

## استفاده از داده‌کاوی در بانکداری الکترونیک با استفاده از روش تاپسیس فازی

محمد عرفان میرزایی<sup>۱</sup>، محمد خلیلی<sup>۲</sup>

<sup>۱</sup> دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات.

<sup>۲</sup> هیئت علمی دانشگاه آزاد اسلامی واحد خمین.

نام نویسنده مسئول:

محمد عرفان میرزایی

### 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 بانک‌ها می‌توانند با بکارگیری روش‌های داده‌کاوی جدید بر اساس روش تاپسیس فازی به کشف تقلب و به پیشگیری و حذف فرایندهای حامی این گونه اعمال بپردازند. نتایج نشان داد که الگوریتم «نایوبیز» در بانکداری الکترونیک بیشترین کاربرد را داشته است. چراکه با مطالعه پیشینه مقالات انجام‌گرفته، این الگوریتم به دلیل کاربرد زیاد و جدید بودن آن نسبت به همه الگوریتم‌ها در بانکداری الکترونیک برتری دارد. 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 مورد استفاده قرار گیرد. الگوریتم نایوبیز دارای چند ویژگی در بانکداری الکترونیک است از جمله خطای کم و یادگیری تدریجی در داده‌کاوی بانکداری الکترونیک نسبت به دیگر الگوریتم‌ها است. 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.

## مقدمه

اخیراً، سازمان‌های پولی و مالی به سمت جامعه اطلاعاتی پیش می‌رود و نقش مؤثر و تحول‌آفرین فناوری اطلاعات در این زمینه باعث شده است بانک‌ها به‌عنوان یک‌نهاد مالی و اعتباری مهم، از فناوری اطلاعات به‌عنوان راهکاری مطلوب، استفاده کنند. از آنجایی که امروزه تغییرات قابل‌توجهی در بخش خدمات مالی رخ داده است، سبب بروز تغییرات در حوزه بانکداری شده و یک فضای به‌شدت رقابتی را پدید آورده است. به همین منظور بسیاری از مؤسسات مالی به‌منظور اجتناب از دست دادن مشتریان باید به توسعه استراتژی‌های مدیریت ارتباط با مشتری همچون ارزش‌گذاری مشتری مبادرت کنند که در نهایت قادر به شناسایی مشتریان وفادار برای شخصی‌سازی خدمات شوند تا با ایجاد روابط طولانی‌مدت با مشتریان، سودآوری سازمان حاصل گردد.

همچنین در گذشته، عموماً استخراج اطلاعات مفید از داده‌های ثبت‌شده، به‌صورت دستی و بر عهده تحلیلگران بوده است. با توجه به اینکه تجزیه و تحلیل 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 بهره‌گرفت 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 با ارزش‌تر 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 تجزیه و تحلیل شوند و برخی از وابستگی‌ها و روابط آن‌ها شناسایی شوند. این امر به‌وسیله مکانیزم‌های داده‌کاوی می‌تواند انجام شود.

## بانکداری الکترونیک

بانکداری الکترونیک یا برخط یا اینترنتی، عبارت است از فراهم آوردن امکاناتی برای کارکنان در جهت افزایش سرعت و کارایی آن‌ها در ارائه خدمات بانکی در محل شعبه و همچنین فرآیندهای بین شعبه‌ای و بین‌بانکی در سراسر دنیا و ارائه امکانات سخت‌افزاری و نرم‌افزاری به مشتریان که با استفاده از آن‌ها بتوانند بدون نیاز به حضور فیزیکی در بانک، در هر ساعت از شبانه‌روز (۲۴ ساعته) از طریق کانال‌های ارتباطی ایمن و با اطمینان عملیات بانکی دلخواه خود را انجام دهند.

به‌عبارت‌دیگر بانکداری الکترونیکی استفاده از فناوری‌های پیشرفته نرم‌افزاری و سخت‌افزاری مبتنی بر شبکه و مخابرات برای تبادل منابع و اطلاعات مالی به‌صورت الکترونیکی است و نیازی به حضور فیزیکی مشتری در شعبه نیست. به مشتریان اجازه می‌دهد تا معاملات اقتصادی را در یک وب‌سایت امن به‌طرقی مثل کارهای خرده بانکی یا بانک مجازی، موسسه مالی و اعتباری یا شرکت‌های ساخت‌وساز انجام دهند [۱].

«کالیس و همکاران»<sup>۱</sup> (۲۰۱۵)، بانکداری الکترونیک را استفاده بانک‌ها از اینترنت برای ارائه خدمات بانکی به مشتریان و استفاده مشتریان از اینترنت برای ساماندهی، کنترل و انجام تراکنش بر روی حساب‌های بانکی خود تعریف می‌کند [۲]. با این حال، برخی از صاحب‌نظران تعریف کلی‌تری ارائه کرده‌اند و استفاده از سایر ابزارها و کانال‌های الکترونیک نظیر تلفن همراه، تلفن و تلویزیون دیجیتال به‌منظور اطلاع‌رسانی، ایجاد ارتباط و انجام تراکنش بانکی را نیز مشمول تعریف بانکداری الکترونیک دانسته‌اند [۲].

## خدمات بانکداری الکترونیک

«پراشار»<sup>۲</sup> (۲۰۱۶)، خدمات بانکداری الکترونیک را از سه جنبه موردتوجه قرار می‌دهد و معتقد است مشتریان بانک‌ها، در سه سطح قادر به دریافت خدمات بانکداری الکترونیک هستند. این سه سطح عبارت‌اند از:

**اطلاع‌رسانی:** این سطح ابتدایی‌ترین سطح خدمات بانکداری الکترونیک است. در این سطح، بانک، اطلاعات مربوط به خدمات و عملیات بانکی خود را از طریق شبکه‌های عمومی یا خصوصی معرفی می‌کند.

1. Çaliş et al

2. Prashar

**تعاملی:** این سطح از خدمات بانکداری الکترونیک، امکان انجام مبادلات بین سیستم بانکی و مشتری را فراهم می‌آورد. ریسک این سطح در خدمات بانکداری الکترونیک بیشتر از شیوه سنتی است و به ابزارهای مناسبی برای کنترل دسترسی کاربران به شبکه بانک نیاز دارد.

**تراکنشی:** در این سطح، مشتری قادر است با استفاده از یک سیستم امنیتی کنترل شده، فعالیت‌هایی از قبیل صدور چک، انتقال وجه و افتتاح حساب را انجام دهد. این سطح از خدمات بانکداری الکترونیک، از بالاترین سطح ریسک برخوردار است [۳].

### «داده‌کاوی»<sup>۳</sup>

داده‌کاوی را می‌توان چنین تعریف کرد که: داده‌کاوی به بهره‌گیری از ابزارهای تجزیه و تحلیل داده‌ها به منظور کشف الگوها و روابط معتبری که تاکنون ناشناخته بوده‌اند اطلاق می‌شود. این ابزارها ممکن است مدل‌های آماری، الگوریتم‌های ریاضی و روش‌های «یادگیرنده»<sup>۴</sup> باشند که کار این خود را به صورت خودکار و بر اساس تجربه‌ای که از طریق «شبکه‌های عصبی»<sup>۵</sup> یا «درخت‌های تصمیم‌گیری»<sup>۶</sup> به دست می‌آورند بهبود می‌بخشد. داده‌کاوی منحصر به گردآوری و مدیریت داده‌ها نبوده و تجزیه و تحلیل اطلاعات و پیش‌بینی را نیز شامل می‌شود. برنامه‌های کاربردی که با بررسی فایل‌های متن یا چندرسانه‌ای به کاوش داده‌ها می‌پردازند پارامترهای گوناگونی را در نظر می‌گیرند که عبارت‌اند از:

«**قواعد انجمنی**»<sup>۷</sup>: الگوهایی که بر اساس آن یک رویداد به دیگری مربوط می‌شود مثلاً خرید قلم به خرید کاغذ.

«**ترتیب**»<sup>۸</sup>: الگویی که به تجزیه و تحلیل توالی رویدادها پرداخته و مشخص می‌کند کدام رویداد، رویدادهای دیگری را در پی دارد مثلاً تولد یک نوزاد و خرید پوشاک.

«**پیش‌بینی**»<sup>۹</sup>: در پیش‌بینی هدف پیش‌بینی یک متغیر پیوسته می‌باشد، مانند پیش‌بینی نرخ ارز یا هزینه‌های درمانی.

«**رده‌بندی یا طبقه‌بندی**»<sup>۱۰</sup>: فرآیندی برای پیدا کردن مدلی است که رده‌های موجود در داده‌ها را تعریف نموده و متمایز می‌کند، با این هدف که بتوان از این مدل برای پیش‌بینی رده رکوردهایی که برچسب رده آن‌ها (متغیر هدف) ناشناخته می‌باشد، استفاده نمود. در حقیقت در رده‌بندی برخلاف پیش‌بینی، هدف پیش‌بینی مقدار یک متغیر گسسته است. روش‌های مورد استفاده در پیش‌بینی و رده‌بندی عموماً یکسان هستند.

«**خوشه‌بندی**»<sup>۱۱</sup>: گروه‌بندی مجموعه‌ای از اعضا، رکوردها یا اشیاء به نحوی که اعضای موجود در یک خوشه بیشترین شباهت را به یکدیگر و کمترین شباهت را به اعضای خوشه‌های دیگر داشته باشند.

