



## رویکرد شبکه عصبی مبتنی بر کلونی زنبور عسل مصنوعی جهت تخمین رتبه اعتباری مشتریان بانک ها

سعید فلاح پور<sup>۱</sup>

رضا راعی<sup>۲</sup>

محمد هندجانی زاده<sup>۳</sup>

تاریخ پذیرش: ۹۲/۰/۰۰

تاریخ دریافت: ۹۲/۰/۰۰

### چکیده

کاهش و کنترل ریسک اعتباری به عنوان یکی از عوامل موثر در بهبود فرآیند اعطای اعتبار و در نتیجه در عملکرد بانک ها مطرح گردیده و نقش اساسی در تداوم ارائه تسهیلات، سودآوری و بقای بانک ها و موسسات مالی ایفا می نماید. در این راستا، پژوهش حاضر سعی در ارائه رویکردی نو برای ارزیابی ریسک اعتباری مشتریان بانکی دارد. روش شبکه عصبی به عنوان طبقه بندی کننده ی اصلی مشتریان تسهیلات بانکی با یک روش انتخاب ویژگی پوشش دهنده به نام کلونی زنبور عسل ترکیب می گردد. روش های دیگر به کار رفته در این پژوهش شبکه عصبی مبتنی بر آنالیز اجزای اساسی به عنوان یک روش فیلتر کننده انتخاب ویژگی و روش شبکه عصبی به تنهایی می باشند. به منظور نشان دادن اثر بخشی روش پیشنهادی از داده های مربوط به دویست و پنجاه تن از تسهیلات گیرندگان حقیقی بانک کارآفرین در یک بازه زمانی سه ساله (۱۳۹۰-۱۳۸۷) به همراه هجده ویژگی مربوط به هریک از آنها استفاده نموده ایم. نتایج رویکرد ارائه شده با روش شبکه مصنوعی و روش شبکه مصنوعی بر پایه آنالیز اجزای اساسی مقایسه گردیده است. یافته های پژوهش دلالت بر آن داشته که در ارزیابی ریسک اعتباری متقاضیان تسهیلات اعتباری، مدل شبکه عصبی بر مبنای کلونی زنبور عسل نسبت به روش شبکه عصبی و مدل شبکه عصبی بر مبنای آنالیز اجزای اساسی از عملکرد خوبی برخوردار است.

**واژه های کلیدی:** کلونی زنبور عسل، شبکه عصبی، ریسک اعتباری، آنالیز اجزای اساسی.

۱- استادیار مدیریت مالی، دانشکده مدیریت دانشگاه تهران،

۲- استاد مدیریت مالی، دانشکده مدیریت دانشگاه تهران،

۳- کارشناس ارشد مهندسی مالی، دانشگاه تهران، (مسئول مکاتبات) hendijani@ut.ac.ir

## ۱- مقدمه

امروزه با جهانی شدن اقتصاد و تشدید رقابت بین بانکها، حاشیه سود فعالیت های سنتی بانکداری کاهش یافته و موجب افزایش ریسک در بانک ها گردیده است. با توجه به مشکلات پیش روی مؤسسات بانکی، مدیریت ریسک بعنوان عاملی که نقش موثری در کسب مزیت رقابتی ایجاد میکند، بایستی در این مؤسسات و سایر مؤسسات مالی و اعتباری به شدت مورد توجه قرار گیرد.

ریسک اعتباری عبارتست از احتمال اینکه بعضی از دارایی های بانک، به ویژه تسهیلات اعطایی از نظر ارزش کاهش یابد و یا بی ارزش شود. باتوجه به اینکه سرمایه بانک ها نسبت به کل ارزش دارایی های آن ها کم است، حتی اگر درصد کمی از وام ها قابل وصول نباشند، بانک با خطر ورشکستگی روبرو خواهد شد. ریسک اعتباری در صنعت بانکداری یا از جانب مشتریان متوجه بانک میشود و یا ممکن است خود بانک با عدم توانایی در پرداخت بدهی ها مواجه گردد (فلاح شمس و تهرانی، ۱۳۸۴). ریسک نکول به عنوان یکی از مهمترین اجزاء ریسک اعتباری محسوب میگردد. ریسک نکول هنگامی رخ میدهد که وام گیرنده به علت عدم توان یا تمایل، به تعهدات خود در مقابل وام دهنده در تاریخ سررسید عمل ننماید. این ریسک از قدیمی ترین و مهم ترین ریسکهایی است که خصوصاً نهادهای پولی و مالی را تحت تأثیر قرار میدهد؛ چراکه نکول تعداد اندکی از مشتریان میتواند زیانهای جبران ناپذیری را به یک سازمان وارد سازد (خانی، ۱۳۸۶). اما بانک ها همواره با این چالش روبرو بوده اند که چگونه و بر اساس چه شاخص ها و روش هایی متقاضیان اعطای اعتبار را مورد ارزیابی قرار دهند؟ انجام این امر بوسیله یک سیستم جامع، ساختار یافته و انتخاب تکنیکهایی با عملکرد پر بازده تر امکان پذیر است. این ارائه نیز در پی پیشنهاد و راهکاری کارشناسانه جهت حل این مساله بوده تا بتواند متقاضیان اعتبارات مالی را به خوبی دسته بندی کرده واز احتمال عدم بازپرداخت تسهیلات اعطایی بکاهد.

تاکنون پژوهش های گوناگونی برای ارائه مدلی کارا، جهت کمک به اتخاذ تصمیمات صحیح اعطای اعتبار از سوی مؤسسات مالی ارائه شده است، که افزایش دقت پیش بینی احتمال نکول متقاضیان تسهیلات، از مهمترین اهداف آن ها بوده است. رویکردهای داده کاوی، یکی از شاخص ترین این روشها می باشد که تلاش میکند دانش نهفته در داده های تاریخی را استخراج کرده و از آن در جهت پیش بینی احتمال نکول مشتری، بهره گیرد. در این پژوهش سعی گردیده است تا با به کارگیری رویکرد های ترکیبی شبکه عصبی بر پایه کلونی زنبور عسل (ABC-NN)، شبکه عصبی بر پایه آنالیز اجزای اساسی (PCA-NN) و شبکه عصبی به تنهایی (NN) اقدام به طبقه بندی مشتریان بانکی به دو دسته مشتریان خوش حساب و بد حساب نموده و دقت طبقه بندی کنندگان بیان شده، استخراج گردد. به طور خلاصه هدف اصلی این مقاله مقایسه عملکرد روش های بیان شده و همچنین مقایسه رویکرد آنالیز اجزای اساسی به عنوان یک روش فیلتر کننده انتخاب ویژگی با رویکرد کلونی زنبور عسل به عنوان یک روش پوشش دهنده انتخاب ویژگی می باشد.

داده های به کار رفته در این پژوهش اطلاعات مربوط دویست و پنجاه نفر از مشتریان بانک کارآفرین در بازه زمانی سه ساله (۱۳۹۰-۱۳۸۷) که هفتاد و پنج نفر از آنها شامل مشتریان بدحساب و صد و هفتاد و پنج

نفر از آنها مشتری خوش حساب محسوب می گردند. هجده متغیر توضیحی در رابطه با هر یک از این مشتریان بیان گردیده است که به عنوان ورودی های روش های انتخاب ویژگی به کار رفته در این تحقیق مورد استفاده قرار می گیرند.

در این پژوهش، پس از مقدمه، توضیحاتی پیرامون مبانی نظری و پیشینه تحقیقات صورت گرفته در زمینه روش های ارزیابی ریسک اعتباری مشتریان بانک ها ارائه می گردد. در بخش بعدی مطالبی در رابطه با روش های به کار رفته و متغیرهای تحقیق تقدیم گردیده و در پایان نیز بعد از بررسی نتایج و یافته های پژوهش، نتیجه گیری و پیشنهاداتی برای تحقیقات آتی بیان می گردند.

## ۲- مبانی نظری و مروری پیشینه پژوهش

ارتباط صحیح بین نظام های مالی و تولیدی هر کشوری از مهمترین عوامل رشد و توسعه اقتصادی محسوب می گردد. بانک ها به عنوان بخش اصلی نظام مالی، نقش اساسی در تامین بخش های تولیدی، تجاری و مصرفی و حتی دولتی بر عهده خواهند داشت. در ایران نیز با توجه به ساختار اقتصادی کشور و به دلایلی همچون توسعه نیافتن بازارهای سرمایه و سایر شبکه های غیر بانکی و قراردادی، تامین مالی بخش های واقعی اقتصاد بر عهده شبکه بانکی کشور است. متأسفانه این بخش نیز در تحقق رسالت های خود چندان موفق نبوده است. در حال حاضر تداوم فعالیت ها و بقای اکثر بانک های کشور ناشی از حمایت های دولتی است (خانی، ۱۳۸۶). حجم قابل ملاحظه تسهیلات اعطایی سوخت شده و یا معوقه های مربوط به بانک ها، خود گویای نبود مدل های مناسب و کارشناسی شده در سیستم های اعتباری شبکه بانکی است. همچنین وقوع بحرانهای بانکی در دهه های اخیر در کشورهای صنعتی و به ویژه در کشورهای در حال توسعه به دلایلی همچون فرار سپرده ها، افزایش مطالبات معوق بانک ها، رکود اقتصادی و غیره باعث اختلال در نظم بازارهای مالی شده و زمینه ورشکستگی بسیاری از بانک ها را فراهم می آورد. طی بررسی های به عمل آمده، علت اصلی وقوع این موضوع عدم کفایت سرمایه بانک ها شناسایی شده است. بنابراین اهمیت مقوله ریسک و مدیریت آن بیش از پیش نمود یافته است (همتی و محبی نژاد، ۱۳۸۸).

