فصلنامه پژوهش‌های پولی_بانکی
سال شانزدهم، شماره ۵۷، پاییز ۱۴۰۲
صفحات ۴۰۹_۴۳۷

ارزیابی اثر شبکهٔ اجتماعی در امتیاز اعتباری افراد با روش‌های یادگیری عمیق ماشین

رضا حیدری [*] **سیدموسی خادمی** [†]

تاریخ دریافت: ۱۴۰۲/۱۱/۲۷ تاریخ پذیرش: ۱۴۰۳/۰۳/۰۶

چکیده

مؤسسات مالی از امتیازدهی اعتباری برای ارزیابی خطر بالقوهٔ نکول وام‌گیرندگان استفاده می‌کنند. در سال‌های اخیر، بسیاری از وام‌دهندگان به پتانسیل خوب وام‌گیرندگان کم‌سابقه یا فاقد سابقهٔ مالی کافی پی برده و به‌دنبال استفاده از انواع داده‌های جایگزین به‌منظور جبران کمبود داده‌های سابقهٔ اعتباری برای محاسبهٔ احتمال نکول و امتیاز اعتباری‌اند. این تحقیق به‌دنبال بررسی اثر متغیرها و داده‌های مرتبط با شبکهٔ اجتماعی افراد بر امتیاز اعتباری آن‌هاست. این کار با یافتن اطلاعات معنی‌دار در مورد داده‌های اجتماعی افراد انجام می‌شود تا چگونگی اثر چنین داده‌هایی در امتیاز اعتباری آن‌ها اندازه‌گیری شود. فرضیهٔ اساسی این تحقیق این است که افراد با امتیاز اعتباری بالا با افرادی مشابه و هم‌سنخ با خود ارتباط اجتماعی دارند. در این تحقیق، مجموعه داده‌ای بالغ بر ۳۰۰ هزار از وام‌هایی که توسط یک بانک ایرانی به افراد حقیقی پرداخت شده است، برای تأیید و توضیح اثر متغیرهای شبکهٔ اجتماعی در امتیاز اعتباری مورداستفاده قرار گرفته است. بررسی‌های انجام‌شده با استفاده از روش رگرسیون لجستیک نشان می‌دهد که از نظر آماری، متغیرهای اجتماعی افراد قابلیت پیش‌بینی احتمال نکول وام آن‌ها را دارد. نتایج الگوریتم‌های یادگیری ماشین نیز نشان می‌دهد که اطلاعات شبکهٔ اجتماعی می‌تواند عملکرد پیش‌بینی نکول وام را به‌طور قابل‌توجهی بهبود بخشد.

واژه‌های کلیدی: امتیازدهی اعتباری، احتمال نکول، شمول مالی، عدم تقارن اطلاعاتی، شبکهٔ اجتماعی، پیوند اجتماعی بد
طبقه‌بندی JEL: C11، C51، C58، D81، G20، G41

[*] کارشناسی ارشد اقتصاد دانشگاه تهران و کارشناسی ارشد مدیریت فنّاوری اطلاعات دانشگاه پیام نور، واحد تهران غرب؛ rheydari2017@gmail.com
[†] عضو هیئت‌علمی دانشگاه پیام نور، واحد تهران غرب؛ sm.khadmi@pnu.ir

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]
[DOR: 20.1001.1.26453355.1402.16.57.2.0]

۱ مقدمه

در سال‌های اخیر، تغییرات و پیشرفت‌های فنّاوری اطلاعات موجب تغییر سبک زندگی اشخاص از سنتی به دیجیتال و افزایش رقابت در کسب‌وکارهای الکترونیک برای درآمدزایی بیشتر شده است. به‌طور خلاصه، تغییرات و پیشرفت‌های فنّاوری به دو طریق در حوزهٔ اعتبارسنجی تأثیر داشته است؛ اول، از طریق ایجاد منابع داده‌ای جدید در مورد مشتری و دوم، از طریق بهبود روش‌ها و قدرت محاسباتی ابزارهای اعتبارسنجی با استفاده از فنّاوری‌های جدید. منظور از منابع داده‌ای جدید کلیهٔ داده‌هایی است که از رفتارهای مشتری در استفاده از بسترهای فنّاوری حاصل می‌شود؛ مانند رفتار خرید اینترنتی وی و/یا فعالیت وی در شبکهٔ اجتماعی (تعداد پست‌ها، لایک‌ها، نظردهی، بازدیدها، و...).

