Shahid Beheshti University
Journal of Financial Management Perspective
No. 31, Autumn 2020
PP: 53-73

P-ISSN: 2645-4637
E-ISSN: 2645-4645

Investigating the Variables Affecting Banks' Legal Customers Credit Risk, Using Support Vectors Machine and Decision Tree

Iman Dadashi*, Sajjad Kordmanjiri, Zahra Khoshnoud ***, Hamidreza Gholamnia Roshan******

Abstract

The increase of non-current debts to lending facilities ratio as an indicator of banks' credit risk can endanger the health of the banking sector, financial system and the real economy. Hence, in this paper, analyzing credit risk through the actual balance of non- performing debts by focusing on a broad set of variables including financial, non-financial, contractual, audit and economic variables in a sample of 677 legal customer facility files of a State Bank for the years 2006- 2017 has been accomplished. Based on the results, the LASSO Algorithm with better performance has identified 10 key financial, economic and audit variables affecting the credit risk. However, training these features by support vector machine and decision tree model, which represent the best results in the Lasso algorithm with the decision tree application, confirms the small significance factor for the audit variables. Therefore, using LASSO algorithm with decision tree and focusing on financial and economic variables can be sufficient for credit risk analysis.

Keywords: Credit Risk, Non-Current Debts, LASSO Algorithm, Decision Tree, Recovery Rate.

Received: 2020.January.02, Accepted: 2020.July.13.
* Assistant Prof, Department of Accounting, Babol branch, Islamic Azad University, Mazandaran, Iran. (Corresponding Author). Email: Idadashi@baboliau.ac.ir
** Ph.D. Candidate in Accounting, Babol branch, Islamic Azad University, Mazandaran, Iran.
*** Associate Prof, Banking Group, Monetary and Banking Research Institute, Tehran, Iran.
****Assistant Prof, Department of Accounting, Babol branch, Islamic Azad University, Mazandaran, Iran.

دانشگاه شهیدبهشتی
نشریه چشم‌انداز مدیریت مالی
شماره ۳۱ ـ پائیز ۱۳۹۹
ص‌ص ۷۳-۵۳

شاپای چاپی: ۲۶۴۵-۴۶۳۷
شاپای الکترونیکی: ۲۶۴۵-۴۶۴۵

بررسی متغیرهای موثر بر ریسک اعتباری مشتریان حقوقی بانک‌ها با استفاده از ماشین بردار پشتیبان و درخت تصمیم

ایمان داداشی*، سجاد کردمنجیری، زهرا خوشنود***، حمیدرضا غلام‌نیا روشن*****

چکیده

افزایش نسبت مطالبات غیرجاری به تسهیلات اعطایی به‌عنوان شاخص ریسک اعتباری بانک‌ها می‌تواند سلامت شبکه بانکی، نظام مالی و اقتصاد حقیقی را به‌خطر اندازد. از این رو در این مقاله، بررسی این ریسک با استفاده از نسبت مانده واقعی مطالبات غیرجاری و با تمرکز بر مجموعه‌ای گسترده از متغیرها شامل متغیرهای مالی، غیرمالی، خصوصیتی قراردادها، حسابرسی و اقتصادی، در نمونه‌ای از ۶۷۷ پرونده تسهیلاتی مشتریان حقوقی یک بانک دولتی برای سال‌های ۱۳۸۵ تا ۱۳۹۶ مورد توجه قرار گرفت. براساس نتایج حاصله، در انتخاب متغیرهای تاثیرگذار بر ریسک اعتباری، الگوریتم لاسو با عملکرد بهتر به شناسایی ۱۰ متغیر کلیدی از گروه متغیرهای مالی، اقتصادی و حسابرسی منتهی شد. با این وجود نتایج آموزش این ویژگی‌ها توسط مدل ماشین بردار و درخت تصمیم که بیانگر بهترین نتایج در قالب الگوریتم لاسو به همراه درخت تصمیم هستند، ضریب اهمیت اندکی را برای متغیرهای حسابرسی در نظر می‌گیرند. از این رو استفاده از الگوریتم لاسو به همراه درخت تصمیم با تمرکز بر متغیرهای مالی و اقتصادی می‌تواند از کفایت لازم برخوردار باشد.

کلیدواژه‌ها: ریسک اعتباری، مطالبات غیرجاری، الگوریتم لاسو، درخت تصمیم، نرخ وصول.

تاریخ دریافت مقاله: ۱۳۹۸/۱۰/۱۲، تاریخ پذیرش مقاله: ۱۳۹۹/۰۴/۲۳.

* استادیار، گروه حسابداری، واحد بابل، دانشگاه آزاد اسلامی، مازندران، ایران (نویسنده مسئول)،
E-Mail: Idadashi@baboliau.ac.ir

** دانشجوی دکتری حسابداری، گروه حسابداری، دانشگاه آزاد اسلامی، مازندران، ایران.

*** استادیار، گروه بانکداری، پژوهشکده پولی و بانکی بانک مرکزی، تهران، ایران.

**** استادیار، گروه حسابداری، واحد بابل، دانشگاه آزاد اسلامی، مازندران، ایران.

۱. مقدمه

امروزه ارزیابی عوامل موثر بر ریسک اعتباری[1] از مهم‌ترین موضوعات در زمینه مدیریت ریسک مالی است که اهمیت آن در شبکه بانکی و به‌خصوص در اقتصادهایی با نظام مالی بانک‌محور، مانند ایران، پررنگ‌تر است. هر چند شبکه بانکی با توجه به ویژگی‌های خاص آن، با انواع گوناگونی از ریسک‌ها و مخاطرات، مانند ریسک‌های اعتباری، نقدینگی[2]، بازار[3] و عملیاتی[4] روبه‌رو است؛ با این وجود به عقیده اکثر کارشناسان بانکی، مهم‌ترین ریسکی که نظام بانکی هر کشوری را تهدید کرده و آسیب‌های فراوانی به صنعت بانکی وارد می‌کند، ریسک اعتباری است[۲۷].

حتی سیر تکاملی و رشد مجموعه رهنمودهای کمیته بال[5]، متمرکز بر ریسک اعتباری است (کیمورا، باسو و کایو[6]، ۲۰۱۵) و این کمیته، مدیریت ریسک اعتباری را نه تنها برای ادامه فعالیت خود بانک‌ها، بلکه برای تمام اقتصاد، امری ضروری می‌داند[۱۴]. ریسک اعتباری که توسط کمیته بال (۲۰۰۰) در قالب «احتمال قصور وام‌گیرنده یا طرف مقابل بانک نسبت به انجام تعهداتش طبق شرایط توافقی» تعریف شده، بانک‌ها را با زیان‌های ناشی از عدم بازپرداخت، یا بازپرداخت با تاخیر اصل یا فرع وام، از سوی مشتری مواجه می‌کند[۸]. بنابراین بانک‌ها در صدد اعطای تسهیلات به اشخاصی هستند که ضمن برخورداری از ریسک پایین، بتوانند بازده متناسب با سود تسهیلات اعطایی ایجاد کنند. این امر زمانی محقق می‌شود که بانک‌ها قادر به شناسایی مشتریان اعتباری خود اعم از حقیقی و حقوقی بوده و بتوانند آنها را براساس توانایی و تمایل، نسبت به بازپرداخت کامل و به‌موقع تعهدات، با استفاده از معیارهای مالی و غیرمالی مناسب طبقه‌بندی کنند. در چنین سازوکاری، تسهیلات به متقاضیانی پرداخت می‌شود که از ریسک اعتباری کمتری برخوردار بوده و احتمال بازپرداخت بدهی آنها در سررسید بیشتر است[۲۴].

البته در سال‌های اخیر، در مباحث مربوط به ریسک اعتباری، مفهوم جدیدتری برای پیش‌بینی این ریسک تحت عنوان نرخ وصول[7] مدنظر قرار گرفته است؛ به زبان ساده‌تر، صرفاً مشتریان براساس میزان ریسک یا احتمال عدم بازپرداخت وام طبقه‌بندی و رتبه‌بندی نمی شوند؛ بلکه به میزان نرخ وصول مطالبات در صورت نکول نیز توجه می‌شود. مسلم است که عدم وصول کامل تسهیلات با درصدی از عدم وصول بسیار متفاوت است و نمی‌توان مشتریانی که تنها قسمتی از تسهیلات آنها نکول شده را با مشتریانی یکسان در نظر گرفت که در بازپرداخت بخش اعظم تسهیلات و یا کل تسهیلات خود قصور کرده‌اند.

---

1. Credit Risk
2. Liquidity
3. Market
4. Operational
5. Basel Committee
6. Kimura, Basso & Kayo
7. Recovery Rate (RR)

به‌دلیل اهمیت موضوع و با توجه به اندک بودن پژوهش‌های انجام شده در مدل‌سازی ریسک اعتباری براساس نرخ وصول و عدم بررسی عوامل موثر بر این ریسک در مطالعات محدود فوق، از این رو در این پژوهش برای اولین بار، در بررسی متغیر ریسک اعتباری در کنار استفاده از پنج گروه متغیر مشتمل بر متغیرهای مالی، غیرمالی، خصوصیتی قراردادها، حسابرسی و اقتصادی در پرونده تسهیلاتی مشتریان حقوقی یک بانک دولتی، به شاخص نرخ وصول نیز توجه شده است.

