مجلـه‌ی علمی پژوهش در علوم رایانه ، شماره ۱۱ ، پاییز ۱۳۹۷، ص ۷۸- ۶۶

ISSN: 2745-3673
http://www.Rcsj.ir

مجلـه‌ی علمی پژوهش در علوم رایانه (سال سوم)
شماره ۱۱ / پاییز ۱۳۹۷ / ص ۷۸- ۶۶

اولویت‌بندی مکانیزم‌های داده‌کاوی در بانکداری الکترونیک با استفاده از روش تاپسیس فازی

محمد عرفان میرزایی [۱]، محمد خلیلی [۲]

[۱] گروه کامپیوتر، دانشکده فنی مهندسی، واحد خمین، دانشگاه آزاد اسلامی، خمین، ایران.

[۲] گروه کامپیوتر، دانشکده فنی مهندسی، واحد خمین، دانشگاه آزاد اسلامی، خمین، ایران.

نام نویسنده مسئول:
محمد خلیلی

چکیده

بانکداری الکترونیک نیازمند بررسی و کشف اطلاعات گوناگونی از میان انبوه اطلاعات انباشته خود بر اساس مکانیزم‌های داده‌کاوی هستند. با افزایش خدمات بانک‌ها در اینترنت و رشد تراکنش‌های برخط توسط مشتریان، میزان بروز جرایم در صنعت بانکداری الکترونیک نیز به سرعت در حال رشد است؛ در این راستا بانک ها می توانند با بکارگیری روش های داده کاوی جدید بر اساس روش تاپسیس فازی به کشف تقلب و به پیشگری و حذف فرایندهای حامی این گونه اعمال بپردازند. لذا هدف از پژوهش حاضر اولویت‌بندی مکانیزم‌های داده‌کاوی در بانکداری الکترونیک با استفاده از تاپسیس فازی است. روش پژوهش حاضر به‌صورت تحلیلی توصیفی و با استفاده از پرسشنامه انجام‌گرفته است. در این پژوهش تمام مقالاتی که از الگوریتم‌های CART،CHAID، C5، SVM، Naive Bayes، Discriminant، QUEST، ANN،Logistic–R و K-mens در بانکداری الکترونیک استفاده کرده‌اند، جمع‌آوری‌شده است. با استفاده از پرسشنامه، نظرات ۵ نفر از کارشناسان و اساتید، جمع‌آوری شد. سپس داده‌ها، با استفاده از معیارهای «نظرات اساتید»، «نظرات کارشناسان»، «تعداد مقالات چاپ شده» و «جدید بودن روش» با روش تاپسیس فازی، الگوریتم‌های داده‌کاوی در بانکداری الکترونیک، اولویت‌بندی شد. نتایج نشان داد که، الگوریتم Naive Bayes در طبقه زیاد قرارگرفته است و الگوریتم‌های SVM و ANN در طبقه متوسط قرارگرفته و سایر الگوریتم‌ها در سطوح پایین‌تر دسته‌بندی شده‌اند. لذا بامطالعه پیشینه مقالات انجام‌گرفته و همچنین نظرات خبرگان، الگوریتم Naive Bayes به دلیل کاربرد زیاد و جدید بودن آن نسبت به همه الگوریتم‌ها در بانکداری الکترونیک برتری دارد. همچنین بر اساس تحقیقات صورت گرفته در مقایسه الگوریتم‌های SVM و ANN، نتایج نشان می‌دهد که الگوریتم SVM نسبت به الگوریتم ANN برتری نسبی دارد.

واژگان کلیدی: داده‌کاوی، تاپسیس فازی، بانکداری الکترونیک، Naive Bayes..

مجلـه‌ی علمی پژوهش در علوم رایانه ، شماره ۱۱ ، پاییز ۱۳۹۷، ص ۷۸- ۶۶

ISSN: 2745-3673
http://www.Rcsj.ir

مقدمه

درگذشته، عموماً استخراج اطلاعات مفید از داده‌های ثبت‌شده، به‌صورت دستی و بر عهده تحلیلگران بوده است. با توجه به اینکه تجزیه‌وتحلیل دستی داده‌ها بسیار کند و گران بوده و هرروز بر پیچیدگی و حجم داده‌ها افزوده می‌شد، تحلیل‌های دستی به سمت تحلیل‌های غیرمستقیم خودکار و استفاده از روش‌های کامپیوتری حرکت کرده است و نیاز مبرمی مبنی بر استفاده از فناوری‌های جدید و ابزارهای خودکار به وجود آمد تا به‌صورت هوشمند، حجم زیاد داده را به اطلاعات و دانش تبدیل کند. در این شرایط ضروری است از فناوری اطلاعات برای استفاده از این دانش بهره گرفت و داده‌کاوی پاسخی مناسب برای استخراج این دارایی است. به‌کارگیری تفکرات و روش‌های مکانیزم‌های داده‌کاوی در بانکداری الکترونیکی، به عاملین این بانک‌ها برای دریافت دانش واقعاً ارزشمند از مقادیر گسترده اطلاعات و تصمیم‌گیری‌های آنان کمک می‌کند. داده‌های مربوط به تراکنش مشتریان در سیستم بانکداری الکترونیکی، اغلب دارای اطلاعات مفیدی در خصوص نحوه تعامل مشتریان با بانک‌ها و استفاده از سرویس‌های مختلف آن‌ها است. کاوش در این داده‌ها به‌منظور بررسی الگوهای رفتاری مشتریان امری ضروری است که با استفاده از علم داده‌کاوی می‌توان رفتار مشتریان را تحلیل نمود و با استفاده از نتایج حاصله آن‌ها را دسته‌بندی کرد و گروه مشتریان باارزش‌تر را جهت برنامه‌ریزی و سیاست‌گذاری‌های آینده مشخص نمود با پیشرفت‌های اخیر در تکنولوژی کامپیوتر، مقادیر زیاد داده‌ها باید قابل جمع‌آوری و ذخیره‌سازی باشند، اما تمامی این داده‌ها زمانی کارآمدتر خواهند بود که تجزیه‌وتحلیل شوند و برخی از وابستگی‌ها و روابط آن‌ها شناسایی شوند. این امر به‌وسیله مکانیزم‌های داده‌کاوی می‌تواند انجام شود. در دنیای دیجیتال امروزی، داده‌ها مانند هر چیز دیگری عمدتاً در قالب فایل‌های ساختار نیافته‌ای افزایش می‌یابند. تکنیک‌های داده‌کاوی بسیاری برای کاوش این فایل‌ها وجود دارند. اما کاوش این داده‌های نیمه ساختاریافته، امری چالش‌برانگیز است[۱].

با افزایش خدمات بانک ها در اینترنت و رشد تراکنش های برخط توسط مشتریان، میزان بروز جرایم در صنعت بانکداری الکترونیک نیز به سرعت در حال رشد است، به طوریکه آهنگ رشد جرائم برخط بین ۸ الی ۹ درصد در سال تخمین زده می شود. در این راستا بانک ها می توانند با بکارگیری روش های داده کاوی جدید بر اساس روش تاپسیس فازی به کشف تقلب و به پیشگری و حذف فرایندهای حامی این گونه اعمال بپردازند. ازآنجایی‌که هیچ ابزاری برای داده‌کاوی کامل نیست، در این تحقیق به بررسی مکانیزم‌های داده‌کاوی در بانکداری الکترونیک و با استفاده از روش تاپسیس فازی اولویت‌بندی این مکانیزم‌ها انجام خواهد شد.

مکانیزم‌های استفاده‌شده در حوزه داده‌کاوی در بانکداری الکترونیک

درخت تصمیم CART

درخت تصمیم‌گیری یک روش تقسیم‌بندی بازگشتی «باینری»[1] است. داده‌ها در این الگوریتم به‌صورت خام استفاده می‌شوند و هیچ‌گونه پاک‌سازی نه نیاز است نه پیشنهاد می‌شود. درختان بدون استفاده از یک قانون متوقف کننده به رشد حداکثری خود می‌رسند و سپس اصلاح می‌شوند. اصلاح تا ریشه ادامه دارد و با اصلاح پیچیدگی کار بالا می‌رود. قسمت بعدی برای اصلاح بخشی است که کمترین کمک را به کارکرد کلی درخت در پردازش اطلاعات می‌کند. هدف مکانیزم درخت طبقه‌بندی رگرسیون تولید، تنها یک درخت نیست بلکه تولید یک سری درختان اصلاح شده تودرتو است که همه آن درختان بهینه داوطلب هستند. درخت با اندازه مناسب یا «درخت درست» به‌وسیله ارزش‌گذاری عملکرد پیش‌گویانه هر درخت در توالی اصلاح، شناخته می‌شود. درخت طبقه‌بندی رگرسیون هیچ‌گونه اندازه عملکرد داخلی برای انتخاب درخت بر اساس پردازش اطلاعات پیشنهاد نمی‌کند زیرا این اندازه‌ها قابل اطمینان نیستند. به‌جای آن عملکرد درخت در آزمایش داده‌های جداگانه (یا از طریق تأیید میانه) اندازه‌گیری می‌شود و انتخاب درخت تنها پس از ارزشیابی آزمایش داده‌ها صورت می‌گیرد. اگر هیچ آزمایش داده‌ای وجود نداشته باشد و تأییدیه میانی انجام نشده باشد درخت طبقه‌بندی رگرسیون نمی‌تواند بهترین درخت توالی را مشخص کند[۲].

