****

****

دانشگاه آزاد اسلامي

واحد تهران شمال

دانشکده فنی مهندسی، گروه صنایع

پایان‌نامه براي دريافت درجه كارشناسي ارشد (M.Sc) / دکتری (Ph.D)

گرايش: **مهندسی مالی**

عنوان:

**تلفیق ماتریس زیان و طبقه بندی ترکیبی مبتنی بر روش بگینگ به‌منظور ارزیابی ریسک اعتباری**

استاد راهنما:

دکتر **شروین اسدزاده**

استاد مشاور:

دکتر (نام استاد مشاور)

نگارش:

**نیما عالم شناس**

**زمستان** **1403**

**A blue and black logo

Description automatically generated**

تعهدنامه اصالت رساله یا پایان‌نامه

این‌جانب نیما عالم شناس دانش‌آموخته مقطع کارشناسی ارشد ناپیوسته/دکترای تخصصی در رشته مهندسی صنایع گرایش مهندسی مالی که در تاریخ .................. از پایان‌نامه/ رساله خود تحت عنوان

"تلفیق ماتریس زیان و طبقه‌بندی ترکیبی مبتنی بر روش بگینگ به منظور ارزیابی ریسک اعتباری"

با کسب نمره ................. و درجه ..................... دفاع نموده‌ام بدین‌وسیله متعهد می‌شوم:

1. این پایان‌نامه/ رساله حاصل تحقیق و پژوهش انجام‌شده توسط این‌جانب بوده و در مواردی که از دستاوردهای علمی و پژوهشی دیگران(اعم از پایان‌نامه، کتاب،مقاله و....) استفاده نموده‌ام، مطابق ضوابط و رویه موجود، نام منبع مورداستفاده و سایر مشخصات آن را در فهرست مربوطه ذکر و درج کرده‌ام.
2. این پایان‌نامه/ رساله قبلاً برای دریافت هیچ مدرک تحصیلی(هم‌سطح،پایین‌تر یا بالاتر) در سایر دانشگاه‌ها و مؤسسات آموزشی عالی ارائه نشده است.
3. چنانچه بعد از فراغت تحصیل، قصد استفاده و هرگونه بهره‌برداری اعم از چاپ کتاب،ثبت اختراع و... از این پایان‌نامه داشته باشم،از حوزه معاونت پژوهشی واحد مجوزهای مربوطه را اخذ نمایم.
4. چنانچه در هر مقطعی زمانی خلاف موارد فوق ثابت شود،عواقب ناشی از آن را می‌پذیرم و واحد دانشگاهی مجاز است با این‌جانب مطابق ضوابط و مقررات رفتار نموده و در صورت ابطال مدرک تحصیلی‌ام هیچ‌گونه ادعایی نخواهم داشت.

نام و نام خانوادگی

تاریخ و امضاء

تقدیم به

پدرم، نماد استقامت و شرافت  
 و مادرم، سرچشمه عشق و مهر

با سپاس از

آنکه با مهر، درخت دانش را بارور کرد  
با تلاش و صبوری، مسیرمان را هموار کرد  
خانواده‌ای که پشتوانه این راه دشوار بودند  
و استادی که فانوس علم را در دستمان سپرد

فهرست مطالب

عنوان صفحه

[چکیده 1](#_Toc188023291)

[فصل اول 2](#_Toc188023292)

[بیان مسئله 3](#_Toc188023293)

[1-2 اهمیت و ضرورت انجام تحقیق 6](#_Toc188023294)

[1-2-1 محدودیت‌های روش‌های سنتی در مدیریت ریسک اعتباری 7](#_Toc188023295)

[1-2-2 ضرورت بهره‌گیری از ماتریس زیان و تکنیک بگینگ در تحلیل ریسک 7](#_Toc188023296)

[1-2-3 مزایای استفاده از بگینگ در کاهش خطاهای تصمیم‌گیری 8](#_Toc188023297)

[1-2-4 ابزارهای کاربردی برای بانک‌ها در مدیریت ریسک 8](#_Toc188023298)

[1-3 جنبه جديد بودن و نوآوري در تحقيق 9](#_Toc188023299)

[1-3-1 ترکیب ماتریس زیان و بگینگ 9](#_Toc188023300)

[1-3-2 تحلیل داده‌های نامطمئن 10](#_Toc188023301)

[1-3-3 مدل‌های پیشرفته در ساختار بگینگ 10](#_Toc188023302)

[1-3-4 کاربردهای عملیاتی و کاهش هزینه‌ها 10](#_Toc188023303)

[1-3-5 پایداری پیش‌بینی‌ها در شرایط ناپایدار 10](#_Toc188023304)

[1-3-6 مقایسه با سایر رویکردها 11](#_Toc188023305)

[1-4 سؤالات تحقیق 12](#_Toc188023306)

[1-5 فرضيه‏هاي تحقیق 12](#_Toc188023307)

[1-6 روش شناسی تحقیق 12](#_Toc188023308)

[1-6-1 جمع‌آوری و پیش‌پردازش داده‌ها 13](#_Toc188023309)

[1-6-2 طراحی و پیاده‌سازی ماتریس زیان 13](#_Toc188023310)

[1-6-3 استفاده از الگوریتم‌های یادگیری ماشین 13](#_Toc188023311)

[1-6-4 رویکرد بگینگ در طبقه‌بندی ترکیبی 14](#_Toc188023312)

[1-6-5 ارزیابی و اعتبارسنجی مدل 14](#_Toc188023313)

[1-7 جامعه آماري 15](#_Toc188023314)

[1-8 مدل مورد استفاده در تحقیق و تعریف متغیرها 16](#_Toc188023315)

[1-8-1 متغیرهای مستقل و تأثیر آن‌ها بر ریسک اعتباری 17](#_Toc188023316)

[1-8-2 نقش ماتریس زیان در بهبود عملکرد مدل 17](#_Toc188023317)

[1-8-3 تحلیل داده‌ها با روش‌های طبقه‌بندی ترکیبی 17](#_Toc188023318)

[1-8-4 قابلیت‌های تطبیق‌پذیری و به‌روزرسانی مدل 18](#_Toc188023319)

[1-8-5 تأثیر مدل پیشنهادی بر مدیریت ریسک بانکی 18](#_Toc188023320)

[1-9 ساختار پایان‌نامه 18](#_Toc188023321)

[فصل دوم: 20](#_Toc188023322)

[مروری بر مطالعات پیشین 20](#_Toc188023323)

[20](#_Toc188023324)

[2-1 مقدمه 21](#_Toc188023325)

[2-2 مبانی نظری 22](#_Toc188023326)

[2-2-1 مفهوم ریسک و مدیریت آن 22](#_Toc188023327)

[2-2-2 ریسک اعتباری 22](#_Toc188023328)

[2-2-3 ماتریس زیان در ارزیابی ریسک 23](#_Toc188023329)

[4-2-2 روش بگینگ در یادگیری ماشین 24](#_Toc188023330)

[5-2-2 ترکیب روش‌های طبقه‌بندی 25](#_Toc188023331)

[2-3 پیشینه تحقیق 26](#_Toc188023332)

[1-3-2 تحقیقات داخلی 26](#_Toc188023333)

[2-3-2 تحقیقات خارجی 28](#_Toc188023334)

[2-4 جمع‌بندی 34](#_Toc188023335)

[فصل سوم: 36](#_Toc188023336)

[روش تحقیق 36](#_Toc188023337)

[(مراحل انجام پژوهش) 36](#_Toc188023338)

[36](#_Toc188023339)

[3-1 عنوان موردنظر 37](#_Toc188023340)

[فصل چهارم: 39](#_Toc188023341)

[شبیه‌سازی و تحلیل نتایج 39](#_Toc188023342)

[(نتایج و بحث) 39](#_Toc188023343)

[4-1 عنوان موردنظر 40](#_Toc188023344)

[فصل پنجم: 42](#_Toc188023345)

[نتیجه‌گیری و پیشنهادها 42](#_Toc188023346)

[42](#_Toc188023347)

[5-1 نتیجه‌گیری 43](#_Toc188023348)

[5-2 پيشنهادها 43](#_Toc188023349)

[پیوست الف (در صورت وجود) 44](#_Toc188023350)

[فهرست 46](#_Toc188023351)

[منابع و مأخذ 46](#_Toc188023352)

[46](#_Toc188023353)

[فهرست منابع و مأخذ 47](#_Toc188023354)

فهرست جدول‌ها

عنوان صفحه

فهرست شکل­ها

عنوان صفحه

چکیده

متن چکیده

کلمات کلیدی: متن کلمات کلیدی.

فصل اول

بیان مسئله

بیان مسئله ریسک اعتباری به‌عنوان یکی از مهم‌ترین چالش‌های مالی، به احتمال نکول وام‌ها یا عدم بازپرداخت تعهدات مالی توسط وام‌گیرندگان اشاره دارد. این موضوع تأثیر مستقیمی بر عملکرد بانک‌ها، مؤسسات مالی و در نهایت اقتصاد کلان دارد. ناکامی در مدیریت این ریسک می‌تواند منجر به افزایش بدهی‌های غیرجاری، کاهش منابع قابل‌تخصیص بانک‌ها، و ایجاد فشار بر توانایی آن‌ها در اعطای وام‌های جدید شود. چنین وضعیتی، نه‌تنها عملکرد مالی مؤسسات را مختل می‌کند، بلکه به کاهش اعتماد مشتریان و سرمایه‌گذاران نیز منجر می‌شود. بنابراین، مدیریت کارآمد ریسک اعتباری، کلید حفظ ثبات مالی بانک‌ها و مؤسسات مالی است.

روش‌های سنتی مدیریت ریسک، عمدتاً مبتنی بر تصمیم‌گیری دو کلاسه بودند که مشتریان را به دو دسته خوش‌حساب و بدحساب تقسیم می‌کنند. این رویکرد ساده‌سازی شده، در مواجهه با داده‌های نامطمئن و ناقص، ناکارآمدی‌هایی از خود نشان می‌دهد. تصمیم‌گیری‌های نادرست در این روش‌ها منجر به افزایش خطاهای تصمیم‌گیری و در نهایت کاهش بهره‌وری بانک‌ها می‌شود. همچنین، در این روش‌ها دسته‌ای از مشتریان که اطلاعات کافی برای قضاوت قاطع درباره وضعیت آن‌ها وجود ندارد، نادیده گرفته می‌شوند. این چالش‌ها ضرورت استفاده از رویکردهای نوین را افزایش داده است.

پیشگیری از نکول مستلزم به‌کارگیری روش‌های پیشرفته ارزیابی و مدیریت ریسک اعتباری است. بانک‌ها می‌توانند با استفاده از مدل‌های پیش‌بینی مبتنی بر داده، مشتریان پرخطر را شناسایی کرده و سیاست‌های اعتباری مناسب را اعمال کنند. این راهکارها به کاهش نکول و افزایش ثبات مالی بانک کمک می‌کند.

نکول وام‌ها تأثیر مستقیمی بر نقدینگی بانک‌ها دارد و می‌تواند منجر به کاهش منابع مالی قابل‌دسترس برای عملیات جاری آن‌ها شود. با عدم بازپرداخت وام توسط مشتریان، بانک‌ها با کاهش جریان‌های نقدی ورودی مواجه شده و در نتیجه، توانایی آن‌ها در تأمین مالی سایر وام‌ها یا پرداخت به سپرده‌گذاران کاهش می‌یابد. این موضوع فشار مضاعفی بر منابع مالی بانک وارد کرده و ممکن است نیاز به تأمین مالی اضطراری از طریق بازارهای سرمایه یا استفاده از ذخایر نقدی را افزایش دهد. در شرایط حادتر، نکول می‌تواند به افزایش بدهی‌های غیرجاری در ترازنامه بانک منجر شود و سلامت مالی آن را به خطر بیندازد. این وضعیت، نه‌تنها هزینه‌های تأمین مالی بانک را افزایش می‌دهد، بلکه اعتماد مشتریان و سرمایه‌گذاران به بانک را نیز تضعیف می‌کند. مدیریت صحیح ریسک اعتباری و اجرای سیاست‌های پیشگیرانه برای کاهش نکول، کلیدی برای حفظ نقدینگی و ثبات مالی بانک‌ها است.

تحلیل جریان نقدی یکی از ابزارهای اساسی در ارزیابی سلامت مالی بانک‌ها و شرکت‌ها است که میزان ورود و خروج وجوه نقد در یک دوره زمانی مشخص را بررسی می‌کند. این تحلیل کمک می‌کند تا توانایی سازمان در تأمین تعهدات مالی، پرداخت بدهی‌ها، و سرمایه‌گذاری در پروژه‌های جدید ارزیابی شود. در بانک‌ها، جریان نقدی مثبت نشان‌دهنده عملکرد مطلوب در مدیریت وام‌ها و بازپرداخت‌ها است، در حالی که جریان نقدی منفی می‌تواند نشان‌دهنده افزایش نکول و مشکلات نقدینگی باشد. این تحلیل به مدیران کمک می‌کند تا استراتژی‌هایی برای مدیریت بهینه منابع مالی و کاهش ریسک نقدینگی تدوین کنند. همچنین، پیش‌بینی جریان نقدی با استفاده از مدل‌های پیشرفته تحلیل داده، می‌تواند تصمیم‌گیری‌های آگاهانه‌تر در تخصیص منابع و سرمایه‌گذاری‌ها را ممکن سازد. بهره‌گیری از تحلیل جریان نقدی، نقش مهمی در تقویت پایداری مالی و اعتماد سرمایه‌گذاران و مشتریان به بانک‌ها و مؤسسات مالی ایفا می‌کند.

بحران‌های بانکی اغلب ناشی از عدم مدیریت صحیح ریسک‌های اعتباری، نقدینگی، و بازار هستند و می‌توانند تأثیرات مخربی بر اقتصاد کلان داشته باشند. کاهش و کنترل این ریسک‌ها نیازمند اجرای چارچوب‌های قوی مدیریت ریسک است که شامل شناسایی، ارزیابی، و پایش مداوم ریسک‌ها می‌شود. بانک‌ها می‌توانند با به‌کارگیری مدل‌های پیش‌بینی ریسک و تحلیل داده‌های تاریخی، احتمال وقوع بحران‌ها را کاهش دهند. از سوی دیگر، تقویت شفافیت در گزارش‌دهی مالی و افزایش اعتماد عمومی از طریق ارتباطات شفاف با سهام‌داران و مشتریان، نقشی حیاتی در مدیریت و کنترل ریسک بحران‌های بانکی دارد. اجرای این رویکردها به حفظ ثبات نظام مالی و جلوگیری از بروز بحران‌های سیستمی کمک می‌کند.

یکی از راهکارهای نوین برای مدیریت بهتر ریسک اعتباری، استفاده از ماتریس زیان است. این رویکرد، به‌جای محدود کردن تصمیم‌گیری به دو گزینه ساده، داده‌ها را به سه دسته خوش‌حساب، بدحساب و مرزی تقسیم می‌کند. دسته‌بندی موارد مرزی فرصتی برای جمع‌آوری اطلاعات بیشتر و اتخاذ تصمیمات دقیق‌تر فراهم می‌آورد. این دسته‌بندی، انعطاف‌پذیری بیشتری در مواجهه با داده‌های نامطمئن ایجاد می‌کند و امکان کاهش ریسک‌های ناشی از تصمیم‌گیری‌های اشتباه را فراهم می‌سازد.

موفقیت در ارزیابی ریسک اعتباری به تعیین دقیق آستانه‌های α و β وابسته است که مرزهای این دسته‌ها را مشخص می‌کنند. اگر این آستانه‌ها به‌درستی تعیین نشوند، دسته‌بندی نادرست می‌تواند منجر به افزایش خطاها و کاهش دقت مدل شود. در این پژوهش، آستانه‌های α و β با استفاده از تحلیل ماتریس زیان و با توجه به پیامدهای تصمیم‌گیری به‌گونه‌ای تعیین می‌شوند که زیان کل سیستم کاهش یابد. این روش به مدل امکان می‌دهد تا تصمیمات مؤثرتری در ارزیابی ریسک اعتباری اتخاذ کرده و دقت آن را ارتقا دهد. برای تحلیل و پیش‌بینی دقیق‌تر دسته موارد مرزی، از رویکرد طبقه‌بندی ترکیبی با روش بگینگ استفاده می‌شود. بگینگ با اجرای چندین مدل پایه به‌صورت موازی و ترکیب خروجی آن‌ها از طریق روش‌هایی مانند رأی‌گیری یا میانگین‌گیری، دقت تحلیل و پیش‌بینی را بهبود می‌بخشد و پایداری مدل را افزایش می‌دهد.

در سال‌های اخیر، با رشد فزاینده تقاضای اعتبارات و توسعه بانکداری دیجیتال، حجم داده‌های مالی به‌طرز چشمگیری افزایش یافته است. این تغییرات موجب شده تا روش‌های سنتی ارزیابی ریسک، توانایی خود را در تحلیل داده‌های حجیم و پیچیده از دست بدهند.

ازاین‌رو، استفاده از رویکردهای داده‌محور و تکنیک‌های یادگیری ماشین در مدیریت ریسک اعتباری اهمیت بیشتری یافته است. این رویکردها به بانک‌ها اجازه می‌دهند تا الگوهای پنهان در داده‌ها را شناسایی کرده و تصمیمات هوشمندانه‌تری اتخاذ کنند. در این میان، ترکیب تکنیک‌های طبقه‌بندی پیشرفته با ماتریس زیان، امکان ارائه چارچوب‌های دقیق‌تر و سازگارتر با شرایط واقعی را فراهم می‌کند. همچنین، رویکردهای سنتی اغلب از هزینه‌های مرتبط با خطاهای طبقه‌بندی چشم‌پوشی می‌کنند و درنتیجه، مدل‌های پیشنهادی فاقد تحلیل عمیق تأثیرات مالی هستند. ماتریس زیان با وارد کردن این هزینه‌ها در فرآیند تصمیم‌گیری، به بانک‌ها کمک می‌کند تا با محاسبه دقیق ریسک‌های مالی، سیاست‌های مؤثرتری اتخاذ کنند. به‌عنوان‌مثال، پیامدهای اقتصادی ناشی از طبقه‌بندی اشتباه یک مشتری بدحساب به‌عنوان خوش‌حساب، می‌تواند منجر به افزایش زیان‌های اعتباری بانک شود. در مقابل، طبقه‌بندی یک مشتری مرزی به‌عنوان بدحساب ممکن است منجر به از دست دادن فرصت‌های سودآور شود. بنابراین، تحلیل دقیق پیامدهای خطاهای تصمیم‌گیری از طریق ماتریس زیان، نقش کلیدی در بهینه‌سازی مدل‌های پیش‌بینی دارد.