در این خصوص، پژوهش حاضر نظر به سهم بالای ریسک اعتباری در کشور در کنار حجم بیشتر تسهیلات اعطایی به مشتریان حقوقی در مقایسه با مشتریان حقیقی؛ با طراحی و مدل‌سازی ریسک اعتباری در بانک نمونه با استناد به رویکرد نوین مورد استفاده در تصریح این متغیر با استفاده از داده‌های مشتریان حقوقی یک بانک دولتی در بازه زمانی ۱۳۹۶-۱۳۸۵ می‌تواند به نتایجی قابل اتکا و تسری به کل شبکه بانکی منتهی شود.

در این راستا در ادامه، ابتدا مبانی نظری و پیشینه پژوهش ارائه شده و در گام بعد پس از معرفی روش پژوهش، ابتدا ویژگی‌های تاثیرگذار، با استفاده از الگوریتم‌های همبستگی[1] و لاسو[2] و براساس میزان بهتر بودن عملکرد الگوریتم، انتخاب؛ سپس این متغیرها با به‌کارگیری رگرسیون ماشین بردار پشتیبان[3] و درخت تصمیم[4] در طراحی مدل ریسک اعتباری ارزیابی شده و به این ترتیب در پایان، یافته‌ها، نتیجه‌گیری و پیشنهادها ارائه می‌شوند.

۲. مبانی نظری و پیشینه پژوهش

افزایش مطالبات غیرجاری نشان‌دهنده افزایش ریسک اعتباری بانک‌ها است و حتی شاخص نسبت تسهیلات غیرجاری به کل تسهیلات- به‌عنوان یکی از شاخص‌های شناسایی ریسک اعتباری بانک‌ها- براساس مقدار بالای این نسبت در چند سال اخیر در کشور، گویای وخامت ترازنامه بانک‌ها و عملکرد ضعیف آنها در مدیریت این ریسک است که برای پایان سال ۱۳۹۶ نسبت این متغیر در نماگرهای اقتصادی منتشره توسط بانک‌مرکزی ۱۰/۳ درصد گزارش شده است[5,6]. از این رو ضرورت توجه بیشتر به این ریسک در شبکه بانکی کشور مشاهده می‌شود.

حتی حجم زیاد تسهیلات غیرجاری نظام بانکی کشور، گویای نبود مدل‌های مناسب اندازه‌گیری ریسک اعتباری و سیستم‌های کارا در مدیریت ریسک در شبکه بانکی است که منجر به کاهش نسبت کفایت سرمایه در شبکه بانکی و وخیم شدن سلامت شبکه بانکی شده است[۲۷]. این

---

1. Correlation Feature Selection (CFS)
2. Least Absolute Shrinkage and Selection Operator (LASSO)
3. Support Vector Regression (SVR)
4. Decision Tree (DT)

۵. نماگرهای اقتصادی بانک مرکزی، ۱۳۹۷.

۶. براساس آمار بانک جهانی، مطالبات معوق باید بین ۲ تا ۵ درصد تسهیلات اعطایی باشد و مقادیر بیشتر این نسبت به عنوان ریسک بالا محسوب می‌شود [۲۶].

مساله در سال‌های اخیر با افزایش نگرانی نسبت به انتشار این وضعیت نامساعد به دیگر بازارهای مالی، ضرورت اتخاذ برنامه‌هایی جهت اصلاح نظام بانکی را نیز به‌دنبال داشته است. بنابراین شناسایی عوامل تاثیرگذار بر مطالبات غیرجاری و ریسک اعتباری نظام بانکی امری ضروری به نظر می‌رسد.

در این راستا با توجه به مبانی نظری موجود در عرصه بین‌المللی، مشاهده می‌شود از زمان معرفی سیستم‌های هوش مصنوعی نظیر شبکه عصبی، الگوریتم ژنتیک، سیستم‌های خبره، ماشین بردار پشتیبان و درخت تصمیم، استفاده از آنها در تحقیقات مالی و رتبه‌بندی اعتباری مرسوم شده و به‌سرعت در حال گسترش است.

به‌عنوان مثال در مطالعات خارجی، ترکیب تحلیل ممیزی و شبکه عصبی (یو[1] و همکاران، ۲۰۰۸)، تحلیل پوششی داده‌ها (مین و لی[2]، ۲۰۰۸)، درخت تصمیم (یانپینگ[3] و همکاران، ۲۰۱۲)، ترکیب درخت تصمیم و شبکه عصبی (کاباری و نواچکوا[4]، ۲۰۱۳) و روش مبتنی بر داده‌های بزرگ با استفاده از شبیه‌سازی مونت‌کارلو[5] و مدل خطی مختلط[6] (پرز-مارتین[7] و همکاران، ۲۰۱۸) در مدل‌سازی ریسک اعتباری موثر واقع شده‌اند.

در تحقیقات داخلی نیز، ماشین بردار پشتیبان (راعی و فلاح‌پور، ۱۳۸۷؛ منصورفر و همکاران، ۱۳۹۴)، شبکه عصبی مبتنی بر کلونی زنبور عسل مصنوعی (فلاح‌پور و همکاران، ۱۳۹۳)، درخت تقسیم و رگرسیون (میرغفوری و امین‌آشوری، ۱۳۹۴)، الگوریتم فرابتکاری و هیبریدی چند معیاره شبکه کلونی مورچگان- عصبی فازی (حری و مهدوی، ۱۳۹۴) و مدل ماشین بردار پشتیبان و مدل هیبریدی الگوریتم ژنتیک (محمدیان‌حاجی‌کُرد و همکاران، ۱۳۹۵) از جمله مدل‌های مبتنی بر هوش مصنوعی و موفق در پیش‌بینی ریسک اعتباری هستند.

با این وجود بسیاری از این مدل‌ها، حوزه جامعی از ویژگی‌ها و عوامل ریسک را مورد توجه قرار نداده و بیشتر مبتنی بر خصوصیات مالی و اقتصادی بوده‌اند. در شناسایی این متغیرها در بیشتر مطالعات داخلی، صرفاً بر مجموعه متغیرهای مالی (میرغفوری و امین‌آشوری، ۱۳۹۴؛ محمدیان‌حاجی‌کُرد و همکاران، ۱۳۹۵؛ دارابی و مشایخی، ۱۳۹۵)، متغیرهای اقتصادی (نوروزی، ۱۳۹۳؛ رستم‌زاده و همکاران، ۱۳۹۷؛ رستمی و همکاران، ۱۳۹۷؛ محققنیا و همکاران، ۱۳۹۷)، متغیرهای خصوصیتی قرارداد (جعفری‌اسکندری و روحی، ۱۳۹۶؛ تاری و همکاران، ۱۳۹۶)، و یا حتی متغیرهای مالی در کنار متغیرهای غیرمالی (صفری و همکاران، ۱۳۹۰؛ ابراهیمی و دریابر، ۱۳۹۱) تاکید شده است. با این وجود متغیرهای حسابرسی در این زمینه مورد

---

[1]. Yu
[2]. Min & Lee
[3]. Yanping
[4]. Kabari & Nwachukwu
[5]. Monte Carlo
[6]. Linear Mixed Model (LMM)
[7]. Pérez -Martín

توجه لازم واقع نشده‌اند[1] و حتی به ضرورت اتخاذ رویکرد جامع و کلان با لحاظ تمامی متغیرهای فوق در ارزیابی مشتریان حقوقی اشاره نشده است. همچنین بیشتر این مطالعات بر مشتریان حقیقی (فلاح‌پور و همکاران، ۱۳۹۳؛ تاری و همکاران، ۱۳۹۶؛ پورکاظمی و همکاران، ۱۳۹۶ و حبیبی و همکاران، ۱۳۹۷) متمرکز بوده است و از این رو ضرورت تمرکز بیشتر بر مشتریان حقوقی با ارزش اسمی بالاتر اعتبار، مشاهده می‌شود.

در مورد نحوه تصریح متغیرهای فوق در مطالعات تجربی، در خصوص متغیرهای اقتصادی اثرگذار بر مطالبات غیرجاری، متغیرهای مختلفی شامل تولید ناخالص داخلی (جعفری اسکندری و روحی، ۱۳۹۶)، نرخ تورم (آقایی و رضاقلی‌زاده، ۱۳۹۵)، نرخ بهره (بخت و التر[2]، ۲۰۱۴) و شاخص قیمت مسکن (حیدری و همکاران، ۱۳۸۹) مورد توجه قرار گرفته‌اند.

در گروه متغیرهای مالی، با توجه به دامنه گسترده این متغیرها، می‌توان به برخی از آنها از جمله نسبت‌های آنی، جاری، نسبت‌های سودآوری، نرخ‌های رشد دارایی و فروش (میرغفوری و امین‌آشوری، ۱۳۹۴، محمدیان‌حاجی‌کُرد و همکاران، ۱۳۹۵، حکیمی‌پور، ۱۳۹۷، اَدو و همکاران، ۲۰۱۸) اشاره کرد.