«سوساک و بنسیک»[2] (۲۰۰۴) در مقاله‌ای به نمره اعتبار کسب‌وکار کوچک: مقایسه رگرسیون لجستیک، شبکه عصبی و مدل درخت تصمیم‌گیری، پرداختند. در این مقاله مدل‌هایی برای ارزیابی اعتبارهای کسب‌وکار کوچک که توسط رگرسیون لجستیک، شبکه‌های عصبی و درخت تصمیم‌گیری رگرسیون در یک مجموعه داده بانکی کرواسی، مقایسه شده است. مدل‌های به‌دست‌آمده از هر سه روش برآورد شد، سپس عملکرد آن‌ها مقایسه شدند. تفاوت قابل ملاحظه‌ای بین بهترین مدل شبکه عصبی، مدل درخت تصمیم و مدل رگرسیون لجستیک وجود دارد. موفق‌ترین مدل شبکه عصبی با الگوریتم احتمالی به‌دست‌آمده است. بهترین مدل استخراج مهم‌ترین ویژگی‌های کسب‌وکار کوچک کسب‌وکار از داده‌های مشاهده شده است[۳].

---

[1] .Binery
[2] .Susac & Bencik

ISSN: 2745-3673
http://www.Rcsj.ir

الگوریتمCHAID

این الگوریتم یک روش درخت تصمیم‌گیری است که بر اساس آزمون معنی‌دار تنظیم شده «آزمایش بونفرونی[۳]» است. این تکنیک در آفریقای جنوبی توسعه یافت و توسط «گوردون واکس»[۴] منتشر شد که پایان‌نامه دکترای خود را در این زمینه به پایان رسانده است. CHAID را می‌توان برای پیش‌بینی استفاده کرد و همچنین برای طبقه‌بندی و تشخیص تعامل بین متغیرها بکار برد. CHAID بر اساس فرمت رسمی «AID»[۵] ایالات‌متحده و «THAID» است. در عمل، CHAID اغلب در زمینه بازاریابی مستقیم برای انتخاب گروهی از مصرف‌کنندگان استفاده می‌شود و پیش‌بینی می‌کند که چگونه آن‌ها به برخی از متغیرها بر سایر متغیرها تأثیر می‌گذارد. مزایای CHAID مانند دیگر درخت‌های تصمیم‌گیری این است که خروجی آن بسیار بصری و قابل تفسیر است. یکی از مزیت‌های مهم CHAID همانند رگرسیون چندگانه است که غیر پارامتری است[۴]. کوثری و همکاران (۱۳۹۲) در پژوهشی به ارائه مدلی جهت کشف رفتارهای مشکوک در بانکداری الکترونیکی با استفاده الگوریتم‌های درخت تصمیم‌گیری، پرداختند. در این پژوهش، ابتدا متغیرهای مؤثر در تولید قوانین رفتاری تعیین‌شده است و درنهایت، روند چهار الگوریتم CHAID، ex_ CHAID، C4.5 و C5.0 مورد مقایسه قرارگرفته است. نتایج پژوهش نشان می‌دهد که الگوریتمCHAID می‌تواند به‌عنوان روش ماشینی مطمئن جهت کشف الگوهای مشکوک موجود روی تراکنش‌های بانک محسوب شود[۵].

الگوریتمC5

الگوریتم C5 یک نوع درخت تصمیم‌گیری تک متغیره و بهبودیافته الگوریتم C4.5 است. این الگوریتم مشابه با درخت طبقه‌بندی رگرسیون ابتدا درختی تقریباً پر ایجاد می‌کند ولی استراتژی هرس آن کامل متفاوت است. این الگوریتم دسته‌بندی را با تقسیم کردن داده‌ها به زیرمجموعه‌هایی که شامل رکوردهای همگن‌تر از والد خود هستند انجام می‌دهد. در C5 تقسیم کردن نمونه‌ها بر اساس فیلدی که بیشترین بهره اطلاعات را دارد صورت می‌گیرد. این الگوریتم روشی افزایشی از هرس کردن درخت را به کار می‌گیرد تا خطای طبقه‌بندی کردن ناشی از نویز یا جزئیات خیلی زیاد را در داده‌های آموزشی کاهش دهد. هرس کردن با جایگزینی گره داخلی با گره برگ رخ می‌دهد که بدان وسیله درصد یا میزان خطا کاهش می‌یابد[۶]. لنگری و همکاران (۲۰۱۴) در مقاله‌ای به‌کارگیری الگوریتم‌های درخت تصمیم‌گیری جهت کشف رفتارهای مشکوک در بانکداری اینترنتی پرداختند. در این مقاله، ابتدا متغیرهای مؤثر در تولید قوانین رفتاری تعیین‌شده است و درنهایت، روند چهار الگوریتم CHAID، ex_ CHAID ، C4.5 و C5.0 مورد مقایسه قرارگرفته است. نتایج پژوهش نشان می‌دهد که الگوریتم C5.0 با دقت ۹۱ درصد می‌تواند به‌عنوان روش ماشینی مطمئن جهت کشف الگوهای مشکوک موجود روی تراکنش‌های بانک محسوب شود[۷].

الگوریتم SVM

الگوریتم ماشین بردار پشتیبان اولیه توسط «ولادیمر واپنیک»[۶] ابداع شد. یکی از روش‌های «یادگیری با نظارت»[۷] است که از آن برای «طبقه‌بندی»[۸] و رگرسیون استفاده می‌کنند. این روش ازجمله روش‌های نسبتاً جدیدی است که در سال‌های اخیر کارایی خوبی نسبت به روش‌های قدیمی‌تر برای طبقه‌بندی ازجمله «شبکه‌های عصبی پرسپترون»[۹] نشان داده است. مبنای‌کاری دسته‌بندی کننده ماشین بردار پشتیبان دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌کنیم خطی را انتخاب کنیم که حاشیه اطمینان بیشتری داشته باشد. الگوریتم ماشین بردار پشتیبان، جز الگوریتم‌های تشخیص الگو دسته‌بندی می‌شود. از الگوریتم ماشین بردار پشتیبان، در هرجایی که نیاز به تشخیص الگو یا دسته‌بندی اشیا در کلاس‌های خاص باشد می‌توان استفاده کرد[۸]. شربیانی مقدوری و همکاران (۱۳۸۸) در مقاله‌ای به طبقه‌بندی متقاضیان تسهیلات اعتباری بانک‌ها با استفاده از تکنیک ماشین بردار پشتیبان پرداختند. در مقاله حاضر مدل طبقه‌بندی مبتنی بر ماشین بردار پشتیبان با رویکرد هوش مصنوعی، به‌منظور پیش‌بینی عملکرد مالی مشتریان حقوقی بانک‌ها ارائه گردیده است. درواقع، در این نوشتار ماشین بردار پشتیبان به همراه دیگر مکانیزم‌ها ازجمله تکنیک‌های جستجوی گریدی جهت طبقه‌بندی

---

[3] . Bonfrey Test
[4] .Gordon Wax
[5] . Interaction Automatic Detection
[6] . Vladimir Vapnik
[7] . Supervised learning
[8] . Classification
[9] . Perceptron neural networks

ISSN: 2745-3673
http://www.Rcsj.ir

متقاضیان تسهیلات اعتباری بانکی و افزایش کارایی مدل استفاده‌شده است. نتایج، حاکی از افزایش صحت طبقه‌بندی است و نشان می‌دهد که ماشین بردار پشتیبان در مقایسه با دیگر مدل‌های طبقه‌بندی دارای صحت بیشتری است[۹].