از سوی دیگر، استفاده از روش‌های ترکیبی مانند بگینگ، مزایایی نظیر افزایش پایداری و کاهش حساسیت مدل‌ها به داده‌های پرت را به همراه دارد. این روش‌ها از طریق ترکیب چندین مدل پایه، تأثیر نویزهای موجود در داده‌ها را کاهش داده و دقت پیش‌بینی‌ها را بهبود می‌بخشند. درواقع، رویکرد طبقه‌بندی ترکیبی به بانک‌ها کمک می‌کند تا با شناسایی الگوهای پیچیده و نامنظم در داده‌های اعتباری، فرآیند ارزیابی ریسک را بهبود دهند. به‌ویژه در شرایطی که بانک‌ها با داده‌های ناقص یا نامطمئن مواجه هستند، استفاده از چنین رویکردهایی ضروری به نظر می‌رسد. در این پژوهش، تمرکز اصلی بر توسعه مدل‌هایی است که ضمن افزایش دقت پیش‌بینی، هزینه‌های مرتبط با تصمیم‌گیری‌های اشتباه را نیز به حداقل برسانند. چارچوب پیشنهادی، بانک‌ها و مؤسسات مالی را قادر می‌سازد تا علاوه بر شناسایی دقیق مشتریان پرخطر، سیاست‌های اعتباری خود را به‌گونه‌ای تنظیم کنند که ضمن کاهش زیان‌های احتمالی، اعتماد مشتریان و سرمایه‌گذاران نیز حفظ شود. این مدل، می‌تواند به‌عنوان ابزاری کارآمد در ارزیابی ریسک اعتباری و اتخاذ تصمیمات بهینه مالی به کار گرفته شود و درنهایت به بهبود بهره‌وری و پایداری سیستم‌های مالی منجر شود.

1-2 اهمیت و ضرورت انجام تحقیق

ریسک اعتباری به‌عنوان یکی از چالش‌های کلیدی در نظام بانکی و مالی، تأثیری عمیق بر پایداری اقتصادی و کارایی مؤسسات مالی دارد. در محیط رقابتی فعلی بین بانک‌های کشور، جذب مشتریان جدید و حفظ مشتریان فعلی به یک اولویت استراتژیک تبدیل شده است. افزایش تعداد مؤسسات مالی و ارائه خدمات متنوع، بانک‌ها را به سمت استفاده از روش‌های نوین مدیریت ریسک و تحلیل داده‌های مشتریان سوق داده است. رقابت فشرده، بانک‌ها را ملزم می‌کند تا نه تنها دقت خود را در ارزیابی ریسک اعتباری افزایش دهند، بلکه با بهره‌گیری از فناوری‌های پیشرفته مانند یادگیری ماشین، مزیت رقابتی خود را در بازار تثبیت کنند.

ریسک اعتباری نه تنها عملکرد داخلی بانک‌ها و مؤسسات مالی را تحت تأثیر قرار می‌دهد، بلکه ابعاد گسترده‌تری مانند اشتغال، سرمایه‌گذاری و رفاه اجتماعی را نیز شامل می‌شود. عدم مدیریت صحیح این ریسک می‌تواند منجر به کاهش اعتماد عمومی، افت سرمایه‌گذاری‌های داخلی و خارجی، و کاهش پایداری اقتصاد کلان شود. افزایش تقاضا برای وام‌ها و رشد پیچیدگی در داده‌های مالی، بانک‌ها را مجبور به استفاده از روش‌های پیشرفته‌تر برای مدیریت ریسک کرده است. روش‌های سنتی که عموماً بر پایه تصمیم‌گیری دوکلاسه بنا شده‌اند، به دلیل ناکارآمدی در تحلیل داده‌های ناقص و ناتوانی در مدیریت مشتریان مرزی، دیگر پاسخگوی نیازهای امروز نیستند. علاوه بر این، بحران‌های مالی و اقتصادی مانند رکود یا همه‌گیری‌های جهانی، ضرورت بهره‌گیری از رویکردهای انعطاف‌پذیر و پویا را افزایش داده‌اند. بانک‌ها نیازمند مدل‌هایی هستند که بتوانند در شرایط عدم قطعیت و نوسانات اقتصادی، پایداری مالی را تضمین کنند و ریسک‌های مرتبط با اعتباردهی را به حداقل برسانند.

1-2-1 محدودیت‌های روش‌های سنتی در مدیریت ریسک اعتباری

این روش‌ها معمولاً منجر به افزایش خطاهای تصمیم‌گیری، اتلاف منابع و کاهش بهره‌وری سیستم مالی می‌شوند. یکی از خلأهای تحقیقاتی موجود در این حوزه، نبود یک چارچوب برای تعیین آستانه‌های تصمیم‌گیری آلفا و بتا می‌باشد. تحقیقات پیشین اغلب به دلیل تمرکز بر مدل‌های غیرتطبیقی، کاستی‌هایی در تحلیل داده‌های پیچیده و متغیر داشته‌اند. این محدودیت‌ها منجر به کاهش دقت در پیش‌بینی و افزایش خطاهای تصمیم‌گیری شده است. همچنین، در تحلیل موارد مرزی و داده‌های ناقص، این مدل‌ها نتوانسته‌اند کارایی کافی ارائه دهند. علاوه بر این، تغییرات ناگهانی در شرایط اقتصادی یا سیاسی می‌تواند بر خروجی این مدل‌ها تأثیر منفی بگذارد و باعث شود که بانک‌ها نتوانند به‌موقع واکنش نشان دهند.

1-2-2 ضرورت بهره‌گیری از ماتریس زیان و تکنیک بگینگ در تحلیل ریسک

دسته‌بندی دقیق موارد مرزی که در این رویکرد نقش مهمی دارند به دلیل نبود روش‌های مؤثر تحلیل، همچنان یک چالش باقی مانده است. تحقیقات پیشین کمتر بر استفاده از ماتریس زیان و طبقه‌بندی ترکیبی مبتنی بر روش بگینگ متمرکز شده‌اند. این تحقیق با هدف ارائه یک مدل ترکیبی مبتنی بر ماتریس زیان و طبقه‌بندی ترکیبی، به دنبال رفع این چالش‌ها است. در این مدل، آستانه‌های تصمیم‌گیری با استفاده از ماتریس زیان تعیین می‌شوند. این رویکرد امکان سازگاری بیشتر با داده‌های متغیر و شرایط پویا را فراهم کرده و به بهبود دقت و کارایی تصمیم‌گیری کمک می‌کند. علاوه بر این، موارد مرزی که در روش‌های سنتی نادیده گرفته می‌شدند، در این تحقیق با استفاده از تکنیک بگینگ تحلیل می‌شوند.

استفاده از ماتریس زیان به بانک‌ها کمک می‌کند تا به جای تمرکز صرف بر کاهش خطاهای نوع اول یا نوع دوم، توازنی میان این دو برقرار کنند و تصمیم‌گیری‌های بهینه‌تری داشته باشند. این رویکرد باعث کاهش خطاهای پرهزینه و افزایش بهره‌وری مالی خواهد شد. همچنین، استفاده از روش بگینگ برای تحلیل موارد مرزی به بانک‌ها کمک می‌کند تا اطلاعات بیشتری از مشتریان بالقوه جمع‌آوری کرده و احتمال بروز زیان‌های غیرمنتظره را کاهش دهند.

1-2-3 مزایای استفاده از بگینگ در کاهش خطاهای تصمیم‌گیری

بگینگ، با اجرای چندین مدل پایه به‌صورت موازی و ترکیب خروجی آن‌ها از طریق روش‌هایی مانند رأی‌گیری یا میانگین‌گیری، دقت تحلیل و پیش‌بینی را بهبود می‌بخشد و پایداری مدل را افزایش می‌دهد. این نوآوری نه‌تنها از نظر عملی به بانک‌ها و مؤسسات مالی در کاهش خطاهای تصمیم‌گیری و بهبود فرایندها کمک می‌کند، بلکه از منظر نظری نیز به توسعه دانش در حوزه مدیریت ریسک اعتباری و ترکیب روش‌های پیشرفته کمک می‌کند. افزایش پایداری مدل‌ها در مواجهه با تغییرات اقتصادی و داده‌های ناقص نیز از دیگر مزایای این تکنیک است.

1-2-4 ابزارهای کاربردی برای بانک‌ها در مدیریت ریسک

یکی از نقاط قوت اصلی استفاده از این مدل‌ها، امکان کاهش زمان مورد نیاز برای تحلیل داده‌ها و اتخاذ تصمیم‌های کارآمدتر است. در محیطی که بانک‌ها به سرعت در حال تغییر و تحول هستند، بهره‌گیری از چنین مدل‌هایی می‌تواند به افزایش چابکی و انطباق‌پذیری آن‌ها کمک کند. این تحقیق می‌تواند به مدیران بانک‌ها و کارشناسان ریسک ابزارهای لازم برای شناسایی سریع‌تر و دقیق‌تر مشتریان پرریسک را ارائه داده و از وقوع بحران‌های مالی جلوگیری کند.

همچنین، این مدل‌ها به بانک‌ها اجازه می‌دهند تا فرآیندهای مدیریت ریسک خود را به‌صورت خودکار و مداوم به‌روزرسانی کنند. این امر باعث کاهش وابستگی به تحلیل‌های دستی و افزایش دقت پیش‌بینی می‌شود. در نتیجه، بانک‌ها می‌توانند منابع خود را به شکل بهینه‌تری تخصیص دهند و از اتلاف منابع جلوگیری کنند.

تحلیل سناریو یکی از ابزارهای کلیدی برای شناسایی ریسک‌های احتمالی و تدوین راهبردهای پیشگیرانه در مواجهه با آن‌ها است. این ابزار به بانک‌ها کمک می‌کند تا به جای واکنش‌های پسینی، رویکردی پیش‌دستانه اتخاذ کنند و از بروز بحران‌های مالی جلوگیری نمایند.

در مجموع، استفاده از مدل‌های ترکیبی مبتنی بر ماتریس زیان و تکنیک بگینگ نه‌تنها می‌تواند دقت و اثربخشی تصمیم‌گیری‌های اعتباری را بهبود بخشد، بلکه به افزایش شفافیت، کاهش رفتارهای تبعیض‌آمیز و تقویت اعتماد عمومی به نظام مالی منجر می‌شود. این رویکرد نوین در مدیریت ریسک اعتباری، زمینه‌ساز ارتقاء سیستم‌های بانکی و افزایش پایداری اقتصادی در شرایط رقابتی و پویا خواهد بود. همچنین، این مدل‌ها با افزایش قابلیت انطباق بانک‌ها با شرایط متغیر و افزایش سطح آگاهی از داده‌های مشتریان، بهبود مستمر در تصمیم‌گیری‌ها را امکان‌پذیر می‌سازند.

1-3 جنبه جديد بودن و نوآوري در تحقيق

در این تحقیق، تعیین آستانه‌ها با بهره‌گیری از ماتریس زیان صورت می‌پذیرد. این رویکرد، روشی نظام‌مند برای تنظیم دقیق آستانه‌ها ارائه می‌دهد. این روش با بهره‌گیری از ماتریس زیان و افزودن دسته‌ای تحت عنوان موارد مرزی به رویکرد سنتی دو‌کلاسه، شرایطی را برای مدیریت بهینه داده‌های نامطمئن فراهم می‌آورد. این روش با ارائه سه دسته خوش‌حساب، بدحساب و مرزی، نه‌تنها خطاهای تصمیم‌گیری را کاهش می‌دهد، بلکه فرصت‌هایی برای جمع‌آوری اطلاعات بیشتر درباره موارد مرزی فراهم می‌آورد. بااین‌حال، موفقیت این رویکرد به تعیین دقیق آستانه‌های تصمیم‌گیری آلفا و بتا وابسته است که امکان تطبیق آستانه‌ها با شرایط متغیر بازار و ویژگی‌های داده‌های ورودی را فراهم می‌کند. برخلاف روش‌های دستی، که به‌طور ثابت و بدون انعطاف تعیین می‌شوند، این رویکرد به مدل اجازه می‌دهد که با داده‌های نامطمئن یا نوسانات بازار بهتر سازگار شود. تعیین این آستانه‌ها نقش اساسی در افزایش دقت طبقه‌بندی موارد مرزی و کاهش خطاهای پیش‌بینی ایفا می‌کند.

1-3-1 ترکیب ماتریس زیان و بگینگ

این رویکرد ترکیبی نه‌تنها نقاط قوت ماتریس زیان را در مدیریت خطاهای تصمیم‌گیری به‌کار می‌گیرد، بلکه با استفاده از ساختار بگینگ، یکپارچگی مدل‌ها را تضمین می‌کند. هم‌افزایی میان این دو روش باعث افزایش تطبیق‌پذیری با شرایط متغیر داده‌ها و کاهش خطاهای ناشی از سوگیری مدل‌ها می‌شود. به‌ویژه، رویکرد تلفیقی ارائه‌شده امکان استفاده از چندین مدل پایه را فراهم می‌آورد که در نهایت دقت پیش‌بینی را بهبود بخشیده و توانایی سیستم در مدیریت ریسک‌های پیچیده را تقویت می‌کند. از آنجا که مدل‌های پایه می‌توانند شامل الگوریتم‌های مختلفی باشند، از جمله شبکه‌های عصبی، ماشین‌های بردار پشتیبان و جنگل‌های تصادفی، انعطاف‌پذیری این رویکرد در مواجهه با انواع داده‌ها افزایش می‌یابد. ترکیب این مدل‌ها با رویکرد بگینگ، منجر به کاهش واریانس و افزایش اعتمادپذیری نتایج می‌شود.

1-3-2 تحلیل داده‌های نامطمئن

توانایی این مدل در تحلیل داده‌های نامطمئن یکی از مهم‌ترین نوآوری‌های آن به شمار می‌رود. در شرایطی که داده‌ها دارای عدم قطعیت بالا یا تنوع گسترده هستند، این رویکرد با ارائه راهکاری ساختاریافته، امکان تحلیل دقیق‌تری را فراهم می‌کند. توانایی تعیین آستانه‌هایی سازگار با تغییرات محیطی، این مدل را به ابزاری قابل اعتماد برای پیش‌بینی ریسک در محیط‌های پرچالش تبدیل کرده است. همچنین، مدل قادر است عدم قطعیت داده‌ها را به اطلاعات معنادار تبدیل کرده و دقت تصمیم‌گیری را افزایش دهد. این توانایی به‌ویژه در مواردی که داده‌های تاریخی کافی در دسترس نیست یا داده‌های ورودی از منابع مختلفی جمع‌آوری شده‌اند، بسیار حائز اهمیت است. در چنین شرایطی، این مدل می‌تواند به‌صورت تطبیقی عمل کرده و تغییرات ناگهانی در داده‌ها را شناسایی کند.

1-3-3 مدل‌های پیشرفته در ساختار بگینگ

یکی از ویژگی‌های برجسته این تحقیق، بهره‌گیری از مدل‌های پیشرفته در ساختار بگینگ است. به‌کارگیری مدل‌هایی نظیردرخت تصمیم و نزدیک‌ترین همسایه باعث کاهش واریانس و بهبود پایداری پیش‌بینی می‌شود. این مدل‌ها با رأی‌گیری وزنی، خروجی نهایی را با دقت بالاتری ارائه می‌دهند، که این امر منجر به بهبود عملکرد سیستم در شرایط پیچیده می‌گردد. علاوه بر این، مدل‌های پیشرفته با بهره‌گیری از الگوریتم‌های بهینه‌سازی، می‌توانند پارامترهای آستانه را به‌صورت پویا تنظیم کنند، که این امر به کاهش خطاهای انسانی و بهبود کارایی مدل منجر می‌شود.

1-3-4 کاربردهای عملیاتی و کاهش هزینه‌ها

علاوه بر دقت بالای پیش‌بینی، رویکرد پیشنهادی از نظر کاربرد عملیاتی نیز برای مؤسسات مالی حائز اهمیت است. این روش، منابع پردازشی کمتری را مصرف کرده و زمان تحلیل را کاهش می‌دهد. به‌این‌ترتیب، بانک‌ها و مؤسسات مالی می‌توانند با استفاده از این رویکرد، بهره‌وری خود را افزایش داده و خطاهای پرهزینه را به حداقل برسانند. این ویژگی به‌ویژه در محیط‌های رقابتی و تحت فشار زمانی، ارزش افزوده قابل توجهی ایجاد می‌کند. کاهش زمان پردازش و استفاده بهینه از منابع، به‌طور مستقیم بر سودآوری و کاهش هزینه‌های عملیاتی مؤسسات مالی تأثیرگذار است.

1-3-5 پایداری پیش‌بینی‌ها در شرایط ناپایدار

یکی از دستاوردهای کلیدی این تحقیق، پایداری پیش‌بینی‌ها حتی در مواجهه با تغییرات ناگهانی در داده‌های ورودی است. این ویژگی به مدل اجازه می‌دهد که در شرایط ناپایدار، مانند تغییرات اقتصادی یا نوسانات بازار، عملکرد دقیقی داشته باشد. همچنین، خطاهای پیش‌بینی به دلیل استفاده از آستانه‌های تطبیق‌پذیر و مدل‌های ترکیبی کاهش یافته و دقت تصمیم‌گیری بهبود می‌یابد. این ویژگی به‌ویژه در شرایط بحرانی، مانند رکود اقتصادی یا تغییرات ناگهانی در سیاست‌های اعتباری، می‌تواند از زیان‌های بزرگ جلوگیری کند.