متغیرهای غیرمالی مورد استفاده در مطالعات نیز برحسب تمرکز بر مشتریان حقیقی و حقوقی متفاوت است. در مورد مشتریان حقیقی، بیشتر تمرکز بر سن و جنس (بلانکو[3] و همکاران، ۲۰۱۳، بخت و التر، ۲۰۱۴)، وضعیت اشتغال (تاری و همکاران، ۱۳۹۶) و وضعیت تاهل و تعداد افراد تحت تکفل (بلانکو و همکاران، ۲۰۱۳، پورکاظمی و همکاران، ۱۳۹۶) است. در مورد مشتریان حقوقی نیز بیشتر بر متغیرهایی مانند نوع، سن شرکت (تقوی و همکاران، ۱۳۸۷) و تعداد کارکنان (صداقت‌پرست و همکاران، ۱۳۹۴) تمرکز شده است.

در گروه متغیرهای بیانگر ویژگی‌های قرارداد بین بانک و مشتری نیز نوع قرارداد، تعداد اقساط، روش تقسیط، ارزش وثیقه (جعفری اسکندری و روحی، ۱۳۹۶)، میزان اعتبار و مدت بازپرداخت (بلانکو و همکاران، ۲۰۱۳، تاری و همکاران، ۱۳۹۶) مورد توجه قرار گرفته‌اند.

در ارتباط با رویکرد مدل سازی و استفاده از روش‌های جدید در گروه هوش مصنوعی و با تمرکز بر دو رویکرد ماشین بردار پشتیبان و درخت تصمیم مورد سوال در این مقاله، در مقام جایگزین روش‌های سنتی شامل تحلیل ممیزی و رگرسیون لجستیک، عموماً شواهد تاییدکننده‌ای در مطالعات تجربی یافت می‌شود.

---

۱. این در حالی است که در کنار این متغیرها، نقش متغیرهای حسابرسی در نکول و ورشکستگی این مشتریان در مطالعات مختلف داخلی نیز به اثبات رسیده است (شورورزی و همکاران، ۱۳۹۰، شریفی‌سفیدی، ۱۳۹۲، برادرسلیمی و همکاران، ۱۳۹۴، اکبری و حاجیها، ۱۳۹۵، مزرعه‌فراهانی، ۱۳۹۵، منصورفر و همکاران، ۱۳۹۷). از این رو ضرورت توجه به این گروه از متغیرها در کنار چهار گروه متغیر فوق برای اعمال رویکری جامع در ارزیابی ریسک مشتریان حقوقی مشاهده می‌شود.

2. Bekhet & Elleter

3. Blanco

به عنوان مثال باستوس[1] (۲۰۱۰) به نتایجی دال بر عملکرد بهتر درخت‌های تصمیم، در مقایسه با رگرسیون و روش‌های پارامتریک در الگوسازی شاخص درصد مطالبات غیرقابل وصول[2] دست یافت[۳]. همچنین چو[3] (۲۰۱۷) با به‌کارگیری انواع مختلف مدل‌ها به نتایجی دال بر دقت پیش‌بینی بالای مدل‌های ماشین بردار پشتیبان (بیش از ۹۵ در صد) در مقایسه با مدل رگرسیون لجستیک (حدود ۶۶,۴ درصد) رسید[4][۴].

در مطالعات داخلی نیز نتایجی دال بر عملکرد بهتر دو روش مورد مطالعه در مقایسه با مدل‌های سنتی و استاندارد مشاهده می‌شود. در این زمینه می‌توان به برتری عملکرد مدل ماشین‌بردار پشتیبان در مقایسه با مدل لاجیت (راعی و فلاح‌پور، ۱۳۸۷)، درخت تصمیم در مقایسه با مدل لاجیت (میرغفوری و امین‌آشوری، ۱۳۹۴)، روش ماشین‌بردار تصمیم بهبود یافته با الگوریتم ژنتیک در مقایسه با مدل لاجیت (جعفری اسکندری و روحی، ۱۳۹۶) اشاره کرد[۲۵، ۱۶، ۱۱].

با این وجود در مورد کارکرد بهتر دو مدل فوق در مقایسه با دیگر رویکردهای هوش مصنوعی الزاماً نمی‌توان به جمع‌بندی مورد انتظار رسید. به‌عنوان مثال ادو و همکاران (۲۰۱۸) در مدل سازی ریسک اعتباری، به پایداری بیشتر مدل‌های مبتنی بر درخت تصمیم در مقایسه با مدل‌های مبتنی بر شبکه‌های عصبی چند لایه دست یافتند[۱]. در طرف‌مقابل، کارا و کریچن[5] (۲۰۱۲) با استفاده از شبکه‌های عصبی و ماشین بردار پشتیبان در اعتبارسنجی و ارزیابی ریسک مشتریان، به نتایجی دال بر کارکرد بهتر مدل شبکه عصبی چند لایه در مقایسه با شبکه ماشین بردار پشتیبان رسیدند[۱۳]. در مطالعات داخلی نیز محمدیان‌حاجی‌کُرد و همکاران (۱۳۹۵) شواهدی دال بر عملکرد بهتر ماشین‌بردار پشتیبان بهبودیافته با الگوریتم ژنتیک در مقایسه با مدل صرف ماشین بردار در شناسایی مشتریان ارائه کرد[۱۹].

البته در مطالعات داخلی متمرکز بر مدل های مورد توجه این تحقیق، صرفاً مطالعه محمدیان‌حاجی‌کُرد و همکاران (۱۳۹۵) در دسترس است که بر مشتریان حقوقی متمرکز است، با این وجود صرفاً دامنه محدودی از متغیرهای مالی را در بر گرفته است[6][۱۹].

---

1. Bastos

۲. در این مقاله نیز به رویکردی مشابه وی در مدل‌سازی ریسک اعتباری توجه شده است.

3. Chow

۴. نتایج مطالعه وی حتی بیانگر ضرورت اضافه کردن متغیرهای مبتنی بر نظر کارشناس در کنار متغیرهای مالی برای افزایش دقت پیش‌بینی مدل است. این مساله تاییدکننده رویکرد این مقاله در گسترش دامنه متغیرهای مورد بررسی در ارزیابی ریسک اعتباری است.

5. Karaa & Krichene

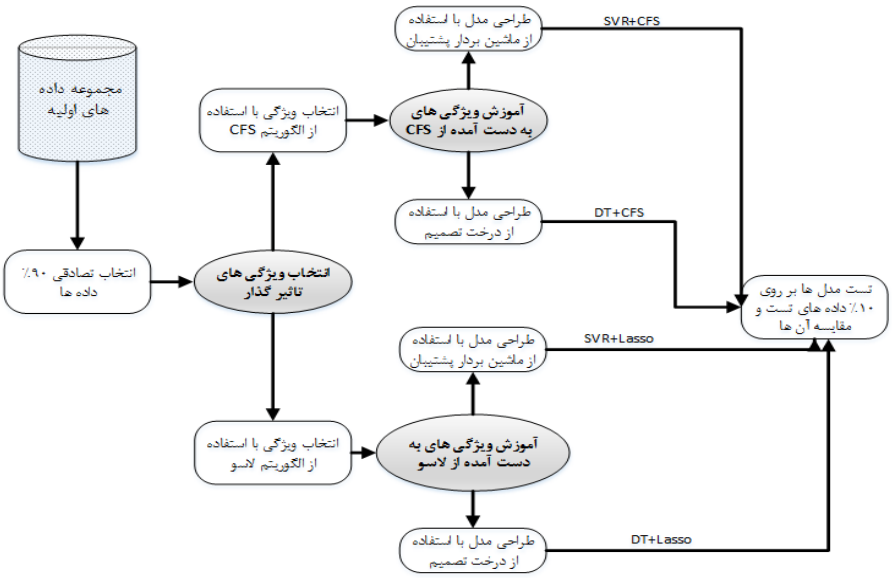
۶. مطالعه پورکاظمی (۱۳۹۶) نیز به عنوان دیگر مطالعه در دسترس، با تمرکز بر متغیرهای غیرمالی در کنار تعداد محدودی متغیر مالی مشتریان حقیقی بانکی خصوصی، به استفاده از شبکه‌های عصبی به پیش‌بینی احتمال نکول مشتریان پرداخته است و از این رو ابعاد مورد انتظار این تحقیق را در بر نمی‌گیرد [۲۴].

بنابراین برای برطرف کردن شکاف‌های موجود در تحقیقات داخلی در کنار پاسخگویی به سوال تحقیق در انتخاب مدل مناسبی در ارزیابی ریسک اعتباری از بین دو روش ماشین بردار پشتیبان و درخت تصمیم، نیاز به اتخاذ رویکردی جامع‌تر با محدود کردن جامعه مشتریان و تمرکز بر مشتریان حقوقی با ارزش بالاتر اعتبار دریافتی در مقایسه با مشتریان حقیقی، در کنار دامنه گسترده‌تر متغیرهای کلیدی مشاهده می‌شود.