الگوریتم «نایوبیز»[10]

یک الگوریتم یادگیری ساده است که از قاعده بیز به همراه فرض محکمی که صفات با توجه به کلاس ازنظر شرطی مستقل هستند، استفاده می‌کند. اگرچه این فرض استقلال در عمل اغلب نقض می‌شود، بااین‌وجود، اغلب شبکه‌های بیزی صحت دسته‌بندی قابل رقابتی ارائه می‌کند. این ویژگی به همراه کارایی محاسباتی و ویژگی‌های مطلوب بسیار دیگری، سبب شده نایوبیز در عمل به‌صورت گسترده مورداستفاده قرار بگیرد. «کریچن»[11] (۲۰۱۷) در پژوهشی به استفاده از یک روش طبقه‌بندی بیزی برای ارزیابی ریسک وام شواهدی از بانک تجاری تونس، پرداخت. نتایج آزمون اعتبار سنجی نشان می‌دهد که میزان طبقه‌بندی خوب ۵۸,۶۶ درصد است، با این وجود، خطاهای نوع I و II نسبتاً بالا به ترتیب ۴۲,۴۲ و ۴۰,۴۷ درصد می‌باشند. منحنی مشخصه عامل گیرنده برای ارزیابی عملکرد مدل طراحی شده است. نتیجه نشان می‌دهد که سطح زیر معیار منحنی، به ترتیب ۶۹ درصد است و نشان از برتری روش طبقه‌بندی بیزین دارد[۱۰].

الگوریتم «تفکیک‌کننده»[12]

«تشخیص خطی فیشر»[13] روش‌های آماری هستند که ازجمله در یادگیری ماشین و بازشناخت الگو برای پیدا کردن ترکیب خطی خصوصیاتی که به بهترین صورت دو یا چند کلاس از اشیا را از هم جدا می‌کند، استفاده می‌شوند. آنالیز تشخیصی خطی بسیار به تحلیل واریانس و تحلیل رگرسیونی نزدیک است، در هر سه این روش‌های آماری متغیر وابسته به‌صورت یک ترکیب خطی از متغیرهای دیگر مدل‌سازی می‌شود. بااین‌حال دو روش آخر متغیر وابسته را از نوع فاصله‌ای در نظر می‌گیرند درحالی‌که آنالیز افتراقی خطی برای متغیرهای وابسته اسمی یا رتبه‌ای به کار می‌رود. ازاین‌رو آنالیز افتراقی خطی به رگرسیون لجستیک شباهت بیشتری دارد. آنالیز تشخیصی خطی همچنین با تحلیل مؤلفه‌های اصلی و تحلیل عاملی هم شباهت دارد، هر دوی این روش‌های آماری برای ترکیب خطی متغیرها به شکلی که داده را به بهترین نحو توضیح بدهد به کار می‌روند یک کاربرد عمده هر دوی این روش‌ها، کاستن تعداد بعدهای داده است. بااین‌حال این روش‌ها تفاوت عمده‌ای باهم دارند: در آنالیز افتراقی خطی، تفاوت کلاس‌ها مدل‌سازی می‌شود درحالی‌که در تحلیل مؤلفه‌های اصلی تفاوت کلاس‌ها نادیده گرفته می‌شود[۱۱]. «آنته و آنا»[14] (۲۰۱۳) در مقاله‌ای به تجزیه‌وتحلیل الگوریتم تفکیک‌کننده برای سطح سوددهی بانک پرداختند. در مجموع، تحلیل تفکیک‌کننده به‌عنوان یک روش آماری مناسب برای حل مسئله تحقیق ارائه‌شده در سوددهی و جلوگیری از ورشکستگی، رتبه‌بندی اعتباری یا مسائل پیش‌فرض در امور مالی است[۱۲].

QUEST

این الگوریتم توسط «لو و شی»[15] برای متغیرهای پاسخ اسمی طراحی شد. درخت رده‌بندی حاصل از این الگوریتم نظیر مدل درخت طبقه‌بندی رگرسیون دارای تقسیمات دوتایی بوده و ملاک تصمیم برای انتخاب متغیرها با استفاده از مقدار P-Value مربوط به آماره F آزمون «آنوا»[16] برای متغیرهای کمی و P-Value آماره کای- دو مربوط به جداول توافقی برای متغیرهای کیفی صورت می‌پذیرد. این الگوریتم با توجه به این‌که از مقدار P-Value برای تصمیم‌گیری استفاده می‌نماید، موجب تشکیل درختی ناریب از متغیرها می‌گردد. این الگوریتم ضمن حفظ دقت برآورد در مدل درخت طبقه‌بندی رگرسیون، از سرعت بالاتری در معرفی یک درخت رده‌بندی نسبت به آن برخوردار است. «کیزیس و همکاران»[17] (۲۰۱۵) در پژوهشی به تلاش برای پایداری بانکداری الکترونیکی با استفاده از الگوریتم Quest پرداختند. نتایج نشان داد که استفاده از الگوریتم QUEST می‌تواند در کاهش ارزش وثیقه وام و همچنین پایداری بانک مؤثر بوده و نسبت به دیگر الگوریتم‌ها از صحت بیشتری برخوردار است[۱۳].

---

[10] . naïve Bayes
[11] .Krichen
[12] . Discriminant
[13] . Fisher's Linear Detection
[14] .Ante & Ana
[15] .Lu & Shi
[16] . ANOVA
[17] .Kizis et al

ISSN: 2745-3673
http://www.Rcsj.ir

الگوریتم شبکه‌های عصبی مصنوعی

شبکه‌های عصبی سیستم‌ها و روش‌های محاسباتی نوین برای یادگیری ماشینی، نمایش دانش و در انتها اعمال دانش به‌دست‌آمده در جهت پیش‌بینی پاسخ‌های خروجی از سامانه‌های پیچیده هستند. ایده اصلی این‌گونه شبکه‌ها تا حدودی الهام گرفته از شیوه کارکرد سیستم عصبی زیستی برای پردازش داده‌ها و اطلاعات به‌منظور یادگیری و ایجاد دانش قرار دارد. عنصر کلیدی این ایده، ایجاد ساختارهایی جدید برای سامانه پردازش اطلاعات است. این سیستم از شمار زیادی عناصر پردازشی فوق‌العاده به‌هم‌پیوسته با نام «نرون» تشکیل شده که برای حل یک مسئله باهم هماهنگ عمل می‌کنند و توسط «سیناپس‌ها»[18] (ارتباطات الکترومغناطیسی) اطلاعات را منتقل می‌کنند. در این شبکه‌ها اگر یک سلول آسیب ببیند بقیه سلول‌ها می‌توانند نبود آن را جبران کرده و نیز در بازسازی آن سهیم باشند. این شبکه‌ها قادر به یادگیری‌اند. مثلاً با اعمال سوزش به سلول‌های عصبی لامسه، سلول‌ها یاد می‌گیرند که به‌طرف جسم داغ نروند و با این الگوریتم سیستم می‌آموزد که خطای خود را اصلاح کند. یادگیری در این سیستم‌ها به‌صورت تطبیقی صورت می‌گیرد، یعنی با استفاده از مثال‌ها وزن سیناپس‌ها به‌گونه‌ای تغییر می‌کند که در صورت دادن ورودی‌های جدید، سیستم پاسخ درستی تولید کند. لطفی (۱۳۸۶) با استفاده از داده‌های مربوط به ۶۴۰ مشتری حقوقی بانک کشاورزی، اقدام به طراحی الگویی به‌منظور رتبه‌بندی اعتباری با بهره‌گیری از ۳ مدل لاجیت، پروبیت و شبکه‌های عصبی شده است. نتایج حاصل تحقیق نشان می‌دهد که از بین ۳ مدل مورد بررسی، مدل لاجیت پیش‌بینی صحیح‌تری از موارد نکول و عدم نکول مشتریان نسبت به دیگر مدل‌ها داشته است[۱۴].

