در پایگاه داده نوین مشارکت است. برای ارائه نتایج توسعه یافته با مدل امتیاز دهی در اینجا از معیارهای مختلف عملکردی مانند خطای میانگین توان ۲، نرخ رگرسیون، امتیاز طبقه بندی درست نادرست استفاده می شود. مدیر پایگاه داده و پایگاه داده های داده کوچک می توانند فاکتور های پایینی را به تراکنشهای دارای پتانسیل رشد بالا اعطا کنند. با این وجود قبل از مشارکت در یک پروژه ، این مدیر پایگاه داده در ارزیابی ریسک پذیری تراکنش با مشکل روبرو هستند.

یکی از راه های کنترل تأثیرات منفی اطلاعات متقارن و هزینه های تراکنش های شبکه های عصبی، استفاده از امتیاز دهی است. امتیاز دهی شامل پیش بینی رفتار تراکنش ها از طریق بررسی سوابق تراکنش ها از پایگاه داده های داده دیگر یا داده های موجود در یک دیتا بیس متمرکز (دیتا سنتر) است. این امر در واقع تراکنش ها مختلف پایگاه داده را با توجه به رفتار آنها در هنگام بازپرداخت در طبقات مختلف طبقه بندی نموده، سپس کلاینت جدید را به یکی از این کلاس ها با داده های ارائه شده مرتبط می کند. اگرچه تعاریفی که به امتیاز دهی نسبت داده می شود از نویسنده ای به نویسنده دیگر متفاوت است، اما به طور کلی پذیرفته شده است که امتیاز دهی به صلاحیت یک ابزار مدیریت ریسک است.

داده های به کار رفته تراکنش ها پذیرفته شده و رد شده در پایگاه داده نوین X است که از قراردادهای مشارکت استفاده می کردند. در اینجا توانایی مدل شبکه عصبی مصنوعی را برای پیش بینی امتیاز دهی به صلاحیت این پایگاه داده نوین تجزیه و تحلیل می شود. در روش مدل سازی پیشنهادی می توان مراحل کار را به صورت زیر فرض نمود:

۱. معیارهای امتیازدهی فاکتورها تعریف می شود.
۲. معیارهای ارزیابی و کلاس بندی امتیازدهی فاکتورها تعریف می شود.
۳. از شبکه MLP در امتیازدهی فاکتورها استفاده می شود.
۴. نتایج بدست آمده با سایر روش های پیش بینی مقایسه می شود.

بخشی از دیتاست مورد استفاده در مقایسه روش ها در شکل زیر آمده است:

age	job	marital	education	default	housing	loan	contact	month
44	blue-collar	married	basic.4y	unknown	yes	no	cellular	aug
53	technician	married	unknown	no	no	no	cellular	nov
28	management	single	university.degree	no	yes	no	cellular	jun
39	services	married	high.school	no	no	no	cellular	apr
55	retired	married	basic.4y	no	yes	no	cellular	aug
30	management	divorced	basic.4y	no	yes	no	cellular	jul
37	blue-collar	married	basic.4y	no	yes	no	cellular	may
39	blue-collar	divorced	basic.9y	no	yes	no	cellular	may
36	admin.	married	university.degree	no	no	no	cellular	jun
27	blue-collar	single	basic.4y	no	yes	no	cellular	apr
34	housemaid	single	university.degree	no	no	no	telephone	may
41	management	married	university.degree	no	yes	no	cellular	aug

55	management	married	university.degree	no	no	no	cellular	aug

جدول ۱ - بخشی از دیتاست مورد استفاده در مقایسه روش ها

در این مقایسه که با استفاده از رابطه زیر و بر اساس روش Bayes رفتار می شود، در حقیقت تزریق یک رفتار جبرانی است که به هر واحد در یک لایه شبکه عصبی کار می کند. این رفتار مستقل از ورودی به هر لایه است. بایسین تأثیر اعمال تحول در مجموع وزنی را دارد:

$$y = \sigma\,(\mathbf{w}^T\,\mathbf{x} + b)$$

اینکه آیا مجموع وزنی ورودی های لایه باعث ایجاد خروجی می شود یا خیر، با تابع فعال سازی فوق تعیین می شود. به طور کلاسیک سه نوع عملکرد فعال سازی قابل بررسی است:

۱) خطی
۲) آستانه (مرحله ورود به لایه)
۳) سیگموئید (مرحله نرم)

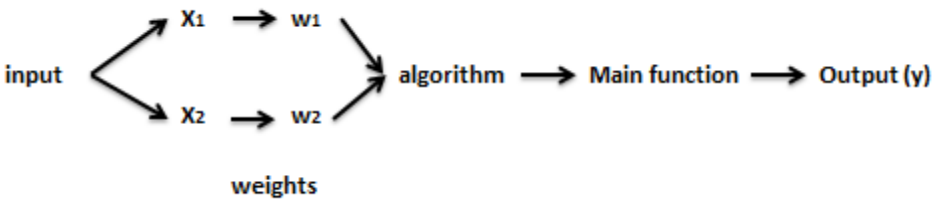
عملکرد سیگموئید حالت خاصی از عملکرد پیش بینی است که با منحنی S شکل آن مشخص می شود. این مکانیزم اغلب مورد استفاده است، زیرا روند غیر خطی بودن را به شبکه اضافه می کند که برای یادگیری وزن به راحتی قابل استخراج است.

بر اساس محدوده خروجی توابع سیگموئید به شکل زیر تقسیم می شود:

الف) سیگموئید لگاریتمی، که از [۰، ۱] عمل می کند.

ب) سیگموئید تماس هذلولی، که در محدوده [۱و۱-] قرار دارد.

نحوه عملکرد الگوریتم فوق و اثر آن بر وزن ها در شکل زیر آمده است:



۴) شکل ۱ عملکرد الگوریتم

برای تجزیه و تحلیل مقایسه ای، نتایج شبیه سازی شبکه عصبی MP با نتایج شبیه سازی امتیازدهی بر اساس رگرسیون خطی و تحلیل تفکیک مقایسه می شود. برای ارزیابی عملکرد یک مدل امتیازدهی از معیارهای ارزیابی عملکرد متفاوت استفاده می شود، مانند

- ماتریس سردرگمی
- نرخ طبقه بندی صحیح
- میانگین خطای مربع

- هزینه طبقه بندی نادرست

برای ارائه نتایج مدل امتیازدهی و ارزیابی عملکرد نیز از دو روش رگرسیون و طبقه بندی استفاده شده است که پیش زمینه اجرایی برای رگرسیون همانا استفاده از خطای مربع میانگین و محاسبه تفاوت بین خروجی و پیش بینی است. همچنین تابع خطا می تواند عملکرد شبکه عصبی را هنگام یادگیری ارزیابی کند. تابع خطا نشانگر چشم و گوش الگوریتم یادگیری است تا با توجه به وضعیت یادگیری فعلی، شبکه کارآمد باشد یا خیر (و بنابراین، چه مقداری باید برای مقادیر وزن آن تنظیم شود).

۶- نتیجه گیری

همانطور که گفته شد، واژه ریسک و ریسک پذیری جز عملکردهای لاینفک سیستم شبکه های عصبی است که با استفاده از روش های مناسب نمی تواند به طور کامل از بین برود، اما قابلیت کاهش مناسبی دارد. یکی از اهداف اصلی در سیستم شبکه های عصبی حفظ یک سیستم صلاحیت دار پایدار و سالم است که از برنامه صلاحیت آغاز می شود و با بسته شدن صلاحیت پایان می یابد. صلاحیت برای اعطای امتیاز، که ضروری ترین شکل ریسک برای پایگاه داده ها است، به شدت با اندازه گیری و مدیریت برتری این روش ارتباط دارد.

عدم قضاوت پایگاه داده ها در مورد درخواست های صلاحیت دار منجر به استفاده ناکارآمد از وجوه می شود. اگر پایگاه داده به یک کلاینت صلاحیت بدهد وانمود می کند که این دسترسی ریسک پذیر نیست، اما اگر مسئله این باشد که هیچ تراکنش بازگشتی وجود نداشته و یا پایگاه داده به دلیل اینکه قصد ندارد مشکل عدم بازپرداخت را در آینده ایجاد کند صلاحیت را به کلاینت اعطا نمی کند. این سناریو باعث ایجاد وضعیتی برای پایگاه داده می شود که ضررهای شدید هدف اصلی آن نیست.