از این رو در این پژوهش سعی شده در تصریح مدل ریسک اعتباری، چهار نوآوری دنبال شود. در گام اول در انتخاب متغیر وابسته، رویکرد متفاوتی نسبت به مطالعات قبلی دنبال شده و در محاسبه نسبت مطالبات غیرجاری، با توجه به نرخ وصول این مطالبات، مانده مطالبات غیرجاری نسبت به تسهیلات اعطایی براساس داده‌های واقعی پرونده‌های مشتریان مورد توجه قرار گرفته است. سپس با استخراج متغیرهای خصوصیتی قراردادها، به همراه متغیرهای غیرمالی و حسابرسی در کنار مجموعه متداول متغیرهای مالی و اقتصادی، بر دامنه گسترده‌تری از انواع متغیرها تمرکز شده؛ و در گام بعد با به‌کارگیری مدل‌های جدید در انتخاب متغیرهای تاثیرگذار، با استفاده از دو نوع الگوریتم؛ در کنار بررسی دقت نتایج این دو الگوریتم با کمک رگرسیون ماشین بردار پشتیبان و درخت تصمیم، رویکرد کاملاً متفاوتی دنبال شده که انتظار می‌رود پاسخگوی نیاز بانک مورد بررسی در این زمینه باشد.

۳. روش‌شناسی پژوهش

این پژوهش بر مبنای هدف از نوع تحقیقات کاربردی و از لحاظ ماهیت و روش، از نوع توصیفی و همبستگی است. به‌منظور بررسی ریسک اعتباری، ابتدا با استفاده از الگوریتم‌های انتخاب ویژگی، ویژگی های مهم و تاثیر گذار از بین دامنه گسترده‌ای از متغیر های مالی، غیر مالی، خصوصیتی قراردادها، حسابرسی و اقتصاد کلان، انتخاب شده است. برای انتخاب ویژگی از دو الگوریتم انتخاب ویژگی همبستگی و لاسو استفاده شده است. هدف از انتخاب ویژگی، گزینش تعداد کمتری از ویژگی‌های مرتبط یا مهم، برای افزایش عملکرد پیش‌بینی است. سپس برای طراحی مدل با توجه به نتایج حاصل از پیشینه پژوهش(شکل ۱)، از الگوریتم‌های ماشین بردار پشتیبان و درخت تصمیم استفاده شده است.



شکل ۱. روش کلی پژوهش

یادداشت- در این شکل CFS بیانگر الگوریتم همبستگی، Lasso الگوریتم لاسو، SVR ماشین بردار پشتیبان، و DT نمایانگر درخت تصمیم است.

جامعه آماری. در این پژوهش از اطلاعات کمی و کیفی ۶۷۷ پرونده تسهیلاتی مشتریان حقوقی در قراردادهای بالای ۱۰ میلیارد ریالی بانکی دولتی استفاده شده است و تمرکز بر این گروه، مدیریت ریسک در گروه هدف با ارزش بالاتر اعتبار در مقایسه با مشتریان حقیقی را پوشش می‌دهد. از آنجا که براساس سیاست‌ها و ضوابط اعتباری بانک ذی‌ربط، پرداخت تسهیلات بیش از ۱۰ میلیارد ریال به متقاضیان، منوط به ارائه صورت‌های مالی حسابرسی شده است؛ بنابراین اطلاعات مالی و حسابرسی این نمونه در کنار دیگر اطلاعات در دسترس قرار گرفت. از این تعداد پرونده تسهیلاتی، ۴۱ درصد بد حساب (غیرجاری) و ۵۹ درصد خوش حساب و سهم شرکت‌های بورسی بیشتر از غیر بورسی است. دوره زمانی این پژوهش ۱۳۹۶-۱۳۸۵ است و دقیقاً مشاهدات از سال شروع طبقه‌بندی دارائی‌ها طبق بخشنامه ذی‌ربط بانک‌مرکزی، مورد توجه قرار گرفته‌اند.

متغیر وابسته این مطالعه که مبتنی بر ریسک اعتباری است، از مانده مطالبات غیرجاری واقعی نسبت به تسهیلات اعطایی هر یک از پرونده‌های تسهیلاتی مشتریان حقوقی بانک نمونه استخراج شده و از این رو رویکردی مشابه باستوس (۲۰۱۰) در مدل‌سازی ریسک اعتباری دنبال شده است[۳]. براساس دستورالعمل بانک‌مرکزی در طبقه‌بندی دارایی‌های بانک‌ها و موسسات اعتباری، مطالبات غیرجاری شامل مطالبات سررسید گذشته، معوق و مشکوک‌الوصول است. با این وجود به دلیل آنکه در نهایت بخشی از مانده مطالبات غیرجاری وصول و بخش دیگری از آن همچنان غیرجاری باقی می‌ماند، بنابراین همه پرونده‌های تسهیلاتی غیرجاری، نباید یکسان در نظر گرفته شوند و استفاده از مانده مطالبات غیرجاری واقعی نسبت به تسهیلات اعطایی هر یک

از پرونده‌های تسهیلاتی مشتریان، تصریح دقیق‌تری از متغیر ریسک را در اختیار قرار می‌دهد و این مهم به‌معنای دقیق بودن میزان این متغیر نسبت به باینری بودن آن است.

در مورد متغیرهای توضیحی و اثرگذار بر مطالبات غیرجاری، براساس مطالعات ادبیات موضوع و پیشینه تحقیق، مجموعه متنوعی از متغیرها باید منظور شوند. در این مطالعه، با توجه به داده‌های موجود در پرونده تسهیلاتی مشتریان حقوقی و در دسترس بودن سایر داده‌ها و به پشتوانه ادبیات نظری و تجربی، از پنج گروه کلی از متغیرهای توضیحی، شامل: ویژگی‌های خصوصیتی قراردادها، نسبت‌های مالی، شاخص‌های غیرمالی، حسابرسی و اقتصادی استفاده شده است. از این رو توجه به دامنه گسترده‌ای از انواع متغیرها در کنار یکدیگر و پرداختن به متغیرهای حسابرسی مستخرج از گزارش حسابرسی، از مهمترین ویژگی‌های متمایزکننده این پژوهش نسبت به مطالعات پیشین است.

به این ترتیب با مطالعه پژوهش‌های پیشین، پرتکرارترین و مرتبط‌ترین متغیرها در شکل‌گیری مطالبات غیرجاری جمع‌آوری؛ و با توجه به گزارش‌های مستمر عملکرد نظام بانکی، داده‌های موجود در بانک، در کنار نظر کارشناسان امور بانکی، متغیرهای پژوهش، نهایی و در قالب جدول (۱) ارائه شد، البته از بین آنها متغیرهای با همبستگی بالا حذف شدند.

به این ترتیب براساس اطلاعات مندرج در جدول ۱، در گروه متغیرهای بیانگر خصوصیات قراردادهای تسهیلاتی منعقد شده با مشتریان حقوقی بانک نمونه، ۱۲ متغیر، گروه شاخص‌های مالی ۲۹ متغیر، گروه شاخص‌های غیرمالی ۱۱ متغیر، گروه شاخص‌های حسابرسی ۴ متغیر، و در گروه شاخص‌های اقتصادی ۵ متغیر مورد توجه قرار گرفته‌اند.