الگوریتم رگرسیون لجستیک

رگرسیون لجستیک یک مدل آماری رگرسیون برای متغیرهای وابسته دوسویی مانند بیماری یا سلامت، مرگ یا زندگی است. این مدل را می‌توان به‌عنوان مدل خطی تعمیم‌یافته‌ای که از تابع لوجیت به‌عنوان تابع پیوند استفاده می‌کند و خطایش از توزیع چندجمله‌ای پیروی می‌کند، به‌حساب آورد. منظور از دو سویی بودن، رخ داد یک واقعه تصادفی در دو موقعیت ممکنه است. به‌عنوان مثال خرید یا عدم خرید، ثبت‌نام یا عدم ثبت‌نام، ورشکسته شدن یا ورشکسته نشدن و ... متغیرهایی هستند که فقط دارای دو موقعیت هستند و مجموع احتمال هر یک آن‌ها درنهایت یک خواهد شد. کاربرد این روش عمدتاً در ابتدای ظهور در مورد کاربردهای پزشکی برای احتمال وقوع یک بیماری مورداستفاده قرار می‌گرفت. لیکن امروزه در تمام زمینه‌های علمی کاربرد وسیعی یافته است. رگرسیون لجستیک می‌تواند یک مورد خاص از مدل خطی عمومی و رگرسیون خطی دیده شود. مدل رگرسیون لجستیک، بر اساس فرض‌های کاملاً متفاوتی (درباره رابطه متغیرهای وابسته و مستقل) از رگرسیون خطی است. تفاوت مهم این دو مدل در دو ویژگی رگرسیون لجستیک می‌تواند دیده شود. اول توزیع شرط $\{y|x\}$: یک توزیع برنولی به‌جای یک توزیع گوسی است چون‌که متغیر وابسته دودویی است. دوم مقادیر پیش‌بینی احتمالاتی: محدود بین بازه صفر و یک و به کمک تابع توزیع لجستیک به دست می‌آید. رگرسیون لجستیک احتمال خروجی پیش‌بینی می‌کند. بی‌نظیر (۱۳۸۴) در تحقیقی با استفاده از مدل رگرسیون لجستیک به تحلیل ریسک اعتباری مشتریان حقوقی بانک کشاورزی پرداخته است. محقق با توجه به فقدان وجود پایگاه اطلاعات کامپیوتری، از طریق روش میدانی و تکمیل پرسشنامه از شعب بانک مورد مطالعه، نمونه ۲۸۵ تایی از داده‌های مشتریان تهیه نموده است و مدل را با ۱۷ متغیر (خصیصه‌های مشتریان) با استفاده از مدل رگرسیون لجستیک تخمین زده است[۱۵].

الگوریتم «کامینز»[19]

روش «کامینز» یکی از روش‌های خوشه‌بندی داده‌ها در داده‌کاوی است. این روش علی‌رغم سادگی آن‌یک روش پایه برای بسیاری از روش‌های خوشه‌بندی دیگر (مانند خوشه‌بندی فازی) محسوب می‌شود. این روش، روشی انحصاری و مسطح محسوب می‌شود. برای این الگوریتم شکل‌های مختلفی بیان‌شده است. ولی همه آن‌ها دارای روالی تکراری هستند که برای تعدادی ثابت از خوشه‌ها سعی در تخمین موارد زیر دارند: به دست آوردن نقاطی به‌عنوان مراکز خوشه‌ها. این نقاط درواقع همان میانگین نقاط متعلق به هر خوشه هستند. نسبت دادن هر نمونه داده به یک خوشه که آن داده کمترین فاصله تا مرکز آن خوشه را دارا باشد. در نوع ساده‌ای از این روش ابتدا به تعداد خوشه‌های موردنیاز نقاطی به‌صورت تصادفی انتخاب می‌شود. سپس درداده‌ها با توجه به میزان نزدیکی (شباهت) به یکی از این خوشه‌ها نسبت داده می‌شوند و بدین ترتیب خوشه‌های جدیدی حاصل می‌شود. با تکرار همین روال می‌توان در هر تکرار با میانگین‌گیری از داده‌ها مراکز جدیدی برای آن‌ها محاسبه کرد و مجدداً داده‌ها را به خوشه‌های جدید نسبت داد. این روند تا زمانی ادامه پیدا می‌کند که دیگر

---

[18] . Synapses
[19] . k-means

ISSN: 2745-3673
http://www.Rcsj.ir

تغییری در داده‌ها حاصل نشود. تابع زیر به‌عنوان تابع هدف مطرح است. «کالیش و همکاران»[20] (۲۰۱۵) به کاربرد داده‌کاوی در بخش بانکداری الکترونیکی با روش خوشه‌بندی و طبقه‌بندی، پرداختند. در این مطالعه که در بخش بانکداری الکترونیکی انجام شد، هدف آن کاهش میزان ریسک در تصمیم‌گیری و به حداقل رساندن وام‌های موجود و ارزیابی عملکرد مشتریان بالقوه با استفاده از روش کامینز بود که یکی از تکنیک‌های خوشه‌بندی و روش درخت تصمیم است. نتایج نشان‌دهنده برتری مدل ارائه‌شده نسبت به دیگر مدل‌ها را دارد[۱۶].

روش تاپسیس فازی

در این مقاله جهت اولویت بندی مکانیزم های داده کاوی، از روش تاپسیس فازی(مثلثی) استفاده می شود. بدین منظور داده‌های مورد نیاز روش پیشنهادی به‌صورت مثلثی فرض می‌نماییم. بنابراین در مسائل واقعی اکثراً مسائل تصمیم‌گیری چند معیاره، بصورت گروهی می‌باشد. در ابتدا گروه خبره که شامل کارشناسان و اساتید هستند با D_1 و D_2 تصمیم‌گیرنده که آشنا با روش‌های داده‌کاوی در بانکداری الکترونیک هستند، تشکیل می دهیم و این مجموعه افراد را E_D می‌نامیم. فرض می‌کنیم که n زیرفاکتور روی مسئله اولویت‌بندی موثرند، بنابراین این مجموعه معیارها را C_j می‌نامیم. درنهایت متناسب با معیارها و مسئله، m گزینه تعیین گردیده و این مجموعه از گزینه‌ها را A_i می‌نامیم.

بنابراین فرض می‌شود که معیارهای ارزیابی و اهمیت وزن‌های فازی هر معیار توسط C به ترتیب به‌صورت: $\tilde{x}_{ijk}=(x^a_{ijc},x^b_{ijc},x^c_{ijc},x^d_{ijc})$ و $\tilde{w}_{jk}=(w^a_{jk},w^b_{jk},w^c_{jk}w^d_{jk})$ $i=1,2,\dots,m$ $j=1,2,\dots,n$ باشد. ازاین‌رو تلفیق نرخ‌های فازی معیارهای هرگزینه $\tilde{x}_{ijc}$ به‌صورت $\tilde{x}_{ij}=(x^a_{ij},x^b_{ij},x^c_{ijc},x^d_{ij})$ محاسبه می‌گردد، به‌طوری‌که $x^a_{ij}=\min_k\{x^a_{ijc}\}$ ، $x^b_{ij}=\frac{1}{K}\sum_{k=1}^{K}x^b_{ijc}$ ، $x^c_{ij}=\frac{1}{K}\sum_{k=1}^{K}x^c_{ijc}$ و $x^d_{ij}=\max_k\{x^d_{ijc}\}$. تلفیق وزن‌های فازی هر معیار $\tilde{C}_j$ نیز به‌صورت $\tilde{C}_j=(C^a_j,C^b_j,C^c_j,C^d_j)$ محاسبه می‌گردد، به‌طوری‌که $C^a_j=\min_k\{C^a_{jk}\}$ ، $C^b_j=\frac{1}{K}\sum_{k=1}^{K}C^b_{jk}$ ، $C^c_j=\frac{1}{K}\sum_{k=1}^{K}C^c_{jk}$ و $C^d_j=\min_k\{C^d_{jk}\}$.

پس با توجه به محاسبات مربوط به تلفیق نرخ‌های فازی معیارهای هر گزینه و تلفیق وزن‌های هر معیار ماتریس‌های تصمیم‌گیری فازی جهت بانکداری الکترونیکی را می‌توان به‌صورت زیر تعریف نمود:

$$\tilde{D}=\begin{bmatrix}\tilde{A}_{11} & \tilde{A}_{12} & \cdots & \tilde{A}_{1n}\\ \tilde{A}_{21} & \tilde{A}_{22} & \cdots & \tilde{A}_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ \tilde{A}_{m1} & \tilde{A}_{m2} & \cdots & \tilde{A}_{mn}\end{bmatrix}, \qquad \tilde{C}=[\tilde{C}_1,\tilde{C}_2,\dots,\tilde{C}_n] \qquad (۱-۳)$$

به‌طوری‌که $\tilde{x}_{ij}=(x^a_{ij},x^b_{ij},x^c_{ij},x^d_{ij})$ و $\tilde{W}_j=(C^a_j,C^b_j,C^c_j,C^d_j)$ $i=1,2,\dots,m$ و $j=1,2,\dots,n$ می‌باشد.

در مسائل تصمیم‌گیری اکثراً فاکتورها، ماهیتشان با یکدیگر در تعارض می‌باشند. عده‌ای از آن‌ها مانند جدید بودن و تعداد مقالات استفاده‌شده به‌صورت هزینه معرفی شده و بعضی دیگر از معیارها مانند اساتید و کارشناسان به‌صورت سود معرفی می‌گردند. بنابراین، برای بی مقیاس کردن ماتریس تصمیم‌گیری فازی را به‌صورت $\tilde{R}=[\tilde{r}_{ij}]_{m\times n}$ نرمال می‌کنیم.

