

Improvement payment loan concerning financial discipline and maximum gain in uncertainly

 <http://orcid.org/0000-0002-7520-7482> ¹

 <http://orcid.org/0000-0003-4327-8320> ²

 <http://orcid.org/0000-0002-4420-3059> ³

 <http://orcid.org/0000-0001-7048-9065> ⁴

1. Roya Cheshmikhani
2. Mohammad Ali Afshar Kazemi
3. Abbas Toluie Ashlaghi
4. Ezatollah Asgharizadeh

Article Info

Article type:

Research Article

Keywords:

Facilities,
Data Mining,
Deep learning,
Clustering,
CNN,
CNN-LSTM.

ABSTRACT

Granting facilities is an important part of every bank's operations. This part of banking activities is economically important. With its operations, banks can provide the transfer of resources from those that have directly invested to those who need money, this repayment makes another people use these resources. Failure to repay facilities on time will cause the bank's resources to stagnate and in the long will cause the country's economic recession. It is important to monitor the correct allocation of resources because if bank resources are used and unfounded payments are made, the banks will not be able to pay the depositors and will become bankrupt. By examining the researches in the field of banking, it was found that most of the researches have focused on the optimal combination of the investment portfolio in the capital market, and less researchers have paid attention to the discussion of the optimal combination in the money market. Most of the researches that used operational and statistical research methods were related to industrial issues, and less in financial issues and operational research discussions were used. In the optimal portfolio of the combination of facilities or investments, mostly the genetic algorithm has been used and less than other fuzzy methods have been used in the conditions of uncertainty. Therefore, in this research, the improvement of facility payment modeling has been addressed by using convolutional neural networks and CNN-LSTM modeling.

بهبود پرداخت تسهیلات با در نظر گرفتن انضباط مالی و حداکثر سودآوری در

شرایط عدم اطمینان

رویا چشمی خانی^۱، محمد علی افشار کاظمی^۲، عباس طلوعی اشلقی^۳، عزت الله اصغری زاده^۴

چکیده

اعطای تسهیلات، بخش مهمی از عملیات هر بانک را تشکیل می دهد و این قسمت از فعالیتهای بانکی از لحاظ اقتصادی حائز اهمیت است. بانک ها با عملیات اعتباری خود موجبات انتقال منابع را از اشخاصی که مستقیماً در بانک سرمایه گذاری کرده اند به کسانی که به پول نیاز دارند فراهم می کنند و این اشخاص با بازپرداخت اقساط خود باعث می شوند گروهی دیگر از مردم بتوانند از این منابع استفاده کنند. عدم بازپرداخت به موقع تسهیلات باعث راکد شدن منابع بانک و در بلندمدت باعث رکود اقتصادی کشور می شود. نظارت بر نحوه ی صحیح تخصیص منابع و بهداشت اعتباری امری بسیار مهم می باشد زیرا در صورت تخصیص نامطلوب منابع بانک و پرداخت بی اصول تسهیلات، بانک ها توان بازپرداخت سپرده های سپرده گذاران را نداشته و با ورشکستگی مواجه خواهند گردید. با بررسی پژوهش های سالهای اخیر در زمینه بانکداری مشخص گردید اکثر پژوهش ها ترکیب بهینه ی سبد سرمایه گذاری بیشتر در بازار سرمایه پرداخته و کمتر پژوهشگری به بحث و بررسی ترکیب بهینه در بازار پولی توجه نموده است. بیشتر پژوهش هایی که از روشهای تحقیق در عملیات و آماری استفاده گردیده در خصوص مسائل صنعتی بوده و کمتر در مسائل مالی از بحث های تحقیق در عملیات استفاده شده است. در پرتفوی بهینه ترکیب تسهیلات یا سرمایه گذاری ها اکثراً از الگوریتم ژنتیک بهره مند شده و کمتر از دیگر روش های فازی در شرایط عدم اطمینان استفاده شده است. بنابراین در این پژوهش به مدلسازی بهبود پرداخت تسهیلات با استفاده از مدلسازی شبکه های عصبی کانولوشنی^۵ و CNN-LSTM^۶ پرداخته شده است.

کلیدواژه ها: تسهیلات، داده کاوی، خوشه بندی، یادگیری عمیق، شبکه عصبی CNN، شبکه عصبی CNN-LSTM

^۱ - دانشجوی دکتری رشته مدیریت صنعتی، گروه مدیریت صنعتی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
roya.khani64@gmail.com

^۲ - استاد تمام رشته مدیریت صنعتی، گروه مدیریت صنعتی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران (نویسنده مسئول)
m_afshrkazemi@iauec.ac.ir

^۳ - استادتمام رشته مدیریت صنعتی، گروه مدیریت صنعتی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
tolouei.research@gmail.com

^۴ - استاد تمام رشته مدیریت صنعتی، گروه مدیریت صنعتی، واحد مدیریت، دانشگاه تهران، تهران، ایران
ASGHARI@UT.AC.IR

^۵ - Convolutional neural network

^۶ - Long-Short Term Memory

مقدمه

به دلیل گسترش مناسب نظام پولی و مالی، عملاً بانک‌ها و شبکه‌ی بانکی به عنوان مهمترین نهاد مالی وظیفه جمع‌آوری و تخصیص منابع اقتصاد را بر عهده گرفته است. بازار سرمایه یکی از ارکان مهم اقتصاد در جامعه است که از آن به عنوان دماسنج اقتصاد یاد می‌شود. بازار سرمایه پلی است که پس‌انداز واحدهای اقتصادی دارای مازاد مانند شرکتها و دولتها را به واحدهای سرمایه‌گذاری که بدان نیازمندند انتقال می‌دهد. در مورد نهادهای موجود در بازار پول نیز این مورد صادق است. بانک‌های تجاری و مؤسسات مالی و اعتباری از یک طرف وجوه را از مردم به عنوان سپرده می‌گیرند و در قبال آن به آنها پاداش پولشان (بهره) می‌دهند. از طرف دیگر این سپرده‌ها توسط این نهادها به دست افرادی که به این وجوه نیاز دارند انتقال خواهد یافت. مکانیزم این انتقال همان فرایند اعطای تسهیلات است که ورودی آن تقاضای مشتریان از بانک‌ها و مؤسسات مالی و اعتباری بوده و خروجی آن اعطای وجوه به آنها است و در این میان باید مدیریت پرتفوی اعتبارات صورت گیرد. در تعیین سیاستهای مربوط به اعطای تسهیلات، مؤسسات مالی باید محورهای اصلی برنامه‌ی اعطای تسهیلات، نحوه تخصیص تسهیلات اعتباری و نیز شیوه مدیریت مجموعه‌ی تسهیلات اعطایی را مورد توجه قرار دهند و پرداخت تسهیلات به بخش‌های مختلف (صنعتی، خدماتی، کشاورزی) بر اساس تأثیرگذاری بر توسعه کسب و کار خرد مورد نظر قرار می‌گیرد. بانک‌ها در اداره کشورها نقش تعیین‌کننده و از مهمترین ابزار رشد و توسعه کشور محسوب می‌شوند. بانک‌های تجاری ایران همواره از مهمترین نهادهای فعال در بازار پول و سرمایه به شمار می‌روند و با توجه به ساختار اقتصادی کشور و دلایلی چون عدم توسعه‌ی بازارهای سرمایه و سایر شبکه‌های غیربانکی، تأمین مالی بخش‌های اقتصادی کشور را بر عهده دارند. امروزه مهم‌ترین دغدغه‌ی بانک‌ها به منظور افزایش سودآوری، پرداخت اصولی تسهیلات به افرادی می‌باشد که ریسک کمتری جهت بازپرداخت اقساط داشته باشند که به ترتیب ریسک اعتباری بانک کاهش یابد و به تبع آن با ریسک نقدینگی و ریسک شهرت کمتری مواجه گردند (Armantier & others, 2015). اکنون بانک‌ها با حجم زیادی از مطالبات غیرجاری مواجه می‌باشند که عمده‌ترین علت آنها پرداخت غیراصولی تسهیلات و عدم رعایت بهداشت اعتباری می‌باشد (Zangeneh, Mohammadi, 2016). این امر باعث بلوکه شدن قسمت عظیمی از پول بانک‌ها و سپرده سپرده‌گذاران دیگر گردیده و باعث افزایش هزینه و کاهش درآمدهای مشاع بانکها گردیده است (Khosroyani, 2022). وجود سیستم مناسب اعطای تسهیلات و ارزیابی فرصت‌های سرمایه‌گذاری و وام‌دهی و اطمینان از وجود رویه‌های مناسب در پذیرش و اعطای تسهیلات موجب جلب اعتماد و کسب شهرت برای مؤسسات مذکور خواهد شد (عسگرزاده, 1386). با طراحی و اجرای تکنیک‌های بهینه‌سازی، مدیریت امور نهاد مالی می‌تواند با توجه به سیاستهای کلان اعتباری و برنامه‌های اعتباری استراتژیک نهاد مالی و در نظر داشتن اولویت‌های تسهیلات از منظر ریسک و بازده به تصمیمات مناسب دست یابد.

در این پژوهش سعی گردیده مدلی جهت بهینه‌سازی پرداخت تسهیلات ارائه گردد بطوریکه با استفاده از این مدل موسسه اعتباری قادر به اعتبارسنجی صحیح و سریع مشتریان درخواست‌کننده تسهیلات بوده و با شناسایی مشتریان اعتباری میزان ریسک و سودآوری مشتری برای مؤسسه قبل از پرداخت تسهیلات مشخص می‌گردد و بانکها و مؤسسات مالی اطمینان می‌یابند که با پرداخت تسهیلات سپرده‌های سپرده‌گذاران دیگر سوخت نمی‌گردد و بانک‌ها متضرر نمی‌شوند در این پژوهش ابتدا ادبیات نظری پژوهش مورد بررسی و تحلیل قرار گرفته و هم زمان با آن، تحلیل جامع مرتبط با موضوع در نظام بانکی نیز بررسی شد سپس با استناد به نتایج به دست آمده از ادبیات نظری و تحلیل‌های نظام بانکی و با توجه به یافته‌های موردانتظار و با استفاده از ابزارهای شبیه‌سازی و داده‌کاوی اقدام به مدل‌سازی برای خوشه‌بندی مشتریان یکی از بانک‌های خصوصی گردید و سپس با استفاده از یادگیری عمیق و تلفیق روش‌های دسته‌بندی فازی از جمله شبکه‌های عصبی کانولوشنی و شبکه عصبی CNN-LSTM به پیش‌بینی رفتار مشتریان در حوزه پرداخت تسهیلات در شبکه بانکی کشور پرداخته شد.

یک شبکه عصبی کانولوشن شامل یک ورودی و یک لایه خروجی و همچنین چندین لایه پنهان است. شبکه عصبی کانولوشن نشان می‌دهد که شبکه از یک عملیات ریاضی به نام کانولوشن بهره می‌برد کانولوشن نوعی تخصصی از عملیات خطی است (Khalila & others, 2021). شبکه‌های کانولوشن مانند شبکه‌های عصبی، ساده هستند که حداقل در یکی از لایه‌های خود از کانولوشن به جای ضرب ماتریس عمومی بهره می‌برند. همچنین معماری CNN-

LSTM شامل CNN برای استخراج ویژگی و LSTM برای یادگیری وابستگی های دراز مدت است (دامی، خیری، ۱۳۹۷). این مدل می تواند توانایی های شبکه CNN و LSTM را مهار کند و نتایج بهتری با دقت ۸۵٪ به دست آورد.