1-3-6 مقایسه با سایر رویکردها

برای ارزیابی عملکرد این مدل، مقایسه‌ای میان روش پیشنهادی و سایر رویکردهای موجود انجام شد. نتایج نشان داد که این مدل با کاهش خطاهای پیش‌بینی و افزایش دقت در تحلیل موارد مرزی، عملکرد بهتری نسبت به روش‌های رایج مانند شبکه‌های عصبی مستقل یا درخت تصمیم سنتی دارد. این ارزیابی تطبیقی برتری روش پیشنهادی در مدیریت ریسک اعتباری را به‌وضوح نشان می‌دهد. علاوه بر این، نتایج نشان داد که استفاده از ترکیب مدل‌ها باعث افزایش مقاومت سیستم در برابر داده‌های نویزی و اطلاعات گمراه‌کننده می‌شود.

در مجموع ، این ساختار از نقاط قوت هر دو بخش بهره می‌برد؛ ماتریس زیان و تعیین آستانه‌ها برای دسته‌بندی اولیه داده‌ها، و روش طبقه‌بندی ترکیبی با رویکرد بگینگ برای تحلیل دقیق‌تر موارد مرزی. این هماهنگی بین دو مرحله به بهبود دقت پیش‌بینی، کاهش خطاهای تصمیم‌گیری و افزایش کارایی مدیریت ریسک اعتباری منجر می‌شود. این رویکرد نه‌تنها قابلیت تحلیل دقیق‌تر را فراهم می‌کند، بلکه با کاهش منابع و زمان پردازش، یک راه‌حل عملی و مؤثر برای بانک‌ها و مؤسسات مالی ارائه می‌دهد. علاوه بر این، این رویکرد می‌تواند به‌صورت خودکار به تغییرات داده‌های ورودی پاسخ دهد و نیاز به بازتنظیم مداوم مدل توسط کارشناسان را کاهش دهد.

همچنین، پیاده‌سازی این رویکرد در عمل می‌تواند به‌طور مستقیم بر سیاست‌های اعطای وام و اعتبار مؤسسات مالی تأثیرگذار باشد. با به‌کارگیری مدل‌های چندگانه و تعیین آستانه‌های تطبیقی، این سیستم قادر است در مواجهه با مشتریان جدید و داده‌های ناشناخته، تصمیمات بهینه‌تری اتخاذ کند. این امر موجب می‌شود که مؤسسات مالی ریسک‌های احتمالی را بهتر شناسایی کرده و منابع خود را به‌صورت مؤثرتری تخصیص دهند. در نهایت، این رویکرد یک چارچوب انعطاف‌پذیر و قابل انطباق با شرایط متغیر بازار ارائه می‌دهد که می‌تواند در کاهش زیان‌های ناشی از پیش‌بینی نادرست و بهبود سودآوری مؤسسات مالی نقش کلیدی ایفا کند. استفاده از این رویکرد در مقیاس‌های بزرگ‌تر می‌تواند به بهبود تحلیل داده‌های کلان و شناسایی الگوهای پنهان در داده‌ها کمک کند و در نتیجه توانایی پیش‌بینی دقیق‌تر در مواجهه با تغییرات پیچیده بازار را فراهم کند.

1-4 سؤالات تحقیق

آیا تلفیق ماتریس زیان و طبقه‌بندی ترکیبی مبتنی بر روش بگینگ می‌تواند منجر به افزایش دقت در پیش‌بینی ریسک اعتباری و بهبود تخصیص منابع مالی شود؟

آیا طبقه‌بندی ترکیبی مبتنی بر روش بگینگ در مدیریت و ارزیابی ریسک اعتباری عملکرد بهتری نسبت به روش‌های طبقه‌بندی پایه ارائه می‌دهد؟

1-5 فرضيه‏هاي تحقیق

تعیین آستانه‌های α و β با استفاده از ماتریس زیان می‌تواند دقت پیش‌بینی ریسک اعتباری را افزایش داده و به بهبود کارایی مدل در مدیریت مشتریان خوش‌حساب، بدحساب و موارد مرزی منجر می شود.

استفاده از طبقه‌بندی ترکیبی با روش بگینگ می‌تواند دقت پیش‌بینی و تحلیل موارد مرزی در مدیریت ریسک اعتباری را به‌طور قابل‌توجهی افزایش می دهد.

مدل تلفیقی مبتنی بر ماتریس زیان و طبقه‌بندی ترکیبی می‌تواند به بهبود عملکرد سیستم‌های مدیریت ریسک اعتباری و کاهش خطاهای تصمیم‌گیری در بانک‌ها و مؤسسات مالی منجرمی شود.

1-6 روش شناسی تحقیق

این تحقیق به‌منظور ارائه یک مدل پیشرفته برای مدیریت ریسک اعتباری، بر مبنای تلفیق ماتریس زیان و روش‌های طبقه‌بندی ترکیبی انجام شده است. هدف از این رویکرد، کاهش زیان‌های ناشی از خطاهای طبقه‌بندی و بهبود دقت پیش‌بینی نکول مشتریان اعتباری است. مدل پیشنهادی در چندین مرحله اجرایی و با بهره‌گیری از داده‌های واقعی جمع‌آوری‌شده از یک پایگاه اطلاعاتی بانکی معتبر پیاده‌سازی می‌شود. این داده‌ها شامل اطلاعات مالی مشتریان، تاریخچه اعتباری، نرخ نکول و سایر ویژگی‌های مرتبط است. داده‌ها به‌صورت عددی و ساختارمند ذخیره شده و برای تحلیل به فرمت‌های استاندارد تبدیل می‌شوند.

1-6-1 جمع‌آوری و پیش‌پردازش داده‌ها

در مرحله نخست، جمع‌آوری داده‌ها از طریق همکاری مستقیم با بانک انجام می‌شود. اطلاعات به‌دست‌آمده شامل متغیرهای کلیدی مانند میزان درآمد، سابقه بازپرداخت وام، تعداد وام‌های دریافت‌شده و وضعیت جاری حساب‌های اعتباری مشتریان است. پس از جمع‌آوری، داده‌ها با استفاده از تکنیک‌های پاک‌سازی داده پردازش می‌شوند تا مشکلاتی نظیر مقادیر گمشده یا داده‌های نادرست برطرف شود. پاک‌سازی داده‌ها نقش مهمی در تضمین صحت تحلیل‌های بعدی ایفا می‌کند. علاوه بر این، داده‌ها به‌طور مداوم بازبینی می‌شوند تا از وجود هرگونه سوگیری یا خطای سیستمی در ورودی‌ها جلوگیری شود.

1-6-2 طراحی و پیاده‌سازی ماتریس زیان

در مرحله بعد، ماتریس زیان طراحی و به مدل اضافه می‌شود. این ماتریس به‌عنوان ابزاری برای کاهش هزینه‌های مرتبط با خطاهای تصمیم‌گیری مورد استفاده قرار می‌گیرد. ماتریس زیان با تخصیص وزن‌های متفاوت به خطاهای نوع اول و دوم، به مدل کمک می‌کند تا اولویت‌های خود را در تشخیص نکول‌های پرریسک تنظیم کند. برای مثال، هزینه عدم شناسایی نکول یک مشتری پرریسک می‌تواند بسیار بالاتر از هزینه اشتباه در طبقه‌بندی یک مشتری خوش‌حساب باشد. استفاده از ماتریس زیان، استراتژی تصمیم‌گیری مدل را بهینه کرده و موجب کاهش زیان‌های کلی می‌شود.

1-6-3 استفاده از الگوریتم‌های یادگیری ماشین

به‌منظور افزایش دقت مدل، الگوریتم‌های یادگیری ماشین پیشرفته‌ای نظیر درخت‌های تصمیم، جنگل تصادفی و روش‌های ترکیبی مورد استفاده قرار گرفته‌اند. این الگوریتم‌ها با استفاده از زبان برنامه‌نویسی پایتون و کتابخانه‌های معروفی نظیر Scikit-learn و TensorFlow توسعه یافته‌اند. هدف اصلی این الگوریتم‌ها، کاهش خطاهای طبقه‌بندی و بهبود توانایی مدل در شناسایی موارد مرزی است. به‌علاوه، تکنیک‌های بهینه‌سازی نظیر الگوریتم‌های ژنتیک و شبکه‌های عصبی نیز برای ارتقای کارایی مدل به کار گرفته شده‌اند.

1-6-4 رویکرد بگینگ در طبقه‌بندی ترکیبی

روش بگینگ به‌عنوان یکی از رویکردهای اصلی طبقه‌بندی ترکیبی در این تحقیق استفاده شده است. بگینگ با کاهش واریانس مدل و بهبود پایداری آن، موجب افزایش دقت پیش‌بینی می‌شود. این روش با ایجاد چندین مدل پایه و ترکیب نتایج آن‌ها، خطر بیش‌برازش را کاهش می‌دهد. در مقایسه با روش‌های دیگر مانند بوستینگ ، بگینگ پیچیدگی کمتری دارد و به دلیل عدم تمرکز بر بهبود تدریجی خطاها، احتمال بیش‌برازش کمتری را به همراه دارد. این ویژگی‌ها بگینگ را به گزینه‌ای مناسب برای مسائل مرتبط با مدیریت ریسک اعتباری تبدیل کرده است.

1-6-5 ارزیابی و اعتبارسنجی مدل

در این رویکرد، مدل‌های پایه‌ای نظیر درخت‌های تصمیم بر روی نمونه‌های مختلف داده‌ها آموزش داده می‌شوند. این مدل‌ها سپس با استفاده از تکنیک بگینگ ترکیب شده و پیش‌بینی‌های نهایی بر اساس نتایج ترکیبی آن‌ها انجام می‌شود. یکی از مزایای بگینگ، توانایی آن در کاهش تأثیر داده‌های نویزی و غیرساختاریافته است. این امر به‌ویژه در پایگاه‌های داده مالی که معمولاً با مقادیر نوسانی و تغییرپذیر مواجه هستند، از اهمیت بالایی برخوردار است.

در فرآیند ارزیابی مدل، معیارهایی نظیر دقت [[1]](#footnote-1)، حساسیت [[2]](#footnote-2)، ویژگی [[3]](#footnote-3) و نرخ خطای طبقه‌بندی محاسبه می‌شوند. همچنین از روش اعتبارسنجی متقابل[[4]](#footnote-4) برای اطمینان از تعمیم‌پذیری مدل بر روی داده‌های جدید استفاده شده است. این تکنیک کمک می‌کند تا مدل پیشنهادی نه‌تنها بر روی داده‌های آموزشی، بلکه در مواجهه با داده‌های واقعی نیز عملکرد مطلوبی داشته باشد. این امر برای کاربرد مدل در محیط‌های عملیاتی بانک‌ها و مؤسسات مالی بسیار حیاتی است.

یکی از چالش‌های اصلی تحقیق حاضر، مدیریت ریسک‌های ناشی از خطاهای الگوریتمی و محدودیت‌های داده‌ای بوده است. برای کاهش این ریسک‌ها، مدل به‌طور مکرر اعتبارسنجی شده و پارامترهای حساس آن با استفاده از تحلیل حساسیت تنظیم شده‌اند. همچنین، بررسی پیوسته تغییرات در داده‌ها و رفتار مشتریان به بهبود مداوم عملکرد مدل کمک کرده است. این فرآیند به‌ویژه در محیط‌های مالی که با تغییرات سریع و پیوسته مواجه هستند، اهمیت زیادی دارد.

در این تحقیق، استفاده از ماتریس زیان در کنار روش بگینگ به‌عنوان یک روش ترکیبی پیشنهاد شده است که انتظار می‌رود منجر به بهبود شناسایی نکول‌ها و کاهش هزینه‌های ناشی از خطاهای طبقه‌بندی شود. در مقایسه با روش‌های سنتی نظیر ماشین بردار پشتیبان و شبکه‌های عصبی مصنوعی [[5]](#footnote-5)، مدل پیشنهادی به دلیل ویژگی‌های ساختاری خود، پتانسیل بهتری در دقت و پایداری دارد. همچنین، پیش‌بینی می‌شود که تحلیل‌های اقتصادی این مدل نشان‌دهنده کاهش زیان‌های ناشی از نکول وام‌ها باشد و بهبود فرایندهای تصمیم‌گیری اعتباری در بانک‌ها را ممکن سازد.

هدف این تحقیق، ارزیابی اثربخشی این مدل ترکیبی در مدیریت ریسک اعتباری و ارائه راهکاری کاربردی برای پیاده‌سازی در سیستم‌های واقعی است. استفاده از روش‌های پیشرفته یادگیری ماشین و طبقه‌بندی ترکیبی در مدل پیشنهادی، امکان سازگاری با تغییرات محیطی و داده‌ای را فراهم کرده و آن را به یک ابزار بالقوه مؤثر برای مدیریت ریسک در مؤسسات مالی تبدیل می‌کند.

1-7 جامعه آماري

در این تحقیق، جامعه آماری شامل داده‌های اعتباری مربوط به مشتریان یکی از بانک‌ کشور است که اطلاعات جامعی در زمینه سابقه اعتباری مشتریان فراهم می‌کند. این جامعه شامل تمامی مشتریانی است که در بازه زمانی مشخصی درخواست وام یا خدمات مالی داشته‌اند و اطلاعات مربوط به وضعیت بازپرداخت وام‌های آن‌ها در دسترس است. داده‌ها بر اساس ویژگی‌های کلیدی مانند میزان وام، مدت بازپرداخت، درآمد ماهانه و سابقه نکول انتخاب می‌شوند تا تحلیل‌های موردنظر تحقیق را پشتیبانی کنند.

روش نمونه‌گیری در این پژوهش به‌صورت تصادفی طبقه‌بندی‌شده انجام می‌شود. این روش امکان انتخاب نمونه‌ای متنوع و نماینده از تمامی زیرگروه‌های جامعه آماری را فراهم می‌کند. به‌این‌ترتیب، داده‌ها به‌گونه‌ای انتخاب می‌شوند که از نظر توزیع ویژگی‌های کلیدی مانند وضعیت اعتباری (ریسک پایین، ریسک بالا و موارد مرزی) تعادل وجود داشته باشد. این تعادل به بهبود دقت تحلیل‌ها و نتایج مدل پیشنهادی کمک می‌کند.

داده‌ها به دو بخش آموزشی و آزمایشی تقسیم می‌شوند تا امکان ارزیابی عملکرد مدل در شرایط مختلف فراهم شود. این تقسیم‌بندی به بهبود دقت پیش‌بینی و کاهش خطاهای تصمیم‌گیری کمک می‌کند.

در نهایت، ترکیب جامعه آماری گسترده، روش نمونه‌گیری تصادفی طبقه‌بندی‌شده و حجم نمونه مناسب، شرایط لازم برای اجرای موفقیت‌آمیز تحقیق و دستیابی به اهداف پژوهش را فراهم می‌آورد.

1-8 مدل مورد استفاده در تحقیق و تعریف متغیرها

در این تحقیق، هدف اصلی طراحی و استفاده از یک مدل مفهومی است که بتواند با دقت بالا روابط میان متغیرها را تحلیل کرده و ریسک اعتباری را پیش‌بینی کند. این مدل بر اساس روش طبقه‌بندی ترکیبی توسعه یافته است تا تحلیل جامع‌تری از عوامل مؤثر بر ریسک اعتباری ارائه کند. متغیر اصلی تحقیق، ریسک اعتباری است که نشان‌دهنده احتمال نکول یا عدم بازپرداخت تعهدات مالی مشتریان است. برای اندازه‌گیری این متغیر از شاخص‌هایی مانند سابقه نکول، میزان تأخیر در بازپرداخت و نرخ بازپرداخت استفاده شده است. داده‌های مربوط به این شاخص‌ها از پایگاه اطلاعاتی بانک‌ها گردآوری شده و پس از آماده‌سازی به صورت کمی تحلیل می‌شوند. این داده‌ها سپس به شکل نمودارها و جداول مقایسه‌ای نمایش داده شده تا تغییرات در طول زمان به وضوح مشخص شود و روندهای مختلف در رفتار مشتریان شناسایی گردد. تحلیل این داده‌ها می‌تواند به بانک‌ها در تنظیم استراتژی‌های دقیق‌تر کمک کرده و اقدامات اصلاحی را در مراحل مختلف ارائه وام تسریع کند.

1-8-1 متغیرهای مستقل و تأثیر آن‌ها بر ریسک اعتباری

متغیرهای مستقل شامل مجموعه‌ای از عوامل مالی، اعتباری و جمعیتی هستند که به طور مستقیم یا غیرمستقیم بر ریسک اعتباری تأثیر می‌گذارند. این متغیرها شامل سابقه اعتباری مشتریان، درآمد ماهانه، نوع وام درخواستی، میزان وام، مدت زمان بازپرداخت و نسبت استفاده از اعتبار هستند. برای تحلیل این متغیرها از روش‌های آماری پیشرفته و الگوریتم‌های یادگیری ماشین استفاده شده است تا روابط پیچیده میان متغیرها به دقت بررسی شود. این فرایند منجر به شناسایی الگوهای خاص و رفتارهای احتمالی مشتریان می‌شود که در پیش‌بینی ریسک بسیار مؤثر است. اطلاعات درآمد ماهانه نیز از اسناد مالی ثبت‌شده مشتریان استخراج شده و تأثیر آن بر ریسک اعتباری بررسی می‌شود. در مرحله پیش‌پردازش، داده‌های گمشده با روش‌های تکمیلی جایگزین و مقادیر نادرست تصحیح می‌شوند.

1-8-2 نقش ماتریس زیان در بهبود عملکرد مدل

در این مدل، آستانه‌های تصمیم‌گیری مانند آلفا و بتا نقش کلیدی دارند و مرز میان طبقه‌های مختلف ریسک را مشخص می‌کنند. این مقادیر بر اساس تحلیل ماتریس زیان تنظیم شده و هدف آن‌ها کاهش خطاهای طبقه‌بندی و بهبود عملکرد مدل است. ماتریس زیان هزینه‌های ناشی از طبقه‌بندی اشتباه مشتریان را در بر می‌گیرد و امکان تصمیم‌گیری دقیق‌تر را برای بانک‌ها فراهم می‌کند. داده‌ها در فرآیند پیش‌پردازش به‌گونه‌ای تنظیم می‌شوند که متغیرهای کیفی به صورت عددی کدگذاری شده و داده‌های پرت حذف شوند. این اقدام باعث می‌شود تا مدل از دقت و اعتبار بالاتری برخوردار شود.