اگر B و C به ترتیب معیارهای سود و هزینه باشند، آنگاه:

$$\tilde{r}_{ij}=\left(\frac{x^a_{ij}}{\theta^*_j},\frac{x^b_{ij}}{\theta^*_j},\frac{x^c_{ij}}{\theta^*_j},\frac{x^d_{ij}}{\theta^*_j}\right) \quad (j\in B)$$
$$\tilde{r}_{ij}=\left(\frac{\varphi_j}{x^d_{ij}},\frac{\varphi_j}{x^c_{ij}},\frac{\varphi_j}{x^b_{ij}},\frac{\varphi_j}{x^a_{ij}}\right) \quad (j\in C) \qquad (۲-۳)$$

که $\theta^*_j(j\in B)$ و $\varphi^-_j(j\in C)$ به ترتیب به‌صورت $\max_i\{x^d_{ij}\}$ و $\min_i\{x^a_{ij}\}$ تعریف می‌گردد.

[20] . Çaliş et al

ISSN: 2745-3673
http://www.Rcsj.ir

روش نرمال سازی فوق، تمام خصوصیات را همچنان حفظ می‌نماید، در حالی که $\tilde{r}_{ij}$ به ازای هر i و j اعداد فازی مثلثی نرمال می-باشند. حال ماتریس تصمیم‌گیری فازی نرمال وزین شده را به‌صورت $\tilde{Q} = [\tilde{q}_{ij}]_{m\times n}$، $i = 1,2,\dots,m$ و $j = 1,2,\dots,n$ به دست می-آوریم، به‌طوری‌که $\tilde{q}_{ij} = \tilde{r}_{ij}(0)\tilde{w}_j$ عناصر ماتریس تصمیم‌گیری فازی نرمال وزین $(\forall i,j)\tilde{q}_{ij} \cong (q^a_{ij}, q^b_{ij}, q^c_{ij}, q^d_{ij})$.

اکنون حل ایده‌آل فازی (F^*) و حل ضد ایده‌آل فازی (F^-) را به‌صورت زیر می‌توانیم تعریف نماییم:

$$F^* = (\tilde{V}^*_1, \tilde{V}^*_2, \dots, \tilde{V}^*_n) \qquad (۳-۳)$$
$$F^- = (\tilde{V}^-_1, \tilde{V}^-_2, \dots, \tilde{V}^-_n)$$

به‌طوری‌که

$$\tilde{V}^*_j = (v^*_j, v^*_j, v^*_j, v^*_j) \quad , \quad \tilde{V}^-_j = (v^-_j, v^-_j, v^-_j, v^-_j) \quad \forall i,j. \qquad (۴-۳)$$
$$v^*_j = \max_i \{q^d_{ij}\} \quad , \quad v^-_j = \min_i \{q^a_{ij}\}$$

حل ایده‌آل فازی یک حل مجازی ایده‌آل است که معیارهای سود را ماکزیمم کرده و معیارهای زیان را مینیمم می‌کند یا به عبارتی دیگر گزینه‌ای است که معیارهای آن در بهترین مقدار خود می باشند، بنابراین در ماتریس تصمیم‌گیری فازی نرمالیزه وزین می‌توانیم گزینه را به‌عنوان حل ایده آل فازی در نظر بگیریم که هر یک از معیارهای آن به‌صورت (۱،۱،۱،۱) می‌باشد. از طرف دیگر، حل ضد-ایده‌آل فازی یک حل مجازی ضد ایده‌آل است که معیارهای سود را مینیمم کرده و معیارهای زیان را ماکزیمم می‌کند یا به عبارتی دیگر گزینه‌ای که معیارهای آن در بدترین مقدار خود می‌باشند، بنابراین، در ماتریس تصمیم‌گیری فازی نرمالیزه وزین می‌توانیم گزینه را به‌عنوان حل ضد- ایده‌آل فازی در نظر بگیریم که هریک از معیارهای آن به‌صورت (۰،۰،۰،۰) می‌باشد. در این پایان‌نامه همانطور که از رابطه (۳-۳) مشاهده می‌گردد، گزینه ایده‌آل و ضد ایده‌آل از بین درایه‌های ماتریس تصمیم‌گیری فازی نرمال وزین شده انتخاب می‌گردند.

فاصله هر گزینه از F^* و F^- بر اساس رابطه (۳-۲) به‌صورتی که سیستم رتبه‌بندی روی $H(R)$ می‌باشد، به ازای αهای گوناگون از رابطه زیر به دست می‌آید:

$$d^*_i(\alpha) = \sum_{j=1}^{n} d\left(\tilde{V}^*_j, \tilde{q}_{ij}\right) = \sum_{j=1}^{n} \left(2v^*_j - q^a_{ij} - q^d_{ij}\right) + \left(\frac{\alpha}{2}\right) \sum_{j=1}^{n} \left(q^a_{ij} + q^d_{ij} - q^b_{ij} - q^c_{ij}\right), \ \forall i, \qquad (۵-۳)$$
$$d^-_i(\alpha) = \sum_{j=1}^{n} d\left(\tilde{q}_{ij}, \tilde{V}^-_j\right) = \sum_{j=1}^{n} \left(q^a_{ij} + q^d_{ij} - 2v^-_j\right) + \left(\frac{\alpha}{2}\right) \sum_{j=1}^{n} \left(q^b_{ij} + q^c_{ij} - q^a_{ij} - q^d_{ij}\right), \ \forall i.$$

توجه کنید که در رابطه فوق $d(0,0)$ و مقدار فاصله بین دو عدد فازی بوده و $\alpha \in [0,1]$ می‌باشد. هدف در این تکنیک تصمیم-گیری این است که گزینه‌ای را انتخاب کنیم که به‌طور همزمان تا حد ممکن به گزینه ایده‌آل فازی نزدیک بوده و از گزینه ضد ایده‌آل فازی دور باشد. بدین منظور یک شاخص ضریب نزدیکی برای رتبه‌بندی تمام گزینه ها با توجه به فاصله حل ایده‌آل فازی (F^*) و حل ضد- ایده-آل فازی (F^-) با توجه به α مربوطه می‌بایست تعریف نمود. بدین منظور شاخص ضریب نزدیکی (CC_i) برای هر گزینه به ازای α خاص به‌صورت زیر محاسبه می‌گردد:

$$CC_i(\alpha) = \frac{d^-_i(\alpha)}{d^*_i(\alpha) + d^-_i(\alpha)}, \quad i = 1,2,\dots,m \qquad (۶-۳)$$

واضح است که از رابطه (۳-۶) اگر F^*F^- آنگاه $CCi(\alpha) = 1$ است، و اگر $F-F^-$ آنگاه $CC_i(\alpha) = 0$ می‌گردد. به عبارت دیگر زمانی که گزینه F_i به F^* نزدیک و F^- دور می‌گردد، با توجه به رابطه (۳-۶) شاخص ضریب نزدیکی $(CC_i(\alpha))$ به عدد ۱ نزدیک می‌شود.

ISSN: 2745-3673
http://www.Rcsj.ir

نتایج

ویژگی‌های داده‌ها

به دلیل فقدان اطلاعات لازم در رابطه با الگوریتم‌های داده‌کاوی نمونه آماری ازلحاظ اساتید و کارشناسان، بسیار کم می‌باشد، نمونه آماری متشکل از ۵ نفر، ۲ نفر کارشناس و ۳ نفر اساتید می‌باشد.

تحلیل داده‌ها

در ابتدا جهت بکار بردن روش تاپسیس فازی لازم است که اطلاعات را از پرسشنامه جمع آوری کرده و سپس آنرا به اعداد فازی تبدیل نماییم.

جدول۱- نتایج ارزیابی مکانیزم‌های داده‌کاوی در بانکداری الکترونیک

الگوریتم‌ها	تعداد مقالات استفاده‌کننده از این روش	نظر کارشناسان	نظر اساتید	جدید بودن روش*
CART	۳	۲	۳	کم
CHAID	۲	۱	۳	ضعیف
C5	۲	۲	۲	کم
SVM	۶	۲	۳	متوسط
Naive Bayes	۵	۲	۳	زیاد
Discriminant	۲	۱	۱	ضعیف
QUEST	۱	۲	۲	کم
ANN	۵	۲	۳	متوسط
Logistic –R	۳	۲	۳	کم
mens-K	۲	۱	۳	کم

*نکته: منظور از جدید بودن روش، تعداد مقالات چاپ شده در سال های بعد از ۲۰۱۰ که از الگوریتم موردنظر استفاده کرده اند و همچنین میانگین نظرات کارشناسان و اساتید که در مورد هر الگوریتم می باشد. بنابراین تعداد مقالاتی که بعد از ۲۰۱۰ و میانگین نظرات کارشناسان و اساتید بیشتر باشد در طبقه زیاد و در غیر این صورت در طبقات متوسط، کم و ضعیف قرار می گیرند.