۱. مرورادبیات

مفهوم تشکیل سبد تسهیلات و تشکیل پرتفوی دارایی ها از دهه ۱۹۵۰ توسط مارکوینز مطرح شده اما از دهه ۱۹۸۰ کاربرد این مفهوم برای دارایی های اعتباری توسعه یافت و به دنبال آن در دهه ۱۹۹۰ مدل های سبد تسهیلات اعتباری معرفی شدند. نظریه مدرن پرتفولیو بر اساس دو فرض بنا نهاده شده است، فرض اول اینکه سرمایه گذاران ریسک گریز هستند و فرض دوم این است که بازده اوراق بهادار دارای توزیع مشترک نرمال هستند، نتیجه این مفروضات آن است که بازده موردانتظار و انحراف استاندارد به طور کامل توزیع بازده اوراق بهادار را توصیف می کند. در مورد سهام، علاقمندی به بازده و در مورد تسهیلات و دارایی های اعتباری علاقمندی به شناسایی زیان های مود انتظار تسهیلات وجود دارد. عمده تحقیقات موجود بر روی بهینه سازی سبدهای سرمایه گذاری بر مبنای الگوها و روش های مختلف انجام گرفته که در غالب این مدل ها بهینه سازی بر مبنای معیارهای مبتنی بر میانگین، واریانس انجام گرفته است (راعی، فرهادی، شیروانی، ۱۳۹۰). در اکثر پژوهش ها به بررسی غیر خطی تصمیمات سرمایه گذاران پرداخته شده است.

۲. پیشینه پژوهش

خلاصه ای از تحقیقات پیشین در خصوص ترکیب بهینه تسهیلات بشرح جدول ذیل می باشد:

جدول (۱) خلاصه تحقیقات پیشین

نام محقق	سال تحقیق	روش تحقیق	ورودی ها	خروجی ها
عسگر زاده	۱۳۸۵	فنون تحقیق در عملیات	تسهیلات	سود
ابزاری، کتابی و عباسی	۱۳۸۴	تحقیق در عملیات	سرمایه گذاری	ریسک
خالدزاده و امیری	۱۳۸۹	الگوریتم ژنتیک	سرمایه گذاری در بورس	ریسک
عبدالعلی زاده شهیر و عشقی	۱۳۹۰	الگوریتم ژنتیک	سرمایه گذاری در بورس	ریسک
نوید، نجومی و میرزاده	۱۳۹۳	الگوریتم ژنتیک	سرمایه گذاری در بورس	ریسک
استیفن لی و پیتز بارنه	۱۹۹۴	شبه سازی	سهام	سود و ریسک
او، کیم و مین	۲۰۰۵	الگوریتم ژنتیک	سرمایه گذاری در صندوق های مشاع	سود
چانگ، یانگ	۲۰۰۹	روشهای آماری	سرمایه گذاری	سود و ریسک
تاتنیا هوبنو و	۲۰۱۴	شبه سازی	تسهیلات	ریسک و بازده
یاسمین بشیر مالیم	۲۰۱۷	شبه سازی	تسهیلات	سود و ریسک

در جدول شماره ۲ محدودیت ها و قابلیت های (نقاط قوت و ضعف) هریک از روش های مورد مطالعه در حوزه های کلان در ادبیات موضوع در مقایسه با روش تحقیق این پژوهش به صورت مختصر ارائه گردیده است:

بهبود پرداخت تسهیلات با در نظر گرفتن انضباط مالی و حداکثر سودآوری در شرایط عدم اطمینان

جدول (۲) نقاط قوت و ضعف روش‌های بررسی شده

روش	نقاط ضعف	نقاط قوت
سیستم های خبره	عدم ثبات، درونی و ذهنی بودن، پرهزینه و زمان بر بودن	می توانند نتایج خود را توضیح دهند، یعنی قدم به قدم یا به طور متوالی منطقی بیاورند.
روش های سنتی آماری	به دلیل الزامات محدودکننده، کاربردهای محدودی دارند.	در محدوده داده های کوچک یا بزرگ، جواب های قابل قبول ارائه می دهند.
روش های داده کاوی	با وجود اینکه به آشکارسازی الگوها و روابط کمک می کند اما اطلاعاتی را درباره ی ارزش یا میزان اهمیت آنها به دست نمی دهد. با وجود توانایی شناسایی روابط بین رفتارها و یا متغیرها لزوماً قادر به کشف روابط علت و معلولی نیست. موفقیت داده کاوی در گرو بهره گیری از کارشناسان فنی و تحلیل گران کار آزموده ای است که از توانایی کافی برخوردار هستند.	به گونه ای گسترش یافته اند که به سادگی می توان آنها را با ابزارهای نرم افزاری و... امروزی تطبیق داده و از اطلاعات جمع آوری شده بهترین بهره را برد.
مدل های رگرسیونی	نیازمند ورودی ها بصورت مقادیر عددی هستند متغیرهای دسته بندی شده باید بصورت متغیرهای عددی تبدیل گردند. هم خطی بین متغیرهای مستقل ورودی، اثر منفی روی نتایج می گذارد.	بسیار پرکاربرد در تدوین نمره دهی اعتباری مشتریان، بی نیازی به فرضیات محدودکننده آماری، در رابطه با متغیرها می توان اهمیت نسبی متغیرهای مربوطه را به دست آورد.
تحلیل ممیزی خطی	نیازمند ورودی ها بصورت مقادیر عددی، فرض توزیع نرمال برای متغیرها و نیز برابر بودن ماتریس کوواریانس برای تمامی گروه ها که در واقع مهمترین نقد وارده بر این روش می باشد.	اگر به دنبال روشی باشیم که ترکیب خطی متغیرها را به گونه ای به دست دهد که هم زمان معیار تفکیک خاصی حداکثر شود، این روش پرکاربرد است. در برخی موارد دیده شده نقض شدن فرض نرمال در داده ها، مانع کارآمدی آن نشده است.
برنامه ریزی خطی	به منظور استفاده از این مدل باید یک حد مرزی تعیین کرد که با توجه به آن	از برنامه ریزی خطی حتی هنگامی که دو گروه به صورت خطی قابل جداسازی

نیستند با استفاده از حداقل کردن حداکثر خطاها یا جمع خطاهای مطلق، می‌توان استفاده کرد.	متقاضیان اعتباری به گروه‌های مختلف تقسیم شوند که این ممکن است به راحتی قابل حصول نباشد.	
می‌توان با فرضیاتی ساده، دانش را از داده‌ها استخراج کرد. خروجی می‌تواند هم امتیاز پیوسته و هم عضویت در کلاس‌ها باشد (Huang, ۲۰۱۸). ماهیت غیرپارامتریک این روش، مدلسازی موارد بی قاعده در تابع ریسک در فضای مشخصات مشتریان را ممکن می‌سازد. این فرآیند کاملاً شهودی بوده و به راحتی قابل توضیح به مدیران است، زیرا به لحاظ مفاهیم و به کارگیری ساده می‌باشد. قابل به کارگیری بصورت پویا و گام به گام بوده و با ورود موارد جدید، می‌تواند خودش را روزآمد کند.	عملکرد این روش در مقایسه با سایر مدل‌های مرسوم، وابسته به تعداد وام‌های بد (تعداد نکول) در داده‌هاست. معمولاً هم احتمال بودن وام‌های خوب و بد در مجموعه یادگیری، کارایی این مدل را بهتر می‌کند؛ که عملاً این گونه نیست.	یادگیری بر مبنای مورد
می‌توانند با فرضیاتی ساده، دانش را از دل داده‌ها استخراج کنند خروجی می‌تواند هم به صورت امتیازات پیوسته و هم به صورت عضویت در کلاس‌ها باشد. روی ورودی‌ها و خروجی‌های غیرخطی هم کار می‌کند که در تحلیل‌های مالی که اغلب متغیرها غیرخطی هستند بسیار مفید است. مناسب برای شرایطی که ساختار مسئله غیرشفاف است و باید نوعی روندیابی یا بازشناسی الگو صورت بگیرد و داده‌های موجود ناقص و تحریف شده‌اند که مدیران مؤسسات مالی و نهادهای سرمایه‌گذاری عملاً با بسیاری از این مسائل مواجه‌اند.	نیازمند ورودی‌ها بصورت مقادیر عددی متغیرهای ورودی ممکن است مقیاس‌های متفاوتی داشته باشند و اثر آنها بر خروجی، به طور نامساوی سنجیده شود، که در این صورت مدل دچار خطا می‌شود. بنابراین مقادیر هر ورودی نیازمند هم مقیاس شدن می‌باشند (اعدادی بین صفر و یک). مشکل بودن فرآیند آموزش در آن که تا حد زیادی کاربرد آن را محدود کرده است. نمی‌توانند نتایج خود را توضیح دهند، یعنی قدم به قدم یا به طور متوالی منطبق بیاورند.	شبکه‌های عصبی

روش خوشه‌بندی k-means	جواب نهایی به انتخاب خوشه‌های اولیه وابستگی دارد. روالی مشخص برای محاسبه اولیه مراکز خوشه‌ها وجود ندارد. اگر در تکراری از الگوریتم تعداد داده‌های متعلق به خوشه ای صفر شد راهی برای تغییر و بهبود ادامه روش وجود ندارد. در این روش فرض می‌شود که تعداد خوشه‌ها از ابتدا مشخص است. اما معمولاً در کاربردهای زیادی تعداد خوشه‌ها مشخص نمی‌باشد.	سادگی روش و کاربرد بودن آن به جهت مقایسه و اعتبارسنجی سایر روش‌های مورد استفاده کم هزینه و کم زمان بودن از آنجا که خوشه‌بندی یک روش یادگیری بدون نظارت محسوب می‌گردد، در موارد بسیاری می‌تواند کاربرد داشته باشد. (Ansari, Riasi 2016.)
ماشین‌های بردار پشتیبان	نیاز به انتخاب تابع کرنل مناسب محدود بودن به دو صفحه خطی و نیاز به محاسبات زیاد، که البته ماشین‌های بردار پشتیبان تقریبی این مشکل را حل کرده‌اند.	SVM ها علی‌الخصوص زمانی که مجموعه داده‌ها بزرگ هستند کاربرد دارند. این مزیت از آنجا ناشی می‌شود که تنها یک زیرمجموعه از داده‌های آموزش (که به آنها بردارهای پشتیبان می‌گویند) برای تعیین تابع تصمیم استفاده می‌شود، در حالیکه می‌توان سایر نقاط را در نظر نگرفت. آموزش نسبتاً آسان، با تعداد کمی داده‌های آموزش به خوبی عمل می‌کند. برای داده‌های با ابعاد زیاد نسبتاً خوب عمل می‌کند.
UTADIS	تخمین توابع ارزش حاشیه‌ای، وزن‌های مشخصه‌ها و مقدار برش با استفاده از تکنیک‌های برنامه‌ریزی خطی محاسبه می‌شوند که می‌تواند شامل معایب تکنیک‌های برنامه‌ریزی خطی گردد.	اجازه مدل‌سازی روابط داده‌های غیرخطی را می‌دهد. مدل ساخته شده می‌تواند متغیرهای کیفی را نیز شامل شود.