بدون شک ارزیابی ریسک اعتباری موضوع مهمی برای تحقیق در زمینه مدیریت ریسک مالی به حساب می آید. در کل ارزیابی درست از ریسک اعتباری می تواند به نوعی استفاده بسیار موثر از سرمایه اقتصادی تبدیل گردد. وقتی بعضی از مشتریان نمی توانند بدهی های خود را بازپرداخت نمایند، نوعی شکست سیستم اقتصادی برای سازمان های مالی وام دهنده رخ می دهد. بنابراین بهبود در امر تصمیم گیری در مورد اعطای اعتبار به مشتریان بانک ها و درجه بندی اعتبار، یکی از مسائل مرتبط با مدیریت ریسک اعتباری بانک ها است. به عبارتی مسئله درجه بندی اعتباری و تخصیص اعتباری به فراخور اعتباری، گریبانگر بسیاری از مراکز تصمیم گیری می باشد. لذا استفاده از مدل های مناسب جهت تخصیص بهینه اعتبار و توزیع اعتبار بانکی میان مشتریانی که از اعتبار بالاتری برخوردارند، اهمیت به سزایی دارند. در بازاری که حاشیه سود بانک ها به دلیل تشدید رقابت به طور مستمر در حال کاهش بوده و همواره فشار

برای کاهش بیشتر هزینه‌ها احساس می‌گردد، مدل‌های ریسک‌آزمایی با پیش‌بینی زیان‌های عدم بازپرداخت وام‌ها نوعی مزیت نسبی برای بانکها و موسسات مالی به وجود می‌آورند (فماندس، ۲۰۰۵). برای اعطای تسهیلات باید درجه اعتبار و قدرت بازپرداخت اصل و سود مبلغ اعتبار را برای دریافت‌کننده تسهیلات تعیین‌گردند. شانس اینکه وام‌گیرنده نتواند وام را بازپرداخت کند، ریسک‌آزمایی یا ریسک عدم بازپرداخت شناخته می‌شود. تعریفی که کمیته بال 1 سویس از ریسک اعتباری ارائه می‌دهد، عبارت است از "امکان بالقوه اینکه قرض‌گیرنده از بانک و یا از طرف حساب وی در اجرای تعهدات خود در مقابل بانک در مدت مشخصی ناتوان شود" (صفری و همکاران، ۱۳۸۹).

در یک تقسیم‌بندی کلی، می‌توان مدل‌های به کار رفته در پیشینه پژوهش در رابطه با ارزیابی ریسک اعتباری مشتریان بانکی را به چهار دسته کلی زیر تقسیم نمود:

الف: مدل‌های آماری مانند تحلیل ممیز، رگرسیون لجستیک، درخت تصمیم‌گیری و پروبیت. مشکل اصلی که در روش‌های آماری وجود دارد برخی فروض محدودکننده که در دنیای واقعی مفروض نمودن آنها دور از ذهن می‌باشد. برای مثال در نظرگرفتن فروضی همچون وجود رابطه خطی بین متغیرها و یا برابر بودن کواریانس زوجی در متغیرهای ورودی این روش‌ها را به عنوان فروض محدودکننده نام برد.

ب: روش‌های تحقیق در عملیات مانند برنامه‌ریزی خطی و برنامه‌ریزی عدد صحیح. در این روش‌ها با تعریف نمودن یک تابع هدف که معمولاً حداقل نمودن سطح ریسک اعتباری بوده و همچنین تعریف کردن یکسری محدودیت، مدل مساله را تعریف می‌نمایند. سپس با به کارگیری یک الگوریتم بهینه‌سازکننده در روش‌های تحقیق در عملیات اقدام به حل مساله مذکور می‌نمایند.

ج: روش‌های هوش مصنوعی: در یک تقسیم‌بندی کلی این روش شامل الگوریتم ژنتیک، شبکه‌های عصبی، سیستم‌های فازی و سیستم‌های هوشمند شده که برتری این روش‌ها نسبت به روش‌های آماری نبودن هیچ‌یک از فروض محدودکننده در روش‌های هوش مصنوعی می‌باشد.

د: روش‌های هیبرید: در این گونه روش‌ها ابتدا با استفاده از یک الگوریتم انتخاب ویژگی، زیرمجموعه‌ای از ویژگی‌ها که باعث افزایش دقت طبقه‌بندی‌کننده اصلی می‌شود، انتخاب شده و این زیرمجموعه بهینه از ویژگی‌ها به عنوان ورودی طبقه‌بندی‌کننده اصلی در نظر گرفته می‌شوند. دو روش از سه روش به کار رفته در این پژوهش نیز از نوع هیبرید بوده که در آن ابتدا با دو روش انتخاب ویژگی کلونی زنبور عسل و آنالیز اجزای اساسی، دو زیرمجموعه بهینه از ویژگی‌ها که باعث کسب بالاترین دقت پیش‌بینی می‌گردند برگزیده شده و دو زیرمجموعه حاصل از هر یک از روش‌ها به عنوان ورودی طبقه‌بندی‌کننده اصلی (شبکه عصبی) در نظر گرفته می‌شوند.

اندازه‌گیری و درجه‌بندی ریسک اعتباری برای نخستین بار در سال 1909 میلادی توسط جان موری بر روی اوراق قرضه انجام شد (گلنتز، ۲۰۰۳). یکی از قدیمی‌ترین مؤسساتی که اقدام به رتبه‌بندی اوراق قرضه نمود، موسسه مودیز است که در سال 1909 تاسیس گردیده است. به دلیل شباهت زیاد اوراق قرضه و تسهیلات اعطایی، درجه‌بندی اعتباری به معنای اندازه‌گیری ریسک عدم پرداخت اصل و بهره سود تسهیلات،

مورد بررسی محققین قرار گرفته است (فلاح شمس، ۱۳۸۷). آلتمن (۱۹۶۸) از دیگر پیشگامان در زمینه اندازه گیری ریسک اعتباری اوراق قرضه شناخته میشود. او در جهت یافتن یک رابطه معنادار بین متغیرهای حسابداری یک شرکت و احتمال عدم توانایی در پرداخت دیون این شرکت در آینده، تلاش بسیاری انجام داد که در نتیجه منجر به ارائه مدل نمره Z توسط وی شد. رهنمای رود پشتی و همکاران (۱۳۸۸) در تحقیق انجام شده با به کارگیری مدل های آلتمن و فالمر به پیش بینی احتمال ورشکستگی شرکت ها پرداخته که نتایج حاصل نشان از محافظه کارانه تر بودن مدل آلتمن نسبت به مدل فالمر بود.

در ارزیابی اعتباری ریسک اعتباری مشتریان موسسات مالی، شبکه های عصبی تبدیل به یکی از دقیق ترین ابزار آنالیز اعتبار در میان سایر ابزار شده است. دیسای و همکارانش (۱۹۹۶) به بررسی توانایی های شبکه های عصبی و تکنیک های آماری متداول نظیر آنالیز ممیزی خطی و آنالیز رگرسیون خطی در ساخت مدل های امتیازدهی اعتباری پرداخته اند. وست (۲۰۰۰) به مقایسه عملکرد الگوریتم شبکه عصبی با روش آنالیز رگرسیون خطی برای رتبه بندی اعتباری مشتریان بانک ها در کشور کرواسی پرداخت. نتایج حاصل حاکی از برتری (دقت پیش بینی بالاتری حاصل گردید) روش شبکه عصبی نسبت به روش آنالیز رگرسیون خطی بود. یکی دیگر از روش های پر کاربرد در صنعت اعتباری، استفاده از ماشین بردار پشتیبان می باشد که هانگ و همکاران (۲۰۰۳) از اولین پژوهشگرانی بودند که با این تکنیک که جزوه روش های یادگیری ماشینی است، اقدام به پیش بینی وضعیت اعتباری مشتریان بانک های ایالات متحده آمریکا نموده و نتایج حاصل دلالت بر دقت پیش بینی نزدیک به ۸۰ درصد برای داده های به کار رفته در تحقیق آن ها داشت.