نتایج روش تاپسیس فازی

با توجه به اینکه یک کمیته خبره از تصمیم‌گیرندگان، D_1،D_2 تشکیل شده است تا الگوریتم‌های داده‌کاوی در بانکداری الکترونیکی را بررسی کند، A_1 ، A_2، A_3، A_4، A_5، A_6، A_7 ، A_8، A_9 و A_{10}، تا مناسب ترین الگوریتم را انتخاب نماید. چهار معیار در نظر گرفته می شود.

C_1:تعداد مقالات استفاده‌کننده از الگوریتم

C_2 :نظر کارشناسان

C_3 :نظر اساتید

C_4 :جدید بودن روش

A_1:CART

A_2:CHAID

A_3 :C5

A_4:SVM

A_5:Naive Bayes

A_6:Discriminant

A_7:QUEST

A_8:ANN

A_9:Logistic –R

A_{10}:K-mens

ISSN: 2745-3673
http://www.Rcsj.ir

فرآیند انتخاب نرم افزار مناسب به وسیله روش تاپسیس فازی ذوزنقه ای شامل مراحل زیر است:

گام اول: عبارت های کلامی استفاده شده برای اولویت های تصمیم گیرندگان و شاخص ها در جدول ۴-۲ آمده است.

جدول۲- ارزش های کلامی اعداد فازی ذوزنقه ای برای عبارت های کلامی

ارزش های کلامی اعداد فازی ذوزنقه ای	متغیرهای کلامی
⟨(1.0, 0.0, 1.0, 0.0)⟩	ضعیف
⟨(2.0, 1.0, 1.0, 0.0)⟩	کم
⟨(2.0, 2.0, 2.0, 1.0)⟩	متوسط
⟨(3.0, 2.0, 2.0, 3.0)⟩	زیاد

هر متخصص ارزش‌های کلامی اعداد فازی ذوزنقه‌ای را به اوزان شاخص‌ها اختصاص می‌دهد. تصمیم‌گیرنده‌گان را به k_1 تا k_5 تقسیم نماییم.

جدول ۳- اوزان شاخص ها تعیین شده به وسیله تصمیم گیرندگان ($c_j^{(k)}$)

K	C_1	C_2	C_3	C_4
k_1	(0.4, 0.4, 0.5, 0.6)	(0.3, 0.4, 0.5, 0.5)	(0.6, 0.4, 0.5, 0.6)	(0.3, 0.4, 0.5, 0.6)
k_2	(0.4, 0.6, 0.7, 0.8)	(0.1, 0.2, 0.3, 0.4)	(0.5, 0.6, 0.7, 0.8)	(0.1, 0.3, 0.3, 0.5)
k_3	(0.5, 0.6, 0.7, 0.8)	(0.1, 0.2, 0.3, 0.4)	(0.3, 0.4, 0.5, 0.6)	(0.3, 0.4, 0.5, 0.6)
k_4	(0.3, 0.6, 0.7, 0.6)	(0.5, 0.4, 0.5, 0.6)	(0.3, 0.4, 0.5, 0.6)	(0.5, 0.6, 0.7, 0.8)
k_5	(0.5, 0.5, 0.7, 0.7)	(0.6, 0.4, 0.4, 0.6)	(0.5, 0.4, 0.3, 0.6)	(0.4, 0.6, 0.5, 0.8)

گام دوم: تعیین اهمیت نسبی نظرات و اوزان تصمیم گیرندگان در جدول ۴ نشان داده شده است.

جدول ۴- اهمیت نسبی نظرات تصمیم گیرندگان

	K	C_1	C_2	C_3	C_4		k	C_1	C_2	C_3	C_4
A_1	۱	۰,۲۷۴	۰,۳۰۹	۰,۲۷۸	۰,۳۰۹	A_1	۱	۰,۱۶۹	۰,۱۶۷	۰,۱۶۷	۰,۱۶۸
A_2		۰,۲۹۷	۰,۳۰۹	۰,۲۷۸	۰,۱۶۸	A_2		۰,۲۵۱	۰,۲۷۸	۰,۱۹۸	۰,۲۱۴
A_3	۲	۰,۲۰۳	۰,۱۶۸	۰,۱۶۷	۰,۲۱۴	A_3	۲	۰,۳۲۹	۰,۲۷۸	۰,۳۱۷	۰,۳۰۳
A_4		۰,۲۲۶	۰,۲۱۴	۰,۲۷۸	۰,۳۰۹	A_4		۰,۲۵۱	۰,۲۷۸	۰,۳۱۷	۰,۱۶۸
A_5	۳	۰,۳۰۹	۰,۳۰۹	۰,۲۵۰	۰,۱۶۸	A_5	۳	۰,۲۷۴	۰,۳۰۹	۰,۲۷۸	۰,۳۰۹
A_6		۰,۲۱۴	۰,۱۶۷	۰,۳۲۱	۰,۲۱۴	A_6		۰,۲۹۷	۰,۳۰۹	۰,۲۷۸	۰,۱۶۸
A_7	۴	۰,۳۰۹	۰,۳۰۹	۰,۲۵۰	۰,۳۰۹	A_7	۴	۰,۲۰۳	۰,۱۶۸	۰,۱۶۷	۰,۲۱۴
A_8		۰,۱۶۸	۰,۲۱۵	۰,۱۷۹	۰,۳۰۳	A_8		۰,۲۲۶	۰,۲۱۴	۰,۲۷۸	۰,۳۰۹
A_9	۵	۰,۳۰۹	۰,۳۰۹	۰,۲۵۰	۰,۱۶۸	A_9	۵	۰,۲۱۵	۰,۲۲۶	۰,۲۷۸	۰,۳۰۹
A_{10}		۰,۲۱۴	۰,۱۶۷	۰,۳۲۱	۰,۲۱۴	A_{10}		۰,۳۰۹	۰,۳۰۳	۰,۲۷۸	۰,۱۶۸
A_1	۱	۰,۳۰۹	۰,۳۰۳	۰,۱۶۷	۰,۱۶۸	A_1	۱	۰,۳۰۹	۰,۳۰۹	۰,۲۵۰	۰,۱۶۸
A_2		۰,۱۶۷	۰,۱۶۸	۰,۲۷۸	۰,۳۰۹	A_2		۰,۲۱۴	۰,۱۶۷	۰,۳۲۱	۰,۲۱۴
A_3	۲	۰,۲۱۵	۰,۲۲۶	۰,۲۷۸	۰,۳۰۹	A_3	۲	۰,۳۰۹	۰,۳۰۹	۰,۲۵۰	۰,۳۰۹
A_4		۰,۳۰۹	۰,۳۰۳	۰,۲۷۸	۰,۱۶۸	A_4		۰,۱۶۸	۰,۲۱۵	۰,۱۷۹	۰,۳۰۳
A_5	۳	۰,۳۰۹	۰,۳۰۹	۰,۲۵۰	۰,۱۶۸	A_5	۲	۰,۲۷۴	۰,۳۰۹	۰,۲۷۸	۰,۳۰۹
A_6		۰,۲۱۴	۰,۱۶۷	۰,۳۲۱	۰,۲۱۴	A_6		۰,۲۹۷	۰,۳۰۹	۰,۲۷۸	۰,۱۶۸

ISSN: 2745-3673
http://www.Rcsj.ir

A_7	۴	۰,۳۰۹	۰,۳۰۹	۰,۲۵۰	۰,۳۰۹	A_7	۴	۰,۲۰۳	۰,۱۶۸	۰,۱۶۷	۰,۲۱۴
A_8		۰,۱۶۸	۰,۲۱۵	۰,۱۷۹	۰,۳۰۳	A_8		۰,۲۲۶	۰,۲۱۴	۰,۲۷۸	۰,۳۰۹
A_9	۵	۰,۲۱۵	۰,۲۲۶	۰,۲۷۸	۰,۳۰۹	A_9	۵	۰,۲۵۱	۰,۲۷۸	۰,۳۱۷	۰,۱۶۸
A_{10}		۰,۳۰۹	۰,۳۰۳	۰,۲۷۸	۰,۱۶۸	A_{10}		۰,۲۷۴	۰,۳۰۹	۰,۲۷۸	۰,۳۰۹

گام سوم: تشکیل ماتریس تصمیم گیری فازی ذوزنقه ای ادغامی برپایه نقطه نظرات هر تصمیم گیرنده مطابق جدول ۵.