«**مصورسازی**»<sup>۱۲</sup>: مصورسازی به هر تکنیکی برای ساختن تصاویر، نمودارها و پویانمایی‌ها گفته می‌شود که بتواند پیامی را منتقل کند.

برنامه‌های کاربردی که در زمینه تجزیه و تحلیل اطلاعات به کار می‌روند از امکاناتی چون «پرس و جوی ساخت‌یافته»<sup>۱۳</sup> که در بسیاری از بانک‌های اطلاعاتی یافت می‌شود و از ابزارهای تجزیه و تحلیل آماری برخوردارند اما برنامه‌های مربوط به داده‌کاوی در عین برخورداری از این قابلیت‌ها از نظر نوع با آن‌ها تفاوت دارند. بسیاری از ابزارهای ساده برای تجزیه و تحلیل داده‌ها روشی بر پایه «راستی آزمایی»<sup>۱۴</sup> را به کار می‌برند که در آن فرضیه‌ای بسط داده شده، آنگاه داده‌ها برای تأیید یا رد آن بررسی می‌شوند. به طور مثال ممکن است این نظریه مطرح شود که فردی که یک چکش خریده حتماً یک بسته میخ هم خواهد خرید. کارایی این روش به میزان خلاقیت کاربر برای ارائه فرضیه‌های متنوع و همچنین ساختار برنامه بکار رفته بستگی دارد. در مقابل در داده‌کاوی روش‌هایی برای کشف روابط بکار برده می‌شوند و به کمک الگوریتم‌هایی روابط چندبعدی بین داده‌ها تشخیص داده شده و آن‌هایی که یکتا یا رایج هستند شناسایی می‌شوند [۴].

3. Data mining

4. Machine Learning Methods

5. Neural Networks

6. Decision Trees

7. Association

8. Sequence

9. Prediction

10. Classification

11. Clustering

12. visualization

13. Structured query

14. verification

## داده‌کاوی در صنعت بانکداری

صنعت بانکداری در جهان بیشترین تغییرات را در نحوه کسب‌وکار خود تجربه کرده است. با پیاده‌سازی‌های اخیر، مقبولیت و استفاده گسترده از بانکداری الکترونیکی، به دست آوردن داده ترانکشی راحت شده و حجم این چنین داده‌ها به‌طور قابل‌ملاحظه‌ای افزایش یافته‌اند. آنالیز این حجم از داده‌های خام برای تبدیل داده به دانش برای سازمان خارج از توان انسان می‌باشد. مقدار زیادی از داده‌ها که بانک طی سال‌ها آن‌ها را جمع‌آوری کرده می‌تواند در موفقیت تلاش‌های داده‌کاوی تأثیر زیادی بگذارد. با استفاده از داده‌کاوی برای آنالیز الگوها و خواسته‌ها، مدیران می‌توانند با دقت بیشتر، واکنش مشتریان به تغییرات نرخ بهره، تمایل مشتریان به قبول خدمات و محصولات ارائه‌شده، اطلاع از اینکه کدام یک از مشتریان در پرداخت بدهی وام از ریسک بالائی برخوردار هستند و چگونگی ایجاد ارتباط مناسب‌تر با مشتریان را پیش‌بینی کنند. صنعت بانکداری در بسیاری از موارد در حال شناسایی اطلاعات مهم درباره مشتریان خود می‌باشد. نظر به اینکه صنعت بانکداری جزء صنعت خدمات می‌باشد، کار نگهداری مؤثر «CRM»<sup>۱۵</sup> بحث بسیار مهمی می‌باشد. برای این منظور، بانک نیاز به سرمایه‌گذاری برای درک بهتر از مشتریان حال و آینده خود دارد. با استفاده از ابزارهای مناسب داده‌کاوی، بانک به دنبال آن می‌تواند محصولات و خدمات سفارشی به مشتریان خود ارائه دهد [۵].

بانکداری الکترونیک به کلیه خدمات بانکی و پولی گفته می‌شود که در بستر الکترونیک و به‌صورت کاملاً متمرکز و حضوری یا غیرحضوری به مشتریان ارائه می‌شوند. به‌عبارت‌دیگر بانکداری الکترونیکی سازمان‌های یکپارچه است که کلیه محصولات و خدمات بانکی و عملیات راهبری و مدیریت آن‌ها را از طریق تجهیزات الکترونیکی به پایگاه داده متمرکز در قالب یک سیستم ارائه می‌کند [۵].

پس از گذشت دو دهه و همزمان با پیشرفت فن‌آوری اطلاعات هر دو سال یک‌بار حجم داده‌ها، دو برابر شد. همچنین تعداد پایگاه داده‌ها با سرعت بیشتری رشد نمود. این در حالی است که تعداد متخصصین تحلیل داده‌ها و آمارشناسان با این سرعت رشد نکرد. حتی اگر چنین امری اتفاق می‌افتاد، بسیاری از پایگاه داده‌ها چنان گسترش یافته‌اند که شامل چند صد میلیون یا چند صد میلیارد رکورد ثبت شده هستند و امکان تحلیل و استخراج اطلاعات با روش‌های معمول آماری از دل انبوه داده‌ها مستلزم چند روز کار با کامپیوترهای موجود است. حال باوجود سیستم‌های یکپارچه اطلاعاتی، سیستم‌های یکپارچه بانکی و تجارت الکترونیک، لحظه‌به‌لحظه به حجم داده‌ها در پایگاه داده‌های مربوط اضافه شده و باعث به وجود آمدن انبارهای (توده‌های) عظیمی از داده‌ها شده است، به‌طوری‌که ضرورت کشف و استخراج سریع و دقیق دانش از این پایگاه داده‌ها را بیش‌ازپیش نمایان کرده است. هم‌اکنون در هر کشور، سازمان‌ها، شرکت‌ها برای امور بازرگانی، پرسنلی، آموزشی، آماری پایگاه داده‌ها ایجاد یا خریداری شده است، به‌طوری‌که این پایگاه داده‌ها برای مدیران، برنامه‌ریزان، پژوهشگران جهت تصمیم‌گیری‌های راهبردی، تهیه گزارش‌های مختلف، توصیف وضعیت جاری خود، می‌تواند مفید باشد. داده‌کاوی یا استخراج و کشف سریع و دقیق اطلاعات باارزش و پنهان از این پایگاه داده‌ها ازجمله اموری است که هر کشور، سازمان و شرکتی به‌منظور توسعه علمی، فنی و اقتصادی خود به آن نیاز دارد [۵].

داده‌کاوی واژه مبهمی است که به فرآیند پیدا کردن اطلاعات جالبی در مخازن بزرگی از داده‌ها اشاره دارد. به‌طور دقیق‌تر، این اصطلاح از الگوریتم‌های خاص بر اساس اصول درست در رشته‌های متعدد ازجمله آمار، هوش مصنوعی، یادگیری ماشین، علوم پایگاه داده و بازیابی اطلاعات استفاده می‌کند.

الگوریتم‌های داده‌کاوی در روش‌های حرفه‌ای مختلف مورد استفاده قرار می‌گیرند که داده‌کاوی، استخراج دانش، کشف داده‌ها از بین رفته و یادگیری استقرایی یا قیاسی نامیده می‌شوند. تکنیک‌های داده‌کاوی را می‌توان در طیف وسیعی از انواع داده‌ها ازجمله پایگاه‌های داده، متن، اطلاعات مکانی، اطلاعات زمانی، عکس‌ها و سایر داده‌های پیچیده بکار گرفت [۶].

به‌طورکلی می‌توان گفت، داده‌کاوی فرآیندی است که با استفاده از ابزارهای تجزیه و تحلیل داده‌ها به دنبال کشف الگوها و روابط معتبری که تاکنون ناشناخته بوده‌اند می‌گردند، این فرآیند منحصر به گردآوری و مدیریت داده‌ها نبوده و تجزیه و تحلیل اطلاعات و پیش‌بینی را نیز شامل می‌شود. فرآیند داده‌کاوی است متشکل از چند مرحله ازجمله انتخاب داده‌ها برای تجزیه و تحلیل، تهیه داده‌ها، استفاده از الگوریتم‌های استخراج از معادن داده و سپس تفسیر و ارزیابی نتایج می‌باشد.

داده‌کاوی از چندین رشته علمی به‌طور همزمان بهره می‌برد نظیر تکنولوژی پایگاه داده، هوش مصنوعی، شبکه‌های عصبی، آمار، سیستم‌های مبتنی بر دانش، بازیابی اطلاعات. در عصر انفجار اطلاعات شرکت‌های فردی هرروزه حجم زیادی از داده را تولید جمع‌آوری خواهند کرد. استخراج اطلاعات مفید از پایگاه داده و تبدیل کردن اطلاعات به نتایج عملی چالش اصلی است که شرکت‌ها با آن روبرو هستند. با توجه به پیشرفت کشور در زمینه فن‌آوری اطلاعات و نگاه‌های ویژه به دولت الکترونیک و نفوذ استفاده از سیستم‌های کامپیوتری

در صنعت و ایجاد بانک‌های اطلاعاتی بزرگ توسط ادارات دولتی، در بانک‌ها و بخش‌های خصوصی نیاز به استفاده از داده‌کاوی به‌طور عمیق احساس می‌شود [۵].