در این میان، مؤسسات اعتبارسنجی با تحلیل داده‌های متنوع و جدید حاصل از رفتارهای افراد از یک‌سو و استفاده از مدل‌هایی با قدرت پیش‌بینی‌کنندگی بیشتر از سوی دیگر، توانستند بینش عمیق‌تری نسبت‌به اعتبار متقاضیان پیدا کرده و مدل‌های دقیق‌تری برای اعتبارسنجی ایجاد کنند. استفاده از روش‌های فنی جدید نیز در اعتبارسنجی به‌طور قابل‌توجهی در سال‌های اخیر افزایش داشته است. در همین راستا، استفاده از روش‌های پیچیده و نوین در سنجش اعتبار ــ مانند هوش مصنوعی ــ و روش‌های پیشرفتهٔ یادگیری ماشین ــ مانند جنگل‌های تصادفی ــ و تقویت گرادیان و شبکه‌های عصبی عمیق جایگزین موارد سنتی مانند تحلیل تمایزی خطی و رگرسیون لجستیک شده‌اند. مدل‌های اعتبارسنجی جدید مبتنی بر ماشین برخلافِ روش‌های سنتی این امکان را فراهم می‌کنند تا روابط بین متغیرها بدون محدودیت در تعداد متغیرها کشف و مورداستفاده قرار گیرد؛ برای مثال، با این روش‌ها می‌توان در دریاچهٔ داده‌های مشتریان با بیش از ۳۰۰۰ متغیر جست‌وجو کرد و هر رابطه‌ای را که مرتبط با فرضیه است، کشف و استفاده کرد. این کار از قدرت و توان روش‌های رگرسیون سنتی خارج است.

۱-۱ بیان مسئله

ذی‌نفعان حوزهٔ اعتبارسنجی را می‌توان به سه دستهٔ کلی مشتریان یا وام‌گیرندگان، تولیدکنندگان گزارش‌های اعتبارسنجی یا آژانس‌های مرجع اعتبار، و مؤسسات تأمین‌مالی و اعتباری تقسیم کرد. با توجه به نیازهای این ذی‌نفعان، اهمیت استفاده از داده‌های اجتماعی به‌منظور بهبود نظام اعتبارسنجی در قالب موارد ذیل قابل طرح است:

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]

مشتریان در فقدان یا ناکارآمدی سیستم‌های اعتبارسنجی، به دام بیماری وثیقه‌محوری بانک‌ها گرفتار شده‌اند. استفاده از داده‌های جایگزین در کنار داده‌های مالی و ارتقای سیستم اعتبارسنجی، کمک بزرگی به برقراری عدالت اجتماعی از طریق ارتقای شمول مالی می‌کند؛ برای مثال، فردی به‌تازگی از رشته‌ای با چشم‌انداز درآمدی خوب فارغ‌التحصیل شده است، اما اعتبارسنجیِ صرفاً با داده‌های مالی، با شرایط وی سازگار نیست و لذا از خدمات بانکی محروم می‌شود. دلیل این است که این شخص تازه وارد بازار کار شده است و حساب جاری و/ یا سابقهٔ اخذ تسهیلات بانکی ندارد و لذا مدل‌های اعتبارسنجی مبتنی بر داده‌های مالی صرف نتیجه‌ای را به‌دنبال نخواهند داشت.

برای وام‌دهندهٔ هوشمند، داده‌هایی که از شبکه‌های اجتماعی افراد به‌دست می آید، در واقع نوعی تأیید برای این است که آنان چه کسی هستند ــ اول از همه، اینکه واقعی هستند و دوم اینکه اگر فرد خوبی‌اند، باید با اطرافیان خود ارتباط خوبی داشته باشند؛ مثلاً باید تعداد معینی دوست داشته باشند، شماره تلفن همراه را برای مدت معینی مالک باشند، و...