جدول ۵- ماتریس تصمیم فازی ذوزنقه ای ادغامی

	C_1	C_2	C_3	C_4
A_1	(0.22, 0.32, 0.42, 0.52)	(0.48, 0.58, 0.68, 0.78)	⟨0.17,0.27,0.37,0.47⟩	⟨0.27,0.34,0.48,0.57⟩
A_2	⟨0.32,0.42,0.52,0.62⟩	⟨0.46,0.56,0.66,0.76⟩	⟨0.33, 0.43, 0.53, 0.63⟩	⟨0.45, 0.58, 0.78, 0.78⟩
A_3	⟨0.27, 0.37, 0.47, 0.57⟩	⟨0.67, 0.77, 0.87, 0.97⟩	⟨0.29, 0.39, 0.49, 0.59⟩	⟨0.58, 0.58, 0.68, 0.78⟩
A_4	⟨0.48, 0.58, 0.68, 0.78⟩	⟨0.32, 0.42, 0.52, 0.62⟩	⟨0.35, 0.45, 0.55, 0.65⟩	⟨0.46, 0.56, 0.66, 0.76⟩
A_5	⟨0.23, 0.32, 0.42, 0.52⟩	⟨0.48, 0.58, 0.68, 0.78⟩	⟨0.17, 0.27, 0.37, 0.47⟩	⟨0.27, 0.34, 0.48, 0.57⟩
A_6	⟨0.32, 0.42, 0.52, 0.62⟩	⟨0.46, 0.56, 0.66, 0.76⟩	⟨0.33, 0.43, 0.53, 0.63⟩	⟨0.45, 0.58, 0.78, 0.78⟩
A_7	⟨0.24, 0.37, 0.47, 0.57⟩	⟨0.67, 0.77, 0.87, 0.97⟩	⟨0.29, 0.39, 0.49, 0.59⟩	⟨0.58, 0.58, 0.68, 0.78⟩
A_8	⟨0.48, 0.58, 0.68, 0.78⟩	⟨0.32, 0.42, 0.52, 0.62⟩	⟨0.35, 0.45, 0.55, 0.65⟩	⟨0.46, 0.56, 0.66, 0.76⟩
A_9	⟨0.22, 0.32, 0.42, 0.52⟩	⟨0.48, 0.58, 0.68, 0.78⟩	⟨0.17, 0.27, 0.37, 0.47⟩	⟨0.27, 0.34, 0.48, 0.57⟩
A_{10}	⟨0.48, 0.58, 0.67, 0.78⟩	⟨0.32, 0.42, 0.50, 0.62⟩	⟨0.35, 0.45, 0.54, 0.65⟩	⟨0.46, 0.56, 0.66, 0.76⟩

گام چهارم: تعیین ارزش وزنی مورد انتظار نرمالیزه شده شاخص ها. بر اساس اهمیت شاخص ها ($\xi_{ij}^{(k)}$) تعیین شده به وسیله تصمیم گیرندهkام، ابتدا ξ_{ij} را محاسبه می کنیم که نشان دهنده ارزش وزنی شاخص c_j برای گزینه o_i می باشد، و سپس ارزش وزنی مورد انتظار نرمالیزه شده ξ'_{ij} می تواند به صورت زیر به دست می آید:

$$(\xi'_{ij}) = \begin{bmatrix} 0.418 & 0.254 & 0.347 \\ 0.411 & 0.363 & 0.345 \\ 0.417 & 0.245 & 0.319 \\ 0.424 & 0.343 & 0.353 \end{bmatrix}$$

گام پنجم: تعیین PIS و NIS .

هزینه نرم افزار و سخت افزار (C_1) تعداد مقالات استفاده‌کننده از الگوریتم، نظر کارشناسان (C_2) و نظر اساتید (C_3) جدید بودن روش(C_4) هستند. بنابرین PIS و NIS به صورت زیر به دست می آیند:

$$R^+ = \{\langle(0.21, 0.42, 0.42, 0.52), (0.16, 0.31, 0.41, 0.61)\rangle, \langle(0.67, 0.77, 0.87, 0.97), (0.65, 0.77, 0.88, 0.98)\rangle, \langle(0.36, 0.46, 0.55, 0.64)\rangle\}$$

$$R^- = \{\langle(0.47, 0.55, 0.61, 0.78), (0.40, 0.58, 0.62, 0.85)\rangle, \langle(0.33, 0.42, 0.52, 0.62), (0.24, 0.42, 0.54, 0.71)\rangle, \langle(0.17, 0.27, 0.37, 0.47)\rangle\}$$

گام ششم: اندازه های فاصله وزنی مثبت و منفی در جدول ۶ نشان داده شده اند.

ISSN: 2745-3673
http://www.Rcsj.ir

جدول ۶- اندازه های فاصله و ضرایب نزدیکی نسبی هر گزینه

	D^+	D^-	U^-
A_1	۰,۱۱۰	۰,۱۴۲	۰,۶۶۵
A_2	۰,۱۰۰	۰,۱۵۳	۰,۳۸۱
A_3	۰,۰۴۰	۰,۲۱۱	۰,۵۳۴
A_4	۰,۱۸۲	۰,۰۶۴	۰,۷۵۲
A_5	۰,۱۱۷	۰,۱۴۵	۰,۸۴۷
A_6	۰,۱۰۵	۰,۱۳۵	۰,۲۵۶
A_7	۰,۰۴۰	۰,۲۱۱	۰,۵۰۹
A_8	۰,۱۱۰	۰,۱۵۳	۰,۷۴۹
A_9	۰,۰۴۵	۰,۲۱۸	۰,۶۳۹
A_{10}	۰,۱۸۷	۰,۰۶۴	۰,۴۸۸

گام هفتم: ضرایب نزدیکی نسبی U_i^- هر گزینه A_i بر اساس جدول ۶ به دست می آید.

گام هشتم: در حالیکه $U_5^- > U_4^- > U_8^- > U_1^-$ باشد، رتبه بندی گزینه‌ها مطابق $A_5 > A_4 > A_8 > A_1$ می باشد، بنابراین بهترین گزینه می باشد.

بنابراین طبق طبق نتایج فوق رتبه بندی ۱۰ الگوریتم به ترتیب:

$A_6 < A_2 < A_{10} < A_7 < A_3 < A_9 < A_1 < A_8 < A_4 < A_5$ خواهد بود. به‌وضوح بهترین حالت A_5 می‌باشد چون بیشترین نزدیکی نسبی را دارد. بنابراین بهترین الگوریتم داده‌کاوی در بانکداری الکترونیک، الگوریتم نایوبیز می باشد.

با توجه به نتایج به‌دست‌آمده، مشاهده می‌گردد الگوریتم نایوبیز ازلحاظ چهار معیار مکانیزم داده‌کاوی در بانکداری الکترونیک که در این پژوهش مطرح گردید، بهتر عمل می‌کند. لذا معیار انتخابی ما برای داده‌کاوی در بانکداری الکترونیک، استفاده از الگوریتم نایوبیز می-باشد. بزرگ‌ترین ویژگی این روش این است که حجم آموزش اندکی برای شروع کار و تخمین پارامترها نیاز دارد. برنامه‌های کاربردی بسیاری هستند که پارامترهای نایوبیز را تخمین می‌زنند، بنابراین افراد بدون سروکار داشتن با تئوری بیز می‌توانند از این امکان به‌منظور حل مسائل موردنظر بهره ببرند. باوجود مسائل طراحی و پیش‌فرض‌هایی که در خصوص روش بیز وجود دارد، این روش برای طبقه‌بندی کردن بیشتر مسائل در جهان واقعی، مناسب است.

مقایسه

همانطور که نتایج این پژوهش نشان داد الگوریتم نایوبیز نسبت به دیگر الگوریتم ها در داده کاوی بانکداری الکترونیک در اولویت اول قرار دارد. نتایج تحقیق ما با نتایج[۱] مقایسه شده است. در مقاله[۱] جهت جمع آوری نمونه های تحقیق از پرسشنامه استفاده شد و اندازه‌ی نمونه تحقیق برابر ۱۰۸۱ بوده است و مجموعا ۳۴ متغیر را تحلیل شده است. در این زمینه، مقاله[۱] روش‌های داده‌کاوی و تکنیک‌های مختلف جهت تعیین اینکه کدام متغیرها مهمترین متغیر برای موسسات مالی هستند را بررسی می‌کند، مؤسسات مالی از مهمترین متغیرها جهت پیش‌بینی سطوح احتمالی اطمینان میان کاربران بانکداری الکترونیکی استفاده می‌کنند. نتایج مقاله [۱] نشان داد که از بین همه الگوریتم‌های داده کاوی، الگوریتم SVM نسبت به دیگر الگوریتم های داده کاوی، بهترین متغیر برای پیش‌بینی سطوح احتمالی اطمینان میان کاربران بانکداری الکترونیکی است. این در حالی است که بر اساس نتایج پژوهش ما، الگوریتم نایوبیز نسبت به دیگر الگوریتم‌ها جدید و دقیق تر است.