به دلیل حجم گسترده و بسیار زیاد داده‌های موجود در بانک و همچنین عدم امکان ذخیره‌سازی آن‌ها و نیاز به پردازش و استخراج پویای اطلاعات و دانش نهفته‌ها در داده‌ها بحث داده‌کاوی جریان داده‌ها در بانک طی سال‌های اخیر بسیار مورد توجه قرار گرفته است. ارتباط با مشتری یکی از مهم‌ترین عوامل سازنده مخصوصاً در سازمان‌های خدماتی می‌باشد از جمله این سازمان‌ها، بانک‌ها و مؤسسات مالی است که مدیریت ارتباط با مشتری در آن‌ها نقش کلیدی را دارا می‌باشد. در بانکداری نوین، بانک‌ها برای شناسایی مشتریان و نیازهای آن‌ها دست به بازاریابی تک‌به‌تک و بازاریابی بر اساس پایگاه داده می‌زنند. بازاریابی تک‌به‌تک، ایجاد و مدیریت رابطه فردی با تک‌تک مشتریان می‌باشد. امروزه به دلیل بهره‌گیری از فناوری کامپیوتری و استفاده از پایگاه داده‌ها و داده‌کاوی این امکان فراهم است که بانک برای تک‌به‌تک مشتریان خود برنامه‌ای داشته باشند و بازاریابی تک‌به‌تک در مقیاس وسیعی قابل اجرا و اقتصادی در نظر گرفته شود [۵].

### مدل «تاپسیس»<sup>۱۶</sup>

روش تاپسیس یکی از شاخه‌های تصمیم‌گیری چند معیاره بوده و یک تکنیک کاربردی و عملی برای رتبه‌بندی و انتخاب تعدادی گزینه می‌باشد. این روش بر این مفهوم استوار است که گزینه انتخاب‌شده باید کوتاه‌ترین فاصله را از راه‌حل ایده‌آل و دورترین فاصله را از راه‌حل ایده‌آل منفی داشته باشد.

تاپسیس از نوع روش‌های جبرانی می‌باشد و در آن معیارها نسبت به هم جبرانی هستند، یعنی نقطه‌ضعف موجود در یک معیار می‌تواند توسط نقطه قوتی در معیار دیگر جبران شود.

برای به دست آوردن یک دید کلی، تعدادی از عملگرها از قبیل نرمال‌سازی، روش‌های محاسبه فواصل و عملگرهای میانگین‌گیری مطابق با هر مرحله از روش تاپسیس ارائه می‌شوند. علاوه بر آن ترجیحات بیش از یک تصمیم‌گیرنده در روش تاپسیس با یکدیگر ادغام می‌شوند [۶].

### تعاریف اولیه

در این بخش تعدادی از تعاریف مجموعه‌های فازی ارائه می‌شود.

**تعریف ۱:** اگر  $X$  یک مجموعه از اشیاء باشد آنگاه یک مجموعه فازی  $\tilde{A}$  در  $X$  یک مجموعه از زوج‌های مرتب به صورت زیر می‌باشد:

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) \mid x \in X\} \quad (1-2)$$

$\mu_{\tilde{A}}(x)$  تابع عضویت یا درجه  $x$  در  $\tilde{A}$  نامیده می‌شود، که متناظر با هر عضو  $x$  در  $X$  یک عدد حقیقی در فاصله  $[0,1]$  اختیار می‌کند.

**تعریف ۲:** عدد فازی  $\tilde{M}$  از نوع  $LR$  گفته می‌شود اگر و تنها اگر:

$$\mu_{\tilde{M}}(x) = \begin{cases} L\left(\frac{m-x}{m-l}\right) & l \leq x \leq m \\ R\left(\frac{x-m}{u-m}\right) & m < x \leq u \\ 0 & \text{سایر نقاط} \end{cases} \quad (2-2)$$

و با  $LR = (l, m, u)$  نشان می‌دهند که  $L$  و  $R$  دو تابع دلخواه باشند.

**تعریف ۳:** یک عدد فازی دوزنقه‌ای  $\tilde{M}$  نوع خاصی از اعداد  $LR$  است که به صورت  $(M_1, M_2, M_3, M_4)$  نمایش داده می‌شود. تابع عضویت این نوع اعداد فازی به صورت زیر تعریف می‌شود:

$$\mu_{\tilde{M}}(x) = \begin{cases} \frac{x-M_1}{M_2-M_1} & M_1 \leq x < M_2 \\ \frac{x-M_4}{M_3-M_4} & M_2 \leq x < M_3 \quad M_3 \leq x < M_4 \\ 1 & \text{سایر نقاط} \end{cases} \quad (3-2)$$

هرگاه در یک عدد فازی دوزنقه‌ای  $M_2 = M_3$ ، آنگاه  $\tilde{M}$  یک عدد فازی مثلثی نامیده می‌شود. عدد غیر فازی  $r$  را می‌توان به صورت  $(r, r, r, r)$  بیان نمود.

بر اساس اصول بیان شده در مجموعه‌های فازی، جمع و تفریق هر دو عدد فازی دوزنقه‌ای نیز یک عدد فازی دوزنقه‌ای می‌باشد، اما ضرب هر دو عدد فازی تنها تخمینی از یک عدد فازی دوزنقه‌ای می‌باشد. دو عدد فازی دوزنقه‌ای مثبت  $\vec{M} = (M_1, M_2, M_3, M_4)$  و  $\vec{N} = (N_1, N_2, N_3, N_4)$  و یک عدد حقیقی  $r$  را در نظر بگیرید، عملیات اصلی روی اعداد فازی دوزنقه‌ای  $\vec{M}$  و  $\vec{N}$  به صورت زیر می‌باشد:

$$\begin{aligned} -\vec{N} &= (-N_4, -N_3, -N_2, -N_1) \\ \vec{M} \oplus \vec{N} &= [M_1+N_1, M_2+N_2, M_3+N_3, M_4+N_4] \\ \vec{M} \ominus \vec{N} &= [M_1-N_4, M_2-N_3, M_3-N_2, M_4-N_1] \\ \vec{M} \otimes_r &= [M_1r, M_2r, M_3r, M_4r] \end{aligned} \quad (4-2)$$

اگر  $\vec{M} > 0$  و  $\vec{N} > 0$  آنگاه:

$$\vec{M} \otimes \vec{N} = [M_1N_1, M_2N_2, M_3N_3, M_4N_4]$$

**تعریف ۴:** برای یک عدد فازی  $\vec{M}$ ، مجموعه  $\alpha$  - برش عدد فازی  $\vec{M}$  با  $\alpha \in [0, 1]$  به صورت زیر تعریف می‌گردد:

$$(\alpha \geq (x)) R[\mu_{\vec{M}}] = \{x \in \alpha [\vec{M}]\} \quad (5-2)$$

برای یک عدد فازی دوزنقه‌ای  $\vec{A} = (a, b, c, d)$ ، فواصل حاصل شده توسط مجموعه  $\alpha$  - برش‌ها را به صورت زیر تعریف می‌نماییم:

$$= [(b-a)\alpha + a, (c-d)\alpha + d] = [a(\alpha), c(\alpha)] A(\alpha) \quad (6-2)$$

برای انتخاب معیارها سعی می‌کنیم ۵ اصل را مدنظر قرار دهیم:

- (۱) کامل بودن
- (۲) قابل استفاده بودن
- (۳) تجزیه پذیر بودن
- (۴) غیر فراوانی
- (۵) دارای اندازه می‌نیم بودن

تمامی درجه‌های اختصاص داده شده به حالت‌ها در ارتباط با هر معیار به شکل ماتریس  $D = (x_{ij})_{n \times m}$  تصمیم نمایش داده می‌شود. به طور خلاصه روش تاپسیس دارای گام‌های زیر است:

۱. نرمال کردن ماتریس تصمیم‌گیری  $D$  توسط یکی از روش‌های نرمال سازی
۲. محاسبه کردن ماتریس تصمیم‌گیری نرمال شده وزن دار  $V = (v_{ij})_{n \times m}$  توسط فرمول زیر:

$$v_{ij} = w_j r_{ij}, i=1, \dots, n, j=1, \dots, m \quad (7-2)$$

که در آن  $w_j$  وزن نسبی معیار  $j$ ام و  $r_{ij}$  مقدار نرمال شده می‌باشد.

۳. محاسبه کردن راه‌حل ایده‌آل و راه‌حل ایده‌آل منفی، راه‌حل ایده‌آل منفی به ترتیب برابرند با:

$$A^* = \{v_1^*, \dots, v_m^*\} = \{(\max_j d_{ij}, j \in \Omega_b), (\min_j a_{ij}, j \in \Omega_c)\}, \quad (8-2)$$

$$A^- = \{v_1^-, \dots, v_m^-\} = \{(\min_j a_{ij}, j \in \Omega_b), (\max_j d_{ij}, j \in \Omega_c)\}, \quad (9-2)$$

که در آن  $\Omega_b$  و  $\Omega_c$  به ترتیب مجموعه معیارهای سود و هزینه می‌باشند.