راعی و فلاح پور (۱۳۸۷) در تحقیق خود به منظور پیش بینی درماندگی مالی شرکت های پذیرفته شده در بورس اوراق بهادار تهران، به مقایسه دو روش ماشین بردار پشتیبان و مدل آماری رگرسیون لجستیک پرداختند. یافته های تحقیق نشان از برتری روش ماشین بردار پشتیبان نسبت به روش رگرسیون لجستیک داشته و نه تنها ماشین بردار پشتیبان از دقت بالاتری برخوردار بود، بلکه توانایی بالاتری نیز در تعمیم پذیری از خود نشان داد. از دیگر روش های اندازه گیری ریسک اعتباری که بر اساس برنامه ریزی خطی می باشد، روش تحلیل پوششی بوده که از سوی کاپوتی و همکاران در سال ۱۹۹۸ در زمینه مالی ارائه شد. مین و لی (۲۰۰۷) با به کارگیری رویکرد تحلیل پوششی داده ها اقدام به امتیاز دهی اعتباری ۱۰۶۱ شرکت کره ای نمودند. آنان دریافتند که رویکرد تحلیل پوششی داده ها بدلیل داشتن کارایی لازم، می تواند به عنوان یک جایگزین امیدوار کننده برای روش های دیگر امتیاز دهی اعتباری مورد استفاده قرار گیرد. ارمشی (۱۳۹۰) در پایان نامه کارشناسی ارشد خود به بررسی رابطه بین ریسک اعتباری مشتریان و برخی متغیرهای مالی و جمعیت شناختی پرداخت. پس از تخمین مدل، نتایج ذیل مورد تأیید قرار گرفت: متغیرهای جنسیت، درآمد، نوع سکونت، تاهل، سن و وضعیت شغلی مشتریان بر احتمال عدم نکول تسهیلات تأثیرگذار بوده اما متغیر درآمد دارای تأثیر منفی می باشد. و متغیرهای اندازه وام و دوره بازپرداخت در نمونه مورد مطالعه بی تأثیر میباشند.

پویان فر و همکاران (۱۳۹۲) با استفاده از روش حداقل مربعات ماشین بردار پشتیبان مبتنی بر الگوریتم ژنتیک اقدام به ارزیابی ریسک اعتباری مجموعه داده های خود که شامل اطلاعات ۱۰۰۰ مشتری (۷۰۰ مشتری خوش حساب، ۳۰۰ مشتری بد حساب و بیست ویژگی مربوط به آنها) مربوط به بانک های آلمان نموده که دقت طبقه بندی کننده پیشنهادی آنها از کلیه طبقه بندی کننده های به کار رفته در پیشینه تحقیق مربوط به این مجموعه از داده ها بهتر بود. کیمیاگری و همکاران (۱۳۹۱) در مطالعه ای با عنوان "مدل ریسک اعتباری بازپرداخت تسهیلات مشتریان اعتباری بانک (موردی حوزه شهرضا بانک ملی ایران)" با استفاده از روش رگرسیون لجستیک یک نمونه ۳۱ تایی از شرکت ها که در سال های ۱۳۸۶ تا ۱۳۹۰ از بانک ملی حوزه شهرضا تسهیلات اعتباری دریافت نموده اند، بررسی کرده اند. که نتایج حاصل نشان از برتری روش رگرسیون لجستیک نسبت به روش لاجیت برای داده های به کار رفته در تحقیق آنها داشت. فلاح شمس و مهدوی راد (۱۳۸۹) با به کارگیری دو مدل اقتصادسنجی لاجیت و پروبیت به بررسی عملکرد هریک از این دو مدل در پیش بینی ریسک اعتباری مشتریان حقیقی تسهیلات لیزینگ نمودند. نتایج حاصل حاکی از کارایی بهتر مدل لاجیت نسبت به مدل پروبیت برای داده های آنها بود.

در این پژوهش سعی بر آن بوده است تا با استفاده از روش های هوش مصنوعی و رویکرد داده کاوی آنها، اقدام به ارائه مدلی نو در زمینه رتبه بندی اعتباری مشتریان موسسات مالی صورت گیرد. رویکرد شبکه عصبی یر مبنای کلونی زنبور عسل برای نخستین بار بوده که در زمینه ارزیابی ریسک اعتباری مشتریان بانک ها مطرح می گردد.

### ۳- روش شناسی و مدل های پژوهش

#### ۳-۱- نگرش انتخاب ویژگی

انتخاب ویژگی فرآیندی است که معمولاً برای مسائلی که داده ها شامل ویژگی های زیادی می باشند استفاده می شود. این فرآیند منجر به کاهش ابعاد بردار ویژگی ها گردیده که با حذف ویژگی های غیرضروری و انتخاب ویژگی های ضروری برای یادگیری مدل، بهبود دقت پیش بینی و افزایش قابلیت توضیحی مدل های پیش بینی کننده را به همراه دارد (اورسکی و همکاران، ۲۰۱۲). انتخاب ویژگی در حقیقت مسئله ای انتخاب یک زیرمجموعه از ویژگی ها است که برای توضیح هدف مورد نظر کافی و ضروری می باشد (اورسکی و همکاران، ۲۰۱۲). وقتی که بردار ویژگی داده ها شامل تعداد زیادی ویژگی باشد به دلیل وجود ویژگی های اضافی و غیرضروری دقت بدست آمده از مدل طبقه بندی کننده کمتر از دقتی خواهد بود که وقتی یک زیرمجموعه مناسب را برای یادگیری مدل انتخاب می کنیم. می توان مشکل انتخاب یک زیرمجموعه از ویژگی ها را برای رسیدن به دقت بهینه یا دقت نزدیک به دقت بهینه با استفاده از یک روش مناسب انتخاب ویژگی حل کرد. از جمله مزایای فرآیند انتخاب ویژگی (Feature Selection) میتوان به افزایش دقت و سرعت در اجرای تکنیک های پیش بینی، حذف داده های غیر مرتبط و اضافه و افزایش قابلیت فهم پذیری مدل های مورد استفاده اشاره نمود (اورسکی و همکاران، ۲۰۱۲).

بطور کلی می‌توان روش‌های انتخاب ویژگی را به دو دسته عمده تقسیم کرد: روش‌های فیلترکننده و روش‌های پوشش‌دهنده. روش‌های فیلترکننده انتخاب ویژگی اشاره به این دارند که یک الگوریتم به جستجو در میان فضای ممکن ویژگی‌ها می‌پردازد و پس از انتخاب هر زیرمجموعه از ویژگی‌ها و اجرای یک تابع فیلترکننده ویژگی‌ها را رتبه‌بندی می‌کند. این رویکرد در انتخاب کردن بهترین زیرمجموعه از ویژگی‌ها توجهی به مدل طبقه‌بندی‌کننده ندارد (لی و همکاران، ۲۰۱۴). در حقیقت روش‌های فیلترکننده به ویژگی کلی بردار ویژگی‌ها و ارزیابی و انتخاب زیر مجموعه‌ها بدون ترکیب شدن با مدل طبقه‌بندی‌کننده اتکا دارد. روش‌های فیلترکننده به دلیل عدم استفاده از مدل طبقه‌بندی‌کننده سرعت بالایی دارند بنابراین این روش برای حل مسائلی که داده‌ها شامل تعداد زیادی ویژگی می‌باشند بسیار مناسب است. علاوه بر آن این روش‌ها به راحتی با انواع مختلف مدل‌های طبقه‌بندی‌کننده قابل اجرا هستند. روش‌های پوشش‌دهنده ابتدا از یک الگوریتم بهینه‌سازی استفاده می‌کنند که در آن ویژگی‌های مختلف را اضافه و یا کم می‌کنند تا زیر-مجموعه‌های متفاوتی را تشکیل دهند، سپس از یک مدل طبقه‌بندی‌کننده برای ارزیابی این زیر مجموعه‌های ساخته شده، استفاده می‌کنند و چون بطور مکرر مدل طبقه‌بندی‌کننده را فراخوانی می‌کنند، سرعت کمتری نسبت به روش‌های فیلترکننده دارند (لی و همکاران، ۲۰۱۴).

از محدودیت‌های روش فیلترکننده این است که کاربران باید تعداد ویژگی‌هایی را که می‌خواهند انتخاب شود و یا معیار توقف را مشخص سازند (یونگ و همکاران، ۲۰۰۸)، ولی در روش‌های پوشش‌دهنده تعداد ویژگی‌های انتخاب شده بر اساس دقتی است که مدل طبقه‌بندی‌کننده به آن دست می‌یابد و زیرمجموعه‌ای از ویژگی‌ها را انتخاب می‌کند که بالاترین دقت را حاصل می‌کند. لازم به ذکر است که تعیین پارامترهای جستجوی مناسب برای روش‌های پوشش‌دهنده از اهمیت بالایی برخوردار است به خصوص در مسائلی که تعداد ویژگی‌ها زیاد می‌باشد که در صورت عدم تنظیم صحیح پارامترها به دلیل تشکیل زیرمجموعه‌های زیاد و محاسبه دقت آنها، زمان زیادی صرف محاسبه خواهد شد.