جدول ۱. متغیرهای پژوهش

طبقه اصلی	طبقه فرعی	نام متغیر	نحوه اندازه‌گیری و توضیح متغیر
شاخص‌های خصوصیتی قراردادها		مرجع تصویب‌کننده	۱. ارکان اعتباری مدیریت منطقه ۲. اداره تسهیلات ۳.هیات مدیره
		نوع عقد	۱. عقود مشارکتی ۲. عقود مبادله‌ای
		نوع قرارداد	۱. مشارکت مدنی ۲. فروش اقساطی ۳. سلف ۴. مرابحه ۵. مضاربه ۶. جعاله
		هدف از پرداخت تسهیلات	۱. تامین سرمایه ثابت ۲. تامین سرمایه در گردش
		مبلغ تسهیلات	مبلغ پرداخت‌شده در هر یک از قراردادها
		نحوه بازپرداخت	۱. پرداخت یکجا ۲. اقساطی
		مدت قرارداد	۱. تا یکسال ۲. یک تا دو سال ۳. دو تا سه سال ۴. بالای سه سال
		تغییر در وضعیت قرارداد	۱. بدون تغییر ۲. تمدید ۳. تقسیط مجدد
		نرخ تسهیلات	نرخ سود قرارداد (براساس بخشنامه‌های بانک)
		وثیقه	۱. اموال غیرمنقول محل فعالیت ۲. اموال غیرمنقول خارج از محل فعالیت ۳. ذمه‌ای ۴. سایر
		مراودات مالی شرکت با بانک	نسبت میانگین مراودات به مبلغ تسهیلات ۱. ضعیف ۲. متوسط ۳. خوب ۴. خیلی خوب
		وضعیت مشتری	رفتار مشتری در بازپرداخت تسهیلات قبلی ۱. ضعیف ۲. متوسط ۳. خوب ۴. خیلی خوب
شاخص‌های مالی	شاخص ایفای تعهدات	نسبت آنی	بدهی‌های جاری / دارایی‌های آنی
		نسبت جاری	بدهی‌های جاری/ دارایی‌های جاری
		دارایی جاری به دارایی	دارایی‌ها / دارایی جاری
		اندازه شرکت	براساس دارایی‌های شرکت (لگاریتم دارایی‌های شرکت)
		سرمایه در گردش به دارایی‌ها	جمع دارایی‌ها / سرمایه در گردش
		سرمایه در گردش به فروش	فروش / سرمایه در گردش
		نسبت پوشش بهره	هزینه بهره / سود قبل از بهره و مالیات
		نسبت مالکانه	جمع بدهی‌ها و حقوق صاحبان سهام / حقوق صاحبان سهام
		نسبت بدهی‌ها	جمع دارایی‌ها / جمع بدهی‌ها
	شاخص پایداری رشد شرکت	نرخ رشد سود	سود خالص دوره قبل / سود خالص دوره قبل- سود خالص دوره جاری
		نرخ رشد دارایی‌ها	دارایی دوره قبل / دارایی‌های دوره قبل- دارایی‌های دوره جاری
		نرخ رشد فروش	فروش دوره قبل / فروش دوره قبل- فروش جاری
		نسبت بقاء	سود خالص / سود انباشته
		سود انباشته به دارایی‌ها	کل دارایی‌ها / سود انباشته
	شاخص‌های تسهیلاتی	تسهیلات به دارایی‌ها	کل دارایی‌ها / مانده تسهیلات
		تسهیلات به بدهی‌ها	کل بدهی‌ها / مانده تسهیلات
		حصه جاری به دارایی جاری	دارایی جاری / حصه جاری تسهیلات
		تسهیلات به فروش	فروش / مانده تسهیلات
	شاخص‌های سودآوری	بازده دارایی‌ها	متوسط کل دارایی‌ها / سود خالص دوره
		بازده حقوق صاحبان سهام	متوسط حقوق صاحبان سهام / سود خالص دوره
		حاشیه سود	کل درآمد عملیاتی / سود خالص
	شاخص‌های جریان نقدی	گردش وجوه نقد	بدهی های جاری / وجه نقد حاصل از عملیات
		بازده نقدی دارایی‌ها	متوسط دارایی‌ها / وجه نقد حاصل از عملیات
		کیفیت سود	سود عملیاتی / وجه نقد حاصل از عملیات
		نقدی به هزینه بهره	هزینه بهره / جریان وجه نقد
	شاخص‌های فعالیت	گردش دارایی‌ها	متوسط دارایی‌های ثابت / فروش
		گردش موجودی	متوسط موجودی کالا/ بهای تمام‌شده کالای فروش رفته
		گردش حساب دریافتنی	متوسط حساب دریافتنی / فروش

	نسبت هزینه بدهی	متوسط بدهی‌ها / هزینه مالی
شاخص‌های غیرمالی	نوع شرکت	۰. سهامی خاص ۱. سهامی عام
	سن شرکت	تعداد سال فعالیت از زمان تاسیس
	نوع صنعت	طبقه‌بندی صنعت براساس سازمان بورس اوراق‌بهادار
	نوع فعالیت	۰. غیرتولیدی ۱. تولیدی
	سابقه بورس	تعداد سال فعالیت در بورس
	اندازه هیات مدیره	تعداد اعضای هیات مدیره
	تنوع محصول	تعداد محصول و یا خدمات
	تعداد کارکنان	تعداد کارکنان شرکت
	شهر محل فعالیت	۱. تهران ۲. مرکز استان ۳. شهرستان
	سابقه حساب	تعداد سال‌های حساب فعال نزد بانک تسهیلات‌دهنده
	سابقه تسهیلات	تعداد سال‌های اخذ تسهیلات از بانک تسهیلات‌دهنده
شاخص‌های حسابرسی	اظهارنظر حسابرسی	۱. مقبول ۲. مشروط ۳. مردود ۴. عدم اظهارنظر
	اظهارنظر حسابرس بابت تداوم فعالیت	۰. ابهام در تداوم فعالیت شرکت ۱. عدم ابهام در تداوم فعالیت
	بندهای اظهارنظر در تعدیل گزارش	تعداد بندهای گزارش حسابرسی در خصوص وضعیت مالی شرکت
	اندازه موسسه حسابرسی	۰. موسسه حسابرسی ۱. سازمان حسابرسی و موسسه مفید راهبر
شاخص‌های اقتصادی	نرخ تورم سالانه	بر اساس آمار استخراجی از بانک‌مرکزی
	شاخص بهای کالای مصرفی	براساس جداول آماری استخراجی از بانک‌مرکزی
	نرخ رشد تولید ناخالص داخلی	براساس جداول آماری استخراجی از بانک‌مرکزی (به تفکیک صنعت)
	شاخص قیمت مسکن	براساس جداول آماری استخراجی از بانک‌مرکزی
	نرخ بهره سالانه	براساس جداول آمار استخراجی از بانک‌مرکزی

٤. تحلیل داده‌ها و یافته‌ها

انتخاب ویژگی. در فرآیند انتخاب ویژگی، تعداد کمتری از متغیرهای مستقل، برای توصیف متغیر پاسخ انتخاب می‌شوند. ویژگی‌های مرتبط، تاثیر شگرفی بر عملکرد، دقت و کارایی مطلوب مدل برآوردشده خواهند داشت و ویژگی‌های نامرتبط یا تا حدودی مرتبط، می‌توانند تاثیر منفی بر عملکرد آن داشته باشند. توجه شود که انتخاب ویژگی، یک فرآیند مهم در مدل‌سازی آماری است زیرا تفسیر مدل را آسان‌تر کرده و متغیرهای اضافی که اطلاعات مفیدی به مدل اضافه نمی‌کنند را حذف می‌کند؛ این فرآیند اندازه مدل را کوچک‌تر، هزینه‌های محاسباتی را کاهش و سرعت پاسخ را افزایش می‌دهد؛ در حالی‌که تعداد بیش از اندازه متغیر مستقل، نتایج مخالفی را همراه خواهد داشت.

در این راستا در مقاله حاضر از دو الگوریتم همبستگی و لاسو در انتخاب ویژگی استفاده شده است. به طوری‌که به ازای هر نمونه جدید، پاسخ الگوریتم به صورت به هنگام[1] است و بار محاسباتی سیستم چندان مهم نیست چون تنها برای یک مرتبه اجرا خواهد شد. پس از این اجرا مدل به‌دست آمده برای نمونه‌های جدید نیز به صورت به هنگام جواب می دهد. در الگوریتم همبستگی، زیرمجموعه‌ای از ویژگی‌ها یک زیرمجموعه خوب را شکل می‌دهند که از یک سو، همبستگی بالایی با ویژگی هدف داشته باشند و از سوی دیگر، با یکدیگر ناهمبسته باشند. در الگوریتم لاسو نیز برای هر متغیری که به مجموعه متغیرهای مدل اضافه می‌شود، جریمه‌ای برای کاهش درجه آزادی مدل و جلوگیری از لحاظ تعداد زیاد متغیرهای مستقل در مدل منظور می شود و به این ترتیب، ضریب برخی متغیرها در مدل به صفر رسیده و می‌توان آنها را از مدل حذف کرد.

با استفاده از این الگوریتم‌ها، ۱۰ متغیر که تاثیرگذاری آنها نسبت به دیگر متغیرها بیشتر بود، انتخاب شدند. براساس نتایج جدول ۲ مشاهده می‌شود که سهم متغیرهای مالی در هر دو روش، نسبت به سایر متغیرها بیشتر است و متغیرهای خصوصیتی قرارداد در هیچ یک از این دو الگوریتم جایگاهی به خود اختصاص نداده‌اند. نقش متغیرهای اقتصادی نیز در الگوریتم لاسو پررنگ‌تر بوده و متغیرهای غیرمالی فقط در الگوریتم همبستگی اثرگذارند. در هر دو الگوریتم مورد استفاده نیز صرفاً یک متغیر حسابرسی متفاوت، به مجموعه متغیرهای فوق راه پیدا کردند.

جدول۲. انتخاب متغیرهای توضیحی اثرگذار

متغیرهای انتخابی با الگوریتم همبستگی		**متغیرهای انتخابی با الگوریتم لاسو**	
گروه	**متغیر**	**گروه**	**متغیر**
مالی	سرمایه در گردش به دارایی‌ها	مالی	رشد سود خالص
	نسبت مالکانه		سود انباشته به داراییها
	سود انباشته به دارایی‌ها		بازده حقوق صاحبان سهام
	بازده دارایی‌ها		گردش وجوه نقد
	گردش وجوه نقد		کیفیت سود
	بازده نقدی دارایی‌ها		گردش موجودی کالا
	گردش دارایی‌ها	اقتصادی	نرخ تورم
غیرمالی	نوع شرکت		نرخ رشد تولید ناخالص
اقتصادی	شاخص بهای کالای مصرفی		نرخ بهره
حسابرسی	اظهارنظر حسابرس	حسابرسی	ابهام تداوم فعالیت در گزارش حسابرس
تعداد گروه‌های پوشش داده‌شده = ۴		تعداد گروه‌های پوشش داده‌شده = ۳	

1.Realtime

هر چند در ادامه از دو رویکرد ماشین بردار پشتیبان و درخت تصمیم[۱] در مدل‌سازی استفاده می‌شود، با این وجود در ارزیابی انتخاب ویژگی‌ها توسط دو الگوریتم همبستگی و لاسو، از درخت تصمیم کمک گرفته می‌شود. از آنجا که در ماشین بردار پشتیبان، تمامی متغیرها با یکدیگر ا ستفاده می شوند و در محا سبات رفت و برگ شتی بر یکدیگر تاثیر می‌گذارند، نمی‌توان برای هر یک از متغیرها ضریب اهمیت را به‌صورت تکی محاسبه کرد. از این رو مقادیر فوق برای به تصویر کشیدن اهمیت نسبی این متغیرها، صرفاً برای درخت تصمیم محا سبه و گزارش، که در جدول ۳ نشان داده شده است.