۴. محاسبه کردن فاصله هر حالت از راه‌حل ایده‌آل و ایده‌آل منفی:

$$D_i^* = \sqrt{\sum_{j=1}^m (v_{ij} - v_j^*)^2}, \quad i=1, \dots, n \quad (10-2)$$

$$D_i^- = \sqrt{\sum_{j=1}^m (v_{ij} - v_j^-)^2}, \quad i=1, \dots, n \quad (11-2)$$

۵. محاسبه کردن «نزدیکی نسبی»<sup>۱۷</sup> هر حالت به راه‌حل ایده‌آل.

نزدیکی نسبی حالت  $A_i$  نسبت به  $A^*$  چنین تعریف می‌شود:

$$RC_i = \frac{D_i^-}{D_i^* + D_i^-} \quad (12-2)$$

۶. رتبه‌بندی حالت‌ها با توجه به نزدیکی نسبی آن‌ها. حالت‌ها را به ترتیب نزولی مرتب می‌کنیم:

تفاوت عمده در روش‌های تاپسیس به‌طور خلاصه در انتخاب روش نرمال‌سازی ماتریس تصمیم‌گیری و انتخاب راه‌حل ایده‌آل مثبت فازی و راه‌حل ایده‌آل منفی فازی و محاسبه فاصله بین اعداد فازی است. در بیشتر روش‌ها در ابتدا یا در میانه یا در آخر کار از یک روش فازی زدایی استفاده می‌شود [۶].

روش‌های متعددی برای رتبه‌بندی اعداد فازی در ادبیات موضوع وجود دارد.

در بسیاری از مقالات مروری جامع از روش‌های موجود می‌باشد. عده‌ای از این روش‌ها با استفاده از مفهوم مقدار میانی و بعضی دیگر

اعداد فازی را با مفهوم مجموعه‌های  $\alpha$ -برش، مقایسه می‌نمایند، علاوه بر آن‌ها تعدادی روش‌های رتبه‌بندی امکان اهداد فازی وجود دارد. روش دیگر رتبه‌بندی فازی که تمرکز ما در این مقاله بر اساس آن می‌باشد، رتبه‌بندی بر اساس فاصله نام دارد.

«سعه و شوازلندر»<sup>۱۸</sup> با استفاده از مفهوم فاصله به مقایسه اعداد فازی پرداخته‌اند آن‌ها برای مرتب کردن اعداد فازی تنها از مقادیر

غیرمنفی استفاده نموده‌اند [۷]. «یائو و وو»<sup>۱۹</sup> در  $[\lambda]$  فاصله  $a^*$  را طوری لحاظ نموده‌اند که  $d^*(a, 0) = a$  و  $d^*(a, b) = a - b$  برای

تمام  $a, b \in R$  آنگاه برای

$H(R)$ ،  $\tilde{A}, \tilde{B} \in H(R)$  خانواده اعداد فازی روی  $R$ ، فاصله را به‌صورت زیر تعریف می‌نمایند:

$$d(\tilde{A}, \tilde{B}) = \frac{1}{2} \int_0^1 ([\tilde{A}]_\alpha^L + [\tilde{A}]_\alpha^U - [\tilde{B}]_\alpha^L - [\tilde{B}]_\alpha^U) d\alpha \quad (13-2)$$

که  $[\tilde{A}]_\alpha^L$ ،  $[\tilde{A}]_\alpha^U$ ،  $[\tilde{B}]_\alpha^L$  و  $[\tilde{B}]_\alpha^U$  به ترتیب نشان دهنده حد پایین و بالا اعداد فازی  $\tilde{A}$  و  $\tilde{B}$  می‌باشند.

بر اساس رابطه (۱۳-۲) آن‌ها ثابت کرده‌اند که سیستم رتبه‌بندی روی  $H(R)$  به‌صورت زیر تعریف می‌گردد: [۷].

$$\tilde{B} < \tilde{A} \text{ if } d(\tilde{A}, \tilde{B}) > 0, \tilde{B} > \tilde{A} \text{ if } d(\tilde{A}, \tilde{B}) < 0, \tilde{B} \approx \tilde{A} \text{ if } d(\tilde{A}, \tilde{B}) = 0 \quad (14-2)$$

## داده کاوی مورد استفاده در بانکداری الکترونیک

### درخت تصمیم CART

این روش که موجب تشکیل یک درخت تصمیم با تقسیمات دوتایی می‌گردد، توسط «بریمان و همکارانش»<sup>۲۰</sup> به‌طور کامل معرفی شد. این روش برای متغیرهای کمی طراحی گردیده ولی قابل استفاده برای هر نوع متغیری است. بر اساس این الگوریتم، نرم‌افزار آماری

<sup>17</sup> Relative Closeness

<sup>18</sup> Saade & Schwarzlander

<sup>19</sup> Yao & Wu

<sup>20</sup> Beriman et al

تحت نام CART نیز ساخته شده است که از شناخته شده ترین برنامه‌ها است. در این روش و برای متغیر پاسخ کیفی، «شاخص جینی»<sup>۲۱</sup> به عنوان معیاری برای انتخاب متغیرهای مناسب، معرفی شده است.

درواقع درخت تصمیم‌گیری یک روش تقسیم‌بندی بازگشتی «باینری»<sup>۲۲</sup> است. داده‌ها در این الگوریتم به صورت خام استفاده می‌شوند و هیچ‌گونه پاک‌سازی نه‌نیاز 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» به وسیله 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. به جای 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 [۶].

خلیلی و پاشازاده (۱۳۹۳) در مقاله‌ای به طبقه‌بندی داده‌های حساب‌های جاری بانک با استفاده از درخت تصمیم، پرداختند. نتایج آنان نشان داد که درخت تصمیم رگرسیون در طبقه‌بندی داده‌های حساب‌های جاری بانک با استفاده از نرم‌افزار «راپیدمینر»<sup>۲۳</sup> دقت آن در مقایسه با سایر روش‌های دسته‌بندی مطلوب می‌باشد [۹].

«پلیها»<sup>۲۴</sup> (۲۰۰۷) در مقاله‌ای به روش‌ها و سیستم‌هایی برای سوددهی مشتریان بانکی جهت مطابقت با رفتار مالی تعیین شده توسط بانک با استفاده از درخت تصمیم‌گیری و روند تصادفی پرداخت. یک سیستم سوددهی مالی با استفاده از درخت تصمیم رگرسیون برای پیش‌بینی رفتار مالی یک مشتری بانکی که دارای یک حساب سپرده مستقیم است، پیشنهاد شد. استفاده از درخت تصمیم رگرسیون برای سیستم نمره دهی مالی می‌تواند در تعیین فعالیت‌های حساب کاربری سودمند و غیر سودآور و گروه‌های مشتری کمک کند. نمره دهی رفتار مالی با استفاده از درخت طبقه‌بندی رگرسیون به بانک‌ها کمک می‌کند تا خرده‌فروشان را جذب کنند تا ابتکارات بازاریابی را با همکاری و حمایت مالی خود بر اساس نمره مالی که برای یک گروه خاص از مشتریان بانک به دست می‌آورند، جذب کند [۱۰].

«سوساک و بنسیک»<sup>۲۵</sup> (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 به دست آمده 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 به دست آمده است. بهترین مدل استخراج مهم‌ترین ویژگی‌های کسب‌وکار کوچک کسب‌وکار از داده‌های مشاهده شده است [۱۱].

با توجه به مرور مقالات و کاربردهای این الگوریتم، می‌توان نتیجه گرفت که الگوریتم CART را می‌توان تقریباً در هر حوزه‌ای یافت ولی به طور وسیع در زمینه‌هایی مانند مهندسی الکترونیک، زیست‌شناسی، مطالعات پزشکی و مباحث اقتصادی یافت می‌شود. ولی بیشتر این الگوریتم در تحقیقات بازاریابی یا اجتماعی، بکار برده شده است.

### الگوریتم CHAID

این الگوریتم یک روش درخت تصمیم‌گیری است که بر اساس آزمون معنی‌دار تنظیم شده «آزمایش بونفرونی»<sup>۲۶</sup> است. این تکنیک در آفریقای جنوبی توسعه یافت و توسط «گوردون واکس»<sup>۲۷</sup> 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»<sup>۲۸</sup> ایالات متحده و «THAID» است. در عمل، CHAID اغلب در زمینه بازاریابی مستقیم برای انتخاب گروهی از

21 . Gini Index

22 . Binery

23 . Rapidminer

24 . Pliha

25 . Susac & Bencik

26 . Bonfrey Test

27 . Gordon Wax

28 . Interaction Automatic Detection

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 [۱۲].

کوثری و همکاران (۱۳۹۲) در پژوهشی به ارائه مدلی جهت کشف رفتارهای مشکوک در بانکداری الکترونیکی با استفاده از الگوریتم‌های درخت تصمیم‌گیری، پرداختند. در این پژوهش، ابتدا متغیرهای مؤثر در تولید قوانین رفتاری تعیین شده است و در نهایت، روند چهار الگوریتم CHAID، ex\_CHAID، C4.5 و C5.0 مورد مقایسه قرار گرفته است. نتایج پژوهش نشان می‌دهد که الگوریتم CHAID می‌تواند به‌عنوان روش ماشینی مطمئن جهت کشف الگوهای مشکوک موجود روی تراکنش‌های بانک محسوب شود [۱۳].