در این پژوهش از کلونی زنبور عسل مصنوعی به عنوان یک روش انتخاب ویژگی پوشش‌دهنده استفاده شده است. در این تحقیق به این روش نماد ABC اطلاق می‌گردد. از میان ۱۸ ویژگی بیان شده برای هر یک از دویست و پنجاه مشتری دریافت‌کننده تسهیلات از بانک کارآفرین، ویژگی‌هایی که باعث افزایش در دقت پیش‌بینی طبقه‌بندی‌کننده اصلی (شبکه عصبی) می‌گردد را بر می‌گزینند. همچنین به عنوان یکی از نمایندگان روش‌های انتخاب ویژگی فیلترکننده، تکنیک آنالیز اجزای اساسی را استفاده کرده تا از میان ۱۸ ویژگی ذکر شده، ویژگی‌هایی که باعث افزایش دقت طبقه‌بندی‌کننده شبکه عصبی می‌گردد، انتخاب گردند.

### ۳-۱-۱- کلونی زنبور عسل مصنوعی

کلونی زنبور مصنوعی یا همان Artificial Bee Colony (ABC)، یک الگوریتم فرا مکاشفه‌ای مبتنی بر زندگی اجتماعی زنبور عسل است که در سال ۲۰۰۵ توسط کارابوگا جهت بهینه‌سازی مسائل عددی معرفی

گردید. در این پژوهش به این روش نماد ABC اطلاق می گردد. الگوریتم کلونی زنبور مصنوعی شامل سه مولفه اساسی است: زنبورهای کارگر، غیر کارگر و منابع غذایی. در این مدل دو روش هدایت گر رفتار وجود دارد که شامل به خدمت گرفتن زنبور جدید کاوشگر برای منابع غذایی غنی، با دریافت بازخورد مثبت و ترک منابع غذایی ضعیف توسط زنبورها با دریافت بازخور منفی، میباشد. در الگوریتم زنبور مصنوعی، کلونی شامل سه گروه از زنبورها می باشد که عبارتند از: زنبوران کارگر مرتبط به منابع غذایی مشخص، زنبوران ناظر که بر انتخاب منابع غذایی نظاره می کنند و زنبوران پیشاهنگ که به صورت تصادفی به دنبال منابع غذایی می گردند. زنبوران ناظر و پیش آهنگ هر دو زنبورهای غیر کارگر نامیده میشوند. ابتدا مکان تمامی منابع غذایی توسط زنبوران پیش آهنگ کشف می گردد، سپس شاهد منابع غذایی توسط زنبوران کارگر و ناظر استخراج می گردد. این بهره برداری پیوسته منجر به تهی شدن آن منبع می گردد. بعد از آن زنبوران کارگر به زنبوران پیشاهنگی تبدیل می گردند تا دوباره به جست و جوی منابع غذایی بیشتری بپردازند. در کلونی زنبور مصنوعی، موقعیت منبع غذایی یک جواب ممکن برای مساله بهینه سازی می باشد و میزان شاهد منبع غذایی متناظر با برازش جواب مربوطه است. تعداد زنبورهای کارگر یا زنبورهای ناظر مساوی با تعداد راه حل های جمعیت می باشد، از این رو هر زنبور فقط و فقط با یک منبع غذایی در ارتباط است. در گام اول، الگوریتم یک جمعیت تصادفی توزیع شده اولیه از SN جواب (مکان های منبع غذا) ایجاد می کند، SN تعداد زنبورهای کارگر یا ناظر می باشد. هر جواب  $x_i$  یک بردار D بعدی می باشد. در اینجا D تعداد پارامترهای بهینه سازی می باشد. در الگوریتم کلونی زنبور مصنوعی، هر دور حلقه اصلی شامل سه گام اصلی است:

در ابتدا زنبورهای کارگر به سمت منابع غذایی ارسال میشوند و میزان شاهد منابع ارزیابی میگردد. بعد از به اشتراک گذاشتن اطلاعات منابع غذایی، انتخاب نواحی منابع غذایی توسط ناظرها انجام میشود و میزان شاهد منابع غذایی جدید استخراج گردیده و مورد ارزیابی قرار میگیرد. سپس زنبورهای پیش آهنگ مشخص می گردند و به صورت تصادفی به سمت منابع غذایی احتمالی ارسال می گردند. این گام ها به تعداد معینی تکرار می گردند که این تعداد بیشترین تعداد دور حلقه (MCN) نامیده میشود. زنبور ناظر مصنوعی، یک منبع غذایی را با توجه به مقدار احتمال ( $P_i$ ) مربوط به آن منبع غذایی انتخاب می کند. این احتمال توسط رابطه زیر محاسبه می گردد:

$$P_i = \frac{Fitness(S_i)}{\sum_{n=1}^{SN} Fitness(S_n)} \quad (1)$$

در این رابطه  $Fitness(S_i)$  مقدار شایستگی جواب  $S_i$  می باشد که متناظر با میزان شاهد منبع غذایی در مکان  $i$  ام است و  $S_n$  تعداد منابع غذایی است که مساوی با تعداد زنبورهای کارگر یا ناظر می باشد. به منظور ایجاد موقعیت غذای کاندید جدید، با توجه به موقعیتی که در حافظه می باشد، کلونی زنبور مصنوعی از عبارت زیر استفاده می کند:

$$v_{ij} = (X_{ij} - X_{kj})\phi_{ij} + X_{ij} \quad (2)$$



در رابطه فوق  $\{1, 2, \dots, SN\}$  و  $k \in \{1, 2, \dots, D\}$  شاخص هایی هستند که به صورت تصادفی انتخاب می گردند. گر چه  $k$  به صورت تصادفی انتخاب می گردد، لیکن مقدار آن با  $i$  فرق دارد.  $\emptyset_{ij}$  یک عدد تصادفی بین منفی یک و مثبت یک می باشد و پارامتری هست که تولید منابع غذایی همسایه حول  $x_{ij}$  را کنترل می کند. از این رابطه متوجه می شویم که زمانی که تفاوت بین  $x_{jk}$  و  $x_{ij}$  کاهش یابد، به معنی کاهش انحراف از موقعیت نقطه اول است. بنابراین زمانی که جست و جو به جواب بهینه نزدیک می گردد، طول گام ها نیز کاهش می یابد. اگر مقدار پارامتری از حد مجاز تجاوز کند، با مقدار قابل قبول تنظیم می گردد. در این تکنیک، زنبور پیش آهنگ منبع غذایی جدید را می تواند با استفاده از رابطه زیر پیدا کند:

$$X_i^j = X_j^{min} + rand[0,1](X_{max}^j - X_{min}^j) \quad (3)$$

بعد از اینکه مکان منبع کاندید یعنی  $V_{ij}$  ایجاد شد و توسط زنبور مصنوعی ارزیابی گردید، کارایی آن نسبت به مکان قبلی ارزیابی شده و اگر غذای جدید شهد بیشتر یا به همان اندازه مکان قبلی داشته باشد، جایگزین قبلی در حافظه میگردد در غیر این صورت قبلی در حافظه می ماند. به بیان دیگر یک انتخاب حریصانه در عملیات میان مکان قبلی و مکان جدید انجام می پذیرد. شکل شماره (۱) نشان دهنده شبه کد الگوریتم کلونی زنبور عسل بوده که در این پژوهش مورد استفاده قرار گرفته است. به طور کلی کلنی زنبور عسل سه فرآیند انتخاب متفاوت زیر رابه کار می گیرد:

الف: فرآیند انتخاب سراسری که مقدار احتمال در این مرحله محاسبه شده و توسط زنبورهای ناظر برای کشف نواحی محتمل به کار می رود.

ب: انتخاب حریصانه برای برگزیدن منبع برتر صورت می پذیرد.

ج: یک انتخاب تصادفی توسط زنبورهای پیش آهنگ انجام می پذیرد.