در مورد اینکه کدامیک از این روش‌ها مناسب‌تر است باید اشاره کرد که عموماً هر محقق مدل خود را بهترین می‌داند. اما در تحقیقات گوناگون، محققین مختلف بین این روش‌ها مقایساتی انجام داده‌اند که بر اساس آنها نتایج کاملاً متناقضی حاصل شده است. شاید یکی از مهمترین دلایل حصول به نتایج متناقض، تفاوت زمان و جامعه آماری است. با این وجود باز هم برخی تلاش می‌کنند مدل‌هایی برای کاربرد عام طراحی کنند، به گونه‌ای که مدل در یک جامعه آماری طراحی و در جامعه آماری دیگری پیاده سازی شود. اگرچه هر کدام از این مدل‌ها بر جنبه‌ای خاص تکیه می‌کنند، ولی تلاش شده از تحقیقاتی استفاده گردد که در مفروضات خود از داده‌های مالی استفاده کرده و نیز از اطلاعات موجود در بازارهای مالی به عنوان داده استفاده می‌کنند.

شایان ذکر است پژوهش حاضر توانسته رفتار مشتریان بانک را بر اساس شاخص های مالی و غیر مالی خوشه بندی نماید و نتایج حاصل از خوشه بندی را در مدل ریاضی تخصیص خدمات و محصولات بگونه ای بکار گرفته تا با ورود مشتریان اهداف از پیش تعیین شده مورد نظر بانک در مناسب ترین حالت خود قرار گیرند. با توجه آنکه جواب های مؤثر^۱ یا راه حل های غیر مسلط^۲ هر کدام از اهداف این تحقیق ممکن است جواب غیر مؤثر (مسلط) هدف دیگری باشند، بنابراین مسئله دارای جواب بهینه مطلق نمی باشد.

۳. شکاف تحقیق و تعریف مسئله

در حال حاضر بانک های تجاری در تحقق رسالت خویش چندان موفق نیستند. بالا بودن ذخایر بانک ها از یکسو و تسهیلات اعطایی سوخت شده و معوقه از سوی دیگر بیانگر عدم توجه کافی به مدیریت ریسک و عدم وجود مدیریت پرتفوی اعتباری در آنهاست (صالحی و اسکندری، ۱۳۹۳). بدیهی است در صورتیکه مدلی برای مدیریت پرتفوی اعتباری طراحی نشود و به عبارتی در پرتفوی تسهیلات اعطایی، تخصیص به درستی صورت نگیرد، بانک در بلندمدت با ورشکستگی مواجه خواهد شد، چرا که بودجه و سرمایه زیادی را به بعضی از تسهیلات اعطایی خود اختصاص می دهد که باعث افزایش ریسک اعتباری و در نتیجه احتمال عدم وصول مطالبات خواهد شد. مسائل و مشکلات مدیریت پرتفوی تسهیلات مهمترین دلیل ورشکستگی یا زیان دهی بانک ها و مؤسسات مالی و اعتباری می باشد (خواجوی و غیوری مقدم، ۱۳۹۲). مدل هایی مثل برنامه ریزی خطی، برنامه ریزی عدد صحیح، برنامه ریزی صفر و یک وجود دارند که می توانند ترکیبی بهینه از عناصر تشکیل دهنده سبد تسهیلات را ارائه دهند اما نمی توان برای رسیدن به چنین هدفی اطلاعات مالی را با در نظر گرفتن تمام شرایط وارد برنامه ریزی ریاضی کرد. بیشترین روش های انجام شده مربوط به چهار مدل مطرح شده با عنوان مدل های پرتفوی اعتباری به سه دسته مدل های ساختاری، مدل های بر مبنای کلان اقتصادی و مدل های مبتنی بر عملکرد گذشته، تقسیم بندی می شوند. مدل های ساختاری از داده ها و اقلام صورت های مالی و تغییرات ارزش دارایی ها در بازار سرمایه استفاده می نمایند. مدل های بر مبنای کلان اقتصادی، ریسک اعتباری را بر مبنای همبستگی و ارتباط صنایع با برخی از متغیرهای کلان اقتصادی محاسبه می کنند. مدل های آماری که همان مدل های مبتنی بر عملکرد گذشته هستند با استفاده از برخی توزیع های آماری، ریسک اعتباری را در بانکها اندازه گیری نموده اند. فلذا بررسی سود و ریسک هریک از مشتریان بانک در صورت اعطای تسهیلات در هیچ یک از مدل ها و روش های پیشین مطرح نبوده است. در صورتیکه در این پژوهش با استفاده از داده کاوی به بهینه سازی مدل چند هدفه تخصیص تسهیلات با استفاده از شبکه عصبی پرداخته شده است.

۴. روش شناسی

۴.۱ ساختار بانک اطلاعاتی سازمان مورد پژوهش

برای بررسی متغیرهای مستقل و تشخیص نوع آنها و تحلیل داده ها از نرم افزار R استفاده شده است. نرم افزار R دارای محیطی شبیه نرم افزار Plus-S بوده و برای انجام هر فعالیت در این نرم افزار باید از دستورها و کدهای مناسب، که اغلب شبیه دستورهای مورد استفاده در نرم افزار Plus-S هستند استفاده کرد. داده ها به طور خلاصه بشرح ذیل می باشند که ۱۶۵۸ سطر و ۱۸ ستون دارا می باشد.

جدول (۳) متغیرهای تحقیق

^۱- Efficient solution

^۲- Non dominated solution

بهبود پرداخت تسهیلات با در نظر گرفتن انضباط مالی و حداکثر سودآوری در شرایط عدم اطمینان

```
> # structure of the data
> str(df)
tibble [1,658 x 18] (S3: tbl_df/tbl/data.frame)
 $ customer_type      : Factor w/ 2 levels "haghighi","hoghughi": 1 1 1 1 1 1 1 1 1 ...
 $ gender              : Factor w/ 3 levels "C","F","M": 3 2 2 3 3 3 3 3 3 ...
 $ loan_type           : Factor w/ 4 levels "morabehe","mosharekat_madani",...: 3 1 1 2 1 2 1 1 1 ...
 $ facility_type       : Factor w/ 2 levels "mobadele1","mosharekat1": 1 1 1 2 1 2 1 1 1 ...
 $ customer_career     : Factor w/ 4 levels "azad","karmand",...: 1 2 2 2 NA 2 2 2 2 ...
 $ customer_income     : Factor w/ 3 levels "daramad_bala",...: 2 3 2 2 NA 2 2 2 2 ...
 $ loan_department     : Factor w/ 5 levels "bazargani","khadamat",...: 1 2 2 2 2 2 2 2 2 ...
 $ collateral_type     : Factor w/ 3 levels "1","2","3": 2 3 3 3 3 3 3 3 3 ...
 $ payment_date        : chr [1:1658] "1394/07/27" "1400/02/05" "1400/06/25" "1398/09/12" ...
 $ current_situation   : Factor w/ 4 levels "faal","mashkukolvusul",...: 2 1 1 1 1 1 1 1 1 ...
 $ period              : num [1:1658] 180 1095 1095 1460 1460 ...
 $ approved_amount     : num [1:1658] 3.5e+08 6.0e+08 1.2e+09 1.0e+09 5.0e+08 1.0e+09 5.0e+08 5.0e+08 2.0e+08 3.0e+08 ...
 $ loan_remaining      : num [1:1658] 3.91e+08 5.21e+08 1.30e+09 5.81e+08 2.98e+08 ...
 $ interest_rate       : num [1:1658] 24 18 18 16 16 16 16 18 18 ...
 $ ghest               : num [1:1658] 62484034 21466922 42933843 28062608 14031304 ...
 $ profit_rate         : num [1:1658] 0.0266 0.0937 0.0937 0.1071 0.1071 ...
 $ income_or_cost      : num [1:1658] 0.9746 0.024 0.9759 0.0237 0.0239 ...
 $ risk                : num [1:1658] 0 0.333 0.639 0.333 0.333 ...
```

همانطور که می بینیم اکثر متغیرها از جنس فاکتور و اعداد می باشند به جز متغیر پرداخت تسهیلات که از جنس کاراکتر می باشد. در زبان برنامه نویسی R، مقدارهای متغیرهای طبقه ای را در یک متغیر عامل یا فاکتور ذخیره می کنند. البته ممکن است چنین مقدارهایی را در متغیرهای پیوسته نیز ذخیره کرد ولی کارایی و مفهوم طبقه ای بودن در چنین حالتی از آنها گرفته می شود.

۴,۲ آماده سازی داده ها

۴,۲,۱ پاک سازی داده^۱

پاک سازی یا تمیز کردن داده ها به فرآیندی جهت تشخیص، حذف و اصلاح داده های نادرست از مجموعه جداول، رکوردها یا بانک های اطلاعاتی، همچنین شناسایی قسمت های ناقص و نادرست داده ها و سپس اصلاح و جایگزینی آنها اشاره دارد. هدف از پاک سازی داده ها استخراج اطلاعات دقیق و درست است، چرا که اطلاعات نادرست می تواند منجر به نتیجه گیری غلط شود. در این مرحله برخی از متغیرهای موجود در مجموعه داده های بدست آمده از بانک مورد نیاز محقق نمی باشد و این داده ها در ادامه فعالیت از مجموعه داده ها حذف گردیدند بطور مثال ستون هایی مانند ردیف، عنوان تسهیلات و ... از داده ها حذف گردیدند.

۴,۲,۲ تبدیل داده^۲

گاهی اوقات برای اینکه دقت تجزیه و تحلیل را بالا ببریم باید در داده های خامی که برای تحلیل در دسترس ما قرار دارند، تغییراتی ایجاد کنیم، یکی از این تغییرات، فرایند تبدیل داده ها است. تبدیل داده ها روش هایی بر پایه ریاضی است که برای متغیرهایی به کار می رود که از شاخص های آماری نرمال بودن، خطی بودن، پراکندگی یکسان و ... پیروی نمی کنند. تبدیل داده نوعی روش تثبیت داده نیز به شمار می رود. در این فاز، داده های انتخاب شده به فرم دیگری تبدیل می شود. در این مرحله متغیرهایی که از جنس فاکتور بودند به کاراکتر تبدیل شدند و نوع وثیقه که کاراکتر می بود به نوع عددی تبدیل شده اند. برای مباحث آمار و برنامه نویسی، متغیر عامل با فاکتور^۳، دارای مقدارهای عددی محدود است. این اعداد و ارقام اصولاً بصورت اعداد صحیح ثبت یا نمایش داده می شوند، از آنجائیکه برخی از این ستون ها فقط بصورت یک متن می باشند با کاراکتر نشان داده شده اند بقیه متغیرها

^۱ - Data Cleaning

^۲ - Data Transformation

^۳ - factor

نیز چون از نوع عددی هستند از نوع عددی^۱ تعریف شده‌اند. قبل از انجام هرگونه عملیات محقق تصمیم گرفته است که متغیرهای کاراکتری به متغیرهای فاکتور با استفاده از تابع as تبدیل گردد. که بشرح ذیل می‌باشد.