«کایونکویل و اوزگلباس»<sup>۲۹</sup> (۲۰۰۹)، در پژوهشی به مدل‌سازی خطر توسط الگوریتم درخت تصمیم CHAID در بانکداری الکترونیکی پرداختند. در این مقاله، یک مدل داده‌کاوی برای تشخیص شاخص‌های ریسک مالی و عملیاتی در بانک با درخت تصمیم‌گیری CHAID ارائه شده است. شناسایی عوامل خطر با روشن شدن رابطه بین متغیرها، کشف دانش از متغیرهای مالی و عملیاتی را مشخص می‌کند. در طول شکل‌گیری مدل، نتایج نشان داد که الگوریتم CHAID مدلی آسان برای تفسیر پیش‌زمینه نظری با کشف روابط ضمنی بین داده‌ها در بانکداری الکترونیکی و شناسایی اثر عامل است [۱۴].

بنابراین می‌توان گفت که این الگوریتم برای استفاده در مورد متغیرهای کیفی است که می‌تواند برای متغیرهای کمی گروه‌بندی شده نیز استفاده شود. در هر گروه، می‌توان بیش از دو تقسیم نیز داشت. در این روش از مقدار P-Value آماره کای-دو مربوط به آزمون استقلال جداول توافق استفاده می‌شود. از بین متغیرهای موجود، متغیری که دارای P-Value کوچک‌تری باشد در مرحله اول برای تقسیمات روی یک گروه در نظر گرفته می‌شود. ضعف این الگوریتم عدم توانایی آن در ایجاد بهترین تقسیمات ممکن بر اساس متغیرهای موجود است. از این الگوریتم بیشتر در پیش‌بینی‌های هواشناسی، مدیریت مالی استفاده می‌شود.

### الگوریتم C5

الگوریتم C5 یک نوع درخت تصمیم‌گیری تک متغیره و بهبودیافته الگوریتم C4.5 است. این الگوریتم مشابه با درخت طبقه‌بندی رگرسیون ابتدا درختی تقریباً بر 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 [۹].

لنگری و همکاران (۲۰۱۴) در مقاله‌ای به کارگیری الگوریتم‌های درخت تصمیم‌گیری جهت کشف رفتارهای مشکوک در بانکداری اینترنتی پرداختند. در این مقاله، ابتدا متغیرهای مؤثر در تولید قوانین رفتاری تعیین شده است و در نهایت، روند چهار الگوریتم CHAID، ex\_CHAID، C4.5 و C5.0 مورد مقایسه قرار گرفته 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 [۱۵].

«سولین»<sup>۳۰</sup> (۲۰۰۹) در مقاله‌ای به الگوریتم طبقه‌بندی C5 و کاربرد آن در ارزیابی اعتبار فردی بانک‌ها پرداختند. این مقاله بر روی ارزیابی اعتباری فردی بانک تجاری تمرکز دارد. پرونده اعتبار فردی شامل داده‌های عددی و غیر عددی می‌باشد. درخت تصمیم‌گیری یک راه‌حل مناسب برای این نوع مسئله است. این الگوریتم C4.5 درخت تصمیم‌گیری محبوب است، اما الگوریتم C5 هنوز در حال تکامل است. در این مقاله، ما تحقیق عمیقی در مورد الگوریتم C5 با جاسازی تکنولوژی «افزایش» در ماتریس هزینه و درخت حساس برای ایجاد یک مدل جدید برای ارزیابی اعتباری فردی بانک تجاری انجام می‌دهیم. ما مدل جدیدی را برای ارزیابی سوابق اعتباری فردی بانک آلمانی اعمال می‌کنیم و نتایج حاصل از مدل درخت تصمیم‌گیری تصحیح‌شده و اصلی را مقایسه می‌کنیم. مقایسه نشان می‌دهد که مدل درخت تصمیم C5 دقیق‌تر است [۱۶].

از این الگوریتم در بسیاری از برنامه‌های کاربردی، مجموعه قوانین ترجیح داده‌شده، استفاده شده است چراکه درک آن‌ها نسبت به درخت‌های تصمیم‌گیری، ساده‌تر است، استفاده از این الگوریتم بیشتر در حوزه‌های پزشکی و مدیریت الکترونیک می‌باشد.

<sup>29</sup> . Koyuncugil & Ozgulbas

<sup>30</sup> .Su-Lin

**الگوریتم SVM**

الگوریتم ماشین بردار پشتیبان اولیه توسط «ولادیمیر واپنیک»<sup>۳۱</sup> ابداع شد. یکی از روش‌های «یادگیری با نظارت»<sup>۳۲</sup> است که از آن برای «طبقه‌بندی»<sup>۳۳</sup> و رگرسیون استفاده می‌کنند. این روش از جمله روش‌های نسبتاً جدیدی است که در سال‌های اخیر کارایی خوبی نسبت به روش‌های قدیمی‌تر برای طبقه‌بندی از جمله «شبکه‌های عصبی پرسپترون»<sup>۳۴</sup> نشان داده است. مبنای کاری 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 هر جایی که نیاز به تشخیص الگو یا دسته‌بندی اشیاء در کلاس‌های خاص باشد می‌توان استفاده کرد [۱۷].

«چنگ و همکاران»<sup>۳۵</sup> (۲۰۰۶) در تحقیقی به مدل‌سازی یک روش «هیبرید»<sup>۳۶</sup> برای اعتبارسنجی پرداخته می‌شود. با این صورت که ترکیب الگوریتم ژنتیک و ماشین بردار پشتیبان هم برای انتخاب خصیصه‌ها و هم برای بهینه‌سازی پارامترهای مدل اعتبارسنجی بکار گرفته می‌شود. مقایساتی نیز بین روش ماشین بردار پشتیبان و طبقه‌بندی شبکه‌های عصبی، برنامه‌ریزی ژنتیکی و درخت تصمیم صورت گرفته که نتایج نسبتاً یکسانی حاصل می‌شود. ولی نتایج تجربی نشان می‌دهد که ماشین بردار پشتیبان، نسبت به روش‌های موجود داده‌کاوی، نتایج امیدبخش‌تری را ارائه می‌دهد [۱۸].

طلوعی اشقلی و همکاران (۱۳۸۸) در پژوهشی به امتیازدهی اعتباری متقاضیان کارت‌های اعتباری بانک‌ها با استفاده از تکنیک ماشین بردار پشتیبان پرداختند. در این مقاله مدل ترکیبی ماشین بردار پشتیبان مبتنی بر دو استراتژی زیر برای امتیازدهی اعتباری متقاضیان کارت‌های اعتباری بکار رفته است: ۱- استفاده از تکنیک F-Score برای انتخاب ویژگی‌ها ۲- استفاده از تکنیک «جستجوی گریدی»<sup>۳۷</sup> برای بهینه کردن مقادیر پارامترها. نتایج نشان‌دهنده برتری ماشین بردار پشتیبان نسبت به دیگر الگوریتم‌ها می‌باشد [۱۹].

شربانی 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، به منظور پیش‌بینی عملکرد مالی مشتریان حقوقی بانک‌ها ارائه گردیده است. در واقع، در این نوشتار ماشین بردار پشتیبان به همراه دیگر مکانیزم‌ها از جمله تکنیک‌های جستجوی گریدی جهت طبقه‌بندی متقاضیان تسهیلات اعتباری بانکی و افزایش کارایی مدل استفاده شده 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 [۲۰].

«شوویی و همکاران»<sup>۳۸</sup> (۲۰۱۳) پیش‌بینی خطر سیستماتیک بانکی بر اساس ماشین بردار پشتیبان، با توجه به خصوصیات غیرخطی پیچیده بانکداری سیستماتیک بانکی، در این مقاله، ما از ماشین بردار پشتیبانی برای پیش‌بینی خطرات سیستماتیک بانکی استفاده کردیم نتایج نشان داد که مدل پیش‌بینی مبتنی بر ماشین بردار پشتیبان یک روش کارآمد در این مورد است [۲۱].

این روش از جمله روش‌های نسبتاً جدیدی است که در سال‌های اخیر کارایی خوبی نسبت به روش‌های قدیمی‌تر برای طبقه‌بندی از جمله شبکه‌های عصبی پرسپترون نشان داده است. الگوریتم ماشین بردار پشتیبان، جز الگوریتم‌های تشخیص الگو دسته‌بندی می‌شود. از الگوریتم ماشین بردار پشتیبان، در هر جایی که نیاز به تشخیص الگو یا دسته‌بندی اشیاء در کلاس‌های خاص باشد می‌توان استفاده کرد.

**الگوریتم «نایوبیز»<sup>۳۹</sup>**

یک الگوریتم یادگیری ساده است که از قاعده بیز به همراه فرض محکمی که صفات با توجه به کلاس از نظر شرطی مستقل هستند، استفاده می‌کند. اگرچه این فرض استقلال در عمل اغلب نقض می‌شود، با این وجود، 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 به صورت گسترده مورد استفاده قرار بگیرد.