با حدس نزدیک به واقع، امروزه نفوذ تلفن همراه در جهان در واقع بیش از ۱۰۰ درصد است. مردم بیش از یک تلفن همراه دارند. آن‌ها همچنین به برخی از سیستم‌هایی که با تلفن‌های همراه ارائه می‌شوند، مانند رسانه‌های اجتماعی، دسترسی دارند. در چنین محیطی، می‌توان به داده‌های بسیار دقیقی در مورد استفادهٔ مردم از تلفن همراه و برنامه‌هایی که از این طریق استفاده می‌شود، دست یافت.

آنچه مسلم است، تولید گزارش‌های اعتبارسنجی در لحظهٔ نیاز مشتری و مؤسسات اعتباری و استفاده از داده‌های بیشتر در مدل آن، در این شرایط سخت رقابتی و کمبود درآمد می‌تواند موجب تسریع در فرایندهای اعطای تسهیلات و اعتبار شده و همچنین شناسایی شاخصه‌ها و رفتارهای بد اعتباری، موجب کنترل و کاهش هزینه‌های ناشی از نکول یا پیگیری و وصول مطالبات شود.

ایجاد سیستم اعتبارسنجی مناسب موجب بهبود رشد و توسعهٔ اقتصادی می‌شود. این موضوع برای مثال با اعطای تسهیلات برای راه‌اندازی کسب و کارهای ساده، تأمین سرمایهٔ در گردش واحدهای کشاورزی، یا آموزش مهارت های خاص امکان‌پذیر می‌شود. دقیقاً به همین دلیل است که نیاز به استفاده از داده‌های غیرمالی در ارزیابی‌های اعتباری مهم است.

ایجاد فرهنگ دفاع از اعتبار خود در افراد جامعه به‌معنی رشد ثروت پنهان آن جامعه است. اینکه افراد چه‌کسی را می‌شناسند، مهم است. شبکهٔ اجتماعی افراد اهمیت دارد. این ایده که افرادی که در شبکهٔ اجتماعی در ارتباط‌اند، ویژگی‌ها و رفتارهای بسیار مشابهی با هم

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]

دارند، مفهوم و فرهنگی بسیار قوی و بسیار قدرتمند است که جامعه را به‌سمت خودکنترلی هدایت می‌کند.

تلاش در جهتِ مدیریت انواع ریسک‌ها موجب ایجاد و رسوب دانش‌های باارزشی در سازمان‌ها و افراد آن‌ها می‌شود.

با توجه به موارد گفته‌شده، مسئلهٔ موردتوجه این تحقیق آن است که در وضع موجود، متغیرهای مورداستفاده توسط نظام اعتبارسنجی کشور که عمدتاً بر داده‌های مالی و رفتار بانکی مشتریان تکیه دارد، برای امتیازدهی اعتباری افراد کافی نیست و به‌دنبال پاسخ به این سؤال است که آیا با افزودن داده‌های حاصل از فعالیت‌های ارتباطی و روابط اجتماعی افراد می‌توان این نظام اعتبارسنجی را بهبود بخشید؟

۱-۲ اهداف تحقیق

به‌طور خلاصه، اهداف اصلی این تحقیق را می‌توان موارد زیر عنوان کرد:

- کاهش عدم تقارن اطلاعاتی[1]. عدم تقارن اطلاعاتی به وضعیتی گفته می‌شود که دو طرف یک قرارداد یا دادوستد مالی، به اطلاعات طرف مقابل خود برای تصمیم‌گیری دسترسی کامل و/یا یکسان نداشته باشند. به‌عبارت دیگر، افراد برای وام‌گرفتن همهٔ اطلاعات خود را ابراز نمی‌کنند، لذا بانک‌ها اطلاعات محدود دارند، و نهایتاً به استعلامات از بانک مرکزی و مرکز ملی اعتبارسنجی (مؤسسهٔ رتبه‌بندی ایران) اکتفا می‌کنند. حال اگر روش‌های دیگری برای دسترسی به اطلاعات افراد وجود داشته باشد، این تقارن اطلاعاتی برای مؤسسات اعتباری کم می‌شود. بانک‌ها با بررسی داده‌های اجتماعی به‌موازات داده‌های مالی، شانس بیشتری برای شناخت مشتریان خواهند داشت و صرفاً به اطلاعات اظهاری آنان اکتفا نمی‌کنند.
- افزایش در شمول مالی[2]. به این معنی که بانک‌ها با استفاده از داده‌های اجتماعی در کنار داده‌های مالی، می‌توانند اطلاعات دقیق‌تری را در مورد مشتریان موجود و افراد با سابقهٔ مالی کم به‌دست بیاورند و در نتیجه، دامنهٔ ارائهٔ خدمات اعتباری خود را گسترش داده و همگانی کنند.

۱-۳ فرضیهٔ تحقیق

فرضیهٔ این تحقیق براساس گزارهٔ خبری ذیل تدوین و پیشنهاد شده است:

---

[1] information asymmetry

[2] financial inclusion

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]

«در فرایند تشکیل شبکهٔ اجتماعی، افراد ترجیح می‌دهند با افرادی دارای خلق‌وخو و شخصیت شبیه به خود روابط اجتماعی برقرار کنند.»

با توجه به گمانه‌زنی‌ها و مصاحبه با خبرگان اعتبارسنجی، فرضیهٔ زیر برای بررسی تأثیر روابط داخل شبکهٔ اجتماعی افراد با احتمال نکول[1] آن‌ها تعریف شد:

«فرضیهٔ صفر عبارت است از کسانی که وام‌ها را نکول می‌کنند (وام‌گیرندگان بد) و روابط اجتماعی بدکمتر یا مساوی با کسانی دارند که وام‌های خود را سر موقع بازپرداخت می‌کنند.»

توضیح اینکه در این تحقیق، در صورت گذشت زمان بیش از ۹۰ روز از سررسید بدهی پرداخت‌نشده، وام‌گیرنده به‌عنوان نکول‌کننده تعریف می‌شود. این زمان برای افراد عادی تا ۳۰ روز و برای افراد تأخیرکننده ۳۰ تا ۹۰ روز در نظر گرفته شده است.

۱-۴ مرور پیشینهٔ تحقیق

وکی و اِن تووئیگا[2] (۲۰۱۶) بر سه مزیت هنگام استفاده از داده‌های رسانه‌های اجتماعی در ارزیابی ریسک‌های اعتباری تأکید کردند: اول، غلبه بر آثار نامطلوب انتخاب با جذب وام‌گیرندگان ارزشمند که سابقهٔ مالی محدودی دارند؛ دوم، درک نیازهای واقعی مشتری و در نتیجه کاهش آثار مخاطرهٔ اخلاقی؛ و در نهایت، تطبیق اطلاعات ارائه‌شده توسط وام‌گیرندگان با پروفایل‌های رسانه‌های اجتماعی آن‌ها برای غلبه بر عدم تقارن اطلاعاتی.

برخی وام‌گیرندگان بر این باورند که وام‌دهندگان واقعاً اطلاعات کاملی را در مورد شرایط و ضوابط وام با هدف کسب سود بیشتر ارائه نمی‌دهند. در طرف مقابل، این شواهد نیز وجود دارد که وام‌گیرنده نیز اطلاعاتی را که نشان از توانایی ضعیف بازپرداخت آن‌هاست، اعلام نمی‌کند. این دو موضوع می‌تواند منجر به وضعیت نامتقارن اطلاعاتی شود که در نهایت باعث انتخاب نامطلوب یا مورد خطر اخلاقی می‌شود. این وضعیت در شکل ۱ توسط یان[3] و همکاران (۲۰۱۵) به تصویر کشیده شده است.