جدول (۴) تبدیل داده‌ها

```
tibble [1,658 x 18] (S3: tbl_df/tbl/data.frame)
 $ customer_type      : Factor w/ 2 levels "haghighi","hoghughi": 1 1 1 1 1 1 1 1 1 ...
 $ gender             : Factor w/ 3 levels "C","F","M": 3 2 2 3 3 3 3 3 3 ...
 $ loan_type          : Factor w/ 4 levels "morabehe","mosharekat_madani",...: 3 1 1 2 1 2 1 1 1 ...
 $ facility_type      : Factor w/ 2 levels "mobadelei","mosharekati": 1 1 1 2 1 2 1 1 1 ...
 $ customer_career    : Factor w/ 4 levels "azad","karmand",...: 1 2 2 2 NA 2 2 2 2 ...
 $ customer_income    : Factor w/ 3 levels "daramad_bala",...: 2 3 2 2 NA 2 2 2 2 ...
 $ loan_department    : Factor w/ 5 levels "bazargani","khadamat",...: 1 2 2 2 2 2 2 2 2 ...
 $ collateral_type    : Factor w/ 3 levels "1","2","3": 2 3 3 3 3 3 3 3 3 ...
 $ payment_date       : chr [1:1658] "1394/07/27" "1400/02/05" "1400/06/25" "1398/09/12" ...
 $ current_situation : Factor w/ 4 levels "faal","mashkukolvusul",...: 2 1 1 1 1 1 1 1 1 ...
 $ period             : num [1:1658] 180 1095 1095 1460 1460 ...
 $ approved_amount    : num [1:1658] 3.5e+08 6.0e+08 1.2e+09 1.0e+09 5.0e+08 1.0e+09 5.0e+08 5.0e+08 2.0e+08 3.0e+08 ...
 $ loan_remaining     : num [1:1658] 3.91e+08 5.21e+08 1.30e+09 5.81e+08 2.98e+08 ...
 $ interest_rate      : num [1:1658] 24 18 18 16 16 16 16 18 18 ...
 $ ghest              : num [1:1658] 62484034 21466922 42933843 28062608 14031304 ...
 $ profit_rate        : num [1:1658] 0.0266 0.0937 0.0937 0.1071 0.1071 ...
 $ income_or_cost     : num [1:1658] -1.95e+08 7.44e+06 -9.47e+06 4.49e+07 2.17e+07 ...
 $ risk              : num [1:1658] 100 0 4.21 0 0 ...
```

۳.۴.۲ داده‌های گمشده^۲

مسئله مقادیر گمشده در علم داده^۳ و بخصوص داده کاوی، زمانی رخ می دهد که یک یا چند مشاهده دارای مقدارهایی ثبت نشده یا ناموجود در ستون های «چارچوب اطلاعاتی»^۴ هستند. در چنین حالت می گوئیم آن مشاهده دارای مقدار گمشده یا مقدار ناموجود است. در این حالت این مجموعه داده را دارای «داده گمشده» یا داده های گمشده می گوئیم. به منظور تحلیل های آماری روی مجموعه های اطلاعاتی با داده گمشده، باید مشخص کنیم که نقش چنین مشاهداتی در محاسبات مربوط به تحلیل آماری چگونه است. برای معرفی داده های گمشده در یک بردار از عبارت NA استفاده می کنیم. برای بررسی داده های گمشده در یک بردار از تابع is.na() استفاده می شود. این تابع برداری منطقی به طول بردار موردنظر ارائه می دهد که مقادیر گمشده متناظر با اعضای بردار در آن با TRUE نشان داده شده اند (Bhandari, ۲۰۲۲).

۴.۲.۴ مدیریت داده های گم شده

روش های مختلفی برای مقابله با داده های از دست رفته وجود دارد، شاید راحت ترین روش، حذف نمونه هایی از داده های از دست رفته باشد. ولی در این روش تعداد نمونه ها کاهش پیدا می کند. همانطور که می دانید در داده کاوی هرچقدر تعداد نمونه ها بیشتر باشد مدلسازی با دقت بالاتری انجام می شود. گاهی می توان مقادیر خالی را با استفاده از الگوریتم های مختلف مقداردهی نمود (Howell, ۲۰۲۱). در این پژوهش از روش نزدیکترین همسایه استفاده شده است. الگو کی- نزدیکترین همسایه^۵ یک متد آمار ناپارامتری است که برای طبقه بندی آماری و رگرسیون استفاده می شود (Li, Du, ۲۰۱۹). در هر دو حالت k شامل نزدیکترین مثال آموزشی در فضای داده ای می باشد و خروجی آن بسته به نوع مورد استفاده در طبقه بندی و رگرسیون متغیر است. در حالت طبقه بندی با توجه به مقدار مشخص شده برای کی، به محاسبه فاصله نقطه ای که می خواهیم برچسب آن را مشخص کنیم با نزدیکترین نقاط

^۱ - numeric

^۴ -missing value

^۵ -data science

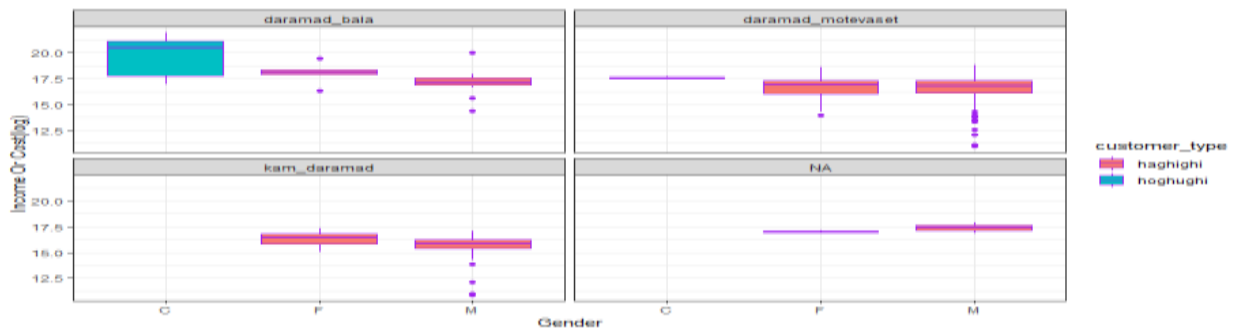
^۶ -data frame

^۵ -K- nearest neighbors algorithm

می‌پردازد و با توجه به تعداد رأی حداکثری این نقاط همسایه، در رابطه با برچسب نقطه موردنظر تصمیم‌گیری می‌کنیم (Huang, ۲۰۱۸). برای محاسبه این فاصله می‌توان از روش‌های مختلفی استفاده کرد که یکی از مطرح‌ترین این روش‌ها، فاصله اقلیدسی است (Estivill, Houle, ۲۰۰۱). در این پژوهش با استفاده از روش نزدیکترین همسایه، مقدار میانگین هر ستون بجای مقادیر گم‌شده جایگزین گردیده است.

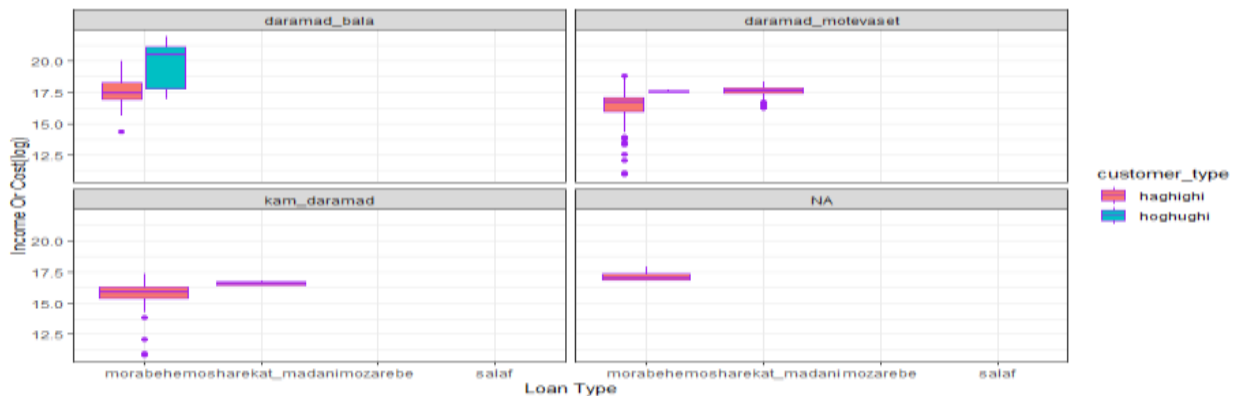
۴,۲,۵ بصری‌سازی متغیرها

جهت تصویرسازی بهتر متغیرها استفاده از نمودار ggplot توصیه می‌گردد، نمودار تأثیر متغیرها روی متغیر وابسته که درآمد - هزینه یا ریسک می‌باشد به شرح زیر و به تصویر کشیده شده‌اند.



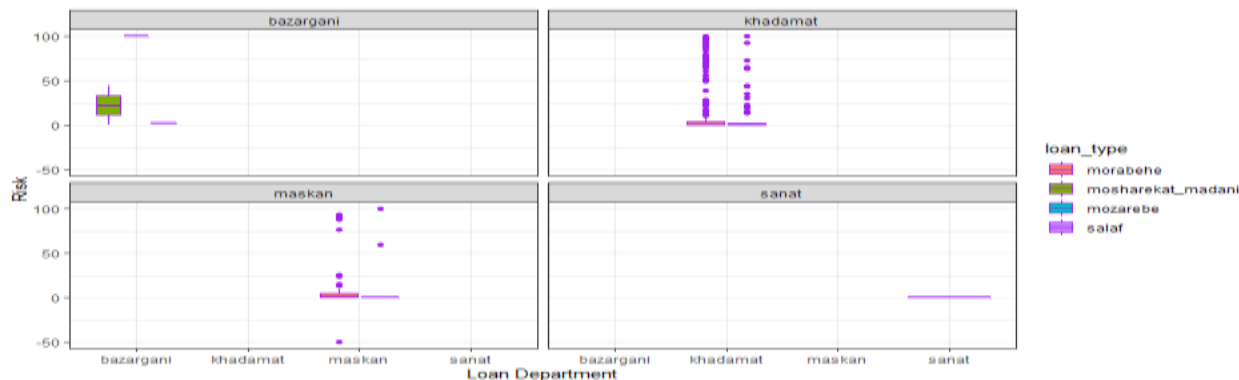
شکل (۱) تأثیر متغیر جنسیت بر متغیر وابسته هزینه/درآمد

در شکل شماره ۱ نشان داده شده است که تسهیلات پرداختی به اشخاص حقیقی با درآمد بالا و درآمد متوسط از نوع مرابحه بیشترین تأثیر را روی درآمد-هزینه بانک داشته است.