«هسیه و همکاران»<sup>۴۰</sup> (۲۰۱۰) در مطالعه‌ای به طبقه‌بندی متقاضیان اعتبار در سه گروه اعتباری خوب، بد و مرزی می‌پردازد. با توجه به اهمیت روش‌های ترکیبی داده‌کاوی در طبقه‌بندی، این پژوهش ابتدا مشتریان را در خوشه‌های همگن قرار داده است. پس از طی نمودن

31 . Vladimir Vapnik

32 . Supervised learning

33 . Classification

34 . Perceptron neural networks

35 . Cheng et al

36 . Hybrid

37 . Grid search

38 . Shuei et al

39 . naïve Bayes

40 . Hsieh et al

گام‌های پیش‌پردازش داده‌ها، متغیرهایی که مقادیر پیوسته دارند را به گسسته تبدیل نموده است. سپس سه روش شبکه‌های عصبی، شبکه بیزین و ماشین بردار پشتیبان را به منظور ترکیب الگوریتم‌ها بکار گرفته است. نهایتاً یک سیستم هیبریدی برای رتبه‌بندی اعتباری مشتریان ارائه نموده است [۲۲].

«آکاروسی و پارلار»<sup>۴۱</sup> (۲۰۱۷)، در پژوهشی به استفاده از داده‌کاوی برای شناسایی ویژگی مهم بازاریابی در بانکداری الکترونیکی، پرداختند. در این مقاله تکنیک‌های داده‌کاوی برای تفسیر و تعریف ویژگی‌های مهم مورد استفاده قرار می‌گیرند. در این مقاله از دو روش انتخاب ویژگی یعنی به دست آوردن اطلاعات و روش‌های «کای اسکوتر»<sup>۴۲</sup> برای انتخاب ویژگی‌های مهم استفاده شد. روش‌ها با استفاده از یک الگوریتم یادگیری دستگاه نظارت نایبیز مقایسه شدند. این نتایج نشان می‌دهد که کاهش مجموعه ویژگی‌های عملکرد می‌تواند طبقه‌بندی را بهبود بخشد [۲۳].

«ویجایا کامبل و همکاران»<sup>۴۳</sup> (۲۰۱۷) در مقاله‌ای به تشخیص وب‌سایت «فیشینگ»<sup>۴۴</sup> بانکداری الکترونیک با استفاده از طبقه‌بندی «نایبیز» پرداختند. این تحقیق رویکردی را برای غلبه بر مشکلات ناشی از حملات یا تقلب و پیچیدگی آنلاین در تشخیص و پیش‌بینی وب‌سایت «فیشینگ» بانکداری الکترونیک ارائه می‌دهد. ما یک مدل مؤثر ارائه دادیم که مبتنی بر استفاده از الگوریتم‌های داده‌کاوی است. از این‌رو، نتایج نشان داد که استفاده از روش طبقه‌بندی «نایبیز» در برنامه‌های واقعی و عملکرد بهتر آن در مقایسه با سایر الگوریتم‌های طبقه‌بندی سنتی بهتر عمل می‌کند [۲۴].

«ودالا و کومار»<sup>۴۵</sup> (۲۰۱۲) در مقاله‌ای به استفاده از طبقه‌بندی «نایبیز» برای امتیاز اعتبار در پلت فرم الکترونیکی اعطای وام پرداختند. نتایج آنان نشان داد که استفاده از روش طبقه‌بندی «نایبیز» برای پیش‌بینی احتمالات پیش‌فرض وام‌گیرندگان مناسب خواهد بود [۲۵].

با وجود مسائل طراحی و پیش‌فرض‌هایی که در خصوص روش بیز وجود دارد، این روش برای طبقه‌بندی کردن بیشتر مسائل در جهان واقعی، مناسب است. در چندین سال پیش «شبکه‌های بیزین» توسط افراد مورد توجه قرار گرفتند و به‌عنوان الگوریتم طبقه‌بندی در بانکداری الکترونیکی توسط افرادی به کار گرفته شدند.

### الگوریتم تفکیک‌کننده»<sup>۴۶</sup>

«تشخیص خطی فشر»<sup>۴۷</sup> روش‌های آماری هستند که از جمله در یادگیری ماشین و بازشناخت الگو برای پیدا کردن ترکیب خطی خصوصیتی 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. باین‌حال 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. از این‌رو 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. باین‌حال این روش‌ها تفاوت عمده‌ای باهم دارند: در آنالیز افتراقی خطی، تفاوت کلاس‌ها مدل‌سازی می‌شود درحالی‌که در تحلیل مؤلفه‌های اصلی تفاوت کلاس‌ها نادیده گرفته می‌شود [۲۶].

«بن‌پام»<sup>۴۸</sup> (۲۰۱۳) در مقاله‌ای به تجزیه و تحلیل با استفاده از الگوریتم تفکیک‌کننده در پیش‌بینی ورشکستگی بانک‌ها در نیجریه پرداخت. نتایج حاکی از آن بود که «مدل MDA»<sup>۴۹</sup> و با استفاده از الگوریتم تفکیک‌کننده همچنان یک ابزار قدرتمند در پیش‌بینی پتانسیل ورشکستگی است. متغیرهای کلیدی در مدل «آلتمن» شاخص‌های مثبت در تجزیه و تحلیل هستند و سازمان‌های نظارتی در اجرای نتایج حاصل از تجزیه و تحلیل دقیق نیستند. در نتیجه تشخیص الگوریتم تفکیک‌کننده نسبت به دیگر روش‌ها از دقت بالاتری برخوردار است [۲۷].

41. Acaroci & Parllar

42. Chi-square

43. Vijaya Kamble et al

44. Phishing

45. Vedala Komar

46. Discriminant

47. Fisher's Linear Detection

48. Ben-Pam

49. Model Driven Architecture

«آنته و آنا»<sup>۵۰</sup> (۲۰۱۳) در مقاله‌ای به تجزیه و تحلیل الگوریتم تفکیک کننده برای سطح سوددهی بانک پرداختند. در مجموع، تحلیل تفکیک کننده به عنوان یک روش آماری مناسب برای حل مسئله تحقیق ارائه شده در سوددهی و جلوگیری از ورشکستگی، رتبه بندی اعتباری یا مسائل پیش فرض در امور مالی است [۲۸].

از این الگوریتم بیشتر در حوزه های مالی استفاده شده است و در دیگر حوزه های مدیریت و مدیریت الکترونیک از دقت کمی برخوردار بوده است.

## QUEST

این الگوریتم توسط «لو و شی»<sup>۵۱</sup> برای متغیرهای پاسخ اسمی طراحی شد. درخت رده بندی حاصل از این الگوریتم نظیر مدل درخت طبقه بندی رگرسیون دارای تقسیمات دوتایی بوده و ملاک تصمیم برای انتخاب متغیرها با استفاده از مقدار P-Value مربوط به آماره F آزمون «آنوا»<sup>۵۲</sup> برای متغیرهای کمی و P-Value آماره کای-دو مربوط به جداول توافقی برای متغیرهای کیفی صورت می پذیرد. این الگوریتم با توجه به این که از مقدار P-Value برای تصمیم گیری استفاده می نماید، موجب تشکیل درختی ناریب از متغیرها می گردد. این الگوریتم ضمن حفظ دقت برآورد در مدل درخت طبقه بندی رگرسیون، از سرعت بالاتری در معرفی یک درخت رده بندی نسبت به آن برخوردار است.

«کیزیس و همکاران»<sup>۵۳</sup> (۲۰۱۵) در پژوهشی به تلاش برای پایداری بانکداری الکترونیکی با استفاده از الگوریتم Quest پرداختند. نتایج نشان داد که استفاده از الگوریتم QUEST می تواند در کاهش ارزش وثیقه وام و همچنین پایداری بانک مؤثر بوده و نسبت به دیگر الگوریتم ها از صحت بیشتری برخوردار است [۲۹].

این الگوریتم ضمن حفظ دقت برآورد در مدل CART، از سرعت بالاتری در معرفی یک درخت رده بندی نسبت به آن برخوردار است. از این الگوریتم در حوزه های پزشکی، مدیریت مالی و کامپیوتر استفاده شده است.

## الگوریتم شبکه های عصبی مصنوعی

شبکه های عصبی سیستم ها و روش های محاسباتی نوین برای یادگیری ماشینی، نمایش دانش و در انتها اعمال دانش به دست آمده در جهت پیش بینی پاسخ های خروجی از سامانه های پیچیده هستند. ایده اصلی این گونه شبکه ها تا حدودی الهام گرفته از شیوه کارکرد سیستم عصبی زیستی برای پردازش داده ها و اطلاعات به منظور یادگیری و ایجاد دانش قرار دارد. عنصر کلیدی این ایده، ایجاد ساختارهایی جدید برای سامانه پردازش اطلاعات است.

ودادی (۱۳۸۳) در رساله ای به مدل سازی رفتار اعتباری مشتریان بانک مسکن با استفاده از شبکه های عصبی جهت تخصیص بهینه تسهیلات خرید و ساخت مسکن می پردازد. محقق در این تحقیق ۳۵۰ نمونه انتخابی را در سه دسته مشتریان خوش حساب، سر رسید گذشته و بد حساب تقسیم می نماید. در این تحقیق مقایسه ای بین قدرت تفکیک مدل های شبکه عصبی و مدل آنالیز ممیزی در مورد داده های مشتریان صورت گرفته است و در نهایت نتایج حاصل از دو مدل نشان داده که مدل های رتبه بندی اعتباری با استفاده از شبکه های عصبی نسبت به مدل آنالیز ممیزی، از قدرت تفکیک (صحت پیش بینی) بیشتری برخوردار است [۳۰].