---

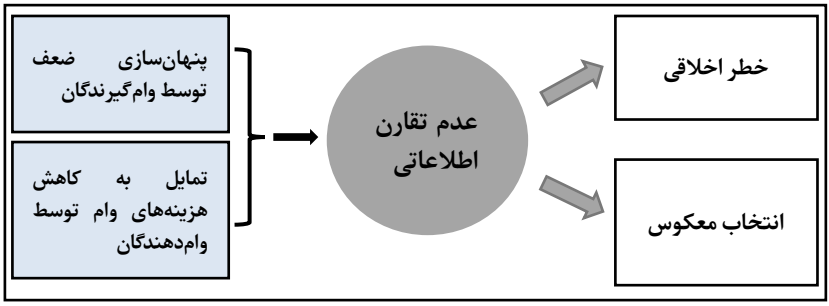
[1] default probability

[2] Weke & Ntwiga

[3] Yan

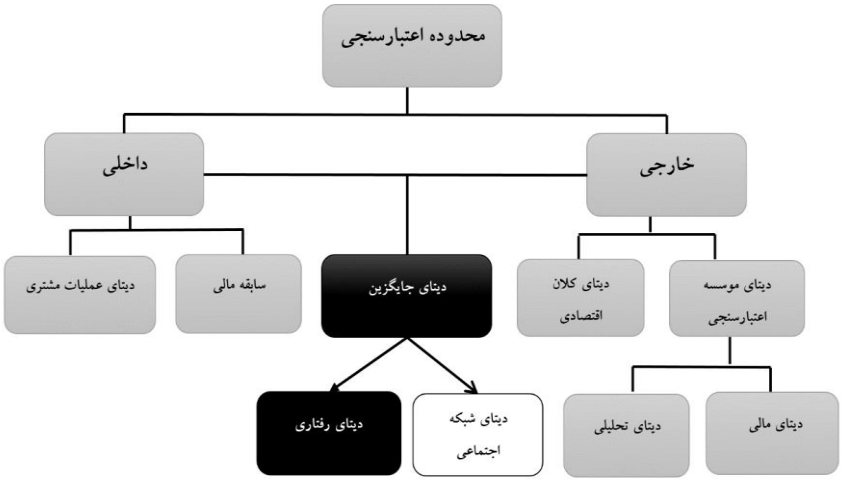
[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]



شکل ۱. علل و آثار عدم تقارن اطلاعاتی در اعتبار

یکی از راه‌های تقویت قدرت پیش‌بینی رفتار اعتباری مشتریان، یافتن و استفاده از منابع جدیدی از داده‌ها و اطلاعات مشتری است. داده‌های رفتاری و اجتماعی در ادبیات مدیریت ریسک به‌عنوان «داده‌های جایگزین»[1] شناخته می‌شود که می‌تواند از منابع داخلی و خارجی سازمان تأمین شود. محدودهٔ اعتبارسنجی در استفاده از انواع داده‌ها را می‌توان در قالب شکل ۲ نمایش داد.



شکل ۲. انواع داده‌ها در حوزهٔ اعتبارسنجی

[1] alternative data

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]
[DOR: 20.1001.1.26453355.1402.16.57.2.0]

در مطالعهٔ ادبیات مرتبط با موضوع اعتبارسنجی، متغیرهای مؤثر در امتیاز اعتباری را می‌توان به دو دستهٔ سنتی (مالی) و جایگزین (رفتاری و اجتماعی) طبقه‌بندی کرد. در جداول ۱ و ۲ به برخی از این مطالعات اشاره می‌شود.