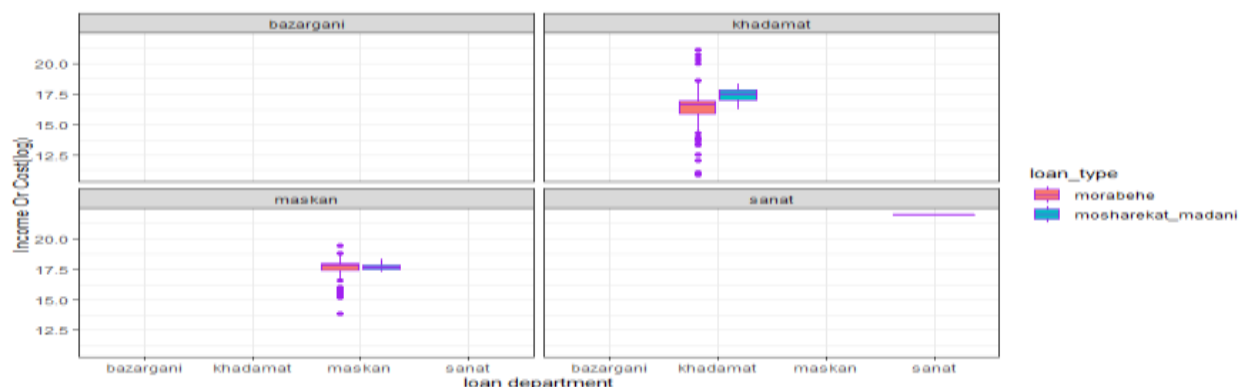
شکل (۲) تأثیر متغیر نوع تسهیلات بر متغیر وابسته هزینه/درآمد

طبق شکل شماره ۲ تسهیلات مرابحه و مشارکت مدنی پرداختی به اشخاص حقوقی بیشترین تأثیر را روی درآمد-هزینه داشته است. در شکل شماره ۳ مشخص می‌باشد که تسهیلات مشارکت مدنی در بخش بازرگانی ریسک بیشتری داشته و مرابحه در بخش مسکن و خدمات به ترتیب ریسک کمتری داشته است.



شکل (۳) تاثیر متغیر بخش تسهیلات بر متغیر وابسته ریسک

همچنان در شکل ۴ مشخص می‌باشد که تسهیلات مراجه در بخش خدمات و مسکن و مشارکت مدنی در بخش خدمات بیشترین تأثیر را روی درآمد-هزینه داشته است.



شکل (۴) تاثیر متغیر بخش تسهیلات بر متغیر وابسته هزینه/درآمد

۶.۲.۴ نرمال سازی داده‌ها

نرمال سازی داده‌ها روشی برای یکنواخت کردن بازه مقادیر مربوط به متغیرهای مختلف پژوهش است و به بی‌مقیاس سازی داده‌ها نیز معروف است. اگر واحد سنجش متغیرهای مورد مطالعه متنوع باشد با استفاده از روش‌های نرمال سازی می‌توان داده‌ها را بی‌مقیاس کرد. مفهوم دیگر از نرمال کردن که به استاندارد کردن نیز موسوم است در تحلیل شبکه‌های عصبی مصنوعی و تحلیل پوششی داده‌ها کاربرد دارد. برای استاندارد کردن یک عنصر باید آن عنصر را منهای میانگین کرده و بر انحراف معیار تقسیم نمایید. استراتژی‌های ماکسیمین، ماکسیماکس و مینیمکس سه رویکرد تصمیم‌گیری تحت عدم اطمینان هستند که بخصوص در مسائل مالی و تصمیم‌گیری‌های اقتصادی، نقش مهمی دارند (Henderi, Wahyuningsih, Rahwanto, ۲۰۲۱). این گونه روش‌های انتخاب و تصمیم، برای اجرای یک فعالیت در استراتژی‌های مالی و بازار سهام نیز قابل استفاده است. برای اطمینان از بکارگیری روش مناسب جهت استاندارد سازی داده‌ها در این تحقیق از روش فازی یا همان روش ماکزیم-مینیم استفاده می‌شود. که این روش برای یک شاخص (X_j) باجنبه مثبت (مانند درآمد هریک از تسهیلات) و یک شاخص با جنبه منفی مانند (ریسک و هزینه هریک از تسهیلات) عبارت است از:

$$n_{ij}^+ = \frac{x_{ij} - x_{jMin}}{x_{jMax} - x_{jMin}}, \quad n_{ij}^- = \frac{x_{jMax} - x_{ij}}{x_{jMax} - x_{jMin}} \quad (1)$$

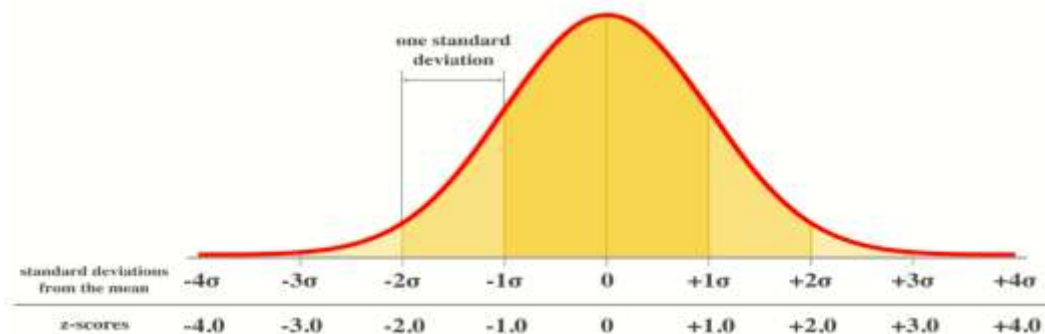
اعداد در این روش نرمال سازی دقیقاً بین صفر و یک خواهند بود بطوریکه فارغ از جنبه مثبت و منفی، صفر برای بدترین نتیجه و یک برای بهترین نتیجه می باشد (اصغرپور، ۱۳۸۸).

۴,۲,۷ حذف داده های پرت با استفاده از رویکرد z-score

یک روش متداول برای شناسایی و حذف داده های پرت استفاده از الگوریتم Z-score می باشد که بر اساس فاصله داده ها از میانگین و استاندارد داده ها محاسبه می شود. برای استفاده از این روش، ابتدا باید میانگین و استاندارد داده های مجموعه داده را محاسبه کرد. سپس برای هر داده، فاصله آن از میانگین را با استفاده از فرمول زیر محاسبه می کنیم:

$$z = (x - \mu) / \sigma \quad (۲)$$

که در این فرمول، X داده مورد نظر، μ میانگین داده ها و σ انحراف معیار داده ها است. زمانی که ما Z-score همه داده ها را محاسبه کردیم، می توانیم به داده های پرت پی ببریم. داده هایی که Z-score بیشتر از یک حداکثر مشخص مثلاً ۳ یا ۴ شده باشد به عنوان داده های پرت شناخته می شوند و می توانیم آنها را از مجموعه داده حذف کنیم. شکل زیر حالت کلی استفاده از این روش می باشد.



شکل (۵) z-score

باتوجه به اینکه تعداد داده های این پژوهش زیاد نمی باشد در حذف داده های پرت محتاط بوده به همین دلیل داده های پرت بزرگتر از ۴ حذف گردیده است.

۴,۲,۸ خوشه بندی متغیرها

خوشه بندی، فرآیندی است که به کمک آن می توان مجموعه ای از اشیاء را به گروه های مجزا افراز کرد. هر افراز یک خوشه نامیده می شود. اعضاء هر خوشه با توجه به ویژگی هایی که دارند به یکدیگر بسیار شبیه هستند و در عوض میزان شباهت بین خوشه ها کمترین مقدار است. در چنین حالتی هدف از خوشه بندی، نسبت دادن برجسب هایی به اشیاء است که نشان دهنده عضویت هر شیء به خوشه است (Bock، ۲۰۰۷). در این پژوهش با استفاده از معیار سیلوئیت تعداد بهینه خوشه با استفاده از فرمول زیر در نرم افزار R-Studio بدست آمده است:

```
avg=function(k){  
  km.res=kmeans(dfn ,centers = k , nstart = 25)  
  ss=silhouette(km.res$cluster, dist(dfn))  
}
```

(۳)

```
return(mean(ss[,3]))
}
k_values=2:6
avg_sil=map_dbl(k_values , avg)
fviz_nbclust(dfn , kmeans , method = "silhouette")
dfn <- df%>%
```

```
dplyr::select(approved_amount,loan_remaining,ghesh,profit_rate,income_or_cost,risk)
```

روش SC^۱ یکی از متداولترین و بهترین روش های اعتبار سنجی الگوریتم خوشه بندی کا-میانه^۲ می باشد. که اولین بار توسط کافمن و روسیو در سال ۱۹۹۰ پیشنهاد شد. که در این رابطه همبستگی $a(x)$ متوسط فاصله (x) نسبت به بردار های دیگر در همان خوشه است، شاخص تفکیک $b(x)$ متوسط فاصله (x) نسبت به سایر خوشه ها را نشان می دهد $S(x)$ نمایش اندازه گیری میزان نزدیکی هر نقطه در یک خوشه است، که امتیاز آنها نسبت به خوشه های مجاورش سنجیده می شود. محدوده امتیاز دهی آن بین ۰ و ۱+ است (Chen, Bose, ۲۰۱۵). به این صورت که اگر این اندازه گیری در محدوده ۱+ باشد نشان دهنده این است که خوشه مورد نظر نسب به خوشه مجاورش بسیار دور است. و حالت صفر نشان می دهد که هیچ گونه تفکیکی بین خوشه مورد نظر و خوشه های همسایه وجود ندارد. و بالاخره حالت ۱- نشان دهنده احتمالات اختصاص خوشه مورد نظر به صورت غلط است. که نتیجه به دست آمده از روش فوق به شکل زیر می باشد:

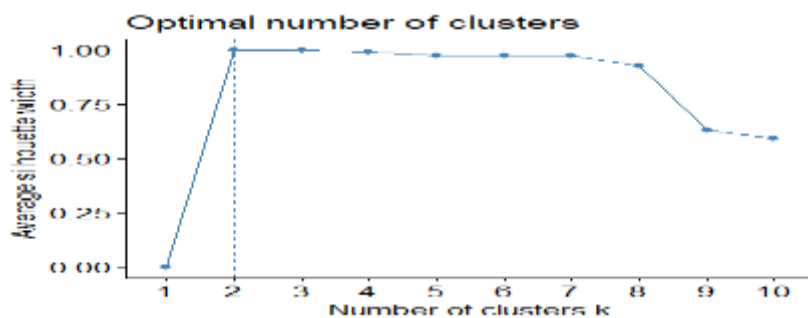
Within cluster sum of squares by cluster

7. 134004e+24 1. 253211e+24

(۴)

(% between_SS / total_SS = 75. 4)

همانگونه که از خروجی مشخص می باشد total-ss میزان ۷۵/۴٪ را به خود اختصاص داده که نشانگر این موضوع است که به درستی خوشه ها انتخاب شده اند. چنانچه از شکل زیر مشخص می باشد بهترین تعداد خوشه دو می باشد.



شکل (۶) تعداد بهینه خوشه

۲،۹ روش خوشه بندی کا-میانه^۳

الگوریتم کا-میانه الگوریتمی مبتنی بر شیء می باشد و نماینده ی خوشه ها را از میان خود داده ها و نه میانگین گیری از آنها انتخاب می کند. در واقع medoids یک خوشه، مرکزی ترین عنصر یک خوشه است. هدف این روش، کم کردن حساسیت نسبت به مقادیر بزرگ در مجموعه داده ها است.