در روش سنتی ارزیابی فاکتور ها، تنها اعطای صلاحیت به تراکنش های است که در تکمیل امتیاز تاخیری ندارند. در این رویه، به دلیل عدم تطابق در سیستم تصمیم گیری صلاحیت دار و عدم ارزیابی هر کلاینت با متغیرهای عینی، به دلیل استفاده از قضاوت ذهنی حرفه ای پذیره نویسی دسترسی، نارضایتی برای تراکنش ایجاد می شود.

با استفاده از این استراتژی، که هدف آن ارزیابی برنامه های صلاحیت دار است که سرعت بیشتر، فرایند های کارآمدتر و دقت قابل ملاحظه ای را می طلبد. سرعت سیستم با ماهیت عینی ارزیابی امتیاز و زمان تصمیم گیری مرتبط است. کارایی و صحت الگوی مورد استفاده در روش ارزیابی فاکتور ها متکی به قدرت و قدسترسی پیش بینی است. امتیاز دهی یکی از مهمترین کاربردهای داده کاوی و مشکلات طبقه بندی است که طی دهه های گذشته توجه بیشتری را به خود جلب کرده است. روش پیشنهادی با استفاده از شبکه های عصبی مصنوعی برای امتیازدهی، طبقه بندی خوبی از تراکنش ها ارائه می دهد. این تحقیق مدل ارزیابی امتیاز را در ارزیابی کاربردهای صلاحیت برای قراردادهای مشارکت ارائه می دهد.

رویکرد شبکه عصبی در امتیازدهی می تواند به مدیر پایگاه داده کمک کند تا ضررهای خود را به میزان قابل توجهی کاهش دهند. تحقیقات نشان می دهد که افزودن معیارهای شراکت بخشی با استفاده از شبکه عصبی MP در امتیاز دهی نتایج بهتری نسبت به امتیاز دهی معمولی از نظر عملکردی به همراه دارد. علاوه بر این، استفاده از شبکه عصبی MP به داشتن یک خطای میانگین خوب و بازگشت مناسب کمک می کند. همچنین مدل شبکه عصبی MP با تکنیک های معمولی مانند رگرسیون خطی و تجزیه و تحلیل تفکیک مقایسه می شود.

۷- منابع

۱. Afolabi, J. A. (2010), "Analysis of loan repayment among small scale farmers in Oyo State, Nigeria", Journal of Social Sciences, Vol. 22, No. 2, pp. 115-119.

۲. Akkoc, S. (2012), "An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: the case of Turkish credit card data", European Journal of Operational Research, Vol. 222, No. 1, pp. 168-178.

۳. Altman, E. I. (1968), "Financial ratios, discriminant analysis and the prediction of corporation bankruptcy", The Journal of Finance, Vol. 23, No. 4, pp. 589-609.

۴. Anderson, R. (2007), The Credit Scoring Toolkit, Oxford University Press, New York.

۵. Bai, C., Shi, B., Liu, F., Sarkis, J. (2019), "Banking credit worthiness: evaluating the complex relationships", Omega, Vol. 83, pp. 26-38.

۶. Bailey, M. (2004), Consumer Credit Quality: Underwriting, Scoring, Fraud Prevention and Collections, White Box Publishing, Bristol.

۷. Bask, A., Merisalo-Ratanen, H., Tinnila, M., Lauraeus, T. (2011), "Towards e-banking: the evolution of business models in financial services", International Journal of Electronic Finance, Vol. 5, No. 4, pp. 333-356.

۸. Bayes, T. (1763), "An essay towards solving a problem in the doctrine of chances", Philosophical Transactions, Vol. 53, pp. 370-418.

۹. Beaver, W. H. (1966), "Financial ratios as predictors of failure", Journal of Accounting Research, Vol. 4, pp. 77-111.

۱۰. Bekhet, H. A., Eletter, S. F. (2014), "Credit risk assessment model for Jordanian commercial banks: neural scoring approach", Review of Development Finance, Vol. 4, No. 1, pp. 20-28.

۱۱. Chen, M. C., Huang, S. H. (2003), "Credit scoring and rejected instances reassigning through evolutionary computation techniques", Expert Systems with Applications, Vol. 24, No. 4, pp. 433-441.

۱۲. Chopde, K., Gosar, P., Kapadia, P., Maheshwari, N., Chawan, P. M. (2012), "A study of classification based credit risk analysis algorithm", International Journal of Engineering and Advanced Technology, Vol. 1, No. 4, pp. 142-144.