1-8-3 تحلیل داده‌ها با روش‌های طبقه‌بندی ترکیبی

برای تحلیل و پیش‌بینی، روش‌های طبقه‌بندی ترکیبی به‌کار گرفته شده‌اند که از ترکیب چندین مدل پایه برای بهبود دقت و کاهش خطاهای احتمالی استفاده می‌کنند. این روش به کاهش خطاهای طبقه‌بندی و افزایش ثبات نتایج کمک کرده و باعث می‌شود تأثیر داده‌های پرت کاهش یابد. عملکرد مدل با معیارهایی نظیر دقت، حساسیت، ویژگی و میانگین مربع خطا ارزیابی شده تا نتایج به دست آمده اعتبار بالایی داشته باشند. مدل‌های پایه به صورت مجزا آموزش داده شده و سپس نتایج آن‌ها در قالب یک سیستم ترکیبی تجمیع می‌شوند. این تجمیع به مدل کمک می‌کند تا از نقاط قوت هر یک از مدل‌های پایه بهره‌برداری کرده و نقاط ضعف آن‌ها را کاهش دهد.

1-8-4 قابلیت‌های تطبیق‌پذیری و به‌روزرسانی مدل

یکی از ویژگی‌های کلیدی این مدل، توانایی آن در انطباق با داده‌های جدید و پویاست. با استفاده از روش‌های یادگیری ماشین و به‌روزرسانی مداوم پارامترهای مدل، می‌توان عملکرد آن را در طول زمان بهبود داد. این قابلیت باعث می‌شود که مدل بتواند به تغییرات محیطی و تغییر رفتار مشتریان واکنش نشان دهد و تصمیم‌گیری‌های دقیق‌تری ارائه کند. علاوه بر این، مدل با استفاده از تحلیل حساسیت، قادر است تأثیر تغییرات هر یک از متغیرها را بر خروجی نهایی بررسی کند. این تحلیل به بانک‌ها امکان می‌دهد تا عوامل بحرانی را شناسایی کرده و سیاست‌های اعتباری خود را به‌طور دقیق تنظیم کنند.

1-8-5 تأثیر مدل پیشنهادی بر مدیریت ریسک بانکی

در نهایت، با استفاده از رویکرد پیشنهادی، بانک‌ها می‌توانند سطح ریسک سبد اعتباری خود را کاهش داده و سودآوری بلندمدت خود را افزایش دهند. این مدل نه‌تنها در ارزیابی ریسک اعتباری مؤثر است، بلکه به بهبود فرایندهای مدیریتی در حوزه‌های مختلف بانکی نیز کمک می‌کند. بهره‌گیری از این رویکرد می‌تواند به ایجاد یک سیستم هوشمند مدیریت ریسک منجر شود که ضمن کاهش ضررهای مالی، به افزایش اعتماد مشتریان نیز کمک کند. به این ترتیب، مدل پیشنهادی می‌تواند به‌عنوان یک ابزار راهبردی در تصمیم‌گیری‌های کلان بانکی و سیاست‌گذاری‌های اعتباری مورد استفاده قرار گیرد.

1-9 ساختار پایان‌نامه

تحقیق حاضر شامل 5 فصل می‌باشد ، در فصل اول به کلیات پژوهش شامل بیان مسأله، اهمیت و ضرورت تحقیق، اهداف و سؤالات پژوهش، و همچنین ساختار کلی پایان‌نامه پرداخته شده است.  
در فصل دوم ادبیات نظری و پیشینه تحقیق مورد بررسی قرار گرفته است. در این فصل، ابتدا مفاهیم مربوط به ارزیابی ریسک اعتباری، ماتریس زیان و روش‌های ترکیبی مبتنی بر بگینگ تشریح شده و سپس مطالعات انجام‌شده در این زمینه در دو بخش داخلی و خارجی ارائه شده است.فصل سوم به روش‌شناسی تحقیق اختصاص دارد. در این فصل، روش‌های جمع‌آوری داده‌ها، جامعه آماری، ابزارهای تحقیق، و روش‌های تجزیه و تحلیل داده‌ها توضیح داده شده است.در فصل چهارم، داده‌های گردآوری‌شده تحلیل شده و الگوی پیشنهادی تحقیق که شامل تلفیق ماتریس زیان و روش طبقه‌بندی ترکیبی مبتنی بر بگینگ است، مورد آزمون قرار گرفته است. همچنین، نتایج آزمون فرضیات و تحلیل‌های مربوط به سؤالات پژوهش ارائه شده است.در نهایت، فصل پنجم شامل نتایج کلی تحقیق، بحث و بررسی یافته‌ها، محدودیت‌های پژوهش، و ارائه پیشنهادهای کاربردی و پژوهشی برای تحقیقات آتی می‌باشد.

فصل دوم:

مروری بر مطالعات پیشین

(مبانی نظری و پیشینه‌ی تحقیق)

2-1 مقدمه

بانک‌ها، به‌عنوان نهادهای اصلی واسطه‌گری مالی، منابع مالی را از سپرده‌گذاران جمع‌آوری کرده و به بخش‌های مختلف اقتصاد تخصیص می‌دهند. این فرآیند باعث رشد تولید، افزایش اشتغال و تقویت زیرساخت‌های اقتصادی می‌شود. در ایران، به دلیل ساختار بانک‌محور نظام مالی، بانک‌ها نقشی حیاتی در تأمین مالی پروژه‌ها و بخش‌های مختلف اقتصادی ایفا می‌کنند. نظام بانکی، در عین حال که محرک اصلی رشد اقتصادی است، می‌تواند در معرض ریسک‌های مختلف، به‌ویژه ریسک اعتباری قرار بگیرد. ضرورت مدیریت ریسک، به‌ویژه در کشورهایی با اقتصاد بانک‌محور ، از اهمیت ویژه‌ای برخوردار است. اگر بانک‌ها در ارزیابی و مدیریت ریسک‌های اعتباری و دیگر ریسک‌های بانکی کوتاهی کنند، پیامدهای منفی آن می‌تواند کل اقتصاد را تحت تأثیر قرار دهد. افزایش نکول و کاهش نقدینگی، از جمله پیامدهای عدم مدیریت صحیح ریسک است که می‌تواند باعث بی‌ثباتی مالی و کاهش اعتماد عمومی به نظام بانکی شود.([Aboutorabi et al., 2021](#_ENREF_2" \o "Aboutorabi, 2021 #12))

مدل‌های مختلفی برای طبقه‌بندی ریسک در نظام بانکی ارائه شده است. همچنین، در بررسی عقود اسلامی، این ریسک‌ها به‌صورت خاص‌تری در دو دسته «ریسک‌های مرتبط با عقود مبادله‌ای» و «ریسک‌های مرتبط با عقود مشارکتی» تحلیل می‌شوند.در این میان برخی ریسک‌ها مانند ریسک اعتباری و ریسک عملیاتی، مشترک بین بانکداری متعارف و اسلامی هستند، اما به‌دلیل ماهیت عقود اسلامی، مدیریت این ریسک‌ها در بانکداری بدون ربا پیچیدگی بیشتری دارد .([ابوالحسنی and حسنی مقدم, 2008](#_ENREF_21" \o "ابوالحسنی, 2008 #13)) .

با توجه به این نکته که ماهیت فعالیت‌های بانکی همراه با پذیرش ریسک است و کاهش آن می‌تواند کارایی و ثبات مالی بانک‌ها را بهبود بخشد . عدم مدیریت صحیح ریسک ممکن است منجر به کاهش درآمد، از دست دادن سرمایه و ایجاد محدودیت در دستیابی به اهداف اقتصادی بانک‌ها شود .([پندار and ویسی, 2020](#_ENREF_22" \o "پندار, 2020 #14))

2-2 مبانی نظری

2-2-1 مفهوم ریسک و مدیریت آن

ریسک را میتوان به نحو های گوناگون تعریف کرد در این میان میتوان ریسک را به‌عنوان احتمال وقوع پیامدهای ناخواسته و نامطلوب که ممکن است بر افراد، اموال یا محیط زیست تأثیر منفی بگذارد، تعریف کرد. ریسک به‌طور معمول به‌عنوان حاصل‌ضرب احتمال وقوع یک رخداد منفی و شدت پیامدهای آن تعریف می‌شود. این تعریف، ریسک را مفهومی چندبعدی می‌داند که نه‌تنها بر احتمال وقوع رویدادهای نامطلوب تأکید دارد، بلکه میزان زیان حاصل از آن‌ها را نیز در نظر می‌گیرد. ریسک معمولاً به‌صورت کمی از طریق حاصل‌ضرب احتمال وقوع یک رویداد و شدت پیامدهای آن ارزیابی می‌شود. به این ترتیب، مفهوم ریسک نه‌تنها شامل شناسایی پیامدهای بالقوه است، بلکه میزان و شدت این پیامدها را نیز در بر می‌گیرد.([Gratt, 1989](#_ENREF_7" \o "Gratt, 1989 #15))

با توجه به اهمیت ریسک مفهوم مدیریت مطرح میشود . مدیریت ریسک به‌عنوان یکی از مؤلفه‌های اساسی تصمیم‌گیری‌های مالی شرکتی مطرح شده که هدف آن تنظیم و کنترل نوسانات ذخایر نقدی و ایجاد تعادل میان هزینه‌های سرمایه‌گذاری و پوشش‌های مالی است. مدیریت ریسک نه‌تنها شامل کاهش نوسانات ذخایر نقدی بلکه شامل تدوین سیاست‌های مالی پویا برای شرکت‌ها هستند. مدیریت ریسک نقش مهمی در تشویق شرکت‌ها به اتخاذ فناوری‌های جدید و سرمایه‌گذاری در پروژه‌های نوآورانه ایفا می‌کند. بااین‌حال، این فرآیند با چالش‌هایی همراه است، ازجمله افزایش هزینه‌های ناشی از نوسانات بالاتر ذخایر نقدی و ریسک‌های مرتبط با تأمین مالی خارجی پرهزینه. ([Luo and Liu, 2024](#_ENREF_13))

2-2-2 ریسک اعتباری

در حوزه بانکداری و امور مالی، ریسک اعتباری یکی از مهم‌ترین انواع ریسک‌هایی است که مؤسسات مالی با آن مواجه می‌شوند. ریسک اعتباری به احتمال عدم بازپرداخت اصل و سود وام از سوی وام‌گیرنده اشاره دارد که می‌تواند منجر به زیان‌های مالی و اختلال در جریان‌های نقدی بانک‌ها شود .([Kamara, 2024](#_ENREF_10" \o "Kamara, 2024 #17))

به‌طور کلی، این ریسک در انواع مختلفی از وام‌ها و اعتبارها وجود دارد، از جمله وام‌های مصرفی، وام‌های تجاری، و حتی اوراق قرضه منتشرشده توسط دولت‌ها و شرکت‌ها. یکی از عوامل کلیدی در مدیریت ریسک اعتباری، ارزیابی اعتبار وام‌گیرنده بر اساس شاخص‌هایی مانند سابقه اعتباری، ظرفیت بازپرداخت، سرمایه، شرایط وام و وثیقه مرتبط است.([Graham and Coyle, 2000](#_ENREF_6" \o "Graham, 2000 #18))

مقاله مذکور تأکید دارد که در دهه گذشته، بانک‌ها به سمت استفاده از مدل‌های پیچیده ارزیابی ریسک اعتباری رفته‌اند که اغلب شامل روش‌های مبتنی بر الگوریتم و داده‌کاوی برای تحلیل ریسک‌های مرتبط هستند. این مدل‌ها به بانک‌ها کمک می‌کنند تا ریسک اعتباری را به‌طور دقیق‌تر کمی‌سازی، تجمیع و مدیریت کنند و به‌ویژه در زمینه شناسایی و کاهش وام‌های غیرجاری مؤثر باشند . در این راستا، بانک‌ها باید به‌صورت منظم داده‌های مرتبط با ریسک اعتباری را تحلیل کرده و از ابزارهای پیشرفته‌ای برای پیش‌بینی اثرات احتمالی ریسک استفاده کنند. ([Kamara, 2024](#_ENREF_10))

2-2-3 ماتریس زیان در ارزیابی ریسک

در روش های گذشته ، تمام سوابق اعتباری تحت شرایط اطلاعاتی یکسان و با استفاده از یک مدل واحد پیش‌بینی می‌شوند و به هر نمونه یکی از دو برچسب بله (قبول) یا خیر (رد) اختصاص داده می‌شود. این فرآیند به عنوان تصمیم‌گیری دودویی شناخته می‌شود. با این حال، همیشه برخی از نمونه‌ها وجود دارند که با اطلاعات موجود نمی‌توان به‌خوبی پیش‌بینی کرد. روش تصمیم‌گیری سه‌گانه یک مکانیزم تأخیر در تصمیم‌گیری به تصمیم‌گیری دودویی سنتی اضافه می‌کند. به این معنا که برای برخی از نمونه‌ها نیازی نیست بلافاصله تصمیم نهایی (قبول یا رد) گرفته شود؛ بلکه پس از دریافت اطلاعات بیشتر، تصمیم‌گیری نهایی انجام می‌شود. این رویکرد ریسک تصمیم‌گیری اشتباه در مورد نمونه‌های دشوار را کاهش می‌دهد. تصمیم‌های مختلف روی یک نمونه می‌تواند منجر به زیان‌های متفاوتی شود. این مقادیر زیان، مبنای تعیین آستانه‌ها هستند.

با این حال، مقادیر زیان [[6]](#footnote-6) به‌سادگی قابل اندازه‌گیری نیستند. به‌عنوان مثال، تعیین میزان زیان در صورتی که یک نمونه به منطقه مرزی اختصاص یابد، کار دشواری است. یکی از روش های موجود آن است که مقادیر زیان را به‌صورت تصادفی تعیین گردند تا آستانه‌ها را محاسبه کنند.([Xu et al., 2023](#_ENREF_18" \o "Xu, 2023 #21))

از جمله روش های دیگر میتوان به روش هایی اشاره کرد که در آن‌ها به‌جای مقادیر زیان، پارامترهای جایگزین میشوند می‌شوند.یک از روش های موجود استفاده از نظریه پشیمانی[[7]](#footnote-7) میباشد و مقادیر پشیمانی، مقادیر خشنودی و میزان ادراک روانی را برای تعیین آستانه‌های تصمیم‌گیری سه‌گانه به کار بردند.([Zhu et al., 2022](#_ENREF_19" \o "Zhu, 2022 #22))

4-2-2 روش بگینگ در یادگیری ماشین

ایده اصلی یادگیری ترکیبی [[8]](#footnote-8) این است که چند الگوریتم ضعیف [[9]](#footnote-9) را با هم ترکیب کرده و یک مدل قوی [[10]](#footnote-10) با عملکرد بهتر ایجاد کنیم.

یکی از رویکردهای مؤثر در افزایش قابلیت تعمیم‌پذیری مدل‌های یادگیری ماشین و کاهش ریسک‌های ناشی از بیش‌برازش [[11]](#footnote-11)، استفاده از الگوریتم بگینگ [[12]](#footnote-12) است. این روش که توسط 1996 معرفی شده است، با ایجاد زیرمجموعه‌های تصادفی از داده‌های آموزشی و آموزش مدل‌های پایه [[13]](#footnote-13) روی این زیرمجموعه‌ها، تلاش می‌کند واریانس مدل را کاهش داده و پایداری پیش‌بینی‌ها را بهبود بخشد. روش بگینگ می‌تواند در مسائل مرتبط با داده‌های نامتوازن، به‌ویژه در محیط‌های پرریسک مانند مالی و بانکی، به بهبود عملکرد طبقه‌بندی کمک کند. این الگوریتم با ترکیب پیش‌بینی‌های مدل‌های پایه از طریق رأی‌گیری اکثریت [[14]](#footnote-14)، باعث افزایش دقت مدل‌ها و کاهش خطاهای ناشی از تصمیم‌گیری‌های اشتباه می‌شود .([Breiman, 1996](#_ENREF_4" \o "Breiman, 1996 #20))

در مسائل مرتبط با ارزیابی ریسک اعتباری، یکی از چالش‌های اساسی، وجود عدم توازن طبقات [[15]](#footnote-15) است. بیشتر داده‌های موجود متعلق به مشتریان خوش‌حساب (طبقه اکثریت) هستند و تعداد کمی از داده‌ها مربوط به مشتریان پرریسک (طبقه اقلیت) است. در چنین شرایطی، روش‌های یادگیری سنتی معمولاً تمایل به طبقه‌بندی داده‌ها در طبقه اکثریت دارند و از شناسایی دقیق طبقه اقلیت بازمی‌مانند. روش بگینگ با ایجاد زیرمجموعه‌های متوازن از داده‌ها می‌تواند به افزایش دقت در شناسایی طبقه اقلیت کمک کند.

استفاده از نسخه‌های تغییریافته از روش بگینگ، به‌ویژه در محیط‌های داده‌ای ناپایدار و نامتوازن، می‌تواند ریسک‌های مالی را کاهش دهد و به بهبود مدیریت ریسک اعتباری منجر شود. این نسخه تغییریافته از بگینگ به‌جای نمونه‌برداری تصادفی کامل، از نمونه‌برداری زیرمجموعه‌های متوازن [[16]](#footnote-16) برای ایجاد داده‌های آموزشی متوازن استفاده می‌کند. در این روش، داده‌های مربوط به مشتریان پرریسک به‌گونه‌ای انتخاب می‌شوند که مدل بتواند الگوهای رفتاری آن‌ها را بهتر شناسایی کند. نتایج این پژوهش نشان می‌دهد که کاهش واریانس مدل‌ها و ترکیب پیش‌بینی‌های مدل‌های پایه به افزایش دقت پیش‌بینی‌ها کمک کرده و می‌تواند در تصمیم‌گیری‌های مالی مؤسسات تأثیرگذار باشد .([Abdoli et al., 2023](#_ENREF_1" \o "Abdoli, 2023 #19))

5-2-2 ترکیب روش‌های طبقه‌بندی

روش‌های ترکیبی طبقه‌بندی با هدف افزایش دقت و کاهش خطای مدل‌های پیش‌بینی از طریق ترکیب چندین طبقه‌بند پایه به‌کار می‌روند. در روش بگینگ، داده‌های آموزشی با جایگزینی به چندین مجموعه تقسیم شده و هر طبقه‌بند پایه به‌طور مستقل بر روی یکی از این مجموعه‌ها آموزش داده می‌شود. این روش به‌ویژه برای کاهش واریانس خطا در مدل‌های پرنویز مؤثر است و برای طبقه‌بندهای ناپایدار نظیر درخت تصمیم به‌کار می‌رود. یکی از ویژگی‌های اصلی بگینگ، کاهش نوسانات پیش‌بینی و بهبود پایداری نتایج است، اما چالش اصلی آن، تنوع پایین طبقه‌بندهای پایه است که می‌تواند دقت نهایی را کاهش دهد. مقاله به راهکارهایی اشاره می‌کند که با انتخاب هوشمند طبقه‌بندهای پایه، تنوع و دقت مدل بهبود می‌یابد. این الگوریتم با حذف طبقه‌بندهایی که کارایی پایینی دارند، عملکرد بگینگ را به‌طور محسوسی ارتقا می‌دهد. همچنین استفاده از تکنیک‌های خوشه‌بندی داده‌ها در فرایند بگینگ به‌منظور افزایش تنوع طبقه‌بندها پیشنهاد شده است که می‌تواند دقت کلی پیش‌بینی‌ها را افزایش دهد.