در این پژوهش کلونی زنبور عسل مصنوعی به عنوان یک روش انتخاب ویژگی پوشش دهنده مورد استفاده قرار گرفته است. تمامی عملیات به کارگیری این روش با استفاده از کدنویسی با زبان برنامه نویسی جاوا صورت پذیرفته است.

```

Function ABC (problem) returns a state that is a local maximum
Input: Population size, Problem size, MCN, Limit
Output: Best Subset( $Subset_{best}$ )
Employ Bee Population  $\leftarrow$  Initialize Population (Population Size, Problem Size);
Evaluate Population (Employed Bee Population);
Best Subset  $\leftarrow$  Get Best Solution (Employed Bee Population);
For each Employed  $Bee_i \in$  Employed Bee Population do
  Employed  $Bee_{i,nc} \leftarrow \tilde{\phantom{x}}$ ;
End
For  $i=1$  to MCN do
  For each Employed  $Bee_i \in$  Employed Bee Population do
     $Subset_i \leftarrow$  Construct Solution(Employ  $Bee_i$ , Employ  $Bee_{random}$ );
    If Fitness ( $Subset_i$ ) > Fitness(Employed  $Bee_i$ ) then
      Employed  $Bee_i \leftarrow Subset_i$ ;
      Employed  $Bee_{i,nc} \leftarrow \tilde{\phantom{x}}$ 
    End
  End
  Onlooker Bee Population  $\leftarrow \emptyset$ ;
  For each Onlooker  $Bee_i \in$  Onlooker Bee Population do
    Onlooker  $Bee_i \leftarrow$  Select Bee (Employed Bee Population);
     $Subset_i \leftarrow$  Construct Solution(Onlooker  $Bee_i$ , Employed  $Bee_{random}$ );
    If fitness( $Subset_i$ ) > fitness (Onlooker  $Bee_i$ ) then
      Onlooker  $Bee_i \leftarrow Subset_i$ 
    End
  Onlooker Bee Population  $\leftarrow$  Onlooker  $Bee_i$ ;
  End
  For each Employed  $Bee_i \in$  Employed Bee Population do
    Employed  $Bee_{i,nc} \leftarrow$  Employed  $Bee_{i,nc}+1$ ;
    If Employed  $Bee_{i,nc} \geq$  Limit then
      Scout  $Bee_i \leftarrow$  Construct Random Solution();
      If Fitness(Scout  $Bee_i$ ) > fitness(Employed  $Bee_i$ ) then
        Employed  $Bee_i \leftarrow$  Scout  $Bee_i$ ;
        Employed  $Bee_{i,nc} \leftarrow 0$ ;
      End
    End
  End
  Bee Population  $\leftarrow$  Employed Bee Population + Onlooker Bee Population;
   $I_{best} \leftarrow$  Get Best Solution(Bee Population);
  If fitness( $I_{best}$ )  $\geq$  fitness( $Subset_{best}$ ) then
     $Subset_{best} \leftarrow I_{best}$ ;
  End
End

```

**Return**  $Subset_{best}$

شکل شماره (۱) شبه کد کلونی زنبور عسل

### ۳-۱-۲- الگوریتم آنالیز اجزای اساسی

روش آنالیز اجزای یا همان principal component analysis (PCA)، یک روش ریاضی بوده که با استفاده از تبدیل متعامد، یک مجموعه‌ای از مشاهدات احتمالا همبسته را به یک مجموعه ارزش از مشاهدات ناهمبسته که به آن مولفه‌های اصلی می‌گویند، تبدیل می‌کند. تعداد مولفه‌های اساسی کوچکتر یا

مساوی تعداد ویژگی‌های ورودی می‌باشد. اولین جزء اصلی به گونه‌ای انتخاب می‌شود که بالاترین واریانس را داشته باشد. هر جزئی که واریانس بالایی داشته باشد، می‌توان آن را ناهمبسته از اجزای دیگر دانست و ویژگی‌ها بر اساس واریانس خود مرتب می‌شوند. به بیان دیگر این تکنیک راهی است برای شناسایی الگوها در مجموعه داده‌ها و نشان دادن داده‌ها به طریقی که شباهت‌ها و اختلاف‌ها را برجسته نماید. به عبارتی این تکنیک در بردارنده‌ی یک رویه ریاضی است که تعدادی متغیر همبسته را تبدیل به تعداد کمتری متغیر ناهمبسته می‌کند که به آنها اجزای اصلی می‌گوییم (جولیف، ۲۰۰۲). در این پژوهش به این روش نماد PCA اتلاق می‌گردد.

اگر  $n$  نشان دهنده تعداد مشتریان باشد، هر مشتری دارای  $m$  ویژگی بوده و در این حالت یک ماتریس  $n \times m$  وجود خواهد داشت:

$$X = \begin{bmatrix} x_1^1 & \dots & x_1^m \\ \vdots & \ddots & \vdots \\ x_n^1 & \dots & x_n^m \end{bmatrix} = \begin{bmatrix} x_1^t \\ x_2^t \\ \vdots \\ x_n^t \end{bmatrix} \quad (4)$$

در رابطه بالا  $x_i$  نشان دهنده بردار هریک از مشتریان و  $t$  نشان دهنده ترانهاده بردارها می‌باشد. در این حالت داده‌ها استاندارد سازی میشوند که بدین منظور برای هر ستون انحراف معیار استاندارد محاسبه شده و بعد از استاندارد سازی کواریانس بین داده‌ها بدست آورده میشود.

$$S = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2} \quad (5)$$

$$X_{norm} = \frac{2 \times (X - X_{min})}{(X_{max} - X_{min})} \quad (6)$$

$$C_x = E[XX^t] \quad (7)$$

بردار  $C_x$  نشان دهنده کواریانس بین متغیرها می‌باشد. از آنجا که ماتریس کواریانس یک مماتریس متقارن بوده، یک پایه متعامد با پیدا کردن مقادیر ویژه و بردارهای ویژه قابل محاسبه است که بردارهای ویژه  $e_k$  و مقادیر ویژه  $\gamma_k$  جواب‌های معادله زیر می‌باشند:

$$C_x e_k = \gamma_k e_k \quad k = 1, 2, \dots, m \quad (8)$$

فضای متعامد ایجاد شده برای یک سری از بردارهای ویژه دارای بیشترین واریانس در یک جهت خاص می‌باشد. لذا جهت‌هایی که در آنها، مجموعه داده‌ها دارای انرژی بیشتری بوده، می‌توانند پیدا شوند. در نتیجه با نشان دادن داده‌ها به صورت چند بردار متعامد زیر مجموعه‌های بهینه از ویژگی‌ها را بدست می‌آوریم. در

این پژوهش، برای به کارگیری روش آنالیز اجزای اساسی از نرم افزار Rapid Miner که جزوه نرم افزار های پر کاربرد داده کاوی می باشد، استفاده کرده ایم.

### ۳-۲- الگوریتم شبکه عصبی

شبکه های عصبی (Neural networks) یا به طور دقیق شبکه های مصنوعی، نوعی فناوری هستند که ریشه در علوم بسیاری از قبیل عصب شناسی، ریاضیات، آمار، فیزیک، علوم کامپیوتر و مهندسی دارد. شبکه های عصبی در حوزه های متعددی همچون ایجاد مدل، تحلیل سری های زمانی، شناخت الگو، پردازش علائم و کنترل کاربرد فراوانی دارند. در این پژوهش نماد NN به این روش اطلاق می گردد. هر شبکه عصبی مجموعه ای از نرون ها می باشد که مانند شبکه های کامپیوتری می توانند به شکل های مختلف سازماندهی گردند. متداول ترین نوع شبکه عصبی که امروزه مورد استفاده قرار می گیرد پرسپترون چند لایه هست که در این تحقیق نیز از آن استفاده گردیده است. پردازش اطلاعات در شبکه های عصبی از سه بخش ورودی ها، خروجی ها و وزن ها تشکیل شده است. ورودی (Input) مربوط به یک صفت منحصر به فرد بوده که ارزش این صفت ها و ورودی های شبکه را تشکیل می دهند. خروجی ها دلالت بر پاسخ مسئله است که در پی حل آن هستیم، می باشد. وزن ها عناصر اصلی شبکه عصبی بوده که قدرت نسبی یا ارزش ریاضی داده های ورودی اولیه و ارتباطات متنوع انتقال دهنده داده ها از یک لایه به لایه دیگر را نشان می دهند (چنگ و همکاران، ۲۰۰۸). رابطه بین فعال سازی درونی و خروجی شبکه می تواند خطی و یا غیر خطی باشد که این روابط به وسیله تابع محرک بیان می گردند که متداول ترین آنها تابع سیگموئید می باشد. شبکه های عصبی مانند انسان از طریق خطا هایش یاد میگیرد. قابلیت یادگیری در شبکه های عصبی به معنای توانایی تنظیم پارامترهای شبکه در مسیر زمان که محیط شبکه تغییر می کند و شبکه شرایط جدید را تجربه می نماید، با این هدف که اگر شبکه برای وضعیت خاصی آموزش دید و تغییر کوچکی در شرایط محیطی آن رخ داد، شبکه بتواند با آموزش مختصر برای شرایط جدید نیز کار آمد باشد (منهاج، ۱۳۸۶). برای موفقیت در استفاده از شبکه های عصبی رعایت چهار نکته ضروری می باشد. اول انتخاب صحیح مجموعه آموزش، سپس ارائه داده ها به صورتی که توانایی تشخیص شبکه حداکثر گردد. در مراحل بعدی تفسیر نتایج شبکه و درک جزئیات درون شبکه همچون هندسه و پارامترهای کنترل یادگیری از اهمیت فراوانی برخوردار هستند (قاسمی دشتکی، ۱۳۸۷). در مورد هندسه شبکه، انتخاب صحیح اندازه لایه پنهان بسیار مهم بوده و هر چه تعداد گره های آن بیشتر باشد، قابلیت شبکه برای تشخیص الگو بیشتر می گردد. تعیین تعداد نرون های لایه پنهان کار ساده ای نبوده و بیشتر با سعی و خطا صورت می گیرد. البته در این رابطه یک سری قواعد سر انگشتی نیز وجود دارد (راعی و فلاح پور، ۱۳۸۳). توضیحات کامل مربوط به شبکه های عصبی مصنوعی، انواع آن، روش های مختلف یادگیری آن ها را میتوان در (منهاج، ۱۳۸۶) یافت.