در این الگوریتم هر خوشه با یکی از داده های نزدیک به مرکز معرفی می شود (Guo & others, ۲۰۰۳). نحوه ی محاسبه کامیانه بصورت زیر می باشد:

```
kmedoid. re<-pam(dfn , k=2, metric = "euclidean" , stand = FALSE)
```

(۵)

^۱ – Silhouette Coefficient

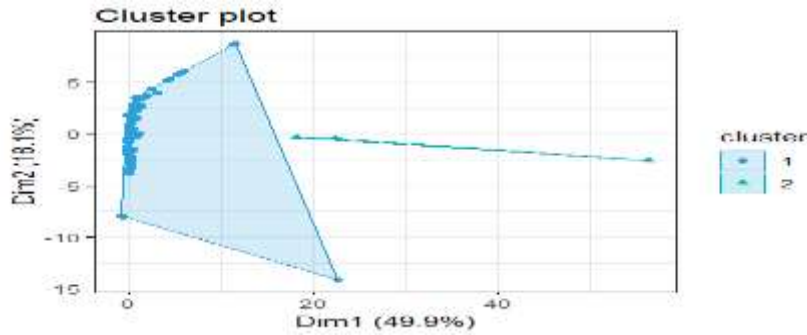
^۲ – K-Means

^۳ – k-medoids

بهبود پرداخت تسهیلات با در نظر گرفتن انضباط مالی و حداکثر سودآوری در شرایط عدم اطمینان

kmedoid. re<-pam(dfn , k=2, metric = "manhattan" , stand = FALSE)

همانطور که مشاهده می شود تعداد بهینه خوشه ۲ بوده و از فاصله اقلیدسی و منتهن برای خوشه بندی استفاده گردیده است که نتیجه هر دو یکسان و به شکل زیر می باشد:



شکل (۷) خوشه بندی کا-میانه

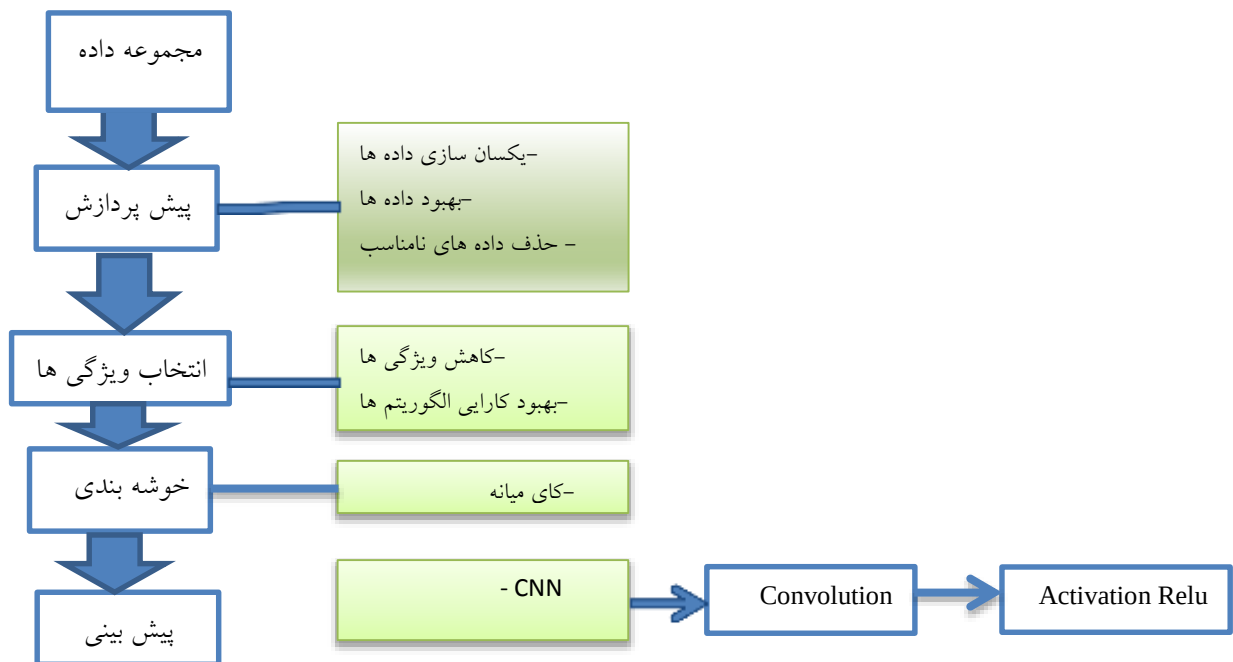
نتیجه خوشه بندی با روش کا-میانه بشرح ذیل می باشد:

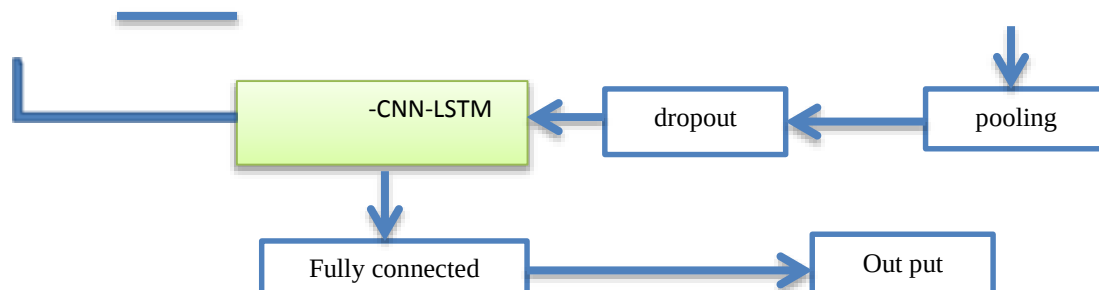
Medoids:

	ID	approved_amount	loan_remaining	ghest	profit_rate	income_or_cost	risk
[1,]	1148	5.000000e+08	4.157526e+08	17889101	0.09369886	7893890	0.00000
[2,]	1424	1.204945e+12	1.600629e+12	93028258643	0.06284257	-24865811247	16.41918

که نشانگر این می باشد که مشتریان در دو گروه مشتریان خوش حساب و مشتریان بد حساب طبقه بندی گردیده اند که مشتریان خوش حساب دارای درآمد مثبت و ریسک صفر بوده و مشتریان بد حساب دارای ریسک ۱۶.۴۲ و درآمد منفی یا بهتر است ذکر کنیم برای بانک دارای هزینه بوده اند.

۵. چارچوب پژوهش





شکل (۸) چارچوب پژوهش

۶. یافته‌های پژوهش

۶.۱ یادگیری عمیق با استفاده از شبکه‌های عصبی

یادگیری عمیق از شبکه‌های عصبی مصنوعی برای انجام محاسبات پیچیده بر روی حجم زیادی از داده‌ها استفاده می‌کند. یادگیری عمیق، نوعی از یادگیری ماشین است (Ngufor & others, ۲۰۱۹)، که بر اساس ساختار و عملکرد مغز انسان کار می‌کند (Adnan, ۲۰۲۰). یادگیری عمیق زیر مجموعه‌ای از هوش مصنوعی است که در آن به جای انسان از ماشین استفاده می‌شود در ابتدا باید بگوییم که مغز انسان از رشته‌های عصبی تشکیل شده است که این عصب‌ها با هم در ارتباط‌اند و پردازش اطلاعات می‌کنند کارهایی که انجام می‌دهیم براساس ورودی‌هایی است که دریافت می‌کنیم و در اصل مغز ما مانند یک تابع رفتار می‌کند که ورودی‌هایی را می‌گیرد و عملیات انجام می‌دهد و خروجی به ما تحویل می‌دهد یادگیری عمیق نیز با استفاده از شبکه عصبی عمیق که مانند مغز انسان عمل می‌کند و پردازش انجام می‌دهد (Al Machot & others, ۲۰۱۵). شبکه عصبی توسط ورودی‌ها آموزش داده می‌شود و شامل سه لایه ورودی و پنهان و خروجی است و هر کدام از عصب‌ها دارای مقدار آستانه و تابع فعال سازی می‌باشند که به ما خروجی می‌دهند نتیجه‌ای که به دست می‌آوریم با خروجی که انتظار داریم مقایسه می‌شود که این دو مقدار باید نزدیک به هم باشند، مدل یاد می‌گیرد که وزن‌ها و مقدار آستانه را طوری تنظیم کند که خروجی درست دریافت کند. لایه‌های شبکه‌ی عصبی، از گره‌ها تشکیل شده‌اند. یک گره، مانند نورون‌های مغز انسان، مکانی برای انجام محاسبات است. مجموعه‌ای از نورون‌های فعال شده منجر به یادگیری می‌شوند (شریف فر، ۱۴۰۱).

۶.۱.۱ شبکه عصبی CNN

CNN از گیرنده‌های چند لایه تشکیل شده است منظور از گیرنده‌های چند لایه شبکه‌های کاملاً متصل هستند، که هر نورون در یک لایه به همه سلول‌های عصبی در لایه بعدی متصل می‌شود اتصال کامل این شبکه‌ها موجب می‌شود که بیش از حد از داده‌ها استفاده کنند CNN ها از الگوی سلسله‌مراتبی در داده‌ها استفاده می‌کنند و با استفاده از الگوهای کوچکتر و ساده الگوهای پیچیده‌تری را بدست می‌آورند (Oquab & others, ۲۰۱۴). ورودی بیشتر مدل‌های یادگیری عمیق مانند شبکه عصبی باید به صورت عدد باشند به همین منظور داده‌های کیفی را باید در مرحله پیش پردازش داده‌ها به صورت عدد تبدیل شوند. در مدل‌های یادگیری عمیق، برای داده‌هایی که به صورت دسته‌ای هستند (مثلاً در این مجموعه داده ایی ویژگی نوع مشتری)، باید قبل از اعمال مدل، آنها را به صورت عددی تبدیل کرد. روش رمزگذاری به صورت برجسب یکی از روش‌های تبدیل داده‌های کیفی به داده‌های عددی می‌باشد (Krizhevsky & others, ۲۰۱۷). در این پژوهش، برای هر دسته یک عدد مخصوص تعیین گردید. به عنوان مثال برای ویژگی نوع مشتری، برای حقیقی، حقوقی به ترتیب می‌توان عدد ۱ و ۲ را نسبت داد. سپس برای هر داده، مقدار عددی مخصوص آن دسته برای آن نمونه انتخاب گردید. همچنین بخش آموزش برای آموزش مدل‌های یادگیری عمیق استفاده می‌شود، بخش آزمایش برای ارزیابی عملکرد مدل‌ها بر روی داده‌های جدید استفاده می‌شود و بخش اعتبارسنجی برای تنظیم پارامترهای مدل و انتخاب بهترین مدل از بین چندین مدل مختلف استفاده می‌شود (Ole Hjelkrem,

Elif de Lange (۲۰۲۳). برای تقسیم مجموعه داده به سه بخش، معمولاً از روش تصادفی استفاده گردیده است. در این تحقیق، ۷۰٪ از داده‌ها به بخش آموزش، ۱۵٪ به بخش آزمایش و ۱۵٪ به بخش اعتبارسنجی تخصیص داده شده است. در این پژوهش مدل به صورت لایه به لایه تعریف شده است. در ابتدا، یک لایه کانولوشن با ۶۴ فیلتر، اندازه کرنل ۳ و تابع فعال‌سازی relu تعریف شده است. سپس یک لایه MaxPooling با اندازه ۲ تعریف شده است که به منظور کاهش ابعاد داده‌ها استفاده شده است. این لایه با انتخاب بزرگترین مقدار در هر دو مقدار داده‌های ورودی، ابعاد داده‌ها را نصف می‌کند. برای جلوگیری از بیش‌برازش، یک لایه Dropout با مقدار ۰/۵ تعریف شده است. این لایه به منظور حذف تصادفی وزن‌های یادگیری مدل استفاده می‌شود و در اینجا مقدار ۰/۵ به این معنی است که در هر دوره ۵۰ درصد از وزن‌های مدل به صورت تصادفی حذف می‌شوند. پس از آن، یک لایه Flatten تعریف شده است که به منظور تبدیل داده‌های ورودی به داده‌های یک بعدی استفاده می‌شود. این لایه به صورت خطی داده‌های ورودی را به یک بعد تبدیل می‌کند تا بتواند به لایه Dense بعدی داده شوند. در نهایت، دو لایه Dense تعریف شده است. لایه اول با ۱۲۸ نورون و تابع فعال‌سازی relu تعریف شده است و لایه دوم با دو نورون و بدون فعال‌سازی تعریف شده است.