در مقابل، روش بوستینگ به‌صورت ترتیبی عمل می‌کند و در هر مرحله داده‌های آموزشی بر اساس خطاهای مدل قبلی وزن‌دهی می‌شوند. این روش به کاهش بایاس و واریانس مدل کمک می‌کند، اما در داده‌های پرنویز می‌تواند منجر به بیش‌برازش شود. مقاله بیان می‌کند که بوستینگ در ترکیب با طبقه‌بندهای ضعیف نظیر درخت‌های تصمیم کوتاه بسیار مؤثر است. هر دو روش بگینگ و بوستینگ در مسائل مالی از جمله ارزیابی ریسک اعتباری قابل‌استفاده هستند؛ زیرا می‌توانند با ترکیب خروجی‌های چندین طبقه‌بند، دقت پیش‌بینی را افزایش داده و عدم‌قطعیت در تصمیم‌گیری را کاهش دهند. همچنین مقاله به اهمیت انتخاب ایستا و پویا در بهبود عملکرد مدل‌های ترکیبی اشاره دارد و نشان می‌دهد که استفاده از رویکردهای ترکیبی می‌تواند خطای پیش‌بینی را در داده‌های نامتوازن و پرنویز کاهش دهد .([Jurek et al., 2014](#_ENREF_9" \o "Jurek, 2014 #25))

2-3 پیشینه تحقیق

1-3-2 تحقیقات داخلی

در سال‌های اخیر، ارزیابی و مدیریت ریسک اعتباری به‌عنوان یکی از چالش‌های اصلی نظام‌های مالی مدرن مطرح شده است. به‌طور سنتی، این ریسک‌ها با استفاده از روش‌های آماری مانند تحلیل تشخیصی خطی و رگرسیون لجستیک اندازه‌گیری می‌شدند، اما این روش‌ها به دلیل محدودیت در پردازش داده‌های حجیم و پیچیده، کمتر مورد استفاده قرار می‌گیرند . با پیشرفت‌های اخیر در حوزه هوش مصنوعی و یادگیری ماشین، الگوریتم‌های جدیدی برای برآورد ریسک اعتباری توسعه یافته‌اند که دقت و کارایی بیشتری نسبت به روش‌های سنتی دارند.

روش‌های مبتنی بر یادگیری ماشین به‌دلیل انعطاف‌پذیری و قابلیت پردازش داده‌های غیرساختاریافته، عملکرد بهتری در مقایسه با روش‌های آماری سنتی دارند. به‌طور خاص، الگوریتم‌هایی مانند ماشین بردار پشتیبان و جنگل تصادفی توانسته‌اند دقت بیشتری در پیش‌بینی ریسک اعتباری ارائه دهند. علاوه بر این، استفاده از روش‌های ترکیبی نظیر بگینگ و بوستینگ نیز به‌عنوان یکی از راهکارهای مؤثر در بهبود دقت پیش‌بینی شناخته شده‌اند. این روش‌ها با ترکیب چندین مدل، توانسته‌اند خطای پیش‌بینی را کاهش دهند و عملکرد کلی سیستم‌های ارزیابی ریسک را بهبود بخشند .یکی از چالش‌های اساسی که در این مقاله به آن اشاره شده است، عدم تعادل داده‌های اعتباری است. داده‌های اعتباری معمولاً شامل تعداد کمی از موارد نکول در مقایسه با موارد غیرنکول هستند که این مسئله می‌تواند دقت مدل‌های یادگیری ماشین را کاهش دهد. برای مقابله با این چالش، استفاده از روش‌های نمونه‌گیری بیش‌ازحد و تولید نمونه‌های مصنوعی از داده‌های اقلیت پیشنهاد شده است. این روش‌ها می‌توانند به بهبود عملکرد مدل‌ها در پیش‌بینی موارد نادر کمک کنند.

چالش دیگر که در این پژوهش مورد بررسی قرار گرفته، شفافیت پایین مدل‌های یادگیری عمیق است. این مدل‌ها به دلیل ساختار پیچیده و تعداد زیاد لایه‌های مخفی، تفسیرپذیری کمتری دارند و این موضوع می‌تواند در تصمیم‌گیری‌های اعتباری چالش‌برانگیز باشد. با وجود این چالش‌ها، نتایج مقاله حاکی از آن است که مدل‌های یادگیری عمیق نظیر شبکه‌های عصبی بازگشتی و حافظه بلندمدت کوتاه‌مدت توانسته‌اند الگوهای پیچیده و روابط پنهان در داده‌های اعتباری را بهتر شناسایی کنند و دقت بالاتری در پیش‌بینی ریسک اعتباری داشته باشند. به‌طور کلی، مقاله تأکید می‌کند که مدل‌های یادگیری عمیق در مقایسه با روش‌های یادگیری ماشین سنتی، عملکرد بهتری در تشخیص ریسک نکول دارند. همچنین استفاده از رویکردهای ترکیبی و پردازش داده‌های نامتعادل، از جمله عواملی هستند که می‌توانند در بهبود دقت و کارایی مدل‌های پیش‌بینی ریسک اعتباری نقش مؤثری داشته باشند. این نتایج می‌تواند به مؤسسات مالی کمک کند تا استراتژی‌های بهتری در مدیریت ریسک اعتباری تدوین کنند و از زیان‌های ناشی از نکول جلوگیری نمایند. ([روحی, 1402](#_ENREF_23))

ریسک اعتباری به‌عنوان یکی از اصلی‌ترین چالش‌های مؤسسات مالی معرفی شده است. این پژوهش بر این نکته تأکید دارد که ریسک اعتباری به معنای عدم توانایی قرض‌گیرنده در بازپرداخت تسهیلات و وام در موعد مقرر می‌باشد و مدیریت این ریسک برای کاهش ضررهای مالی مؤسسات ضروری است. با توجه به افزایش داده‌های مشتریان در بانکداری الکترونیکی، روش‌های سنتی اعتبارسنجی دیگر کارآمد نبوده و نیاز به رویکردهای هوشمند بر اساس تحلیل داده‌های واقعی احساس می‌شود.در این تحقیق الگوریتم‌های مختلف یادگیری ماشین برای پیش‌بینی ریسک اعتباری بررسی شده‌اند. از جمله چالش‌های این پژوهش، انتخاب ویژگی‌های مؤثر و مدلسازی بر روی کلاس‌های نامتوازن بوده است. پژوهشگران از سه روش تحلیل واریانس [[17]](#footnote-17)، اطلاعات متقابل [[18]](#footnote-18) و حذف بازگشتی ویژگی[[19]](#footnote-19) برای انتخاب ویژگی‌ها استفاده کرده‌اند و سپس این ویژگی‌ها را در مدل‌های مختلف یادگیری ماشین اعمال نموده‌اند.

در این مطالعه، الگوریتم‌های حساس به هزینه مانند درخت تصمیم و ماشین بردار پشتیبان و همچنین روش‌های جمعی مانند جنگل تصادفی و تقویت گرادیان سبک [[20]](#footnote-20) مورد استفاده قرار گرفتند. این پژوهش نشان داد که الگوریتم تقویت گرادیان سبک عملکرد بهتری نسبت به سایر روش‌ها داشته و بهترین نتایج مدلسازی زمانی حاصل شده که ویژگی‌ها با روش تحلیل واریانس انتخاب شده‌اند.

در این پژوهش استفاده از روش‌های نوین حوزه یادگیری ماشین مانند الگوریتم تقویت گرادیان سبک مورد بررسی قرار گرفته است. نتایج نشان داده‌اند که این روش با دستیابی به معیار سطح زیر منحنی مشخصه عملکرد[[21]](#footnote-21) (0.724) و معیار دقت وزنی (0.788) بالاترین دقت را در پیش‌بینی ریسک اعتباری داشته است. در نهایت، این پژوهش بر اهمیت بهره‌گیری از سیستم‌های هوشمند و الگوریتم‌های بهینه برای کاهش زیان‌های ناشی از اعطای وام‌های پرخطر تأکید دارد و نشان می‌دهد که چنین سیستم‌هایی می‌توانند نقش مؤثری در بهبود فرآیندهای تصمیم‌گیری و مدیریت ریسک در نهادهای مالی ایفا کنند.

2-3-2 تحقیقات خارجی

در حوزه امتیازدهی اعتباری، همواره انتخاب الگوریتم‌های مناسب برای بهبود دقت پیش‌بینی و کاهش ریسک‌های مالی، یک چالش اساسی بوده است. پیش از این، مدل‌های طبقه‌بندی پایه به‌تنهایی قادر به ارائه دقت مطلوب در پیش‌بینی امتیاز اعتباری نبودند و نیاز به روشی مؤثرتر احساس می‌شد. در این راستا، پژوهشی با استفاده از مجموعه داده‌های اعتباری استرالیا، پس از پیش‌پردازش داده‌ها و جایگزینی مقادیر گمشده، اقدام به کاهش ابعاد داده‌ها با تحلیل مؤلفه‌های اصلی نمود. سپس، الگوریتم‌های طبقه‌بندی پایه مانند رگرسیون لجستیک، درخت تصمیم، ماشین بردار پشتیبان، نزدیک‌ترین همسایه، شبکه عصبی چندلایه و بیز ساده به‌کار گرفته شدند. علاوه بر آن، الگوریتم‌های گروهی نظیر جنگل تصادفی، درخت تصمیم بسته‌بندی‌شده، طبقه‌بندی هایی مبنتی بر روش های تقویتی، و رأی‌گیری گروهی مورد ارزیابی قرار گرفتند. نتایج نشان داد که الگوریتم‌های گروهی، به‌ویژه جنگل تصادفی و طبقه‌بندی هایی مبنتی بر روش های تقویتی ، با دقت بالاتر و خطای کمتر، عملکرد بهتری در پیش‌بینی امتیاز اعتباری دارند. ([Parvin and Saleena, 2020](#_ENREF_15))

مطالعات پیشین بر روی ارزیابی اعتبار مشتریان نشان می‌دهد که مدل‌های رایج همچون رگرسیون لجستیک و تحلیل تمایز خطی به دلیل فرضیات غیرواقعی مانند تفکیک خطی و توزیع نرمال داده‌ها عملکرد ضعیفی داشتند. به کارگیری روش‌هایی مانند شبکه عصبی و ماشین بردار پشتیبان در تلاش بود تا دقت مدل‌های ارزیابی اعتبار را بهبود بخشد، اما این مدل‌های تک‌طبقه‌بند در شناسایی ویژگی‌های خاص مشتریان و کاهش تأثیر نمونه‌های نادر به مشکل برخوردند. برای رفع این محدودیت‌ها، برخی تحقیقات از روش‌های ترکیبی شامل طبقه‌بندی هایی با رای‌گیری های وزن‌دار استفاده کردند، اما این روش‌ها همچنان در مدیریت داده‌های نامتوازن با مشکل مواجه بودند. در این پژوهش، از نقشه‌های خودسازمان‌ده برای خوشه‌بندی چندلایه داده‌ها بهره گرفته شده و سپس از ترکیب طبقه‌بندهایی چون کا-نزدیک‌ترین همسایگی و ماشین بردار پشتیبان با رای‌گیری نرم استفاده شده است. به منظور افزایش دقت و کاهش تأثیر داده‌های نامتوازن، بهینه‌سازی وزن‌ها با الگوریتم بهینه‌سازی سیلفیش[[22]](#footnote-22) انجام شده که الگوریتمی متاهیوریستیک با سرعت همگرایی بالا و قابلیت جستجوی فراگیر است. این ترکیب از خوشه‌بندی خودسازمان‌ده و الگوریتم بهینه‌سازی سیلفیش نه تنها به طبقه‌بندی دقیق‌تر و شناسایی مؤثر نمونه‌های نادر کمک می‌کند، بلکه پردازش کارآمدتری نیز در مقایسه با روش‌های پیشین دارد. ([Singh et al., 2024](#_ENREF_17))

در تحقیقات گذشته، مسئله‌ مهمی که در مدل‌سازی ریسک اعتباری با روش خوشه‌بندی وجود داشت، انتخاب متغیرهای مناسب برای ایجاد خوشه‌های دقیق بود، زیرا انتخاب نادرست متغیرها می‌توانست منجر به خوشه‌های غیرقابل‌تشخیص و ناموثر در پیش‌بینی شود. برخی از پژوهش‌ها با استفاده از تحلیل مؤلفه‌های اصلی به کاهش بعد متغیرها پرداختند اما این روش پیچیدگی خوشه‌بندی را افزایش می‌داد و قابلیت تفسیر مدل را کاهش می‌داد. روش‌های دیگر نظیر حذف تدریجی متغیرها با استفاده از رگرسیون خطی نیز به کار گرفته شده‌اند، اما این روش‌ها به دلیل انتخاب‌های تصادفی، دقت خوشه‌بندی و کیفیت پیش‌بینی را تحت تأثیر قرار می‌دادند. مقاله حاضر با معرفی یک روش مبتنی بر یادگیری ماشین، از مقادیر شاپلی[[23]](#footnote-23) برای سنجش اهمیت متغیرها استفاده می‌کند و با انتخاب متغیرهای اصلی در مرحله خوشه‌بندی، سعی دارد دقت پیش‌بینی مدل را افزایش دهد. این رویکرد، از الگوریتم‌های جنگل تصادفی و درخت‌های تقویت‌شده بهره برده و با استفاده از الگوریتم کاپروتوتایپ‌ها [[24]](#footnote-24)و تکنیک‌های اعتبارسنجی، خوشه‌بندی را طوری انجام می‌دهد که نتایج هم از نظر دقت پیش‌بینی و هم از نظر تفسیرپذیری مورد تایید قرار گیرد.([Zöllner and Gürtler, 2023](#_ENREF_20" \o "Zöllner, 2023 #3))

مطالعات پیشین در زمینه ارزیابی ریسک اعتباری عمدتاً بر استفاده از روش‌های خوشه‌بندی و طبقه‌بندی داده‌ها تمرکز داشته‌اند. روش‌های رایج مانند کا میانگین[[25]](#footnote-25) و سلسله‌مراتبی[[26]](#footnote-26)، علی‌رغم سادگی و تفسیرپذیری، در داده‌های چندبعدی و غیرخطی به محدودیت‌هایی از جمله کاهش دقت در خوشه‌بندی و ناکارآمدی در جداسازی داده‌های نویزی مواجه شده . در این مقاله، یک رویکرد ترکیبی ارائه شده است که خوشه‌بندی کا میانگین و سلسله‌مراتبی را با استفاده از مختصات فاکتوریل استخراج‌شده از تحلیل مکاتبات چندگانه[[27]](#footnote-27) تلفیق می‌کند. این روش، ابتدا با پیش‌پردازش داده‌ها شامل نرمال‌سازی و کاهش نویز داده‌ها ، ساختار داده ها را برای تحلیل آماده می‌سازد. سپس، تحلیل مکاتبات چندگانه به منظور کاهش ابعاد و استخراج ویژگی‌های اساسی داده‌ها به کار گرفته می‌شود. در ادامه، خوشه‌بندی کا میانگین و سلسله‌مراتبی به صورت ترکیبی اعمال شده و کیفیت خوشه‌بندی بر اساس تجزیه اینرسی ارزیابی می‌شود. این روش امکان مدیریت دقیق‌تر داده‌ها، کاهش نویز و بهبود دقت مدل‌های یادگیری ماشین در تحلیل ریسک اعتباری را فراهم کرده است. ([Jadwal et al., 2022](#_ENREF_8))

در زمینه ارزیابی ریسک اعتباری، انتخاب ویژگی به عنوان یکی از چالش‌های اصلی در بهبود دقت پیش‌بینی و کاهش پیچیدگی مدل‌ها شناخته شده است. پیش از این، روش‌های غربالگری مانند رتبه‌بندی همبستگی[[28]](#footnote-28) یا اطلاعات متقابل[[29]](#footnote-29) به دلیل سوگیری در انتخاب ویژگی‌ها و کاهش دقت پیش‌بینی، محدودیت‌هایی داشتند. همچنین، روش‌های ترکیبی دو‌مرحله‌ای مانند استفاده از فیلترهای اولیه و سپس الگوریتم‌های جستجوی تکاملی، اگرچه در برخی موارد به بهبود دقت کمک می‌کردند، اما باعث افزایش پیچیدگی محاسبات و کاهش پایداری مدل‌ها می‌شدند. در این مقاله، رویکردهای یک‌مرحله‌ای مانند حذف بازگشتی ویژگی [[30]](#footnote-30) با مدل‌های جنگل تصادفی [[31]](#footnote-31) و ماشین بردار پشتیبان [[32]](#footnote-32) بررسی شده‌اند. این روش‌ها ویژگی‌ها را به طور پویا در فرآیند یادگیری انتخاب کرده و امکان پایداری بیشتر انتخاب ویژگی را فراهم می‌کنند. همچنین، استفاده از رگرسیون لجستیک جریمه‌ای [[33]](#footnote-33) به منظور کاهش ابعاد داده‌ها و کنترل همبستگی متغیرها بررسی شده است. روش حذف بازگشتی ویژگی با جنگل تصادفی در معیارهایی مانند دقت پیش‌بینی و پایداری انتخاب ویژگی عملکرد بهتری نسبت به روش‌های دیگر داشته و الگوریتم حذف بازگشتی ویژگی با ماشین بردار پشتیبان با پایداری بالاتر، گزینه‌ای مناسب برای مجموعه داده‌های با ابعاد بزرگ است. ([Atif and Salmi, 2022](#_ENREF_3))