جدول ۳. ضریب اهمیت ویژگی‌های اثرگذار الگوریتم همبستگی و لاسو به‌همراه درخت تصمیم

الگوریتم همبستگی		الگوریتم لاسو	
نام ویژگی	**ضریب اهمیت ویژگی**	**نام ویژگی**	**ضریب اهمیت ویژگی**
سود انباشته به دارایی‌ها	۰/۳۰۲۵	بازده حقوق صاحبان سهام	۰/۶۲۸۱
گردش وجوه نقد	۰/۲۳۶۴	نرخ تورم	۰/۳۱۲۷
سرمایه در گردش به دارایی‌ها	۰/۱۳۲۱	نرخ رشد تولید ناخالص	۰/۲۸۵۶
گردش دارایی‌ها	۰/۰۹۹۲	رشد سود خالص	۰/۲۲۶۸
نسبت مالکانه	۰/۰۵۴۹	کیفیت سود	۰/۱۲۳۲
نوع شرکت	۰/۰۴۵۸	گردش موجودی کالا	۰/۰۸۹۱
شاخص بهای کالای مصرفی	۰/۰۴۴۴	گردش وجوه نقد	۰/۰۸۷۵
بازده دارایی‌ها	۰/۰۳۹۳	نرخ بهره	۰/۰۸۲۴
بازده نقدی دارایی‌ها	۰/۰۰۷۲	سود انباشته به دارایی‌ها	۰/۰۷۹۹
اظهارنظر حسابرس	۰/۰۰۰۰	ابهام تداوم فعالیت در گزارش حسابرس	۰/۰۰۷۷

با توجه به مقادیر به دست آمده از جدول بالا، تمامی متغیرهای حاصل از الگوریتم لاسو در پیش‌بینی درخت ت صمیم تاثیرگذار بوده‌اند و تنها ویژگی ابهام تداوم فعالیت در گزارش ح سابرس تاثیرگذاری اندکی داشته است. از این رو اگر تنها از ۹ ویژگی دیگر استفاده کنیم، نتیجه به دست آمده از درخت تصمیم چندان تفاوتی نخواهد کرد. همچنین از بین مجموعه متغیرهای مورد بررسی، نقش متغیرهای مالی چشمگیر است و بعد از آن، متغیرهای غیرمالی و اقتصادی از ضریب اهمیت ویژگی بالاتری برخوردارند.

همچنین در ویژگی‌های حا صل از الگوریتم همبستگی، متغیر بازده نقدی دارایی‌ها، ضریب تاثیر اندکی دارد و ویژگی اظهارنظر ح سابرس، از ضریب تاثیر صفر برخوردار ا ست و به عبارتی عدم لحاظ این ویژگی حتی در نتیجه مدل تاثیری ندارد. از این رو در قالب این الگوریتم، متغیرهای مالی و اقتصادی ضریب اهمیت ویژگی بالایی دارند.

---

[۱]. در مدل‌سازی درخت تصمیم از الگوریتم C4.5 استفاده شده است.

با این حال چون در این پژوهش، هدف محاسبه و مقایسه میزان عملکرد الگوریتم‌های یادگیر ماشین بردار پشتیبان و درخت تصمیم با توجه به دامنه گسترده‌ای از متغیرهای اثرگذار است، در هر دو الگوریتم از ده متغیر به دست آمده استفاده شده است. به بیان دقیق‌تر، از حذف ویژگی اظهارنظر حسابرس از مجموعه ویژگی‌های انتخابی در قالب الگوریتم لاسو در حالت به‌کارگیری درخت تصمیم صرف‌نظر شده است.

بنابراین شواهد فوق بیانگر عدم اثرگذاری ویژگی‌های مبتنی بر خصوصیات قرارداد، اهمیت فوق العاده اندک متغیرهای مستخرج از گزارش‌های حسابرسی، و نقش اندک متغیرهای غیرمالی در ارزیابی شایستگی اعتباری مشتریان حقوقی نمونه توسط الگوریتم‌های مورد استفاده در قالب مدل درخت تصمیم بوده است.

از این رو نتایج حاصله تاییدکننده بسیاری از مطالعات متمرکز بر متغیرهای مالی (در تطابق با مطالعات میرغفوری و امین‌آشوری، ۱۳۹۴؛ محمدیان حاجی‌کُرد و همکاران، ۱۳۹۵؛ دارابی و مشایخی، ۱۳۹۵؛ کارا و کری چن، ۲۰۱۲، وانگ و ما، ۲۰۱۲، آدو و همکاران، ۲۰۱۸، پرز–مارتین و همکاران، ۲۰۱۸)، و اقتصادی (در تطابق با مطالعات نوروزی، ۱۳۹۳؛ رستم‌زاده و همکاران، ۱۳۹۷؛ رستمی و همکاران، ۱۳۹۷؛ محقق‌نیا و همکاران، ۱۳۹۷؛ بخت و التر، ۲۰۱۴) در پیش‌بینی این ریسک در مقایسه با سه گروه دیگر متغیر مورد مطالعه است.

انتخاب الگوریتم کارا. با توجه به تنوع روش‌های مورد استفاده در تصریح مدل ریسک اعتباری در مطالعات مبتنی بر انواع مختلفی از مدل‌های مبتنی بر هوش مصنوعی، به‌دلیل تاکید بر ضرورت به‌کارگیری مدل‌های ناپارامتریک مبتنی بر درخت تصمیم (آدو و همکاران، ۲۰۱۸؛ باستوس، ۲۰۱۰؛ میرغفوری و امین‌آشوری، ۱۳۹۴) و ماشین بردار پشتیبان (یااو[1] و همکاران، ۲۰۱۵؛ وانگ و ما، ۲۰۱۲؛ جعفری‌اسکندری و روحی ۱۳۹۶؛ دارابی و مشایخی، ۱۳۹۵؛ محمدیان‌حاجی‌کُرد و همکاران، ۱۳۹۵)، در کنار ضرورت پاسخ به سوال مقاله در خصوص مقایسه کارکرد بهتر مدل ماشین بردار پشتیبان و درخت تصمیم، در ادامه از این دو الگوریتم در آزمون ویژگی‌ها در قالب انتخاب مدل کارا استفاده می‌شود.

به این ترتیب با ترکیب دو الگوریتم انتخاب ویژگی همبستگی و لاسو، در کنار الگوریتم‌های ماشین بردار پشتیبان و درخت تصمیم در مدل‌سازی ریسک اعتباری و آزمون ویژگی‌ها، در نهایت با چهار الگوی ترکیبی منتخب و مندرج در جدول ۴ در انتخاب ویژگی‌ها و آزمون آنها در قالب مدل‌سازی ریسک اعتباری مواجه می‌شویم.

حال برای ارزیابی دقت این الگوها در دو نمونه آموزش و آزمون، از معیارهای متداول مبتنی بر میانگین مربعات خطا (MSE[2]) و میانگین مطلق خطا (MAE[3]) که معیارهایی برای تشخیص

---

[1] . Yao

[2]. Mean Square Error (MSE)

[3]. Mean Absolut Error (MAE)

تفاوت بین مقادیر پیش‌بینی شده و واقعی هستند و همچنین ضریب تعیین (R^2) استفاده شده است. البته در کنار موارد فوق، به مقدار احتمال معناداری مدل (P- value) در دو ستون آخر جدول ۴ نیز اشاره می‌شود.

از آنجا که داده‌های آموزش و آزمون به‌صورت تصادفی انتخاب می‌شوند، هر یک از روش‌ها ۱۰ مرتبه تکرار شده و میانگین این ۱۰ مرتبه تکرار به عنوان نتیجه نهایی ثبت شده است. براساس نتایج حاصله مشاهده می شود برای داده‌های آزمون، با استفاده از ویژگی‌های به‌دست آمده از الگوریتم لاسو و درخت تصمیم، معیارهای (MSE) و (MAE) کمترین مقادیر را دارا هستند. همچنین مقدار ضریب تعیین به‌دست آمده در درخت تصمیم با استفاده از ویژگی‌های انتخابی توسط الگوریتم لاسو برای داده‌های آزمون، برابر با ۰,۹۱۲۳ و بیشتر از نتایج حاصله برای مورد مشابه در دیگر حالات مدل و الگوریتم انتخاب ویژگی است.

همچنین در ستون آخر این جدول، مقادیر احتمال (P- value) حاصل از آزمون آماری t با درصد اطمینان ۹۵ درصد گزارش شده است. از آنجا که مقادیر (P- value) کمتر از ۰,۰۵ نشان‌دهنده وجود ارتباط معنادار برای نتایج حاصله بین متغیرهای مستقل و متغیر وابسته است، بنابراین نتایج بیانگر تایید ارتباط معنادار در تمامی حالات پیاده‌سازی مدل است.