جدول ۱

مطالعاتی در مورد متغیرهای سنتی

متغیرهای سنتی	
(۲۰۱۶) لئو و کروک[1] میانگین ارزش تراکنش‌های بانکی، تعداد برداشت وجه نقد، تغییرات محدودیت اعتباری، نسبت ماه‌های معوقه، مبلغ بازپرداخت‌شده، و ماندهٔ باقیمانده	(۱۹۸۹) استین‌وَکر و خووارس [2] درآمد، وضعیت تأهل، ملیت، جنس، تعداد فرزندان، سن، حرفه، بخش، وضعیت سکونت، نوع اشتغال، شغل، و ویژگی‌های خاص وام مانند مدت، مبلغ، و هدف وام‌گیری
(۲۰۱۳) روسلی[3] تحصیلات و اشتغال	(۲۰۱۳) کروپا و همکاران[4] سایر بدهی‌ها یا تعهدات در زمان درخواست وام و عملکرد تاریخی اعتبارگیرنده و رفتار پرداخت وی در گذشته
(۲۰۱۹) اسکاردوتیر[5] و همکاران سن و مبلغ مخارج در ماه منتهی به درخواست وام	(۲۰۰۰) وست[6] طول عمر حساب، سابقهٔ اعتباری، طبقه‌بندی اشتغال، عملکرد (گردش) حساب، دارایی تحت مالکیت، سال اقامت، سایر وام‌های موجود، نوع سکونت، میزان وام، هدف از دریافت وام، سال استخدام، وضعیت حساب پس‌انداز

[1] Leow & Crook
[2] Steenackers & Goovaerts
[3] Rusli
[4] Kruppa
[5] Óskarsdóttir
[6] West

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]
[DOR: 20.1001.1.26453355.1402.16.57.2.0]

جدول ۲

مطالعاتی در مورد متغیرهای جایگزین

متغیرهای جایگزین	
(۲۰۱۵)ماسیوتین[1]	(۲۰۰۷)بروکت و گلدن[2]
داده‌های موجود در رسانه‌های اجتماعی (مثل وضعیت تأهل: متأهل، در عشق، پیچیده، و غیره)	داده‌های قبض‌های تلفن از ارائه‌دهندگان خدمات مخابراتی
فعالیت (تعداد زیاد ویدیوهای ارسال‌شده)	زمان مکالمه به‌ازای هر تماس، زمان استفاده از تلفن همراه
متوسط (۳ـ۷) تعداد اشتراک صفحه و غیره	
(۲۰۱۵) بیورکگرن و گریسن[3]	(۲۰۱۱) هان[4] و همکاران
حجم و توزیع تماس‌ها در طول زمان، میزان شارژ نقدی، و اعتباری موبایل	متون و فعالیت‌ها، تصاویر، ویدئوها، وبلاگ‌ها، علایق، و پیوندهای شبکه‌های اجتماعی افراد
(۲۰۱۶)سوزا[5] و همکاران	(۲۰۱۳) روسلی
طلاق، بیکاری، بیماری	تعداد دنبال‌کنندگان، دوستان و نوع (های) آن‌ها
(۲۰۱۹)اسکاردوتیر	–
مدت زمان تماس تلفنی و تعداد تماس‌های تلفنی دریافتی از افراد بدحساب و دارای تأخیر پرداخت	

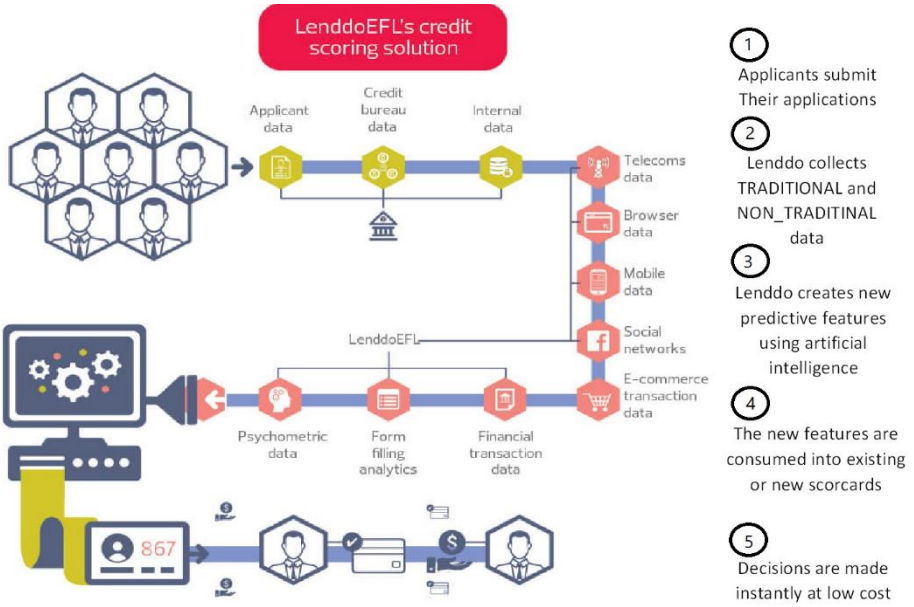
افرم[6] فین‌تکی است که مدل خود برای ارزیابی ریسک اعتباری افراد را از طریق شماره تلفن آن‌ها طراحی کرده است و از این طریق به داده‌های رسانه‌های اجتماعی و اطلاعات بازاریابی در تلفن‌های هوشمند افراد دست یافته است. گزارش شده است که افرم بیش از ۷۰،۰۰۰ ویژگی اطلاعاتی در مورد افراد به‌دست آورده است که می‌تواند نتیجهٔ اعتبار آنان را پیش‌بینی کند (ردراپ[7]، ۲۰۱۷).