«آلیس و همکاران»<sup>۵۴</sup> (۲۰۰۰) در مقاله ای به استفاده از داده کاوی برای پایگاه داده بازاریابی آنلاین «بانک گارانتی»<sup>۵۵</sup> پرداختند. بانک گارانتی دو برنامه عمده را در این مقاله انجام داد: تقسیم بندی مشتری و به ثمر رساندن پایگاه داده برای اهداف بازاریابی. «یک انبار داده»<sup>۵۶</sup> مرکزی که به خوبی نگهداری می شود، امکان تهیه داده ها را به گونه ای کارآمد برای برنامه های کاربردی فراهم می کند. تقسیم بندی مشتری با الگوریتم جمعیت شناختی انجام شد. پایگاه داده با استفاده از تکنیک های مختلف طبقه بندی و تکنیک های مدل سازی شده اجرا شد. درخت تصمیم گیری، شبکه های عصبی و توابع شعاعی با استفاده از پایگاه داده، مورد استفاده قرار گرفت. نتایج حاکی از آن است که با داده کاوی در بانک می توان تصمیماتی به دنبال استراتژی بازاریابی تهاجمی با نتایج تجزیه و تحلیل داده ارائه داد [۳۱].

<sup>50</sup> .Ante & Ana

<sup>51</sup> .Lu & Shi

<sup>52</sup> . ANOVA

<sup>53</sup> .Kizis et al

<sup>54</sup> .Allis et al

<sup>55</sup> . Bank Guarantee

<sup>56</sup> . Datawarehouse

«گریس»<sup>۵۷</sup>(۲۰۱۶) به بررسی شبکه عصبی بر اساس الگوریتم یادگیری پس انتشار و سیستم استنتاج فازی ممدانی بر اساس مدل به منظور ارزیابی سطح ریسک متقاضیان وام پرداخت. تجزیه و تحلیل مقایسه از اجرای هر دو سیستم انجام شد نتایج تجربی نشان می‌دهد که شبکه‌های عصبی با دقت پیش‌بینی کلی ۹۶٫۸۹٪ با ۹۴٫۴۴٪ بهتر از روش منطق فازی انجام شده است [۳۲].

شبکه‌های عصبی مصنوعی برای تخمین و تقریب، کارایی بسیار بالایی از خود نشان داده‌اند. گستره کاربرد این مدل‌های ریاضی برگرفته از عملکرد مغز انسان، بسیار وسیع می‌باشد که به عنوان چند نمونه کوچک می‌توان استفاده از این ابزار ریاضی در پردازش «سیگنال‌های بیولوژیکی»<sup>۵۸</sup>، مخابراتی و الکترونیکی تا کمک در نجوم و فضا نوردی را نام برد.

### 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 به عنوان مدل خطی تعمیم یافته‌ای که از تابع لوجیت به عنوان تابع پیوند استفاده می‌کند و خطایش از توزیع چندجمله‌ای پیروی می‌کند، به حساب آورد. منظور از دو سویی بودن، رخ داد یک واقعه تصادفی در دو موقعیت ممکنه است. به عنوان مثال خرید یا عدم خرید، ثبت نام یا عدم ثبت نام، ورشکسته شدن یا ورشکسته نشدن و ... متغیرهایی هستند که فقط دارای دو موقعیت هستند و مجموع احتمال هر یک آن‌ها در نهایت یک خواهد شد. کاربرد این روش عمدتاً در ابتدای ظهور در مورد کاربردهای پزشکی برای احتمال وقوع یک بیماری مورد استفاده 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 به جای یک توزیع گوسی است چون که 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.

ذکاوت (۱۳۸۱) در تحقیقی، بر اساس یک نمونه ۱۲۰ تایی از مشتریان حقوقی بانک صادرات صورت داده است. پس از بررسی‌های صورت‌های مالی مشتریان، ۵ نسبت مالی به عنوان متغیرهای مؤثر در رفتار اعتباری، انتخاب شده و با استفاده از مدل‌های تحلیل ممیزی گام به گام و رگرسیون لجستیک، مشتریان از نظر ریسک اعتباری تفکیک شده‌اند. یافته‌های تحقیق نشان می‌دهد که متغیر «نسبت جاری» بیشترین قدرت تفکیک مشتریان خوش حساب و بد حساب را از میان بقیه متغیرها دارد. از سوی دیگر روش‌های تحلیل ممیزی و رگرسیون لجستیک در رابطه با دسته‌بندی مشتریان از نظر ریسک اعتباری، نتایج مشابهی را ارائه می‌دهند [۳۳].

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 [۳۴].

رگرسیون منطقی یک روش آماری برای مدل‌سازی‌هایی که نتایج دودویی دارند، است. برای شرایطی که هدف برآورد مفاهیمی چون «رخ دادن» یا «رخ ندادن» است مانند نتیجه‌ی یک مسابقه فینال که دو حالت بیشتر ندارد. ممکن است مقادیر پیش‌بینی شده بی‌معنی باشد و یا در عمل تفسیرپذیر نباشند یا امکان مقایسه مقادیر پیش‌بینی باهم نباشد که ممکن است از نتایج الگوریتم‌های درخت تصمیم و رگرسیون خطی باشد در چنین شرایطی برای رفع ابهامات از این الگوریتم استفاده می‌شود. از این مدل برای به دست آوردن نتایج بهینه پیش‌بینی استفاده می‌شود. از رگرسیون لجستیک می‌توان به عنوان تنوع دیگری از الگوریتم شبکه‌های عصبی نام برد. رگرسیون منطقی یک مدل آماری رگرسیون برای متغیرهای باینری است.

### الگوریتم «کامینز»<sup>۵۹</sup>

روش «کامینز» یکی از روش‌های خوشه‌بندی داده‌ها در داده‌کاوی است. این روش علی‌رغم سادگی آن یک 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 بیان شده است. ولی همه آن‌ها دارای روالی تکراری هستند که برای تعدادی ثابت از خوشه‌ها سعی در تخمین موارد زیر دارند:

57. Grace

58. Biological signals

59. k-means

به دست آوردن نقاطی به‌عنوان مراکز خوشه‌ها. این نقاط در واقع 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 در داده‌ها 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 تغییری در داده‌ها حاصل نشود. تابع زیر به‌عنوان تابع هدف مطرح است.

در الگوریتم کامینز ابتدا  $k$  عضو (که  $k$  تعداد خوشه‌ها است) به‌صورت تصادفی از میان  $n$  عضو به‌عنوان مراکز خوشه‌ها انتخاب می‌شود. سپس  $n - k$  عضو باقیمانده به نزدیک‌ترین خوشه تخصیص می‌یابند. بعد از تخصیص همه اعضا مراکز خوشه مجدداً محاسبه می‌شوند و با توجه به مراکز جدید به خوشه‌ها تخصیص می‌یابند و این کار تا زمانی که مراکز خوشه‌ها ثابت بماند ادامه می‌یابد.

«دانوتا»<sup>۶۰</sup> (۲۰۰۷) در تحقیقی بر پایه ترکیب طبقه‌بندی با ناظر و بدون ناظر می‌باشد. به این صورت که در گام اول با استفاده از الگوریتم خوشه‌بندی «کامینز»، مشتریان به گروه‌هایی با ویژگی‌های مشابه، تقسیم گردیده‌اند و در گام بعدی برای هر گروه از مشتریان درخت تصمیم ساخته شده است. روش پیشنهادی در این تحقیق اجازه استفاده از قوانین متفاوت برای مشتریان مختلف که ریسک‌های بالایی دارند را به بانک می‌دهد [۳۵].

روش میانگین «کامینز» در عین سادگی یک روش بسیار کاربردی و پایه چند روش دیگر مثل خوشه‌بندی فازی و «الگوریتم خوشه-بندی توزیع شده نیمه صحیح»<sup>۶۱</sup> می‌باشد. روش کار به این صورت است که ابتدا به تعداد دلخواه نقاطی به‌عنوان مراکز خوشه در نظر گرفته می‌شود. سپس با بررسی هر داده، آن را به نزدیک‌ترین مرکز خوشه نسبت می‌دهیم. پس از اتمام این کار با گرفتن میانگین در هر خوشه می‌توانیم مراکز خوشه و به دنبال آن خوشه‌های جدید ایجاد کنیم. (با تکرار مراحل قبل) از جمله مشکلات این روش این است که بهینگی آن وابسته به انتخاب اولیه مراکز بوده و بنابراین بهینه نیست. مشکلات دیگر آن تعیین تعداد خوشه‌ها و صفر شدن خوشه‌ها می‌باشد. خوشه‌بندی در واقع یک عملیات غیر نظارتی می‌باشد. این عملیات هنگامی استفاده می‌شود که ما به دنبال یافتن گروه‌هایی از داده‌های مشابه می‌باشیم بدون این‌که از قبل پیش‌بینی در مورد شباهت‌های موجود داشته باشیم. خوشه‌بندی برای اکثر روش‌های داده‌کاوی مثل درخت تصمیم‌گیری و شبکه‌های عصبی، با یک مجموعه آموزشی شروع کرده و به کمک این مجموعه سعی می‌شود یک مدل برای بخش‌بندی داده‌ها، ایجاد گردد. سپس از آن مدل برای پیش‌بینی داده‌های جدید استفاده شود. استفاده از کامینز در همه حوزه‌ها بکار برده شده است.