در این تحقیق از نرم افزار Rapid Miner که جزو نرم افزارهای پر کاربرد داده کاوی می باشد، برای تحقق رویکردهای شبکه عصبی و شبکه عصبی مصنوعی مبتنی بر آنالیز اجزای اساسی (PCA) استفاده شده

است. همچنین برای تحقق رویکرد شبکه عصبی مبتنی بر کلونی زنبور عسل از کد نویسی در زبان برنامه نویسی جاوا بهره جسته ایم.

۱۸ ویژگی هریک از دوایست و پنجاه مشتری به عنوان ورودی هریک از روش های آنالیز اجزای اساسی و کلونی زنبور عسل در نظر گرفته شده بودند که زیر مجموعه انتخابی به وسیله هریک از روش ها به عنوان ورودی شبکه عصبی به عنوان طبقه بندی کننده اصلی در نظر گرفته شد. جدول ۱ نشان گر متغیرهای به کار رفته در شبکه عصبی می باشد

جدول (۱) پارامترهای به کار رفته برای الگوریتم شبکه عصبی

تنظیمات	نام متغیر
پرسپترون چند لایه	نوع شبکه
الگوریتم پس انتشار خطا	الگوریتم یادگیری شبکه
۱	تعداد لایه های پنهان
۱۲	حجم لایه های پنهان
سیگموئید	نوع تابع محرک

### ۳-۳- توصیف داده ها

همان طور که گفته شد، این پژوهش سعی در ارائه رویکردی جدید برای ارزیابی ریسک اعتباری مشتریان بانک ها و موسسات مالی داشته و به کارگیری الگوریتم شبکه عصبی مبتنی بر کلونی زنبور عسل برای نخستین بار در زمینه ارزیابی ریسک اعتباری مطرح می گردد. هجده ویژگی به عنوان ورودی های الگوریتم های انتخاب ویژگی کلونی زنبور عسل و آنالیز اجزای اساسی برگزیده شده تا این الگوریتم ها با انتخاب ویژگی های بهینه با توجه به معیار دقت پیش بینی، آن ها را به عنوان ورودی روش طبقه بندی کننده شبکه عصبی قرار دهند. جدول نام این متغیرها و توضیحات مربوط به آنها را نشان می دهد. با توجه به عملیات پیش پردازش ، در نرم افزار Rapid Miner کلیه داده ها طبق رابطه (۹) به بازه [0,1] تبدیل گشتند. جدول ۲ در بردارنده این ۱۸ ویژگی و توضیحات مربوط به آنها می باشد. داده های مربوط به واحد مرکزی مدیریت ریسک بانک کارآفرین در شهر تهران در بازه زمانی سه ساله ۱۳۹۰/۰۱/۱-۱۳۸۷/۰۱/۰۱ می باشند. این اطلاعات برای دوایست و پنجاه تن از مشتریان بانک کارآفرین بوده که صد و هفتاد و پنج مشتری خوش حساب و هفتاد و پنج مشتری بد حساب در نظر گرفته شده اند.

$$x_i = \frac{x_i - \bar{x}_i}{6_i} \quad (9)$$

## جدول (۲) نام ویژگی ها و توضیحات مربوط به آنها

نوع متغیر	توضیح متغیر	نماد متغیر
دودویی	جنسیت	X1
عدد صحیح	سن	X2
رسته ای	تحصیلات متقاضی تسهیلات	X3
عدد حقیقی	سابقه کار متقاضی	X4
رسته ای	نوع وثیقه	X5
عدد حقیقی	ارزش وثیقه	X6
عدد حقیقی	درآمد متقاضی تسهیلات	X7
عدد حقیقی	زمان ارتباط متقاضی با بانک	X8
عدد حقیقی	میانگین موجودی حساب بانکی متقاضی	X9
رسته ای	سابقه متقاضی در گرفتن تسهیلات	X10
عدد حقیقی	مقدار اعتبار تخصیص یافته	X11
رسته ای	نوع شغل متقاضی	X12
رسته ای	هدف متقاضی از گرفتن تسهیلات	X13
عدد صحیح	تعداد افراد ضامن برای متقاضی	X14
عدد حقیقی	قسمتی از درآمد متقاضی که برای پرداخت اقساط پرداخت می گردد	X15
عدد صحیح	مدت زمان بازپرداخت تسهیلات اعطایی	X16
عدد صحیح	تعداد چک ها برگشت خورده متقاضی	X17
دودویی	صاحب خانه و یا مستاجر بودن گیرنده تسهیلات	X18

## روش اعتبارسنجی مقطعی

برای بدست آوردن دقت هر یک از رویکرد های پیشنهادی از روش اعتبارسنجی مقطعی استفاده کرده ایم. این عملگر داده ها را به دو قسمت، داده های آموزشی و داده های آزمایشی تقسیم می کند به طوری که ابتدا داده ها را به K زیرمجموعه ی مساوی تقسیم می کند، پس از آن هر بار K-1 زیرمجموعه از داده ها را به عنوان داده های آموزشی و فقط یک زیرمجموعه را به عنوان داده های تست در نظر گرفته و این کار را K بار انجام می دهد به طوری که هر زیرمجموعه یک بار به عنوان داده ی تست قرار می گیرد و دقت پیش-بینی برای آن زیرمجموعه محاسبه می گردد و دقت کلی مدل از میانگین دقت های بدست آمده محاسبه می-شود. در این پژوهش مقدار K برابر با ده در نظر گرفته شده است.

#### ۴- فرضیه‌های پژوهش

با توجه به مبانی علمی توضیح داده شده و در جهت پاسخ به دو سوال اصلی تحقیق در ارائه رویکردی نو برای ارزیابی ریسک اعتباری مشتریان موسسات مالی فرضیه های تحقیق به صورت زیر بیان میگردند:

**فرضیه اول:** " از لحاظ آماری تفاوت معناداری بین روش شبکه عصبی و روش شبکه عصبی مبتنی بر آنالیز اجزای اساسی وجود دارد."

**فرضیه دوم:** " از لحاظ آماری تفاوت معناداری بین روش شبکه عصبی مبتنی بر کلونی زنبور عسل با روش شبکه عصبی مبتنی بر آنالیز اجزای اساسی وجود دارد."

#### ۵- نتایج پژوهش

در این قسمت از پژوهش به بررسی نتایج تجربی حاصل از پیاده سازی سه روش شبکه عصبی، شبکه عصبی مبتنی بر کلونی زنبور عسل و شبکه عصبی مبتنی بر آنالیز اجزای اساسی می پردازیم. جدول (۳) نشان دهنده دقت کلی پیش بینی هر یک از سه روش به کار رفته در این پژوهش می باشد. دقت کلی پیش بینی در این پژوهش از روش اعتبارسنجی مقطعی محاسبه گردیده است که در هر مرحله با تقسیم کردن داده ها به ده زیرمجموعه و قرار دادن یک زیر مجموعه به عنوان داده های تست و سایر داده ها به عنوان داده های آموزشی عمل می نماید. این عمل به تعداد ده بار انجام شده تا همه ی داده ها یکبار در زیر مجموعه داده های تست قرار گیرند. دقت کلی از طریق میانگین گیری این ده بار حاصل شده است. جدول شماره ۴ نیز نشان دهنده ویژگی های انتخابی توسط دو رویکرد شبکه عصبی مبتنی بر کلونی زنبور عسل و شبکه عصبی مبتنی بر آنالیز اجزای اساسی می باشد. ویژگی های انتخابی با نماد (\*) نشان داده شده اند.