۱۳. Chorowski, J., Wang, J., Zurada, M. J. (2014), "Review and comparison of SVM- and ELM-based classifiers", Neurocomputing, Vol. 128, pp. 506-516.

۱۴. Crook, J. N., Edelman, D. B., Thomas, L. C. (2007), "Recent developments in consumer credit risk assessment", European Journal of Operational Research, Vol. 183, No. 3, pp.1447-1465.

۱۵. Curran, K., Orr, J. (2011), "Integrating geolocation into electronic finance applications for additional security", International Journal of Electronic Finance, Vol. 5, No. 3, pp. 272-285.

۱۶. Danenas, P., Garsva, G. (2012), "Credit risk evaluation modelling using evolutionary linear SVM classifiers and sliding window approach", Procedia Computer Science, Vol. 9, pp. 1324-1333.

۱۷. Danenas, P., Grasva, G. (2015), "Selection of support vector machines based classifiers for credit risk domain", Expert System with Application, Vol. 42, No. 6, pp. 3194-3204.

۱۸. Dhaiya, S., Handa, S. S., Singh, N. (2016), "Impact of bagging on MLP classifier for credit evaluation", in Hoda, M. N. (ed.), 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), 16-18 March, IEEE, New Delhi, pp. 3794-3799.

۱۹. Dinh, T. H. T., Kleimeier, S. (2007), "A credit scoring model for Vietnam's retail banking market", International Review of Financial Analysis, Vol. 16 No. 5, pp. 471-495.

۲۰. Durand, D. (1941), Risk Elements in Consumer Installment Financing, NBER, New York.

۲۱. Emel, A. B., Oral, M., Reisman, A., Yolalan, R. (2003), "A credit scoring approach for the commercial banking sector", Socio-Economic Planning Sciences, Vol. 37, No. 2, pp. 103-123.

۲۲. Feelders, A. J. (2000), "Credit scoring and reject inference with mixture models", Intelligent Systems in Accounting, Finance & Management, Vol. 37, No. 1, pp. 1-8.

۲۳. Finlay, S. (2010), "Credit scoring for profitability objectives", European Journal of Operational Research, Vol. 202, No. 2, pp. 528-537.

۲۴. Hájek, P. (2011), "Municipal credit rating modelling by neural networks", Decision Support Systems, Vol. 51, No. 1, pp. 108-118.

۲۵. Hájek, P., Michalak, K. (2013), "Feature selection in corporate credit rating prediction", Knowledge-Based Systems, Vol. 51, pp. 72-84.

۲۶. Hamid, A. J., Ahmed, T. M. (2016), "Developing prediction model of loan risk in banks using data mining", Machine Learning and Applications: An International Journal, Vol. 3, No. 1, pp. 1-9.

۲۷. Harris, T. (2013), "Quantitative credit risk assessment using support vector machines: broad versus narrow default definitions", Expert Systems with Applications, Vol. 40, No. 11, pp. 4404-4413.

۲۸. Hearst, M., Dumais, S. T., Osuna, E., Platt, J., Scholkopf, B. (1998), "Support vector machines", IEEE Intelligent Systems and their Applications, Vol. 13, No. 4, pp. 18-28.

۲۹. Huang, B. G., Zhu, Q. Y., Siew, C. K. (2006a), "Extreme learning machine: theory and application", Neurocomputing, Vol. 70, No. 1-3, pp. 489-501.

۳۰. Huang, B. G., Chen, L., Siew, C. K. (2006b), "Universal approximation using incremental constructive feedforward networks with random hidden nodes", IEEE Transaction on Neural Networks, Vol. 17, No. 4, pp. 879-892.

۳۱. Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., Wu, S. (2004), "Credit rating analysis with support vector machines and neural network: a market comparative study", Decision Support System, Vol. 37, No. 4, pp. 543-558.

۳۲. Hui, X., Gang, Y. S. (2011), "Using clustering-based bagging ensemble for credit scoring", 2011 International Conference on Business Management and Electronic Information, 13-15 May, IEEE, Guangzhou, pp. 369-371.

۳۳. Hwang, R. C. (2013), "Forecasting credit ratings with the varying-coefficient model", Quantitative Finance, Vol. 13, No. 2, pp. 1-19.