<sup>60</sup> . Danuta

<sup>61</sup> . Segment-wise distributional clustering Algorithm

## نتیجه‌گیری

نتایج حاصل از یافته‌های تحقیق نشان داد که الگوریتم «نایوبیز» در بانکداری الکترونیک بیشترین کاربرد را داشته است. چراکه با مطالعه پیشینه مقالات انجام‌گرفته، این الگوریتم به دلیل کاربرد زیاد و جدید بودن آن نسبت به همه الگوریتم‌ها در بانکداری الکترونیک برتری دارد. 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 مورد استفاده قرار گیرد. الگوریتم نایوبیز دارای چند ویژگی در بانکداری الکترونیک است از جمله خطای کم و یادگیری تدریجی در داده‌کاوی بانکداری الکترونیک 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 مورد استفاده در این پژوهش دارد، بنابراین دانش استخراج شده از این الگوریتم به‌عنوان مورد اعتمادترین دانش داده‌کاوی مورد بررسی می‌باشد و قوانین حاصل از الگوریتم نایوبیز به‌عنوان دانش سازمانی استخراج شده از داده‌کاوی در بانکداری الکترونیک مورد مطالعه تلقی می‌گردند.

## پیشنهادات

- آموزش کارکنان شعب در جهت آشنایی هر چه بیشتر آنان در کاربرد الگوریتم‌های SVM و نایوبیز در بانکداری الکترونیک
  - تهیه و تنظیم سیستمی کارآمد جهت بکار بردن این الگوریتم‌ها، به‌منظور پیش‌بینی رفتار مشتریان جدید در استفاده از خدمات اینترنت بانک که می‌تواند نوع خدمات و سرویس‌های ارائه‌شده به مشتریان را شخصی‌سازی کرده و زمینه رضایت مشتریان را فراهم کند.
  - استفاده از مدل ترکیبی الگوریتم‌های SVM و نایوبیز در طبقه‌بندی مشتریان بانکداری الکترونیک
- انجام پژوهشی مشابه روی شبکه‌های دیگر بانکداری مدرن، از جمله همبانک که به‌تازگی وارد حوزه بانکی شده است.

## منابع و مراجع

- [۱] سلیمانی، ع؛ ایمانی، س. پیشرفت های فناوری اطلاعات در بانکداری الکترونیک، مجله مدیریت، شماره ۵۹، ۱۳۹۱، ص ۵۲-۲۲.
- [2] 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.
- [3] 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.
- [۴] ناظمی، ع؛ مشکانی، ع. داده کاوی کاربردی، دانشگاه آزاد اسلامی واحد نیشابور، ۱۳۸۸.
- [5] Liébana-Cabanillas, F., Nogueras, R., Herrera, L.J., Guillén, A. Analysing user trust in electronic banking using data mining methods. Expert Systems with Applications, 2013. 40(14), 5439-5447.
- [6] Chattamvelli, R., Data mining Algorithm. Alpha science, 2011.
- [7] Karahoca, A., Bilgen, O., Karahoca, D. Churn Management of E-Banking Customers by Fuzzy AHP, Financial and Banking Crisis Prediction through Early Warning Systems, 2016, 1-21.
- [8] Yao, J-S., Kweimei, W., Ranking fuzzy numbers based on decomposition principle and signed distance. Fuzzy Sets and Systems, 2010. 116(2), 275-288.
- [۹] پاشازاده، س؛ خلیلی، ش. طبقه‌بندی داده‌های حساب‌های جاری بانک با استفاده از درخت تصمیم. دومین کنفرانس دانش‌پژوهان کامپیوتر و فناوری اطلاعات، ۱۳۹۳.
- [10] Pliha, R., Systems and methods for scoring bank customers direct deposit account transaction activity to match financial behavior to specific acquisition, performance and risk events defined by the bank using a decision tree and stochastic process. EspacenetGlobal DossierDiscuss, 2007, 7(2), 40-91.
- [11] Zekic-Susac, M. Sarlija, N., Bencic, M. Small business credit scoring: a comparison of logistic regression, neural network, and decision tree models, International Conference on Information Technology Interfaces, 2004.
- [12] 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.
- [۱۳] 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، شماره ۲۸، دوره ۳، ۱۳۹۳، صص ۷۰۰-۶۸۱.
- [14] Koyuncugil, A., Ozugulbas, N. Risk modeling by CHAID decision tree algorithm. ICCES, 2009. 11(2), 39-46.
- [۱۵] لنگری، ر. ارائه مدلی جهت کشف رفتارهای مشکوک در بانکداری الکترونیکی با استفاده از الگوریتم‌های درخت تصمیم‌گیری، پایان نامه کارشناسی ارشد، دانشگاه پیام نور تهران: تهران، ۱۳۸۹.
- [16] Pang, S, Gong, J. C5.0 Classification Algorithm and Application on Individual Credit Evaluation of Banks. Systems Engineering — Theory & Practice, 2009. 29(12), 94-104.
- [17] Ying, C., Ying, Y. Learning with Support Vector Machines. Synthesis Lectures on Artificial Intelligence and Machine Learning, 2011, 95(4), 95-106.
- [18] Cheng-Lung, Chen, M., Wang, Ch. credit scoring with a data mining approach based on support vector machines. journal of expert systems with applications, 2006, 33(4), 847-856.
- [۱۹] طلوعی اشلقی، ع؛ مقدوری شربیانی، ف؛ دانشگر، ف. امتیاز دهی اعتباری متقاضیان کارتهای اعتباری بانکها با استفاده از تکنیک ماشین بردار پشتیبان، دومین کنفرانس بین المللی شهر الکترونیک، تهران، پژوهشکده فناوری اطلاعات و ارتباطات جهاد دانشگاهی، شهرداری تهران، ۱۳۸۸.

- [۲۰] شریبانی مقدوری، ف؛ نیکومرام، ه؛ طلوعی اشلقی، ع. طبقه بندی متقاضیان تسهیلات اعتباری بانکها با استفاده از تکنیک ماشین بردار پشتیبان، آینده پژوهی مدیریت، دوره ۲۱، شماره ۱، ۱۳۸۹، صص ۱۹-۱.
- [21] Shouwei, L., Wang, M., Jianmin, H. Prediction of Banking Systemic Risk Based on Support Vector Machine. *Mathematical Problems in Engineering*, 5(4), 2013,15-20.
- [22] Hsieh, N., Lun-Ping, H. data driven ensemble classifier for credit scoring analysis. *Expert Systems with Applications*, 2010, 37(1), 534-454.
- [23] Acaravci, S., Parlar, T. Using Data Mining Techniques for Detecting the Important Features of the Bank Direct Marketing Data. *International Journal of Economics and Financial Issues*, 2017. 7(2), 692-696.
- [24] Vijaya, K., Prajakta, W., Deepti, Y., Pooja, G. Detecting E-Banking Phishing Website Using Navie Baye's Classifier. *INTERNATIONAL JOURNAL FOR RESEARCH IN EMERGING SCIENCE AND TECHNOLOGY*, 2017, 1(4), 94-97.
- [25] Vedala, R., Kumar, B. An application of Naive Bayes classification for credit scoring in e-lending platform. *Confrence Data Science & Engineering (ICDSE)*, 2012.
- [26] McLachlan, G.J. *Discriminant Analysis and Statistical Pattern Recognition*. Wiley Series in Probability and Statistics, 2004.
- [27] Ben Pam, W. Discriminant Analysis and the Prediction of Corporate Bankruptcy in the Banking Sector of Nigeria. *International Journal of Finance and Accounting*, 2013, 2(6), 319-325.
- [28] Ante, R., Ana, K. DISCRIMINANT ANALYSIS OF BANK PROFITABILITY LEVELS. *Croatian Operational Research Review*, 2013, 11(3), 39-49.
- [29] 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.
- [۳۰] ودادی، ا. بررسی رفتار اعتباری مشتریان تسهیلات مصرفی با استفاده از شبکه های عصبی امتیازبندی اعتباری، پایان نامه کارشناسی ارشد، دانشگاه آزاد اسلامی واحد علوم و تحقیقات: تهران، ۱۳۸۳.
- [31] Alls, O.F., Ertan, K., Piero, M. Data Mining for Database Marketing at Garanti Bank. *WIT Transactions on Information and Communication Technologies*, (2000), 25(12) 14-16.
- [32] Grace, A.M. Comparative Analysis of Neural Network and Fuzzy Logic Techniques in Credit Risk Evaluation. *International Journal of Intelligent Information Technologies (IJIT)*, 2016. 12(1), 1-16.
- [۳۳] ذکاوت، م. مدل‌های ریسک اعتباری مشتریان بانک صادرات ایران، موسسه عالی بانکداری ایران، ۱۳۸۱.
- [۳۴] بی نظیر، ع. امتیاز دهی اعتباری مشتریان حقوقی بانکها با استفاده از روش آلتمن مطالعه موردی بانک ملت. ۱۳۸۸، موسسه عالی بانکداری: تهران.
- [35] Danuta, Z. On Integrating UnSupervised and Supervised Classification for Credit Risk Evaluation. *journal of Information Technology and Control*, 2007. 36(1), 124-129.