ISSN: 2745-3673
http://www.Rcsj.ir

نتیجه‌گیری

کل فرآیند این تحقیق به دنبال پاسخ به این سؤال بود که چگونه می‌توان با کمک روش تاپسیس فازی، مکانیزم‌های داده‌کاوی در بانکداری الکترونیک را اولویت‌بندی نمود. در فصل سوم و چهارم مراحل چگونگی این موضوع را با استفاده از روش‌های طبقه‌بندی با الگوریتم‌های داده‌کاوی تشریح گردیده است.

نتایج حاصل از یافته‌های تحقیق نشان داد که بر اساس اولویت‌بندی الگوریتم‌های داده‌کاوی، الگوریتم‌های CHAID و Discriminant در طبقه‌بندی «ضعیف» قرارگرفته‌اند زیرا این الگوریتم‌ها باتوجه به مقالات چاپ‌شده در این زمینه، قدیمی می‌باشد و همچنین این الگوریتم‌ها از دیدگاه خبرگان در حد ضعیف بوده و در بانکداری الکترونیک ازلحاظ تئوری و عملی، کاربردی نمی‌باشد.

بر اساس نتایج به‌دست‌آمده، اولویت‌بندی الگوریتم‌های CART، C5، QUEST، Logistic–R و K-mens در بانکداری الکترونیک در طبقه «کم» قرارگرفته است. چراکه این الگوریتم‌ها با توجه به مطالعه پیشینه مقالات انجام‌گرفته، در بانکداری الکترونیک به تعداد کم می‌باشد و میانگین نظرات خبرگان ازلحاظ آشنایی و کاربرد این الگوریتم‌ها در بانکداری الکترونیک، در حد کم دانسته‌اند.

بر اساس نتایج به‌دست‌آمده، اولویت‌بندی الگوریتم‌های SVM و ANN، در بانکداری الکترونیک در طبقه «متوسط» قرارگرفته است. با توجه به مطالعه پیشینه مقالات انجام شده، این دو الگوریتم به دلیل کاربرد بیشتر و جدید بودن آن نسبت به دیگر الگوریتم‌ها در بانکداری الکترونیک برتری دارد و همچنین از دیدگاه خبرگان، این الگوریتم‌ها در سطح متوسط قرارگرفته است. تحقیقات نشان داده است که الگوریتم SVM ازجمله روش‌های جدیدی است که در سال‌های اخیر کارایی خوبی نسبت به دیگر روش‌های قدیمی‌تر برای طبقه‌بندی، از خود نشان داده است. باتوجه به اینکه این الگوریتم در هرجایی که نیاز به تشخیص الگو یا دسته‌بندی اشیا در کلاس‌های خاص باشد می‌توان استفاده نمود. با توجه به مطالعه پیشینه تحقیقات انجام‌گرفته، الگوریتم ANN برای تخمین و تقریب در حوزه بانکداری الکترونیک نسبت به دیگر الگوریتم‌ها، کارایی بسیار بالایی از خود نشان داده است. بر اساس تحقیقات صورت گرفته در رابطه با الگوریتم‌های SVM و ANN، نشان می‌دهد که الگوریتم SVM نسبت به الگوریتم ANN کاربردی‌تر و برتری نسبی دارد.

درنهایت نتایج حاصل از اولویت‌بندی الگوریتم‌های داده‌کاوی، نشان داد که الگوریتم «نایوبیز» در بانکداری الکترونیک در طبقه «زیاد» قرارگرفته است. چراکه بامطالعه پیشینه مقالات انجام‌گرفته، این الگوریتم به دلیل کاربرد زیاد و جدید بودن آن نسبت به همه الگوریتم‌ها در بانکداری الکترونیک برتری دارد و همچنین از دیدگاه خبرگان، این الگوریتم‌ها در سطح زیاد قرارگرفته است. الگوریتم نایوبیز یک الگوریتم یادگیری است که دارای ویژگی‌هایی به همراه کارایی محاسباتی بسیار مطلوب نسبت به دیگر الگوریتم‌ها، سبب شده تا الگوریتم نایوبیز در عمل به‌صورت گسترده مورداستفاده قرار گیرد. الگوریتم نایوبیز دارای چند ویژگی در بانکداری الکترونیکی است ازجمله خطای کم و یادگیری تدریجی در داده‌کاوی بانکداری الکترونیکی نسبت به دیگر الگوریتم‌ها است. نتایج مقالات انجام شده و نتایج حاصل از فصل قبل نشان می‌دهد که الگوریتم نایوبیز بیشترین کاربرد نسبت به کلیه الگوریتم‌های طبقه‌بندی داده‌کاوی مورداستفاده در این پژوهش دارد، بنابراین دانش استخراج شده از این الگوریتم به‌عنوان مورد اعتمادترین دانش داده‌کاوی موردبررسی می‌باشد و قوانین حاصل از الگوریتم نایوبیز به‌عنوان دانش سازمانی استخراج شده از داده‌کاوی در بانکداری الکترونیک مورد مطالعه تلقی می‌گردند.

ISSN: 2745-3673
http://www.Rcsj.ir

منابع و مراجع

[1] Prashar, S., Comparing Predictive Ability of Classifiers in Forecasting Online Buying Behaviour: An Empirical Study. International Journal of Strategic Decision Sciences, 2016, 6(4), 5-18.

[۲] ناظمی، ع؛ مشکانی، ع. داده کاوی کاربردی، دانشگاه آزاد اسلامی واحد نیشابور، ۱۳۸۸.

[3] Zekic-Susac, M. Sarlija, N., Bensic, M. Small business credit scoring: a comparison of logistic regression, neural network, and decision tree models, International Conference on Information Technology Interfaces, 2004.

[4] Antipov, E., Elena, P. Applying CHAID for logistic regression diagnostics and classification accuracy improvement. Journal of Targeting, Measurement and Analysis for Marketing, 2010. 18(2), 109-117.

[۵] کوثری لنگری، ر؛ مقدم، ن؛ وحدت، د. معرفی یک مدل برای تشخیص رفتارهای مشکوک در بانکداری الکترونیک با استفاده از الگوریتم های تصمیم گیری درخت، مجله پردازش اطلاعات و مدیریت، شماره۲۸، دوره۳، ۱۳۹۳، صص۷۰۰-۶۸۱.

[6] Chattamvelli, R., Data mining Algorithm. Alpha science, 2011.

[۷] لنگری،ر. ارائه مدلی جهت کشف رفتارهای مشکوک در بانکداری الکترونیکی با استفاده الگوریتم‌های درخت تصمیم‌گیری، پایان نامه کارشناسی ارشد، دانشگاه پیام نور تهران: تهران، ۱۳۸۹.

[8] Ying, C., Ying, Y. Learning with Support Vector Machines. Synthesis Lectures on Artificial Intelligence and Machine Learning, 2011, 95(4), 95-106.

[۹] شربیانی مقدوری، ف؛ نیکومرام،ه؛ طلوعی اشلقی، ع. طبقه بندی متقاضیان تسهیلات اعتباری بانک ها با استفاده از تکنیک ماشین بردار پشتیبان، آینده پژوهی مدیریت، دوره۲۱، شماره۱، ۱۳۸۹، صص۱۹-۱.

[10] Krichene, A., Using a naive Bayesian classifier methodology for loan risk assessment: Evidence from a Tunisian commercial bank. Journal of Economics, Finance and Administrative Science, 2017, 22(42), 451-462.

[11] McLachlan, G.J. Discriminant Analysis and Statistical Pattern Recognition. Wiley Series in Probability and Statistics, 2004.

[12] Ante, R., Ana, K, DISCRIMINANT ANALYSIS OF BANK PROFITABILITY LEVELS. Croatian Operational Research Review, 2013, 11(3), 39-49.

[13] Kizys, R., Paltalidis, N., Vergos, K. The Quest for Banking Stability in the Euro Area: The Role of Government Interventions. Journal of International Financial Markets, Institutions and Money, 12(40), 2015,111-133.

[۱۴] لطفی، ل. مدل سازی ریسک اعتباری در بانک کشاورزی؛ رویکرد مدل¬های لاجیت، پروبیت و شبکه های عصبی، پایان نامه کارشناسی ارشد، دانشگاه علامه طباطبایی، ۱۳۸۶.

[۱۵] بی نظیر، ع. امتیاز دهی اعتباری مشتریان حقوقی بانکها با استفاده از روش آلتمن مطالعه موردی بانک ملت. ۱۳۸۸، موسسه عالی بانکداری: تهران.

[16] Çaliş, A., Boyaci, A., Kasım. Data mining application in banking sector with clustering and classification methods. International Conference on Industrial Engineering and Operations Management Dubai, United Arab Emirates, 2015.