جدول (۳) میانگین دقت کلی پیش بینی سه روش به کار رفته

نام روش ترکیبی	میانگین دقت کلی پیش بینی
ABC-NN	76.1%
PCA-NN	73.2%
NN	72.4%

جدول (۴). ویژگی های انتخابی (\*) به وسیله هر روش

نام متغیر	نام طبقه بندی کننده	
	PCA-NN	ABC-NN
X1	*	
X2		
X3	*	*
X4		

نام متغیر	نام طبقه بندی کننده	
	PCA-NN	ABC-NN
X5	*	
X6		*
X7	*	*
X8	*	
X9	*	*
X10	*	
X11		
X12		
X13		
X14	*	*
X15		
X16	*	
X17		*
X18	*	*

برای بررسی فرضیه های پژوهشی مبنی بر عملکرد بهتر روش شبکه عصبی مبتنی بر آنالیز اجزای اساسی نسبت به روش شبکه عصبی تنها و همچنین عملکرد بهتر روش شبکه عصبی مبتنی بر کلونی زنبور عسل نسبت به روش شبکه عصبی بر پایه آنالیز اجزای اساسی، از آزمون مقایسات زوجی استفاده گردیده است. در این آزمون عملکرد بهتر روش اول از نظر آماری بررسی شده است. جداول شماره ۵ و ۶ نشان-دهنده نتایج آزمون مقایسه زوجی می باشد. در این آزمون نتایج هر یک از ده خروجی دقتی حاصل از به کارگیری روش اعتبار سنجی مقطعی برای هر یک از سه روش به عنوان ورودی های این آزمون در نظر گرفته شده اند. آزمون مقایسات زوجی در این پژوهش از طریق نرم افزار متلب صورت پذیرفته است.

#### جدول شماره (۵) نتیجه آزمون مقایسات زوجی برای روش های ذکر شده

NN	PCA-NN	
72.4%	73.2%	میانگین دقت کلی
۰/۰۰۰۴۱۸		p-value
فرضیه پذیرفته می شود		بررسی فرضیه اول پژوهشی



جدول شماره (۶) نتیجه آزمون مقایسات زوجی برای روش های ذکر شده

ABC-NN	PCA-NN	
76.1%	73.2%	میانگین دقت کلی
۰/۰۰۳۵۶		p-value
فرضیه پذیرفته می شود		بررسی فرضیه دوم پژوهشی

نتایج جدول شماره ۵ نشان می دهد (با توجه به p-value حاصل) که دقت کلی مدل شبکه عصبی بر پایه آنالیز اجزای اساسی (PCA-NN) در سطح اطمینان ۹۵٪ از دقت کلی مدل شبکه عصبی (NN) بهتر است. بنابراین فرضیه اول پژوهشی پذیرفته می شود. همچنین نتایج جدول ۶ دلالت بر آن داشته که فرضیه دوم پژوهشی نیز پذیرفته می شود و می توان گفت که در سطح اطمینان ۹۵٪ دقت کلی روش شبکه عصبی مبتنی بر کلونی زنبور عسل (ABC-NN) از دقت کلی مدل شبکه عصبی بر پایه آنالیز اجزای اساسی (PCA-NN) بهتر می باشد. در نتیجه برای داده های به کار رفته در این پژوهش مدل ABC-NN از خود بهترین عملکرد را نشان داد.

طراحی و اجرای سیستم مدیریت امور اعتباری نهادهای مالی با توجه به سیاست های کلان اعتباری و استراتژی های بلند مدت این موسسات می تواند نقش به سزایی در اتخاذ تصمیمات مناسب و اثر بخش ایفا نماید. در این راستا، هدف محوری این پژوهش ارائه نمودن یک طبقه بندی کننده کارا برای ارزیابی ریسک اعتباری مشتریان بانکی با توجه به دقت خروجی حاصل از تقسیم نمودن مشتریان به دو دسته خوش حساب و بد حساب، تعریف گردیده بود. با توجه به نتایج خروجی حاصل از دقت پیش بینی کلی و آزمون فرض های صورت پذیرفته، می توان از رویکرد شبکه عصبی بر پایه کلونی زنبور عسل که در مقایسه با روش های رقیب خود عملکرد بهتری از خود نشان داد، به عنوان یک طبقه بندی کننده ترکیبی پیشنهادی برای بررسی ریسک اعتباری متقاضیان تسهیلات از موسسات مالی استفاده نمود. نکته بسیار مهم در محاسبه دقت پیش بینی برای هر روش در نظر گرفتن هر دو حالت خطای آماری در بدست آوردن این دقت ها می باشد. از دید مالی مشتری خوش حساب به فردی اطلاق می گردد که معوقه نداشته و به کلیه تعهدات خود عمل نموده است. حال اگر این مشتری خوش حساب به اشتباه از طریق به کار گیری طبقه بندی کننده جزوه مشتریان بد حساب در نظر گرفته شده، از دید آماری خطای نوع اول رخ داده است. از سوی دیگر اگر مشتری بد حسابی که در باز پرداخت تسهیلات اعطایی کوتاهی نموده به نادرستی از دید طبقه بندی کننده جزوه مشتریان خوش حساب قرار گیرد، خطای نوع دوم از نگاه آماری صورت پذیرفته که هر دو این خطاها در محاسبه دقت پیش بینی کلی در این تحقیق مد نظر گرفته شده اند. همچنین هر سه مدل بیان گردیده در این پژوهش هم از نظر مالی و اقتصادی معنی دار بوده زیرا نتایج بدست آمده از پیش بینی ها در خصوص تسهیلات گیرندگان با نتایج عملکرد واقعی آنها در پایان دوره مقایسه گردیده و

دقت کلی پیش بینی استخراج گردیده است. نتایج حاصل به خوبی بر اهمیت به کار گیری رویکرد انتخاب ویژگی که در آن دو روش کلونی زنبور عسل به عنوان یک روش پوشش دهنده و آنالیز اجزای اساسی به عنوان یک روش فیلتر کننده نمایندگان آن بودند، دلالت داشته و هر دو روش نسبت به طبقه بندی کننده شبکه عصبی از دقت پیش بینی بالاتری برخوردار بودند. همچنین انتخاب ویژگی هایی همچون تحصیلات تسهیلات گیرنده، درآمد فرد، تعداد افراد ضامن، مستاجر یا صاحب خانه بودن و همچنین میانگین وجودی حساب فرد متقاضی، توسط هر دو روش انتخاب ویژگی حاکی از اهمیت بالا این ویژگی ها در ارزیابی ریسک اعتباری مشتریان برای داده های به کار رفته در این پژوهش می باشد.

#### ۶- نتیجه گیری و بحث

یکی از عوامل مهم ورشکستگی بانک ها ریسک اعتباری (احتمال قصور از پرداخت تسهیلات اعطایی از سوی مشتریان) است. اگر مشتری به موقع تعهدات خود را بازپرداخت نکند، این تسهیلات به صورت مطالبات معوق بانکی درآمده و موجب اختلال در سیستم بانکی و در نتیجه اقتصاد کشور می گردد. هدف اصلی این تحقیق نیز، استفاده از ابزارهای داده کاوی در جهت ارزیابی ریسک اعتباری مشتریان موسسات مالی بوده تا در نهایت منجر به کاهش معوقات بانکی و تصمیم گیری بهتر موسسات مالی در اعطای تسهیلات گردد.

روش اصلی به کار رفته در این تحقیق، مدل شبکه عصبی مبتنی بر کلونی زنبور عسل بوده که عملکرد دقتی این طبقه بندی کننده با روش های شبکه عصبی و روش شبکه عصبی مبتنی بر آنالیز اجزای اساسی برای داده های استفاده شده در این پژوهش که شامل دویست و پنجاه نفر (۱۷۵ نفر خوش حساب و ۷۵ نفر بدحساب) از تسهیلات گیرندگان بانک کارآفرین با توجه به هجده ویژگی مربوط به آن ها در یک بازه زمانی سه ساله می باشد، مقایسه گردیده است.

نتایج حاصل دلالت بر بهتر بودن روش شبکه عصبی مبتنی بر کلونی زنبور عسل نسبت به دو روش دیگر داشت. همچنین استفاده از کلونی زنبور عسل به عنوان یک روش انتخاب ویژگی پوشش دهنده نقش مهمی در افزایش دقت و سرعت در اجرای تکنیک پیش بینی، حذف داده های غیر مرتبط و افزایش قابلیت فهم پذیری ایفا نمود و همچنین در مقایسه با عملکرد روش انتخاب ویژگی فیلتر کننده آنالیز اجزای اساسی، کلونی زنبور عسل به عنوان یک روش پوشش دهنده در ترکیب با شبکه عصبی دقت پیش بینی بالاتری از خود نشان داد.