در زمینه ارزیابی ریسک اعتباری رویکردهای اولیه معمولاً از مدل‌های پایه استفاده می‌کردند که به دلیل نبود ترکیب مناسب الگوریتم‌ها و عدم استفاده از تکنیک‌های پیش‌پردازش، دقت پایین‌تری داشتند. برخی تحقیقات نیز از روش‌های ترکیبی مانند ادغام یادگیری نظارت‌شده و بدون نظارت یا استفاده از شبکه‌های عصبی مصنوعی بهره بردند، اما این روش‌ها به دلیل نیاز به محاسبات با حجم بالا یا عدم مدیریت مناسب ویژگی‌های غیرمؤثر، در محیط‌های واقعی کارایی مطلوبی نداشتند. استفاده از روش‌های انتخاب ویژگی بر اساس همبستگی یا مدل‌های خطی نیز در برخی مطالعات مطرح شده، اما این روش‌ها قادر به شناسایی روابط غیرخطی بین ویژگی‌ها نبودند. در این میان، مطالعاتی که از روش‌های ترکیبی چندلایه استفاده کردند، توانستند با ادغام خروجی مدل‌های مختلف، عملکرد بهتری ارائه دهند. این روش‌ها به‌ویژه در مدیریت عدم توازن داده‌ها و بهبود معیارهایی نظیر دقت و امتیاز F1 نقش مؤثری ایفا کردند. با این حال، چالش انتخاب ویژگی‌های بهینه و کاهش پیچیدگی مدل همچنان به قوت خود باقی بود. پژوهش حاضر با ارائه رویکردی چندلایه و ترکیب آن با انتخاب ویژگی مبتنی بر اطلاعات متقابل، توانست با بهینه‌سازی ویژگی‌های ورودی و کاهش پیچیدگی محاسباتی، نتایج قابل توجهی در پیش‌بینی ریسک اعتباری ارائه دهد. ([Emmanuel et al., 2024](#_ENREF_5))

روش‌های سنتی ارزیابی ریسک اعتباری مانند رگرسیون لجستیک و ماشین بردار پشتیبان با وجود دقت قابل‌قبول، در مدیریت نمونه‌های پیچیده و اطلاعات ناقص ضعف دارند و معمولاً با تصمیم‌گیری دوکلاسه، ریسک خطا را افزایش می‌دهند. رویکردهای ترکیبی مانند تقویت گرادیانی پیشرفته [[34]](#footnote-34) با وجود بهبود نسبی، به دلیل استفاده از آستانه‌های ثابت و ناتوانی در مدیریت داده‌های با هزینه بالا، محدودیت‌هایی داشته‌اند. تصمیم‌گیری سه‌ کلاسه با افزودن تأخیر در تصمیم‌گیری، به کاهش خطا کمک کرده است، اما تنظیم آستانه‌ها در روش‌های پیشین مبتنی بر فرضیات کلی بوده و داده‌های مالی واقعی را نادیده گرفته‌اند. مقاله حاضر با ترکیب تصمیم‌گیری سه کلاسه و استفاده از داده‌های جریان نقدی برای تنظیم آستانه‌ها و بهره‌گیری از مدل بهینه‌سازی چندهدفه، این محدودیت‌ها را برطرف کرده و با ترکیب الگوریتم ماشین تقویت گرادیانی سبک [[35]](#footnote-35) امکان تصمیم‌گیری دقیق‌تر برای نمونه‌های مرزی و پرهزینه را فراهم ساخته و کارایی مدیریت ریسک اعتباری را افزایش داده است. ([Li et al., 2024](#_ENREF_12))

پیش‌بینی ریسک اعتباری عمدتاً بر روش‌های یادگیری ماشین متکی بوده‌اند که شامل تکنیک‌هایی مانند شبکه‌های عصبی، ماشین بردار پشتیبان و رگرسیون لجستیک می‌شوند. این مدل‌ها با وجود توانایی بالا در تحلیل داده‌ها، به دلیل اتخاذ تصمیم‌گیری دو کلاسه، دچار خطاهایی در دسته‌بندی داده‌های مرزی شده‌اند. برای حل این مشکل، رویکردهای سه‌کلاسه معرفی شده‌اند که با تقسیم داده‌ها به سه منطقه مثبت، منفی و مرزی، تصمیم‌گیری برای داده‌های نامطمئن را به تأخیر می‌اندازند. روش‌های مختلفی برای یادگیری آستانه در تصمیم‌گیری سه‌کلاسه استفاده شده است، از جمله استفاده از بهینه‌سازی چندهدفه و مدل‌های مبتنی بر کاهش هزینه‌های تصمیم‌گیری. با این حال، اکثر این رویکردها نیازمند تنظیم دستی پارامترها بوده‌اند که قابلیت کاربرد آن‌ها را محدود کرده است. مقاله حاضر با بهره‌گیری از یک چارچوب دو مرحله‌ای و استفاده از الگوریتم بهینه‌سازی ازدحام ذرات برای یادگیری خودکار آستانه‌ها، رویکردی جامع‌تر ارائه کرده است. این روش با ترکیب بهینه اطلاعات و کاهش هزینه تصمیم‌گیری، ضمن افزایش دقت در دسته‌بندی داده‌های مرزی، امکان استفاده از داده‌های تکمیلی در مراحل بعدی را فراهم می‌کند و محدودیت‌های رویکردهای پیشین را برطرف می‌سازد. ([Li and Sha, 2024](#_ENREF_11))

در ارزیابی ریسک اعتباری، روش‌های سنتی مانند رگرسیون لجستیک و تحلیل تمایز خطی به دلیل سادگی، محبوبیت بالایی داشتند، اما عدم توانایی در مدیریت داده‌های پیچیده و نامطمئن از معایب اصلی آنها بود. مدل‌های پیشرفته‌تر مانند شبکه‌های عصبی مصنوعی و درخت‌های تصمیم‌گیری تقویتی با قدرت تشخیص بالاتر، توانستند این محدودیت را کاهش دهند، اما بیشتر این مدل‌ها تنها از طبقه‌بندی دوتایی استفاده کرده و در مواجهه با اطلاعات ناکافی یا مبهم دچار ضعف می‌شوند. مدل سه‌ کلاسه تصمیم‌گیری با ارائه دسته مرزی برای موارد نامطمئن گامی در مدیریت بهتر این چالش‌ها برداشت، اما این روش نیز به دلیل وابستگی زیاد به تنظیمات آن انعطاف‌پذیری کمی داشت. مقاله حاضر با ترکیب مدل سه‌کلاسه تصمیم‌گیری و روش‌های یادگیری ماشین مانند تقویت‌کننده گرادیان، و بهینه‌سازی با الگوریتم ژنتیک، یک چارچوب بهینه‌سازی ترتیبی ارائه داده است که بدون نیاز به تنظیمات دستی پارامترها آستانه‌های تصمیم‌گیری را بهینه می‌کند و با افزودن تدریجی اطلاعات به موارد مرزی، عدم قطعیت را کاهش داده و عملکرد کلی سیستم را بهبود می‌بخشد. ([Shen et al., 2022](#_ENREF_16))

تصمیم‌گیری سه‌کلاسه ، بر اساس نظریه مجموعه‌های راف احتمالی، به‌عنوان راهکاری مؤثر برای مدیریت عدم قطعیت معرفی شده است. این رویکرد، به‌جای تصمیم‌گیری قطعی در مدل‌های دو‌کلاسه ، با افزودن یک منطقه مرزی، امکان تصمیم‌گیری منعطف‌تر را فراهم می‌کند. پیش از این، مدل‌های دو‌کلاسه مانند رگرسیون لجستیک یا درخت تصمیم برای ارزیابی اعتباری استفاده می‌شدند که مستلزم جمع‌آوری اطلاعات جامع از تمامی متقاضیان و افزایش منابع پردازشی بود. مقاله حاضر با استفاده از مجموعه‌های راف احتمالی، یک فرآیند دو مرحله‌ای را پیشنهاد می‌کند. در مرحله اول، با اطلاعات اولیه مانند فرم‌های درخواست، مشتریان به سه دسته تأیید، رد و مرزی تقسیم می‌شوند. مرحله دوم، که مختص موارد مرزی است، شامل جمع‌آوری اطلاعات تکمیلی از طریق مصاحبه و تحلیل مالی است. تحلیل دقیق در این مرحله با استفاده از رگرسیون لجستیک انجام می‌شود که احتمال بازپرداخت را بر اساس داده‌های تکمیلی پیش‌بینی می‌کند. این فرآیند با مدیریت بهینه منابع ، از جمع‌آوری اطلاعات غیرضروری جلوگیری کرده و دقت و کارایی تصمیم‌گیری را بهبود می‌بخشد. ([Maldonado et al., 2020](#_ENREF_14))

2-4 جمع‌بندی

در فصل دوم این پژوهش به بررسی مبانی نظری و پیشینه تحقیق در زمینه ارزیابی ریسک اعتباری مشتریان بانکی پرداخته شد. این فصل در دو بخش اصلی سازماندهی شده است: بخش اول به مبانی نظری و مفاهیم پایه مرتبط با ریسک و مدیریت ریسک اعتباری اختصاص دارد و بخش دوم به مرور پیشینه مطالعات انجام‌شده در این حوزه می‌پردازد.

در بخش مبانی نظری، مفاهیم ریسک، انواع آن در حوزه بانکی، به‌ویژه ریسک اعتباری، و روش‌های مختلف مدیریت ریسک مورد بررسی قرار گرفت. این بخش نشان داد که ریسک اعتباری یکی از مهم‌ترین چالش‌های نظام بانکی است که به‌طور مستقیم بر بازپرداخت تسهیلات بانکی و سلامت مالی بانک‌ها تأثیر می‌گذارد. همچنین، به اهمیت طبقه‌بندی مشتریان بر اساس ریسک اعتباری و نقش این طبقه‌بندی در بهبود مدیریت تسهیلات و کاهش زیان‌های ناشی از عدم بازپرداخت تسهیلات اشاره شد.

در بخش پیشینه تحقیق، مطالعات مختلف داخلی و خارجی مرتبط با ارزیابی ریسک اعتباری بررسی شد. این مطالعات نشان دادند که رویکردهای متنوعی برای ارزیابی و پیش‌بینی ریسک اعتباری به کار گرفته شده است. برخی تحقیقات از مدل‌های سنتی نظیر تحلیل نسبت‌های مالی و امتیازدهی اعتباری استفاده کرده‌اند، در حالی که مطالعات جدیدتر بر استفاده از روش‌های داده‌کاوی و یادگیری ماشین متمرکز بوده‌اند.

تحقیقات پیشین تأکید کرده‌اند که استفاده از مدل‌های ترکیبی و طبقه‌بندی‌های چندگانه می‌تواند دقت پیش‌بینی ریسک اعتباری را افزایش دهد. به‌ویژه، روش‌های بگینگ با ایجاد مجموعه‌ای از مدل‌های پیش‌بینی و ترکیب نتایج آن‌ها می‌توانند به کاهش خطای طبقه‌بندی و افزایش دقت تصمیم‌گیری منجر شوند. این روش‌ها در مقایسه با مدل‌های منفرد، به دلیل بهره‌گیری از رویکردهای چندگانه، کارایی بالاتری در شناسایی مشتریان پرخطر و کاهش زیان‌های احتمالی بانک‌ها داشته‌اند.

مطالعات پیشین همچنین نشان دادند که عوامل مختلفی مانند درآمد مشتری، سابقه بازپرداخت تسهیلات، نوع مالکیت، و وضعیت مالی شرکت‌ها تأثیر مستقیمی بر ریسک اعتباری دارند. بررسی این عوامل در قالب شاخص‌های کمی و کیفی می‌تواند در طبقه‌بندی دقیق‌تر مشتریان و تخصیص بهینه تسهیلات بانکی مؤثر باشد.

در مجموع، بررسی پیشینه تحقیق نشان داد که هرچند مطالعات متعددی در زمینه ارزیابی ریسک اعتباری انجام شده است، اما کمتر به یکپارچه‌سازی روش‌های سنتی و نوین در قالب یک چارچوب ترکیبی توجه شده است. این پژوهش تلاش دارد با استفاده از **تلفیق ماتریس زیان و روش‌های ترکیبی مبتنی بر بگینگ** به بهبود ارزیابی ریسک اعتباری و ارائه مدلی جامع‌تر برای طبقه‌بندی مشتریان بانکی بپردازد.

فصل سوم:

روش تحقیق

(مراحل انجام پژوهش)

**3-1 مقدمه**

پژوهش علمی فرآیندی نظام‌مند و هدفمند است که در آن از روش‌های علمی برای کشف روابط پنهان میان متغیرها استفاده می‌شود. این فرآیند به پژوهشگر امکان می‌دهد تا با رویکردی منطقی و تجربی، فرضیات خود را مورد آزمون قرار داده و به نتایجی معتبر و قابل اطمینان دست یابد. روش تحقیق در هر پژوهشی ابزاری است که به کمک آن می‌توان از خطاهای احتمالی کاسته و با دقت بیشتری به اهداف پژوهش نزدیک شد.

تحقیقات مرتبط با ارزیابی ریسک اعتباری به دلیل اهمیت بالای این موضوع در تصمیم‌گیری‌های مالی و بانکی، نیازمند به‌کارگیری روش‌های پیشرفته و ترکیبی است. در پژوهش حاضر نیز با ترکیب ماتریس زیان و الگوریتم طبقه‌بندی ترکیبی مبتنی بر روش بگینگ، تلاش شده است تا مدلی ارائه شود که دقت بیشتری در پیش‌بینی ریسک اعتباری داشته باشد. انتخاب این روش‌ها با توجه به ماهیت داده‌ها و اهداف تحقیق انجام شده است.

در فصل‌های پیشین به مبانی نظری، پیشینه تحقیقات مرتبط و مفاهیم پایه‌ای پرداخته شد. در این فصل، روش تحقیق مورد استفاده برای دستیابی به اهداف پژوهش شرح داده خواهد شد. این فصل شامل تشریح جامعه آماری، قلمرو پژوهش، ابزارهای گردآوری داده‌ها، متغیرهای تحقیق و روش‌های تحلیل داده‌ها است. همچنین فرآیند اجرای مدل ترکیبی و نحوه استفاده از ماتریس زیان در تحلیل داده‌ها مورد بحث قرار می‌گیرد. هدف از این فصل، ارائه چارچوبی منسجم و روشن برای درک روش‌شناسی تحقیق و اطمینان از دقت و اعتبار یافته‌ها است.

این ساختار به گونه‌ای طراحی شده که تمامی مراحل تحقیق به صورت نظام‌مند و قابل ارزیابی تشریح شوند تا نتایج نهایی از پشتوانه علمی و عملی کافی برخوردار باشند.

**3-2 روش کلی تحقیق**

در این بخش، ساختار کلی تحقیق با تأکید بر هدف، روش استنتاج و طرح تحقیق به صورت نظام‌مند و با بهره‌گیری از اصول علمی ارائه می‌شود. هدف از این بخش، ارائه توضیحات جامع درباره مبانی نظری و عملی تحقیق به منظور تضمین شفافیت در روش‌شناسی پژوهش است.

3-2-1 هدف پژوهش

این پژوهش از نوع کاربردی است. هدف اصلی آن ارائه راهکارهای عملیاتی برای ارزیابی و کاهش ریسک اعتباری از طریق استفاده از مدل‌های ترکیبی پیشرفته نظیر ماتریس زیان و الگوریتم بگینگ است. ماهیت کاربردی این پژوهش بر نیاز به حل مشکلات واقعی و ارائه پیشنهادهایی مبتنی بر شواهد تجربی تأکید دارد.

3-2-2 روش استنتاج

روش استنتاج این تحقیق توصیفی-تحلیلی است. در بخش توصیفی، داده‌های جمع‌آوری شده از جامعه آماری، ساختاردهی و بررسی می‌شوند. در بخش تحلیلی، این داده‌ها با استفاده از روش‌های آماری پیشرفته و الگوریتم‌های یادگیری ماشین تحلیل شده و روابط میان متغیرها مورد ارزیابی قرار می‌گیرد. این ترکیب روش‌ها، امکان دستیابی به تحلیل‌های دقیق و علمی را فراهم می‌کند.

3-2-3 طرح تحقیق

طرح کلی این پژوهش از نوع پس‌رویدادی یا گذشته‌نگر است. این طرح به دلیل استفاده از داده‌های تاریخی و تحلیل سوابق اعتباری مشتریان بانک‌ها انتخاب شده است. داده‌های مورد استفاده شامل اطلاعات مرتبط با عملکرد اعتباری مشتریان طی دوره زمانی مشخص است که امکان ارزیابی جامع و علمی را فراهم می‌آورد

**3-3 جامعه آماری و قلمرو تحقیق**

3-3-1 جامعه آماری

جامعه آماری این تحقیق تمامی مشتریان حقیقی است که در بازه زمانی مورد مطالعه از بانک مورد مطالعه این پژوهش تسهیلات دریافت کرده‌اند. این جامعه به دو گروه خوش‌حساب و بدحساب تقسیم‌بندی شده است. مشتریان خوش‌حساب، افرادی هستند که به موقع تعهدات مالی خود را انجام داده‌اند، در حالی که مشتریان بدحساب شامل افرادی می‌شوند که با تأخیر بیش از 6 ماه در بازپرداخت تسهیلات مواجه بوده یا از انجام تعهدات خود ناتوان بوده‌اند. این تفکیک به منظور تحلیل بهتر رفتار اعتباری مشتریان انجام شده است.