با این وجود در مجموع ارزیابی نتایج به‌دست آمده با در نظر گرفتن معیارهای MSE ، MAE و R^2، بیانگر اخذ نتیجه بهتر (به ترتیب ۰/۰۲۹۷ ، ۰/۰۸۱۲ و ۰/۹۱۲۳) در حالت استفاده از الگوریتم لاسو به همراه درخت تصمیم در داده‌های آزمون است. از این رو نتایج، تاییدکننده قابلیت پیش‌بینی بهتر ریسک اعتباری توسط درخت تصمیم در مقایسه با رویکرد ماشین بردار پشتیبان در مطالعات پیشین (اَدو و همکاران، ۲۰۱۸؛ باستوس، ۲۰۱۰؛ میرغفوری و امین‌آشوری، ۱۳۹۴) است. حتی مقدار میانگین مربعات خطا در این مدل در قالب به‌کارگیری الگوریتم لاسو و درخت تصمیم (۰/۰۲۹۷) در مقایسه با نتایج حاصله در مدل به نسبت مشابه در مطالعات قبلی (مطالعه جعفری‌اسکندری و روحی (۱۳۹۶) که مقدار ۰/۰۳۲۰ را برای این متغیر گزارش کرده) بیانگر کارایی به نسبت مناسب این مدل در مقایسه با مطالعات داخلی در پیشینه تحقیق است.

جدول ۴. نتایج روش‌های پیاده‌سازی‌شده بر مجموعه داده‌های آموزش و آزمون

الگوریتم	MSE		MAE		R^2		P- value	
	آموزش	آزمون	آموزش	آزمون	آموزش	آزمون	آموزش	آزمون
SVR+CFS	۰/۰۵۲۹	۰/۰۷۳۳	۰/۱۲۵۰	۰/۱۵۳۸	۰/۷۷۴۱	۰/۹۲۶۷	۰/۰۰۴۳	۰/۰۰۰۶
SVR+Lasso	۰/۰۲۵۸	۰/۰۳۱۸	۰/۱۰۶۷	۰/۱۲۷۲	۰/۸۱۹۷	۰/۷۴۵۲	۰/۰۰۰۶	۰/۰۰۰۱
DT+CFS	۰/۰۱۹۵	۰/۰۳۴۳	۰/۰۵۳۶	۰/۰۸۸۶	۰/۹۰۳۷	۰/۸۹۷۰	۰/۰۰۶۵	۰/۰۰۷۱
DT+Lasso	۰/۰۱۷۲	۰/۰۲۹۷	۰/۰۵۸۷	۰/۰۸۱۲	۰/۹۴۲۴	۰/۹۱۲۳	۰/۰۰۴۶	۰/۰۰۲۷

در جدول ۴، CFS بیانگر الگوریتم همبستگی، Lasso الگوریتم لاسو، SVR ماشین بردار پشتیبان، و DT نمایانگر درخت تصمیم است.

علاوه‌براین با توجه به دامنه متغیرهای انتخاب‌شده در قالب الگوریتم لاسو (جدول ۲) و ضریب اهمیت این متغیرها در حالت درخت تصمیم (جدول ۳)، در نهایت مشاهده می‌شود که با وجود لحاظ دامنه گسترده‌ای از انواع متغیرها در مدل‌سازی ریسک اعتباری مشتریان حقوقی در مدل، متغیرهای غیرمالی و خصوصیتی قراردادها نقشی در ارزیابی ریسک اعتباری این گروه از مشتریان نداشته و حتی میزان اهمیت متغیرهای حسابرسی در این گروه در حد ناچیز و قابل اغماض است. بنابراین همانند مطالعات سنتی قبلی، نقش کلیدی متغیرهای مالی و اقتصادی، در صورت به‌کارگیری درخت تصمیم در مدل‌سازی ریسک اعتباری در مقایسه با گروه دیگر متغیرها به اثبات می‌رسد.

۵. بحث و نتیجه‌گیری

در بازارهای مالی یکی از ریسک‌های مهم و مطرح که شناسایی، اندازه‌گیری و کنترل آن می‌تواند نقش به سزایی در پایین آوردن هزینه‌ها، افزایش درآمدها و در نتیجه سودآوری بانک‌ها داشته باشد، ریسک اعتباری است. از این رو مدیریت این ریسک در قالب طراحی مدلی برای اندازه‌گیری آن، در ارتقاء اثربخش عملکرد بانک‌ها و رشد پایدار آنها، به‌ویژه در شبکه بانکی کشور با نرخ بالای مطالبات غیرجاری، نقشی اساسی بازی می‌کند. این مساله در بانک سپه که از سال ۱۳۹۶ با ادغام در پنج بانک و موسسه اعتباری به یکی از بزرگترین بانک‌های کشور تبدیل شده، از اهمیت بیشتری به دلیل ریسک سیستمی در نظام مالی برخوردار است.

بنابراین در این مقاله، مدل‌سازی این ریسک در این بانک با دسترسی به ۶۷۷ پرونده تسهیلاتی مشتریان حقوقی با رویکردی جامع‌تر در مقایسه با مطالعات قبلی، با تمرکز بر اطلاعات ۱۳۸۵–۱۳۹۶ و استفاده از نسبت واقعی مطالبات غیرجاری به‌عنوان نماد ریسک اعتباری دنبال شد. در این رویکرد جهت پاسخ به این سوال که کدام‌یک از مدل‌های ماشین بردار پشتیبان و درخت تصمیم دقت بیشتری در ارزیابی ریسک اعتباری دارند، دامنه گسترده‌تری از متغیرها مورد توجه قرار گرفته و در کنار متغیرهای متعارف مالی و اقتصادی، متغیرهای غیرمالی، خصوصیتی قراردادها، و حسابرسی نیز در شناسایی عوامل اثرگذار بر ریسک اعتباری در نظر گرفته شده است.

با این وجود نتایج حاصل از به‌کارگیری الگوریتم‌های لاسو و همبستگی در انتخاب ویژگی‌های تاثیرگذار بر ریسک اعتباری، الگوریتم لاسو عملکرد بهتری نسبت به الگوریتم همبستگی داشته و به انتخاب ۱۰ متغیر کلیدی تاثیرگذار از گروه متغیرهای مالی، اقتصادی و حسابرسی منتهی شده است. با این حال الزاماً نه تنها در هیچ یک از این الگوریتم‌ها اثربخشی متغیرهای خصوصیتی قراردادها تایید نشده؛ بلکه بیانگر اثربخشی متغیرهای غیرمالی نیز نبوده و صرفاً متغیرهای مالی،

اقتصادی و حسابرسی در هر دو الگوریتم انتخاب شده‌اند که نقش متغیرهای حسابرسی نیز قابل اغماض بوده است.

در ارزیابی ریسک اعتباری با الگوریتم‌های یادگیری ماشین‌بردار پشتیبان و درخت تصمیم نیز، در آزمون ویژگی‌ها و با توجه به عملکرد بهتر الگوریتم لاسو به‌همراه درخت تصمیم، نه تنها به نتایجی تاییدکننده قابلیت پیش‌بینی بهتر ریسک اعتباری توسط درخت تصمیم در مقایسه با رویکرد جایگزین مبتنی بر ماشین بردار پشتیبان در مطالعات پیشین (اَدو و همکاران، ۲۰۱۸؛ باستوس، ۲۰۱۰؛ میرغفوری و امین‌آشوری، ۱۳۹۴) منتهی شده است؛ بلکه نتایج این الگو بیانگر کفایت تمرکز بر متغیرهای مالی (در تطابق با مطالعات میرغفوری و امین‌آشوری، ۱۳۹۴؛ محمدیان‌حاجی‌کُرد و همکاران، ۱۳۹۵؛ دارابی و مشایخی، ۱۳۹۵)، و اقتصادی (در تطابق با مطالعات نوروزی، ۱۳۹۳؛ رستم‌زاده و همکاران، ۱۳۹۷؛ رستمی و همکاران، ۱۳۹۷؛ محقق‌نیا و همکاران، ۱۳۹۷) در پیش‌بینی این ریسک در مقایسه با دیگر متغیرهای غیرمالی و خصوصیت قراردادها و حتی حسابرسی در استفاده از این الگوریتم است.

٦. پیشنهادها و محدودیت‌ها

براساس نتایج، استفاده از الگوریتم لاسو در انتخاب ویژگی‌ها، و درخت تصمیم در آزمون ویژگی‌ها جهت مدل‌سازی ریسک اعتباری مشتریان حقوقی به کارشناسان اعتباری و ارکان اعتباری بانک‌ها پیشنهاد می‌شود. انتظار می‌رود این رویکرد بتواند در کاهش ریسک اعتباری بانک‌ها و مدل‌سازی کارای این متغیر در بخش اعظمی از مشتریان در مقایسه با ماشین‌بردار پشتیبان اثرگذار باشد.