[1] Masyutin
[2] Brockett & Golden
[3] Bjorkegren & Grissen
[4] Han
[5] Sousa
[6] Affirm
[7] Redrup

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]
[DOR: 20.1001.1.26453355.1402.16.57.2.0]

شرکت برنامه‌های ارتباطی وی‌چت[1]، ۶۰۰ میلیون کاربر در سرتاسر چین دارد. این شرکت از طریق سرویسی به نام ویلیدای[2] تا سقف ۳۰،۰۰۰ دلار وام ارائه می‌کند و هنگام تصمیم‌گیری اعتباری، به اطلاعات و محتوای تولیدشده توسط کاربر در برنامه متکی است. این شرکت به‌دلیل تصمیمات سریع خود مشهور است (وید و شانو مک تیگ[3]، ۲۰۱۶).

در مثالی دیگر، لندو[4] تخمین زد که تعداد کاربران تلفن همراه بیشتر از افراد بالای ۱۶ سال دارای حساب بانکی است (به‌ترتیب ۴/۸ میلیارد تلفن همراه در مقابل ۳/۴ میلیارد نفر)؛ بنابراین، با استفاده از مزیت داده‌های بزرگ مدلی طراحی کرد که از داده‌های شخصیتی (روان‌سنجی)، داده‌های رفتاری، داده‌های مرورگر، داده‌های ایمیل، داده‌های شبکهٔ اجتماعی، و داده‌های تلفن همراه استفاده می‌کند (شکل ۳).



شکل ۳. مدل کسب‌وکار Lenddo ـ ردراپ (۲۰۱۷)

[1] Wechat
[2] Weilidai
[3] Wade, Shan & McTeague
[4] lenddo

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]
[DOR: 20.1001.1.26453355.1402.16.57.2.0]

علاوه بر داده های رسانه‌های اجتماعی، داده‌های مخابراتی نیز توسط بسیاری از وام‌دهندگان به عنوان داده های شبکه‌های اجتماعی استفاده شده است. در ابتدا، از ترکیب داده‌های قبض‌های تلفن در کنار داده های مالی برای تجزیه‌وتحلیل الگوهای پرداخت وام گیرندگان استفاده می‌شد. در کنار این متغیر و در ادامه، متغیرهای دیگری مانند دقیقه به‌ازای هر تماس، ساعات استفاده از تلفن همراه و کل مدت‌زمان مکالمات موجب شکل‌گیری مدلهای قابل‌اعتمادی بر اساس شبکه‌های اجتماعی شدند (بروکت و گلدن، ۲۰۰۷).