3-3-2 قلمرو مکانی

قلمرو مکانی این پژوهش شامل تمامی شعب بانک‌های تجاری منتخب در ایران است. بانک‌های منتخب بر اساس معیارهایی همچون گستردگی جغرافیایی، تنوع در ارائه تسهیلات و دسترسی به داده‌های قابل اطمینان انتخاب شده‌اند. این تنوع مکانی امکان تعمیم‌پذیری نتایج پژوهش را بهبود می‌بخشد.

3-3-3 قلمرو زمانی

قلمرو زمانی این تحقیق بازه زمانی پنج ساله از ابتدای سال 1395 تا پایان سال 1399 را شامل می‌شود. انتخاب این بازه زمانی بر مبنای دسترسی به داده‌های کافی و قابلیت تحلیل رفتار مشتریان در یک دوره زمانی مناسب صورت گرفته است. این محدودیت زمانی موجب افزایش دقت و اعتبار نتایج تحقیق می‌شود.

**3-4 روش جمع‌آوری داده‌ها و ابزارها**

داده‌های این پژوهش از طریق دسترسی دقیق به پایگاه‌های داده بانکی جمع‌آوری شده‌اند. فرآیند جمع‌آوری داده‌ها به گونه‌ای طراحی شده است که جامعیت، صحت و قابلیت اطمینان اطلاعات تضمین شود. روش مورد استفاده به شرح زیر است:

دسترسی به داده‌های بانکی: این روش شامل استخراج داده‌های دقیق و جامعی از مشتریان دریافت‌کننده کارت‌های اعتباری در بانک‌های تجاری منتخب است. داده‌ها شامل اطلاعات کلیدی نظیر تاریخ صدور کارت، میزان اعتبار اختصاص‌یافته، وضعیت بازپرداخت، سوابق اعتباری و رفتارهای مالی مشتریان بوده‌اند. مراحل گردآوری این اطلاعات با دقت بالا و از طریق بررسی سیستماتیک پایگاه‌های داده داخلی بانک انجام شده است. برای اطمینان از دقت داده‌ها، فرآیند پالایش داده‌ها شامل شناسایی و حذف داده‌های ناقص، اصلاح تناقضات، و بررسی همخوانی میان مجموعه داده‌های مختلف صورت گرفته است. این داده‌های پالایش‌شده سپس به صورت طبقه‌بندی‌شده برای استفاده در تحلیل‌های پژوهشی آماده شدند.

2-4-3 ابزارهای مورد استفاده

جمع‌آوری، پالایش و تحلیل داده‌های این پژوهش با استفاده از ابزارهای پیشرفته و متنوعی صورت گرفته است. این ابزارها نقش کلیدی در سازمان‌دهی و تحلیل داده‌ها ایفا کرده‌اند. جزئیات ابزارهای مورد استفاده به شرح زیر است:

سامانه‌های اطلاعات اعتباری بانک: این سامانه‌ها که برای مدیریت اطلاعات اعتباری طراحی شده‌اند، داده‌های مرتبط با مشتریان را به صورت کاملاً ساختاریافته و دقیق ارائه داده‌اند. از طریق این سامانه‌ها اطلاعات کلیدی نظیر تاریخ صدور کارت، مبلغ اعتبار، وضعیت بازپرداخت و سوابق اعتباری مشتریان استخراج شده است. این ابزارها امکان پالایش، فیلترگذاری و تحلیل سریع داده‌ها را با دقت بالا فراهم کرده‌اند.

نرم‌افزارهای مدیریت پایگاه داده: برای مدیریت داده‌ها از ابزارهای پیشرفته‌ای مانند SQL استفاده شده است. این نرم‌افزارها قابلیت‌هایی نظیر جستجوهای شرطی، ادغام داده‌ها، پالایش اطلاعات و ایجاد پایگاه‌های داده ساختاریافته را فراهم کرده‌اند. این ابزارها با سرعت بالا و دقت مطلوب امکان استخراج و آماده‌سازی داده‌های موردنیاز را فراهم کرده‌اند.

ابزارهای تحلیل آماری و مدل‌سازی: نرم‌افزارهایی نظیر SPSS و Python برای انجام تحلیل‌های پیچیده آماری و مدل‌سازی داده‌ها به کار گرفته شده‌اند. این ابزارها امکان پردازش داده‌های حجیم، اعمال مدل‌های پیشرفته پیش‌بینی و تحلیل روندها را فراهم کرده‌اند. همچنین، با استفاده از این ابزارها روابط میان متغیرها به‌دقت بررسی و مدل‌سازی‌های متناسب با اهداف پژوهش انجام شده است.

ابزارهای بصری‌سازی داده‌ها: برای ارائه یافته‌های پژوهش و تفسیر بهتر داده‌ها از ابزارهای بصری‌سازی نظیر Tableau و Matplotlib استفاده شده است. این ابزارها با نمایش گرافیکی داده‌ها، روندها و روابط میان متغیرها را به‌صورت واضح‌تر به نمایش گذاشته‌اند و به تحلیل‌گران کمک کرده‌اند تا تصمیمات دقیق‌تری بگیرند.

3-4-3 منابع داده‌ها

منابع داده‌های مورد استفاده در این پژوهش شامل مجموعه‌ای از پایگاه‌های داده داخلی و اطلاعات مکمل بوده است. هر یک از این منابع با دقت انتخاب و بررسی شده‌اند تا از صحت و جامعیت اطلاعات اطمینان حاصل شود. جزئیات این منابع عبارتند از:

بانک اطلاعاتی داخلی کارت‌های اعتباری: این منبع شامل اطلاعات جامعی از مشتریان دریافت‌کننده کارت‌های اعتباری است. داده‌هایی نظیر تاریخچه بازپرداخت، نوع کارت‌های اعتباری، میزان اعتبار اختصاص‌یافته و سوابق اعتباری مشتریان به‌صورت کامل در این پایگاه داده ثبت شده است. این اطلاعات به‌عنوان منبع اصلی پژوهش استفاده شده و مبنای بسیاری از تحلیل‌های آماری و پیش‌بینی‌ها بوده است.

گزارش‌های سالانه و مستندات بانک‌ها: اطلاعات ارائه‌شده در گزارش‌های عمومی و سالانه بانک‌ها که شامل تحلیل‌های کلی عملکرد اعتباری و فرآیندهای صدور کارت بوده است، به‌عنوان منابع مکمل برای اعتبارسنجی و تکمیل داده‌ها مورد استفاده قرار گرفته‌اند. این منابع، بینش‌های مفیدی درباره روندهای اعتباری و مدیریت ریسک در سیستم بانکی ارائه داده‌اند.

ابزارها و نرم‌افزارهای تخصصی مدیریت داده: نرم‌افزارهایی که برای پردازش و تحلیل داده‌های پژوهش استفاده شده‌اند، نقش مهمی در ساختاردهی و آماده‌سازی داده‌ها برای تحلیل ایفا کرده‌اند. این ابزارها داده‌ها را به فرمت‌هایی استاندارد و قابل تحلیل تبدیل کرده و استفاده از مدل‌های پژوهشی را تسهیل کرده‌اند.

در این بخش، روش‌های دقیق و ابزارهای پیشرفته‌ای که برای جمع‌آوری و تحلیل داده‌ها استفاده شده‌اند، معرفی گردید. این روش‌شناسی با تأکید بر جامعیت و دقت، داده‌های معتبری را برای پژوهش فراهم آورده است. ترکیب داده‌های بانکی داخلی با اطلاعات تکمیلی و استفاده از ابزارهای پیشرفته تحلیل، چارچوبی محکم برای بررسی فرضیات پژوهش و ارائه نتایجی با کیفیت بالا ایجاد کرده است. این رویکرد علمی، زمینه‌ساز انجام تحلیل‌های دقیق و مدل‌سازی‌های پیشرفته در فصل‌های آتی خواهد بود.

**3-5 مدل تحقیق و متغیرها**

مدل تحقیق حاضر بر پایه یک روش ترکیبی مبتنی بر الگوریتم بگینگ (Bagging) طراحی شده است. بگینگ، به عنوان یکی از تکنیک‌های یادگیری جمعی (Ensemble Learning)، با ایجاد چندین مدل ضعیف (Weak Learners) و ترکیب نتایج آن‌ها به منظور کاهش واریانس و افزایش دقت پیش‌بینی عمل می‌کند. در این پژوهش، الگوریتم بگینگ با هدف بهبود ارزیابی ریسک اعتباری مشتریان، مورد استفاده قرار گرفته است. این مدل از طریق انتخاب نمونه‌های تصادفی از داده‌های اولیه و ایجاد مدل‌های مستقل، قدرت پیش‌بینی و انعطاف‌پذیری تحلیل را افزایش می‌دهد. این رویکرد، در کنار ویژگی‌های ترکیبی، توانسته است از خطاهای طبقه‌بندی کاسته و نتایجی با قابلیت اطمینان بالا ارائه دهد.

1-5-3 معرفی مدل ترکیبی (بگینگ)

مدل تحقیق حاضر بر پایه یک روش ترکیبی مبتنی بر الگوریتم بگینگ (Bagging) طراحی شده است. بگینگ، به عنوان یکی از تکنیک‌های یادگیری جمعی (Ensemble Learning)، با ایجاد چندین مدل ضعیف (Weak Learners) و ترکیب نتایج آن‌ها به منظور کاهش واریانس و افزایش دقت پیش‌بینی عمل می‌کند. در این پژوهش، الگوریتم بگینگ با هدف بهبود ارزیابی ریسک اعتباری مشتریان، مورد استفاده قرار گرفته است. این مدل از طریق انتخاب نمونه‌های تصادفی از داده‌های اولیه و ایجاد مدل‌های مستقل، قدرت پیش‌بینی و انعطاف‌پذیری تحلیل را افزایش می‌دهد. این رویکرد، در کنار ویژگی‌های ترکیبی، توانسته است از خطاهای طبقه‌بندی کاسته و نتایجی با قابلیت اطمینان بالا ارائه دهد.

2-5-3 نقش ماتریس زیان در مدل

ماتریس زیان (Loss Matrix) به‌عنوان یکی از ابزارهای کلیدی در این پژوهش، برای مدیریت و کاهش هزینه‌های مرتبط با خطاهای طبقه‌بندی به کار گرفته شده است. در فرآیند ارزیابی ریسک اعتباری، دو نوع خطای اصلی مطرح هستند:

خطای نوع اول (False Positive): طبقه‌بندی مشتری بدحساب به عنوان خوش‌حساب که می‌تواند به زیان مالی قابل توجهی برای بانک منجر شود.

خطای نوع دوم (False Negative): طبقه‌بندی مشتری خوش‌حساب به عنوان بدحساب که می‌تواند موجب از دست رفتن فرصت‌های سودآور شود.

ماتریس زیان در مدل تحقیق با وزن‌دهی مناسب به این دو نوع خطا، تاثیر بسزایی در بهینه‌سازی تصمیمات اعتباری داشته است. این ماتریس به‌گونه‌ای طراحی شده است که بیشترین تمرکز بر کاهش خطای نوع اول باشد، چرا که زیان مالی مرتبط با آن برای بانک‌ها اهمیت بیشتری دارد.

3-5-3 متغیرهای مستقل و وابسته

در این پژوهش، متغیرهای مستقل و وابسته به شرح زیر تعریف شده‌اند:

متغیر وابسته: ریسک اعتباری مشتریان که بر اساس مدل ترکیبی بگینگ ارزیابی می‌شود. این متغیر به‌صورت عددی و به‌عنوان شاخصی برای سنجش احتمال بازپرداخت یا رفتار اعتباری مشتریان محاسبه شده است و به‌جای طبقه‌بندی دودویی، یک مقیاس پیوسته ارائه می‌دهد که نمایانگر میزان ریسک اعتباری هر مشتری است.

متغیرهای مستقل: متغیرهای مستقل شامل ویژگی‌های مالی و کیفی مشتریان است که در تصمیم‌گیری اعتباری تاثیرگذار هستند. این متغیرها به دو دسته زیر تقسیم می‌شوند:

متغیرهای مالی: نظیر مبلغ وام‌های دریافت‌شده، میزان بدهی جاری مشتری، مجموع تراکنش‌های حساب، تعداد اقساط بازپرداخت‌شده، و میانگین مانده حساب ماهانه.

متغیرهای کیفی: شامل اطلاعات قابل‌دسترس از پایگاه داده‌های بانک نظیر تعداد تراکنش‌های ماهانه، تعداد تسهیلات دریافتی پیشین، میزان استفاده از تسهیلات در دوره‌های گذشته، و تعداد دفعات درخواست خدمات مالی.

3-5-4 روش اندازه‌گیری متغیرها

برای اندازه‌گیری متغیرهای تحقیق، از روش‌های استاندارد و مبتنی بر داده‌های موجود در پایگاه‌های اطلاعاتی بانک استفاده شده است. جزئیات اندازه‌گیری متغیرها به شرح زیر است:

ریسک اعتباری (متغیر وابسته): به عنوان یک شاخص کمی تعریف شده است که بیانگر احتمال نکول یا عدم بازپرداخت تعهدات مالی مشتریان است. این متغیر بر اساس تاریخچه بازپرداخت، میزان دیرکردها، و وضعیت فعلی تسهیلات دریافتی محاسبه می‌شود و به صورت یک مقدار عددی در مقیاس پیوسته ارائه می‌گردد که نمایانگر سطح ریسک اعتباری هر مشتری است.

نسبت بدهی به دارایی: از تقسیم کل بدهی‌های مشتری بر کل دارایی‌های وی محاسبه شده است.

نسبت جاری: از تقسیم دارایی‌های جاری به بدهی‌های جاری محاسبه شده است.

سابقه اعتباری: بر حسب تعداد ماه‌های همکاری مشتری با بانک اندازه‌گیری شده است.

نوع فعالیت اقتصادی: بر اساس نوع فعالیت اقتصادی مشتری، مانند بازرگانی، خدماتی یا تولیدی، به صورت کدگذاری عددی (1، 2، 3) مشخص شده است.

در این بخش، مدل تحقیق و متغیرهای مرتبط با آن تشریح شدند. ترکیب الگوریتم بگینگ و ماتریس زیان، چارچوبی قدرتمند برای بهبود ارزیابی ریسک اعتباری فراهم آورده است. در فصل‌های بعدی، نتایج به‌دست‌آمده از اجرای مدل و تحلیل‌های مربوط به آن ارائه خواهد شد.

**3-6 روش تحلیل داده‌ها**

تکنیک بگینگ (Bagging)، مخفف عبارت Bootstrap Aggregating، یکی از روش‌های یادگیری جمعی است که در این پژوهش برای ارزیابی ریسک اعتباری به کار گرفته شده است. این روش با ایجاد چندین مجموعه داده تصادفی از طریق نمونه‌گیری بوت‌استرپ و آموزش مدل‌های متعدد بر روی این نمونه‌ها، دقت پیش‌بینی را افزایش می‌دهد. ترکیب این مدل‌ها با روش بگینگ، امکان کاهش واریانس پیش‌بینی و بهبود پایداری مدل را فراهم می‌کند.

2-3-6 نحوه استفاده از ماتریس زیان در ارزیابی مدل

ماتریس زیان (Loss Matrix) یکی از ابزارهای کلیدی برای کاهش هزینه‌های مرتبط با خطاهای طبقه‌بندی در ارزیابی ریسک اعتباری است. در این پژوهش، ماتریس زیان به گونه‌ای طراحی شده است که به خطاهای طبقه‌بندی اهمیت نسبی اختصاص دهد. برای مثال، زیان ناشی از طبقه‌بندی اشتباه یک مشتری بدحساب به عنوان خوش‌حساب، معمولاً بسیار بیشتر از عکس این حالت است. با وزن‌دهی مناسب به انواع خطاها در ماتریس زیان، مدل قادر خواهد بود تصمیماتی اتخاذ کند که به حداقل‌سازی زیان کل منجر شوند. این رویکرد باعث افزایش دقت مدل در پیش‌بینی رفتار مشتریان با ریسک بالا شده است.

3-6-3 روش‌های مقایسه و اعتبارسنجی مدل‌ها

برای اطمینان از کارایی مدل‌های استفاده شده، روش‌های مختلفی برای مقایسه و اعتبارسنجی به کار گرفته شده‌اند:

روش اعتبارسنجی متقابل (Cross-Validation): در این روش، داده‌ها به چند بخش تقسیم شده و مدل بر روی بخشی از داده‌ها آموزش داده شده و بر روی بخش دیگری اعتبارسنجی می‌شود. این فرآیند برای تمامی بخش‌ها تکرار شده و میانگین نتایج به عنوان معیار نهایی دقت مدل محاسبه می‌شود.

شاخص‌های ارزیابی مدل: برای ارزیابی دقت مدل، از معیارهایی مانند دقت کلی (Accuracy)، حساسیت (Sensitivity)، ویژگی (Specificity)، و معیار F1 استفاده شده است. این شاخص‌ها اطلاعات جامعی درباره عملکرد مدل ارائه می‌دهند.

مقایسه با مدل‌های جایگزین: به منظور اطمینان از کارایی روش بگینگ، نتایج آن با مدل‌های ساده‌تر مانند درخت تصمیم‌گیری منفرد و رگرسیون لجستیک مقایسه شده است. این مقایسه نشان می‌دهد که بگینگ با ترکیب چندین مدل، عملکرد بهتری ارائه

**3-7 مراحل اجرایی تحقیق**

در گام اول، داده‌های مورد نیاز از پایگاه داده‌های بانکی استخراج شده و برای تحلیل آماده‌سازی شدند. این فرآیند شامل پاک‌سازی داده‌ها، حذف داده‌های ناقص، و استانداردسازی مقادیر متغیرها برای سازگاری با مدل‌های پیش‌بینی بود. همچنین، داده‌ها به صورت تصادفی به دو مجموعه آموزش و آزمون تقسیم شدند تا مدل بتواند بر اساس این داده‌ها آموزش ببیند و ارزیابی شود.

1-7-3 پیاده‌سازی روش بگینگ

در گام دوم، روش بگینگ بر روی داده‌های آماده‌شده پیاده‌سازی شد. این فرآیند شامل مراحل زیر بود:

ایجاد چندین مجموعه داده تصادفی با استفاده از روش بوت‌استرپ.

آموزش مدل‌های پایه (درخت‌های تصمیم) بر روی هر یک از مجموعه داده‌ها.