۶,۱,۲ مدل CNN-LSTM

این مدل شامل چند لایه است. ابتدا، یک لایه کانولوشن با ۸ فیلتر، اندازه کرنل ۳ تعریف شده است. سپس یک لایه LSTM با ۶۴ نورون و بازگشت دنباله‌های خروجی تعریف شده است. این لایه به منظور یادگیری الگوهای طولانی مدت در داده‌های ورودی استفاده می‌شود. تنظیم return_sequences=True به LSTM می‌گوید که برای تمام داده‌های ورودی، خروجی تولید کند. پس از آن، یک لایه MaxPooling با اندازه ۲ تعریف شده است. این لایه با انتخاب بزرگترین مقدار در هر دو مقدار داده‌های ورودی، ابعاد داده‌ها را نصف می‌کند. سپس، یک لایه Flatten تعریف شده است که به منظور تبدیل داده‌های ورودی به داده‌های یک بعدی استفاده می‌شود. در نهایت، یک لایه Dense با دو نورون و بدون فعال‌سازی تعریف شده است.

۶,۲ پارامترهای کامپایل برای مدل‌های CNN و CNN-LSTM

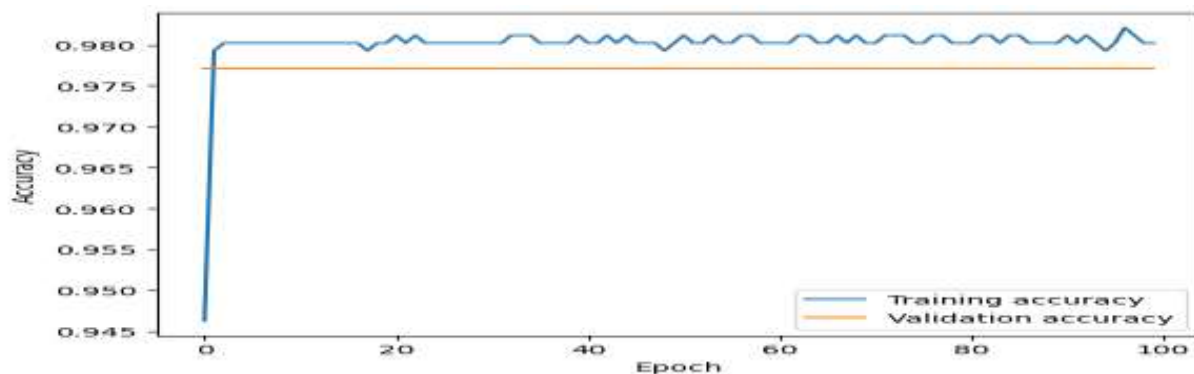
سه پارامتر بهینه ساز، ضرر و متریک برای هر دو مدل CNN و CNN-LSTM یکسان در نظر گرفته شده است تا مقایسه دقت‌های دو مدل در شرایط یکسان بررسی شود.

- پارامتر بهینه‌ساز، الگوریتم بهینه‌سازی است که در فرایند آموزش مدل‌ها مورد استفاده قرار می‌گیرد. در اینجا، الگوریتم بهینه‌سازی آدام انتخاب شده است.
- پارامتر ضرر، تابع هزینه است که در فرایند آموزش مدل مورد استفاده قرار می‌گیرد. در اینجا، تابع هزینه میانگین مربعات خطا انتخاب شده است. این تابع هزینه برای مسائل رگرسیون مناسب است.
- پارامتر متریک، معیارهایی است که در فرایند آموزش مدل جهت ارزیابی عملکرد مدل استفاده می‌شوند (Dormann, ۲۰۱۳). در اینجا، معیار دقت انتخاب شده است. این معیار برای مسائل دسته‌بندی مناسب است و نشان می‌دهد که چه درصد از داده‌های آزمون به درستی دسته‌بندی شده‌اند.

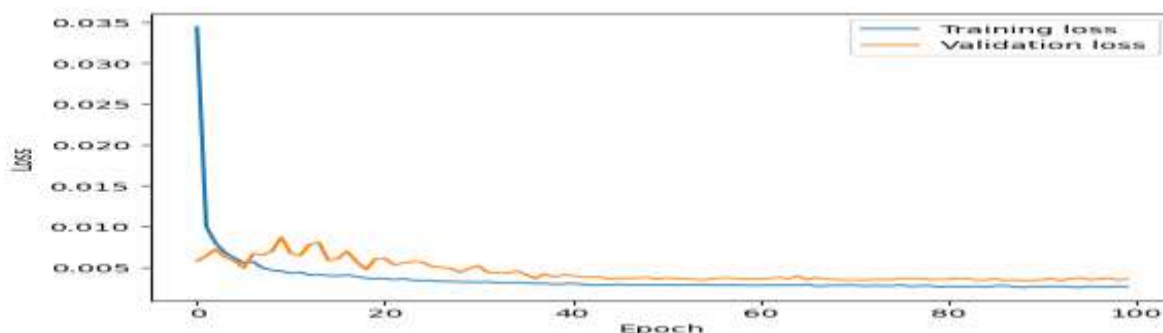
۷. تحلیل یافته‌ها

نمودارهای دقت و خطا مدل‌های CNN و CNN-LSTM به شرح زیر می‌باشد. این نمودارها به ما کمک می‌کند که فرآیند آموزش مدل به چه صورت انجام گردیده است. تعداد مراحل آموزش برای هر دو مدل برابر با ۱۰۰ در نظر گرفته شده است. شکل‌های شماره ۸ و ۹ به ترتیب مربوط به نمودارهای دقت و خطا برای مدل CNN در فرآیند آموزش می‌باشد. نمودار آبی رنگ مربوط به میزان دقت و خطا مدل برای داده‌های آموزش و نمودار نارنجی برای دقت و خطا مدل بر روی داده‌های اعتبارسنجی می‌باشد. همانطور که از نمودار دقت مشخص است در همان مراحل ابتدایی مدل توانسته به برازش نسبتاً کاملی از داده‌ها دست پیدا کند. به طوری که هم دقت مدل بر روی داده‌های آموزش و هم

داده‌های اعتبارسنجی در هر مرحله از دقت بالای ۹۷/۵٪ برخوردار هستند و از طرفی میزان خطا هم در همان مراحل ابتدایی به زیر ۱٪ رسیده است. در نتیجه پس با قطعیت می‌توانیم بگوییم که در فرآیند آموزش مدل CNN بیش برآزش رخ نداده است و مدل با موفقیت توانسته داده‌ها را برازش کند.



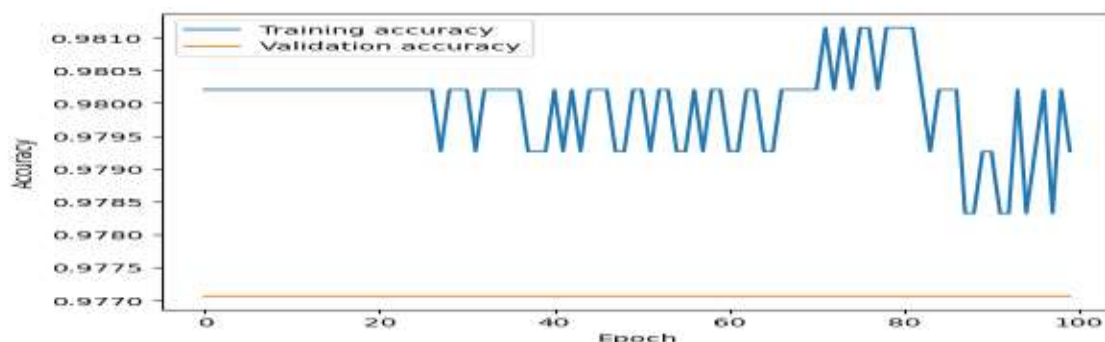
شکل (۹) نمودار دقت مدل CNN



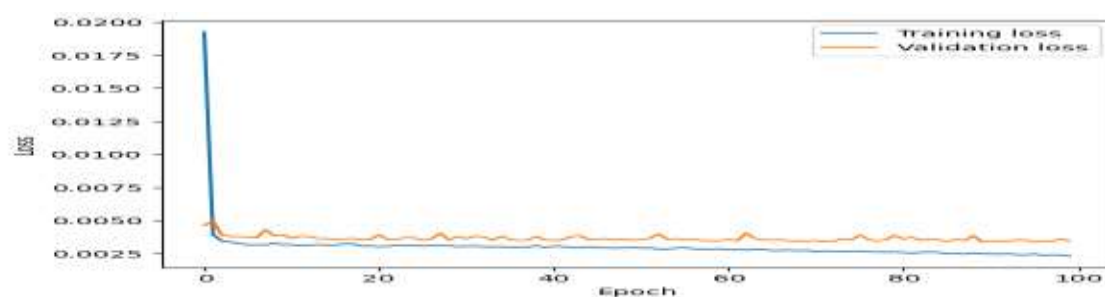
شکل (۱۰) نمودار خطای مدل CNN

همچنین نمودار دقت و خطای مدل CNN-LSTM به شکل زیر می‌باشد. همانگونه که مشخص است، شکل‌های زیر به ترتیب مربوط به نمودارهای دقت و خطای مدل CNN-LSTM در فرآیند آموزش می‌باشد. نمودار آبی رنگ مربوط به میزان دقت و خطای مدل برای داده‌های آموزش و نمودار نارنجی برای دقت و خطای مدل بر روی داده‌های اعتبارسنجی می‌باشد. مدل CNN-LSTM برخلاف مدل CNN در همان نتوانسته به صورت پایدار به برازش نسبتاً کاملی از داده‌ها دست پیدا کند و بعد از مرحله ۱۲۰ام آموزش دقت مدل بر روی داده‌های هر چند مرحله یک بار افزایش و کاهش پیدا می‌کند به نحوی که حتی در مراحل آخر آموزش دقت مدل بر روی داده‌های آموزش به کاهش بیشتری پیدا می‌کند. اما در مجموع مدل موفقیت آمیز بوده است چرا که کمترین میزان دقت بر روی داده‌های آموزش و اعتبارسنجی به ترتیب ۹۷/۸۵٪ و ۹۷/۷۰٪ بوده است. از طرفی دیگر نیز خطای مدل همواره رو به کاهش بوده است به طوری که میزان خطا در همان مراحل ابتدایی به زیر ۰/۵٪ رسیده است. این امر با توجه به میزان دقت مدل نشانگر این موضوع است که مدل دچار بیش برازش نشده است و در نهایت مدل نسبتاً کاملی از داده‌ها برازش شده است.