با توجه به آنکه داده های به کار رفته در این پژوهش منحصر به فرد بوده و در تحقیقات دیگر از این داده ها استفاده نگردیده، به صورت دقیق نمی توان مقایسه ای بین دقت پیش بینی حاصل از به کارگیری رویکرد های این پژوهش با سایر پژوهش ها انجام داد. ولی اشاره به این نکته ضروری بوده که استفاده از کلونی زنبور عسل به عنوان یک روش انتخاب ویژگی پوشش دهنده و ترکیب آن با شبکه عصبی در زمینه ارزیابی ریسک اعتباری بانکی برای نخستین بار در این پژوهش مطرح گردیده است و نتایج این تحقیق به خوبی نشان از برتری این رویکرد نسبت به رویکرد شبکه عصبی می باشد. در پژوهشی که توسط اورسکی و

همکاران (۲۰۱۲) با استفاده از رویکرد ترکیبی شبکه عصبی مبتنی بر الگوریتم ژنتیک برای داده های کشور کرواسی با توجه به سی ویژگی تعریف شده برای هفتصد و پنجاه مشتری بانکی صورت پذیرفت، دقت پیش بینی کلی حاصل از به کارگیری طبقه بندی کننده ترکیبی پیشنهادی برابر با ۷۱٪ حاصل گردیده بود. همچنین در پژوهشی مشابه، مین ولی (۲۰۰۷) با توجه به داده های مشتریان حقوقی خود با استفاده از رویکرد تحلیل پوششی داده ها، اقدام به ارزیابی ریسک اعتباری ۱۰۶۱ شرکت کره ای نموده که دقت پیش بینی طبقه بندی کننده غیر ترکیبی آنها برابر با ۶۷٫۵٪ محاسبه گردید. دقت پیش بینی طبقه بندی کننده ترکیبی پیشنهادی در این پژوهش از دقت پیش بینی کلی هر دو تحقیق بیان شده بیشتر می باشد. با توجه به تأیید کارایی روش های هوش مصنوعی جهت حل مسائل پیچیده، توصیه میشود که در عمل، استفاده از این روشها در سنجش ریسک اعتباری مشتریان مد نظر قرار گیرد، اما به دلیل این که ساختار داده های مشتریان تأثیر بسزایی در عملکرد مدل خواهد داشت، بهتر است در مراحل اولیه حتماً از یکی از روش های انتخاب ویژگی به خصوص روش های پوشش دهنده استفاده شده و با ساخت یک طبقه بندی کننده ترکیبی اقدام به تخمین رتبه اعتباری مشتریان تسهیلات در بانک ها نماییم.

در تحقیقات آتی میتوان از سایر روش های انتخاب ویژگی همچون الگوریتم بهینه سازی ذرات تجمعی، الگوریتم مورچگان، الگوریتم کرم شب تاب، الگوریتم استعماری و یا روش های پیش بینی دیگر همچون لجستیک رگرسیون، ماشین بردار پشتیبان با توجه به توابع کرنل مختلف آن و روش مجاورت نزدیکترین همسایه بهره جست. رویکرد به کار رفته در این تحقیق می تواند کمک شایانی به موسسات مالی و اعتباری در چگونگی تخصیص اعتبارات خود به مشتریان نموده و همچنین ایجاد نرم افزارهایی هوشمند برای تعیین رتبه اعتباری متقاضیان با توجه به ویژگی آن ها از دیگر نکات قابل ذکر برای کاربرد پژوهش انجام شده می باشد.

#### فهرست منابع

- \* ارمشی، مینا (۱۳۹۰) شناسایی متغیرهای مؤثر بر ریسک اعتباری مشتریان حقیقی بانک سامان (شعب شمال کشور)، پایان نامه کارشناسی ارشد، دانشکده علوم انسانی
- \* پویان فر، ا.، فلاح پور، س.، عزیزی، م. (۱۳۹۲)، رویکرد حداقل مربعات ماشین بردار پشتیبان مبتنی بر الگوریتم ژنتیک جهت تخمین رتبه اعتباری مشتریان بانک ها، مجله مهندسی مالی و مدیریت اوراق بهادار، شماره ۱۷
- \* خانی سرو علیایی، م. (۱۳۸۶). طراحی و تبیین مدل ریک اعتبارات در نظام بانکی کشور (پایان نامه ارشد مدیریت بازرگانی، دانشگاه آزاد تهران مرکز)
- \* راعی، رضا، فلاح پور، سعید، (۱۳۸۳)، پیش بینی درماندگی مالی شرکت ها با استفاده از شبکه های مصنوعی، تحقیقات مالی، شماره ۱۷

- \* راعی،ر، فلاح پور، س. (۱۳۸۷). کاربرد ماشین بردار پشتیبان در پیش بینی در ماندگی مالی شرکت ها با استفاده از نسبت های مالی. بررسی های حسابداری و حسابرسی، دوره ۱۵
- \* رهنمای رودپشتی،ف،علی خانی،ر،و مران جوری،م.(۱۳۸۸). بررسی کاربرد مدل های پیش بینی ورشکستگی آلتمن و فالمر در شرکت های پذیرفته شده در بورس اوراق بهادار تهران، بررسی های حسابداری و حسابرسی، ۱۹-۳۴
- \* صفری، س، ابراهیمی شقاقی، م. و شیخ،م.(۱۳۸۹). مدیریت ریسک اعتباری مشتریان حقوقی در بانک های تجاری با رویکرد تحلیل پوشش داده ها (رتبه بندی اعتباری)، پژوهش های مدیریت در ایران، دوره ۱۴، شماره ۴
- \* فلاح شمس،م،و مهدوی راد،ح.(۱۳۸۹). طراحی مدل اعتبار سنجی و پیش بینی ریسک اعتباری مشتریان تسهیلات لیزینگ، مهندسی مالی و مدیریت پرتفوی، ۱-۲۲
- \* فلاح شمس، م.، تهرانی،ر.(۱۳۸۴). طراحی و تبیین مدل ریسک اعتباری در نظام بانکی کشور، مجله علوم اجتماعی و انسانی دانشگاه شیراز، شماره ۴۳.
- \* فلاح شمس، م.(۱۳۸۷). مدل های اندازه گیری ریسک اعتباری در بانک ها و موسسه های اعتباری، نشریه تازه های اقتصاد، ۲۸-۲۲
- \* قاسمی دشتکی، مهسا(۱۳۸۷)، پیش بینی قیمت گذاری عرضه های عمومی اولیه، ترکیب شبکه های عصبی و الگوریتم ژنتیک، پایان نامه کارشناسی ارشد، رشته حسابداری دانشگاه تهران
- \* کیمیاگری، محمدعلی، امینی، محمدجعفر، طباطبایی آرانی، حمید، حسینی، لیلا، (۱۳۹۱) مدل ریسک اعتباری بازپرداخت تسهیلات مشتریان اعتباری بانک (موردی حوزه شهرضا بانک ملی ایران)، نهمین کنفرانس بین المللی مهندسی صنایع
- \* منهای، محمد باقر،(۱۳۸۶)، مبانی شبکه های عصبی (هوش محاسباتی)، تهران، دانشگاه صنعتی امیرکبیر، چاپ چهارم
- \* همتی، عبدالناصر و محبی تژاد، شادی، (۱۳۸۸)، ارزیابی تاثیر متغیرهای کلان اقتصادی بر ریسک اعتباری بانک ها، پژوهشنامه اقتصادی ؛ (۶ ویژه نامه بانک): ۳۳-۵۹.
- \* Altman, E.(1968).Financial ratios discriminate analysis and the prediction of corporate Bankruptcy. Journal of Finance
- \* Caauette,J.-B,Altman.(1998);Managing Credit Risk the next great financial Challenge, Journal of Finance
- \* Chung,C,H,Chen,T,L,Jong Teoh,H,Chiang, C,H.(2008)'Fuzzy time-series based on adaptive expectation model for TAIEX forecasting' Expert System with Applications" 34:1126-1132
- \* D. Karaboga.(2005), An idea based on honey bee swarm for numerical optimization, Technical Report, Computer Engineering Department. Engineering Faculty. Erciyes University.
- \* Desai V.S.(1996):" A comparison of neural networks and linear Rating models in the credit union environment" European journal of Operational Research;Vol.95
- \* Femandes,E,(2005),"Corporate Credit Risk Modeling", Quantitative Rating System,www.defaultrisk.com

- \* Huang, C-L., Chen, M-C., & Wang, C-J. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications*, 847-85
- \* Jolliffe, I.T. (2002). *Principal Component Analysis*, second edition (Springer)
- \* Li, H., Li, C.J., Wu, J. (2014). Statistics-based wrapper for feature selection: An implementation of financial distress identification with support vector machine, *Applied Soft computing*, 19, 57-67
- \* Oreski, S, Oreski, D., & Oreski, G. (2012). Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. *Expert systems with applications*, 39 (16), 12605-12617.
- \* N.g, Yeung, D.S, Firth, M. et al. (2008). Feature selection localized generalization error for supervised classification problems using RBFNN, *Pattern Recognition*, 41, 3706-3719
- \* Min JH, Lee YC.; (2007); "A practical approach to credit Rating"; *Journal of Expert Systems With Applications*; Volume 9
- \* West D.; (2000) "Neural network credit Rating models"; *Journal of Computer & Operation Research*, Volume. 27