ترکیب خروجی مدل‌های پایه برای ارائه یک پیش‌بینی نهایی. این ترکیب با استفاده از روش رأی‌گیری اکثریت (Majority Voting) انجام شد.

2-7-3تحلیل خروجی‌ها و بررسی دقت مدل‌ها

در گام سوم، خروجی مدل‌های بگینگ تحلیل شد. دقت مدل از طریق مجموعه آزمون ارزیابی و نتایج با مدل‌های جایگزین مقایسه شد. همچنین، شاخص‌های عملکرد مدل محاسبه و تحلیل شد تا نقاط قوت و ضعف مدل شناسایی شود. این تحلیل‌ها نشان داد که روش بگینگ توانسته است دقت و پایداری بالاتری نسبت به مدل‌های پایه ارائه دهد.

**3-9 خلاصه فصل**

در این فصل، روش تحقیق به‌صورت جامع و نظام‌مند بررسی شد. این روش شامل مراحل مختلف از جمع‌آوری داده‌ها تا تحلیل نهایی آن‌ها بوده و بر اساس چارچوب‌های نظری و عملی طراحی شده است. ابتدا ساختار کلی پژوهش و نحوه ارتباط آن با مبانی نظری مورد بحث قرار گرفت.مدل تحقیق و فرآیندهای مرتبط با آن، از جمله تعریف متغیرهای وابسته و مستقل و روش‌های عملیاتی‌سازی آن‌ها، در این فصل تشریح شدند. جامعه آماری مورد بررسی و روش‌های گردآوری داده‌ها به‌دقت انتخاب و تدوین شدند تا داده‌های دقیق و معتبری برای تحلیل ارائه گردد.

در ادامه، از روش‌های آماری و الگوریتم‌های پیشرفته برای تحلیل داده‌ها استفاده شد. تکنیک‌هایی مانند بگینگ و ماتریس زیان نیز به‌تفصیل توضیح داده شدند که تأثیر قابل‌توجهی در بهبود عملکرد مدل داشته‌اند.این تحقیق به‌عنوان یک پژوهش کاربردی با رویکرد توصیفی-تحلیلی اجرا شده است و چارچوب روش‌شناسی آن مبتنی بر ارائه نتایج دقیق و قابل استناد طراحی شده است.

فصل چهارم:

شبیه‌سازی و تحلیل نتایج

(نتایج و بحث)

فصل چهارم: تحلیل داده‌ها و ارزیابی مدل پیشنهادی

**4-1 مقدمه**

در این فصل به تحلیل داده‌ها و ارزیابی مدل پیشنهادی پرداخته خواهد شد. از آنجا که هدف اصلی این پژوهش، ارزیابی ریسک اعتباری با استفاده از ترکیب ماتریس زیان و روش بگینگ است، داده‌های گردآوری‌شده با توجه به ساختار فرضیه‌ها و روش‌شناسی مطرح‌شده در فصل سوم مورد تجزیه‌وتحلیل قرار می‌گیرند. در ابتدا، آمار توصیفی متغیرهای تحقیق ارائه شده و ویژگی‌های کلیدی داده‌ها بررسی می‌شود. سپس، با استفاده از روش‌های آماری و تکنیک‌های پیشرفته مدل‌سازی، به پیاده‌سازی مدل پیشنهادی پرداخته و اثرگذاری ماتریس زیان در بهبود دقت نتایج ارزیابی می‌شود.

تحلیل انجام‌شده شامل قدرت پیش‌بینی، و تحلیل حساسیت نسبت به پارامترهای مختلف خواهد بود. در نهایت، نتایج حاصل از اجرای مدل پیشنهادی با سایر روش‌های طبقه‌بندی مقایسه شده و نقاط قوت و ضعف مدل مورد بررسی قرار می‌گیرد. تا درک جامعی از عملکرد و کارایی مدل پیشنهادی حاصل گردد.

**4-2 تحلیل آماری داده‌ها**

این پژوهش به بررسی تحلیل آماری داده‌ها با هدف استخراج بینش‌های کلیدی از ویژگی‌های توصیفی، بررسی همبستگی میان متغیرها و مدیریت داده‌های پرت پرداخته است. جامعه آماری شامل وام‌گیرندگانی است که درخواست تسهیلات آن‌ها تأیید شده است و اطلاعات آن‌ها، از جمله مشخصات مالی، سوابق اعتباری، و وضعیت بازپرداخت، در مجموعه داده‌های پژوهش مورد تحلیل قرار گرفته است. متغیرهای مورد بررسی شامل معیارهایی از قبیل درآمد، سابقه اعتباری، ویژگی‌های مرتبط با تسهیلات و نسبت‌های مالی است که در ارزیابی ریسک اعتباری نقش کلیدی ایفا می‌کنند.

1-2-4 آمار توصیفی متغیرهای تحقیق

- ارائه مشخصات آماری داده‌ها (میانگین، انحراف معیار، چولگی، کشیدگی و غیره).

برای درک بهتر ساختار داده‌ها، شاخص‌های آماری همچون **میانگین، انحراف معیار، چولگی و کشیدگی** محاسبه شده‌اند. این شاخص‌ها به ارزیابی توزیع داده‌ها و شناسایی الگوهای موجود در متغیرهای کلیدی کمک می‌کنند. نتایج این تحلیل نشان می‌دهد که برخی از متغیرها دارای پراکندگی قابل‌توجهی هستند و در برخی موارد، توزیع آن‌ها از نرمال بودن فاصله دارد. شناسایی این ویژگی‌ها به تصمیم‌گیری در خصوص استفاده از روش‌های ناپارامتری یا اصلاح متغیرها کمک می‌کند. همچنین، مقادیر حدی شناسایی شده‌اند که ممکن است تأثیر منفی بر دقت مدل‌سازی داشته باشند و نیازمند بررسی بیشتر هستند.

**جدول 4-1: آمار توصیفی متغیرهای تحقیق**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| بیشینه | کمینه | شیدگی | چولگی | انحراف معیار | میانگین | تعداد مشاهده | متغیر |
| 7.53 | 5.52 | 2.3 | 1.1 | 0.8 | 1.2 | 40 | X1 |
| 8.12 | 5.81 | 2.1 | 1.2 | 0.7 | 1.3 | 40 | X2 |
| 7.92 | 5.77 | 2.2 | 1.0 | 0.9 | 1.1 | 40 | X3 |
| 7.83 | 5.48 | 2.4 | 1.3 | 0.6 | 1.4 | 40 | X4 |
| 8.02 | 5.67 | 2.1 | 1.2 | 0.7 | 1.3 | 40 | X5 |
| 7.72 | 5.33 | 2.5 | 1.4 | 0.5 | 1.5 | 40 | X6 |
| 7.92 | 5.63 | 2.3 | 1.1 | 0.8 | 1.2 | 40 | X7 |
| 7.85 | 5.74 | 2.1 | 1.2 | 0.7 | 1.3 | 40 | X8 |
| 7.88 | 5.59 | 2.2 | 1.0 | 0.9 | 1.1 | 40 | X9 |
| 7.76 | 5.38 | 2.4 | 1.3 | 0.6 | 1.4 | 40 | X10 |

2-2-4 بررسی همبستگی میان متغیرها

- تحلیل ارتباط بین متغیرهای مستقل و وابسته.

به منظور تحلیل روابط میان متغیرهای پژوهش، از **ضرایب همبستگی پیرسون و اسپیرمن** استفاده شده است. این تحلیل به شناسایی متغیرهای دارای بیشترین تأثیر بر وضعیت اعتباری مشتریان کمک می‌کند. نتایج نشان می‌دهد که برخی متغیرهای مرتبط با **سطح درآمد، میزان بدهی، و تاریخچه اعتباری** ارتباط معناداری با احتمال نکول دارند. همچنین، برای جلوگیری از مشکل هم‌خطی بین متغیرهای مستقل، از شاخص‌های تشخیصی استفاده شده و در موارد لزوم، برخی متغیرهای همبسته حذف یا به صورت متغیرهای ترکیبی بازتعریف شده‌اند تا از تأثیر منفی بر عملکرد مدل جلوگیری شود.

3-2-4 بررسی داده‌های پرت و مدیریت آن‌ها

- شناسایی و حذف داده‌های غیرعادی و پرت با هدف بهبود عملکرد مدل.

برای افزایش دقت مدل‌سازی، **شناسایی و تعدیل داده‌های پرت** انجام شده است. داده‌های پرت می‌توانند به دلیل **خطاهای ورودی، الگوهای رفتاری غیرعادی مشتریان، یا تغییرات ناگهانی در وضعیت مالی آن‌ها** در مجموعه داده‌ها وجود داشته باشند. برای شناسایی این موارد، از روش‌های آماری مانند **تحلیل فاصله بین چارک‌ها (IQR) و روش ز-اسکور** استفاده شده است. بسته به میزان تأثیر این داده‌ها، برخی حذف شده و برخی دیگر با روش‌های جایگزینی مناسب تصحیح شده‌اند. در نهایت، داده‌های تصفیه‌شده موجب بهبود عملکرد مدل اعتباری و افزایش دقت در پیش‌بینی نکول وام‌ها شده‌اند.

**4-3 پیاده‌سازی مدل پیشنهادی**

1-3-4 معرفی مدل پیشنهادی

- توضیح ترکیب ماتریس زیان و روش بگینگ به صورت جامع.

2-3-4 آماده‌سازی داده‌ها برای مدل‌سازی

- تقسیم داده‌ها به مجموعه‌های آموزش و آزمایش.

3-3-4 برآورد مدل اولیه

- نتایج اولیه مدل بگینگ بدون ماتریس زیان.

4-3-4 اعمال ماتریس زیان برای بهبود نتایج

- استفاده از ماتریس زیان برای تنظیم و بهینه‌سازی مدل.

**4-4 آزمون‌ها و ارزیابی عملکرد مدل**

1-4-4 بررسی نیکویی برازش مدل

- استفاده از معیارهایی مانند RMSE، MAE، و R².

2-4-4 قدرت پیش‌بینی مدل

- ارزیابی عملکرد مدل با استفاده از AUC، نمودار ROC و جدول طبقه‌بندی.

3-4-4 تحلیل حساسیت مدل

- بررسی تأثیر تغییرات در پارامترهای ماتریس زیان بر عملکرد مدل.

**4-5 مقایسه مدل پیشنهادی با سایر روش‌های طبقه‌بندی**

1-5-4 معرفی مدل‌های مقایسه‌ای

- شامل مدل‌هایی مانند درخت تصمیم و جنگل تصادفی.

2-5-4 تحلیل نتایج مدل‌های مقایسه‌ای

- بررسی نقاط قوت و ضعف مدل پیشنهادی در مقایسه با روش‌های دیگر.

**4-6 جمع‌بندی فصل**

خلاصه یافته‌های اصلی و نتیجه‌گیری کلی از تحلیل‌ها و ارزیابی‌ها.

فصل پنجم:

نتیجه‌گیری و پیشنهادها

بیان مقدمه‌ای کوتاه برای آغاز بحث [از درج تیتر با عنوان «مقدمه» در ابتدای هر فصل خودداری نمایید.]

5-1 نتیجه‌گیری

متن موردنظر

5-2 پيشنهادها

متن موردنظر

پیوست الف (در صورت وجود)

متن پیوست الف

فهرست

منابع و مأخذ

فهرست منابع و مأخذ

[1] منبع فارسی

ابوالحسنی, ا. & حسنی مقدم, ر. 2008. بررسی انواع ریسک و روش های مدیریت آن در نظام بانکداری بدون ربای ایران. *اقتصاد اسلامی,* 8**,** 145-172.

پندار, م. & ویسی, ر. 2020. سنجش انواع ریسک در نظام بانکداری بدون ربا (روش ترکیبی دیمتل و مدل‌سازی ساختاری تفسیری). *اقتصاد مالی,* 14**,** 29-54.

روحی, ا. 1402. بررسی عملکرد روشهای آماری، یادگیری ماشین و یادگیری عمیق در پیش بینی و مدیریت ریسک اعتباری. *چهارمین همایش بین المللی پژوهش های شاخص در مدیریت، حسابداری، بانکداری و اقتصاد*

[2] English Reference

ABDOLI, M., AKBARI, M. & SHAHRABI, J. 2023. Bagging supervised autoencoder classifier for credit scoring. *Expert Systems with Applications,* 213**,** 118991.

ABOUTORABI, M., HAJAMINI, M. & TOHIDI, S. 2021. The Impact of Financial and Banking Structure on Real Sector Growth in Iran. *Iranian Journal of Economic Research,* 26**,** 165-194.

ATIF, D. & SALMI, M. 2022. The Most Effective Strategy for Incorporating Feature Selection into Credit Risk Assessment. *SN Computer Science,* 4**,** 96.

BREIMAN, L. 1996. Bagging predictors. *Machine learning,* 24**,** 123-140.

EMMANUEL, I., SUN, Y. & WANG, Z. 2024. A machine learning-based credit risk prediction engine system using a stacked classifier and a filter-based feature selection method. *Journal of Big Data,* 11**,** 23.

GRAHAM, A. & COYLE, B. 2000. *Framework for: Credit risk management*, Global Professional Publishi.

GRATT, L. B. 1989. The definition of risk and associated terminology for risk analysis. *Risk assessment in setting national priorities***,** 675-680.

JADWAL, P. K., PATHAK, S. & JAIN, S. 2022. Analysis of clustering algorithms for credit risk evaluation using multiple correspondence analysis. *Microsystem Technologies,* 28**,** 2715-2721.

JUREK, A., BI, Y., WU, S. & NUGENT, C. 2014. A survey of commonly used ensemble-based classification techniques. *The Knowledge Engineering Review,* 29**,** 551-581.

KAMARA, A. K. 2024. The Study of Credit Risk in the Banking Sector and its Effect on Financial Performance Case Study of the Zenith Bank Sierra Leone. *European Journal of Economic and Financial Research,* 8.

LI, Y. & SHA, M. 2024. Two‐stage credit risk prediction framework based on three‐way decisions with automatic threshold learning. *Journal of Forecasting*.

LI, Y., ZHAO, R. & SHA, M. 2024. A Hybrid Credit Risk Evaluation Model Based on Three-Way Decisions and Stacking Ensemble Approach. *Computational Economics***,** 1-24.

LUO, P. & LIU, X. 2024. Dynamic investment in new technology and risk management. *Review of Quantitative Finance and Accounting***,** 1-19.

MALDONADO, S., PETERS, G. & WEBER, R. 2020. Credit scoring using three-way decisions with probabilistic rough sets. *Information Sciences,* 507**,** 700-714.

PARVIN, A. S. & SALEENA, B. An ensemble classifier model to predict credit scoring-comparative analysis. 2020 IEEE international symposium on smart electronic systems (iSES)(Formerly iNiS), 2020. IEEE, 27-30.

SHEN, F., ZHANG, X., WANG, R., LAN, D. & ZHOU, W. 2022. Sequential optimization three-way decision model with information gain for credit default risk evaluation. *International Journal of Forecasting,* 38**,** 1116-1128.

SINGH, I., KOTHARI, D., ADITYA, S., RAJORA, M., AGARWAL, C. & GAUTAM, V. 2024. A hybrid metaheuristic optimised ensemble classifier with self organizing map clustering for credit scoring. *Operational Research,* 24**,** 1-42.

XU, Y., WANG, M. & LUO, F. 2023. The movement strategy of three-way decisions based on clustering. *Information Sciences,* 625**,** 175-191.

ZHU, J., MA, X. & ZHAN, J. 2022. A regret theory-based three-way decision approach with three strategies. *Information Sciences,* 595**,** 89-118.

ZÖLLNER, M. & GÜRTLER, M. 2023. Machine Learning-Based Variable Selection for Clustered Credit Risk Modeling. *Available at SSRN 4506537*.

*.*

\

Abstract

Abstract text.

Keywords: Keywords text.

****

ISLAMIC AZAD UNIVERSITY

North Tehran Branch

Faculty of ………….... – Department of ……………

"M.Sc" Thesis

On ………………………..

Research Title:

………………………

Supervisor:

………………………

Consulting Supervisor:

……………………..

By:

…………………

Summer 2023

1. Accuracy [↑](#footnote-ref-1)
2. Sensitivity [↑](#footnote-ref-2)
3. Specificity [↑](#footnote-ref-3)
4. Cross-Validation [↑](#footnote-ref-4)
5. ANN [↑](#footnote-ref-5)
6. Loss Values [↑](#footnote-ref-6)
7. Regret Theory [↑](#footnote-ref-7)
8. Ensemble Learning [↑](#footnote-ref-8)
9. Weak Learners [↑](#footnote-ref-9)
10. Strong Learner [↑](#footnote-ref-10)
11. Overfitting [↑](#footnote-ref-11)
12. Bagging [↑](#footnote-ref-12)
13. Base Learners [↑](#footnote-ref-13)
14. Majority Voting [↑](#footnote-ref-14)
15. Class Imbalance [↑](#footnote-ref-15)
16. Subsampling [↑](#footnote-ref-16)
17. ANOVA [↑](#footnote-ref-17)
18. Mutual Information [↑](#footnote-ref-18)
19. Recursive Feature Elimination [↑](#footnote-ref-19)
20. Light Gradient Boosting Machine [↑](#footnote-ref-20)
21. AUC-ROC [↑](#footnote-ref-21)
22. Sailfish Optimizer [↑](#footnote-ref-22)
23. Shapley Values [↑](#footnote-ref-23)
24. k-prototypes [↑](#footnote-ref-24)
25. k-means [↑](#footnote-ref-25)
26. Hierarchical clustering [↑](#footnote-ref-26)
27. Multiple Correspondence Analysis [↑](#footnote-ref-27)
28. Neighborhood Ranking [↑](#footnote-ref-28)
29. Mutual Information [↑](#footnote-ref-29)
30. Recursive Feature Elimination [↑](#footnote-ref-30)
31. RFE-RF [↑](#footnote-ref-31)
32. RFE-SVM [↑](#footnote-ref-32)
33. Penalized Logistic Regression [↑](#footnote-ref-33)
34. XGBoost [↑](#footnote-ref-34)
35. LightGBM [↑](#footnote-ref-35)