بهبود پرداخت تسهیلات با در نظر گرفتن انضباط مالی و حداکثر سودآوری در شرایط عدم اطمینان



شکل (۱۱) نمودار دقت مدل CNN-LSTM



شکل (۱۲) نمودار خطای مدل CNN-LSTM

نتیجه گیری

مؤسسات مالی در ساختار اقتصادی جوامع از اهمیت بالایی برخوردارند. این مؤسسات به عنوان قلب اقتصاد در دو بازار سرمایه و پول فعالیت داشته و باعث ایجاد جریان نقدینگی می گردند. در اقتصاد ایران حتی، اهمیت این مؤسسات دو چندان می شود چرا که آنها علاوه بر واسطه‌ی وجوه بودن در بازار پول، به دلیل عدم توسعه کافی بازار سرمایه، نقش اساسی در تأمین مالی برنامه‌های میان مدت و بلندمدت اقتصادی کشور ایفا می کنند. بنابراین مبحث سودآوری در این نوع مؤسسات نه تنها برای خود آنها که برای کل اقتصاد حائز ارزش می گردد. این پژوهش در ابتدا به دنبال شناسایی عوامل مؤثر بر سودآوری بانک‌ها و کاهش ریسک داشته و با جداسازی تأثیرات عوامل مختلف، به طور خاص بر کارکرد آنها تمرکز کرده و تأثیر عوامل کارکردی شعب را بر روی سودآوری و ریسک بررسی نموده است. برای رسیدن به هدف مذکور ابتدا از تکنیک‌های خوشه‌بندی برای جدا کردن مشتریان به دو گروه مشتریان خوش حساب و بدحساب استفاده گردید و سپس به بررسی عوامل مؤثر بر افزایش سودآوری و کاهش ریسک پرداخته شده است. هدف این پژوهش طراحی یک مدل یادگیری عمیق CNN و CNN-LSTM بوده است. سپس به برآورد ریسک و سود با مدل موردنظر پرداخته شده است. در مقایسه دقت مدل‌های CNN و CNN-LSTM بر روی داده‌های آزمایش می توان گفت میزان دقت مدل CNN بر روی داده‌های آزمایش برابر با ۹۸/۲۳٪ می باشد. دقت مدل CNN-LSTM بر روی همان داده‌های آزمایش برابر با ۹۷/۷۸٪ می باشد. با توجه به اینکه هر دو مدل در شرایط یکسانی آموزش دیده‌اند و پارامترهای یادگیری هر دو دقیقاً باهم برابر بوده است و با توجه به پایداری بیشتر مدل CNN در فرآیند آموزش و دقت بالاتر آن بر روی داده‌های آزمایش می توان نتیجه گرفت که مدل CNN بهتر از مدل CNN-LSTM عمل کرده است. اما در کل هر دو مدل توانسته اند دقت خوبی بر روی داده‌های آزمایش بدست بیاورند.

پیشنهادهای کاربردی

با توجه به بکارگیری عوامل مؤثر در پرداخت تسهیلات، در بانک‌ها می‌توان با داشتن هریک از پارامترهای مشتریان آنها را در هریک از دو دسته مشتریان بد حساب یا خوش حساب خوشه‌بندی نمود علاوه بر آن می‌توان پیش‌بینی نمود مشتری با ویژگی‌های شخصیتی و درآمدی و با توجه به نوع و مبلغ وام آیا توان باز پرداخت تسهیلات را دارا می‌باشد یا خیر؟ و اگر به مشتری وامی پرداخت شود ریسک و سود مربوط به هر وام چقدر می‌باشد.

تحقیقات آتی

پژوهشگران آتی می‌توانند از بقیه مدل‌های یادگیری عمیق نیز برای مباحث پیش‌بینی استفاده نمایند. همچنین آنها می‌توانند از یافته‌های این تحقیق در بقیه مسائل مالی و مدیریتی استفاده نمایند.

منابع

- اصغر پور، حسین، (۱۳۸۸). اثر بی ثباتی سیاسی بر رشد اقتصادی ایران: رهیافت غیر خطی APARCH.
- خواجه‌ای، شکراله، غیوری مقدم، علی، (۱۳۹۲). بررسی کاربرد تحلیل پوششی داده‌ها در مقایسه و ارزیابی عملکرد واحدهای تجاری.
- دامی، سینا، خیری، فرشته، (۱۳۹۷). شناسایی تقلب در پرداخت کارت‌های اعتباری با شبکه عصبی حافظه طولانی کوتاه مدت (LSTM). سومین کنفرانس ملی فناوری در مهندسی برق و کامپیوتر
- راعی، رضا، فرهادی، روح‌الله، شیروانی، امیر، (۱۳۹۰). رابطه در گذر زمان بین بازده و ریسک: شواهدی از الگوی قیمت گذاری دارایی سرمایه ای در گذر زمان
- ICAPM، فصلنامه چشم انداز مدیریت مالی دانشگاه شهید بهشتی، تهران.
- شریف فر، امیر، خلیلی عراقی، مریم، رئیسی وانانی، ایمان، فلاح شمس، میر فیض، (۱۴۰۱). کاربرد معماری‌های یادگیری عمیق در پیش‌بینی قیمت سهام (رویکرد شبکه عصبی پیچشی CNN)، دوره ۱۰، شماره ۳، صفحه ۱-۲۰.
- صالحی، فهمیه، صالحی، مجتبی، اسکندری، میثم، (۱۳۹۳). بهینه سازی سبد تسهیلات اعطایی موسسات مالی با استفاده از برنامه ریزی ریاضی و الگوریتم ژنتیک (مطالعه موردی بانک تجارت)، فصلنامه توسعه مدیریت پولی و بانکی، دوره ۲، شماره ۳.
- عسگرزاده، مجید، و دیگران، (۱۳۸۶). بررسی ریسک عملیاتی بانکداری الکترونیکی یکپارچه و ارائه راهکارهای کاهش مخاطرات آن، اولین کنفرانس جهانی بانکداری الکترونیک.

References

- Ansari, A., Riasi, A., (2016). Customer Clustering Using a Combination of Fuzzy C-Means and Genetic Algorithms.
- Adnan et al, R.M., (2020). Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs, J. Hydrol.
- Armantier, O., Ghysels, E., Sarkar, A., Shrader, J., (2015). Discount window stigma during the 2007-2008 financial crisis. J. Financ. Econ. 118 (2), 317-335.
- Al Machot, F. et al., (2019). A deep-learning model for subject-independent human emotion recognition using electrodermal activity sensors.
- Bock, H., (2007). Clustering methods: A history of k-means algorithms Selected Contributions in Data Analysis and Classification, . 161-172.
- Bhandari, P., (2022). Missing Data | Types, Explanation, & Imputation. Revised on November 11, 2022.
- Bernanke, B., (2009). The federal reserve's balance sheet: an update. In: A Speech at the Federal Reserve Board Conference on Key Developments in Monetary Policy, Washington, D.C.
- Bose, I., Chen, X., (2015). Detecting the migration of mobile service customers using fuzzy clustering.

- Chen, X.D., (2017). *Analysis and research of common clustering algorithm in data mining* Digital Technol. Appl., pp. 151-152.
- Du S., Li J., (2019). *Parallel processing of improved knn text classification algorithm based on Hadoop*. In: 2019 7th International Conference on Information, Communication and Networks (ICICN), 167-170.
- Dormann et al, C.F., (2013). *Collinearity: a review of methods to deal with it and a simulation study evaluating their performance*.
- Estivill-Castro V., Houle M.E., (2001). *Robust distance-based clustering with applications to spatial data mining*, Algorithmica, 30 (2) , 216-242.
- Guo,G., Wang, H., Bell, D., Greer, Y., (2003). *KNN model-based approach in classification* OTM Confederated International Conferences on the Move to Meaningful Internet Systems, 986-996.
- Howell, E., (2021). *4 Techniques To Deal With Missing Data in Datasets*, Simple methods that can nullify the effects of missing values, Published in Towards Data Science, Sep 17.
- Huang, X.Y. , (2018). *An improved KNN algorithm and its application in real-time car-sharing prediction.*, Dalian University of Technology, Daian, China ,M.S. thesis.
- Huang et al., Huang J., Wei Y., Yi J., Liu M., (2018). *An improved KNN based on class contribution and feature weighting*. In: 2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), 313-316.
- Henderi, T., Wahyuningsih, T., Rahwanto, E., (2021). *Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (KNN) Algorithm to The Accuracy of Type Breast Cance*, International Journal of Informatics and Information System Vol.4, NO.1, 13-20.
- Khosroyani, M., Heydarpoor, F., Yaghoob-nezhad, A., & Poorzamani, Z., (2022). *An artificial neural network model for predicting the liquidity risk of Iranian private banks*. Int. J. Nonlinear Anal. In Press, 1–11 ISSN: 2008-6822 (electronic); <http://dx.doi.org/10.22075/ijnaa.2022.29118.4071>. [In Persian].
- Khalila, S., Amritb, CH, Kocha, T, Dugundjia, E., (2021). *Forecasting public transport ridership: Management of Information system by using CNN and LSTM Architecture*, The 12th International Conference on Ambient Systems, Networks and Technologies (ANT).
- Krizhevsky, A., Sutskever, I., Hinton, G.E., (2017). *Imagenet classification with deep convolutional neural networks*. Communications of the ACM 60, 84 – 90. URL: <https://search.ebscohost-com.proxy.uba.uva.nl:2443/login.aspx?direct=true&db=buh&AN=123446102&site=ehost-live&scope=site>.
- Mohammadi, N., Zangeneh, M., (2016). *Customer credit risk assessment using artificial neural networks* IJ Information Technology and Computer Science, 8 (3) , 58-66.
- Ngufor, C., Van Houten, H., Caffo, B.S., Shah, N.D., McCoy, R.G., (2019). *Mixed Effect Machine Learning: a framework for predicting longitudinal change in hemoglobin A1c* J. Biomed. Inform, 56-67
- Ole Hjelkrem, L., Eilif de Lange, P., (2023). *Explaining Deep Learning Models for Credit Scoring with SHAP: A Case Study Using Open Banking Data*, J. Risk Financial Management., 16(4), 221; <https://doi.org/10.3390/jrfm16040221>.
- Oquab, M., Bottou, L., Laptev, I., Sivic, J., (2014). *Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks*, in: IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, United States. URL: <https://hal.inria.fr/hal-00911179>. conference version of the paper.