۳۴. Islam, J. M., Wu, J. Q. M., Ahmadi, M. (2007), "Investigating the performance of Naïve-Baysed and K- Nearest NEIGHBOR Classifiers", in Na, Y. J., Ko, F. I. S., Lee, J., Truex III, D. P., Sohn,

Sungwon, Nguyen, N. T. (Eds.), International conferences on convergence information technology ICCIT 2007, 21-23 November, IEEE, Gyeongju, pp. 1541-1546.

۳۵. Jain, A., Kumar, A. M. (2007), "Hybrid neural network models for hydrologic time series forecasting", Applied Soft Computing, Vol. 7, No. 2, pp. 585-592.

۳۶. Kao, L. J., Chiu, C. C., Chiu, F. Y. (2012), "A Bayesian latent variable model with classification and regression tree approach for behaviour and credit scoring", Knowledge-Based Systems, Vol. 36, pp. 245-252.

۳۷. Karan, M. B., Ulucan, A., Kaya, M. (2013), "Credit risk estimation using payment history data: a comparative study of Turkish retail stores", Central European Journal of Operations Research, Vol. 21, No. 2, pp. 479-494.

۳۸. Karminsky, A. M., Khromova, E. (2016), "Extended modelling of banks' credit ratings", Procedia Computer Science, Vol. 91, pp. 201-210.

۳۹. Kim, K. J., Ahn, H. (2012), "A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach", Computers & Operations Research, Vol. 39, No. 8, pp. 1800-1811.

۴۰. Kinda, O., Achonu, A. (2012), "Building a credit scoring model for the savings and credit mutual of the Potou Zone (MECZOP)/Senegal", The Journal of Sustainable Development, Vol. 7, No. 1, pp. 17-31.

۴۱. Laplace, P.-S. (1812), Théorie Analytique des Probabilités (Analytical Theory of Probabilities), Courcier, Paris.

۴۲. Lewis, E. M. (1992), An introduction to credit Scoring, Athena Publishing, San Rafael.

۴۳. Malhotra, R., Malhotra, D. K. (2003), "Evaluating consumer loans using neural networks", Omega, Vol. 31, No. 2, pp. 83-89.

۴۴. McAdams, L. (1980), "How to anticipate utility bond rating changes", The Journal of Portfolio Management, Vol. 7, No. 1, pp. 56-60.

۴۵. Nisha, N. (2017), "An empirical study of the balanced scorecard model: evidence from Bangladesh", International Journal of Information Systems in the Service Sector, Vol. 9, No. 1, pp. 68-84.

۴۶. Olafsson, S., Li, X., Wu, S. (2008), "Operations research and data mining", European Journal of Operational Research, Vol. 187, No. 3, pp. 1429-1448.

۴۷. Pandey, T. N., Jagadev, A. K., Choudhury, D., Dehuri, S. (2013), "Machine learning-based classifiers ensemble for credit risk assessment", International Journal of Electronic Finance, Vol. 7, No. 3-4, pp. 227–249.

۴۸. Pandey, T. N., Jagadev, A. K., Mohapatra, S. K., Dehuri, S. (2017), "Credit risk analysis using machine learning classifiers", 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing ICECDS, IEEE, Chennai, pp. 1850-1854.

۴۹. Petropoulos, A., Chatzis, S. P., Xanthopoulos, S. (2016), "A novel corporate credit rating system based on Student's-t hidden Markov models", Expert Systems with Applications, Vol. 53, pp. 87-105.

۵۰. Pinches, G. E., Mingo, K. A. (1973), "A multivariate analysis of industrial bond ratings", The Journal of Finance, Vol. 28, No. 1, pp. 1-18.

۵۱. Pishbahar, E., Ghahremanzadeh, M., Ainollahi, M., Ferdowsi, R. (2015), "Factors influencing agricultural credits repayment performance among farmers in East Azarbaijan province of Iran", Journal of Agricultural Science and Technology, Vol. 17, No. 5, pp. 1095-1101.

۵۲. Saberi, M., Mirtalaie, M. S., Hussain, F. K., Azadeh, A., Hussain, O. K., Ashjari, B. (2013), "A granular computing-based approach to credit scoring modelling", Neurocomputing, Vol. 122, pp. 100-115.