نظر به اینکه مشتریان بانک‌ها شامل اشخاص حقیقی و حقوقی است و پژوهش حاضر صرفاً بر ویژگی‌های اشخاص حقوقی متمرکز است، بنابراین پیشنهاد می‌شود در پژوهش‌های آتی با محوریت این پژوهش، جامعه آماری نمونه به گروه اشخاص حقیقی نیز گسترش یابد و حتی محدود به گروه فوق نیز بشود تا الگوی ارزیابی ریسک اعتباری در دو گروه مشتریان شبکه بانکی به صورت جامع‌تری آزمون شود. این رویکرد می تواند به نحو بهتری اثر متغیرهای خصوصیتی قراردادها را در ارزیابی ریسک اعتباری آزمون کند که در این نمونه به عنوان متغیرهای اثرگذار شناسایی نشدند.

از آنجا که داده‌های اقتصادی در نماگرهای اقتصادی بانک مرکزی در زمان انجام این تحقیق تا آبان‌ماه ۱۳۹۷ به روزرسانی شده بود و اطلاعات پایان سال ۱۳۹۷ این متغیرها در دسترس نبود، امکان بسط دوره نمونه به پایان این سال فراهم نشد و سال ۱۳۹۶ به عنوان سال پایانی نمونه منظور شد. این مساله از جمله محدودیت‌های این تحقیق به شمار می رود و انتظار می‌رود با افزایش سرعت به روز رسانی داده‌ها در نهاد ذی‌ربط، این محدودیت آماری برطرف شود.

منابع

1. Addo, P. M., Guegan, D., & Hassani, B., (2018). Credit Risk Analysis Using Machine And Deep Learning Models. Risks, 6(2). 38.

2. Aghaii, M., Rezagholizadeh. M. (2016). Investigating the Factors Affecting the Volume of Overdue and Overdue Receivables of Selected Branches of Sepah Bank. Quarterly Journal of Islamic Finance and Banking Studies. 2(3). 95-111. (in Persian).

3. Bastos, J. A. (2010). Forecasting Bank Loans Loss-Given-Default. Journal of Banking & Finance, 34 (10), 2510-2517.

4. Chow, J. K. (2017). Analysis of Financial Credit Risk Using Machine Learning, Master of Business Administration Dissertation, Aston Business School, Aston University, Birmingham, United Kingdom.

5. Darabi, R. & Mashayekhi, GH. (2016). The Impact of Financial Intelligence on Credit Risk Prediction Using the Support Vector Machine Model. Accounting and Auditing Research. 9 (30). 1-22. (in Persian).

6. Ebrahimi, M. & Daryabar, A. (2011). Credit Risk Management in the Banking System - Data Envelopment Analysis and Logistic and Neural Network Approach. Journal of Investment Knowledge, 1(2):35-62. (in Persian).

7. Falahpour, S., Raai, R. & Hendijani, M. (2014). Neural Network Approach Based on Artificial Bee Colony for Estimating Credit Rating of Bank Clients. Journal of Financial Engineering and Bond Management.5 (21). 33-53. (in Persian).

8. General Management of Banks and Credit Institutions Supervision. (2005). Principles on Credit Risk Management. Banking Studies and Regulations Office, Central bank of Islamic Republic of Iran. (in Persian).

9. Habibi, R., Kouhi, H. & Baaidi, H. (2018). Bank Facilitation Decisions Using Genetic Algorithm Method (Case Study: Sepah Bank's Real Customers), Islamic Financial and Banking Studies Quarterly. 4 (9). 33-71. (in Persian).

10. Hori, M. & Kaveh, M. (2015). Designing a Model to Predict the Credit Score of Banking Customers Using Multi-criteria Hyper-Algorithm and Ant-Fuzzy-Colony Neural Network Algorithm (Case Study of Tehran Bank Post Branches). Iranian Journal of Management Research. 19 (1). 91-116. (in Persian).

11. Jafari Eskandari. M. & Rouhi, M. (2017). Credit Risk Management of Banking Customers Using Support Vector Machine Optimized by Genetic Algorithm with Data Mining Approach. Journal of Asset Management and Financing. 5(4). 17-32. (in Persian).

12. Kabari, L. G., & Nwachukwu, E. O. (2013). Credit Risk Evaluating System Using Decision Tree – Neuro Based Model. International Journal of Engineering Research & Technology, 2(6), 2738-2745.

13. Karaa, A., Krichene, A. (2012). Credit–Risk Assessment Using Support Vector Machine and Multilayer Neural Network Models: A Comparative Study - Case of Tunisian Bank. Journal of Accounting and Management Information Systems. 11(4), 587-620.

14. Kimura, H., Basso, L. & Kayo, E. (2015). Decision Models in Credit Risk Management. Decision Models in Engineering and Management, Springer, pp.57-73.

15. Mansourfar, Gh., Piri, P., Alikhani, Z. & Asadi, M. (2018). Predicting Financial Distress Given the Moderating Effects of the Independent Auditor's Report. 16th Iranian National Accounting Conference, Isfahan. (in Persian).

16. Mirghafouri, H. & Amin Ashouri, Z. (2015). Credit Risk Assessment of Bank Customers. Two Business Management Quarterly.7 (13). 247-266. (in Persian).

17. Min, J. H. & Lee, Y-C. (2008). A Practical Approach to Credit Scoring. Expert Systems with Applications, 35, 1762–1770.

18. Mohagheghneia, M., Dehghan Dehnavi, M. & Baai, M. (2019). The Impact of Internal and External Factors on the Banking Credit Risk in Iranian Banking Industry. Journal of Financial Economics.13 (46). 127-144. (in Persian).

19. Mohammadiyan Haji Kord, A., Asgharzadeh Zaafarani, M. & Emamdoost, M. (2016). Credit Risk Assessment of Corporate Customers Using Support Vector Machine and Genetic Algorithm Hybrid Model - a Case Study of Tejarat Bank. 7 (27). 17-32. (in Persian).

20. Norouzi, P. (2013). The Impact of Large Variables on Credit Risk of Banks in Iran. . Journal of Monetary& Banking Research, 7 (20). 237-257. (in Persian).

21. Office of Economic Research and Policies. (2006-2017). Economic Indicators. Central bank of Islamic Republic of Iran. (in Persian).

22. Office of Banking Studies and Regulations. (2006). Guidelines for Classifying the Assets of Credit Institutions. Central bank of Islamic Republic of Iran. (in Persian).

23. Pérez-Martín, A., & Pérez-Torregrosa, A., & Vaca, M. (2018). Big Data Techniques to Measure Credit Banking Risk in Home Equity Loans. Journal of Business Research, 89, 448-454.

24. Pourkazemi, M., Sedaghatparrast, E. and Dehpanah, R. (2017). Estimation of the Bank's Real Customer Failure Estimation Using Neural Networks Method. (Case Study: Pasargad Bank). Islamic Financial and Banking Studies Quarterly, 6.1-24. (in Persian).

25. Raai, R. & Falahpour, S. (2008) Application of Support Vector Machine in Predicting Corporate Financial Distress by Using Financial Ratios, Accounting and Auditing Reviews. 15 (53). 17-34. (in Persian).

26. Rostami, M., Nabizade, A & Shahi, Z. (2018). Factors Affecting Credit Risk of Commercial Banks of Iran with Emphasis on Banking and Macroeconomic Specific Factors. Journal of Asset Management and Financing.6 (4). 79-92. (in Persian).

27. Rostamzadeh, P., Shahnazi, R. & Neisani, M. (2018). Identification of Factors Affecting on Credit Risk in the Iran Banking Industry of Iran Using Stress Test. Journal of Research in Economic Modeling.8 (32). 91-128. (in Persian).

28. Safari, S., Ebrahimi, M. & Taheri, M. (2011). Credit Risk Management in the Banking System Comparative Approach to Data Envelopment Analysis and Neural Network. Scientific-Research Journal of Shahed University. 9(47). 121-140. (in Persian).

29. Taghipour, M., Saghaiy, A. & Bagheri, M. (2015). Investigating the Factors Affecting the Credit Risk Assessment of Bank Clients Using a Combined Approach to Data Mining Techniques. Conference on Industrial Engineering, Management and Accounting. (in Persian).

30. Tari. F., Ebrahimi, A., Mousavi, J. & Kalantari, M. (2017). Comparison Between Neural Network, Genetic Algorithm and Logit Models in Evaluating Consumer Credit Risk. Journal of Monetary & Banking Research, (34). 657-680. (in Persian).

31. Yao, X. Crook, J, & Andreeva, G. (2015). Support Vector Regression for Loss Given Default Modelling. European Journal of Operational Research. 240(2), 528-538.

32. Yanping, Y., Zhengming, Q., Min, Y., Rui, G., Liting, F., Penghui, G. (2012). Research on the Application of Decision Tree to the Analysis of Individual Credit Risk. Information Technology, 25, 209-214.

33. Yu, L., Wang, S. Y., Lai, K. K. (2008). Credit Risk Assessment with a Multistage Neural Network Ensemble Learning Approach. Expert Systems with Applications. 34(2), 1434–1444

34. Wang, G., Ma, J. (2012). A Hybrid Ensemble Approach for Enterprise Credit Risk Assessment Based on Support Vector Machine. Expert Systems with Applications. 39(5), 5325–5331.