۵۳. Shi, B., Chen, N., Wang, J. (2016a), "A credit rating model of microfinance based on fuzzy cluster analysis and fuzzy pattern recognition: empirical evidence from Chinese 2,157 small private businesses", Journal of Intelligent & Fuzzy Systems, Vol. 31, No. 6, pp. 3095-3102.

۵۴. Shi, B., Wang, J., Qi, J., Cheng, Y. (2015), "A novel imbalanced data classification approach based on logistic regression and Fisher discriminant", Mathematical Problems in Engineering, Vol. 2015, Article 945359.

۵۵. Shi, B., Yang, H., Wang, J., Zhao, J. (2016b), "City green economy evaluation: empirical evidence from 15 sub-provincial cities in China", Sustainability, Vol. 8, No. 6, pp. 1-39.

۵۶. Sousa, M. R., Gama, J., Brandão, E. (2016), "A new dynamic modelling framework for credit risk assessment", Expert Systems with Applications, Vol. 45, No. 1, pp. 341-351.

۵۷. Sudhakar, M., Reddy, K. (2014), "Credit evaluation model of loan proposals for banks using data mining", International Journal of Latest Research in Science and Technology, pp. 126-131.

۵۸. Thomas, L. C., Edelman, D., Crook, J. N. (2002), Credit Scoring and its Applications, Society for Industrial and Applied Mathematics, Philadelphia.

۵۹. Van-Gool, J., Verbeke, W., Sercu, P., Baesens, B. (2012), "Credit scoring for microfinance: is it worth it? International Journal of Finance & Economics, Vol. 17, No. 2, pp. 103-123.

۶۰. Wang, Y., Wang, S., Lai, K. K. (2005), "A new fuzzy support vector machine to evaluate credit risk", IEEE Transaction on Fuzzy System, Vol. 13, No. 6, pp. 820-831.

۶۱. Zhang, Y., Chi, G. (2018), "A credit rating model based on a customer number bell-shaped distribution", Management Decision, Vol. 56, No. 5, pp. 987-1007.

۶۲. Zhou, H., Lan, Y., Soh, Y. C., Huang, G. B., Zhang, R. (2012), "Credit risk evaluation using extreme learning machine", 2012 IEEE International Conference on Systems, Man, and Cybernetics SMC, 14-17 October, IEEE, Seoul, pp. 1064-1069.

۶۳. A Review of Algorithms for Credit Risk Analysis, Salihu, Armend; Shehu, Visar, 2020

۶۴. Credit scoring models for the microfinance industry using neural networks, Antonio Blanco, Rafael Pino-Mejías, Juan Lara, Salvador Rayo, 2013

65. BRACHMAN, R.J., McGUINESS, D.L., Knowledge Representation, Connectionism and Conceptual Retrieval, Proc. SIGIR'88, Grenoble, June 1988

Software Engineering, 18, 7, pp. 551-562, 1992

68. HEILEMAN, G., et al., A general framework for concurrent simulation of neural networks models, IEEE Trans.

66. FISHMAN, D.H., et al., Overview of the IRIS Database Management System, in Object Oriented Concepts,
Databases and Applications (eds. W. Kim and F.H. Lochovsky), Addison -Wesley, 1989

67. GOERZ, G., Database support for knowledge representation, in [11]

69. LUDWIG, T., WALTER, B., Database concepts for the support of knowledge-based systems, in

70. PAO, Y.-H., SOBAJIC, D.J., Neural networks and Knowledge Engineering, IEEE Knowledge and Data Engineering, 3, 2, pp. 185 - 192, June 1991

71. SCHIKUTA, E., The role of neural networks in knowledge based systems, Proc. Int. Symp. on nonlinear theory and applications, Hawaii, 1993

72. STONEBRAKER, M., et. al., Extending a Database system with Procedures, ACM Transactions Systems, Sept. 1987 on Database

73. P.Manivannan, K.Balasubramanian, D.Prabha, Deep neural network: Recognize Data Management of Artificial Intelligence in Retail ISSN: 2278-3075 (Online), Volume-8 Issue-10, August 2019, DOI: 10.35940/ijitee.J9779.0881019

74. M. Yazdani-Asrami Roadmap on artificial intelligence and big data techniques for superconductivity To cite this article: Mohammad Yazdani-Asrami et al 2023, https://doi.org/10.1088/1361-6668/acbb34