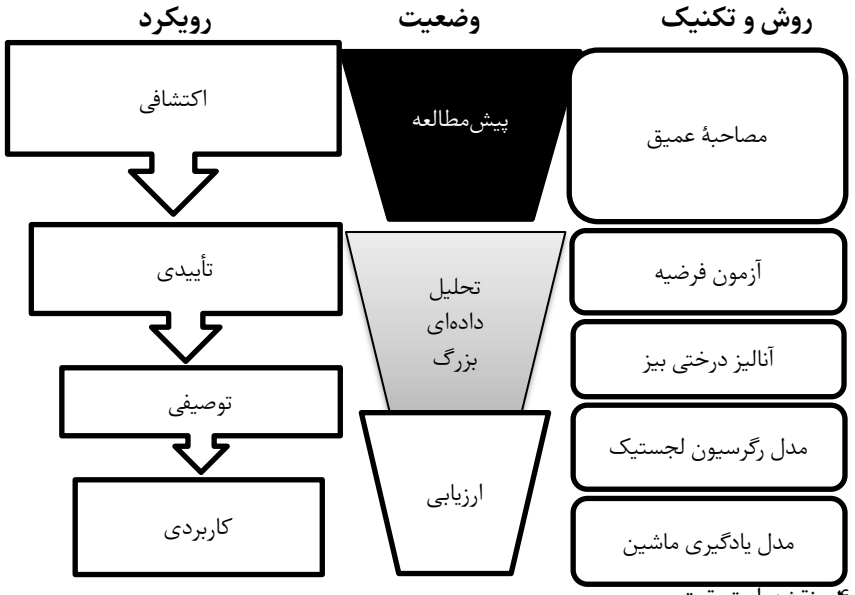
۲ نتایج

۲-۱ روش اجرای پژوهش

برای تعیین متغیرهای تحقیق، با تعدادی از متخصصان بانکی و افراد خبره در حوزهٔ اعتبارسنجی مصاحبهٔ عمیق صورت گرفت و در انتها متغیرها در سه طبقهٔ مالی، رفتاری، و اجتماعی تعیین و طبقه‌بندی شد. برای آزمایش این فرضیه، ابتدا از روش‌های آمار توصیفی، در ادامه از روش رگرسیون لجستیک، و در نهایت از انواع مدل‌های رگرسیون مبتنی بر یادگیری عمیق ماشین ازجمله گرادیان استفاده شده است. چهارچوب مطالعات و نتیجه‌گیری در این تحقیق به‌صورت خلاصه در شکل ۴ نشان داده شده است.

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]



شکل ۴. نقشه‌راه تحقیق

همان‌طور که در شکل ۴ مشاهده می‌شود، با استفاده از روش مصاحبهٔ عمیق کیفی با متخصصان امور بانکداری و اعتبارسنجی، ابتدا متغیرهای اجتماعی کاندیدای بررسی مشخص شدند. پس از شناسایی متغیرهای اجتماعی و تأیید آن‌ها توسط کارشناسان صنعت اعتبارسنجی، مجموعه داده‌ای توسط بانک خصوصی در مورد وام‌های پرداختی به اشخاص حقیقی در پنج سال گذشته در بازار ایران تأمین شد. تمام داده‌ها برای حفاظت از حریم خصوصی به‌صورت رمزگذاری و هش‌شده دریافت شدند. مجموعه دادهٔ این تحقیق دارای ۳۰۳,۱۱۱ وام پرداختی (مشاهده) است. در مرحلهٔ بعد، فیلدها یا ویژگی‌های هریک از مشاهده‌ها براساس یافته‌های حاصل از مصاحبهٔ کیفی که به شرح طبقات زیر خلاصه می‌شود، مشخص شد:

- طبقهٔ وام: مقدار و نوع وام‌های دریافتی و در نهایت هدف از وام؛
- طبقهٔ دارایی متقاضی: مالکیت دارایی، مالکیت خودرو، درآمد، و نوع درآمد؛
- طبقهٔ شرایط زندگی: تحصیلات، وضعیت تأهل، تعداد فرزندان، و تعداد افراد تحت تکفل دیگر؛
- جمعیت‌شناسی: سن و جنسیت؛

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]
[DOR: 20.1001.1.26453355.1402.16.57.2.0]

- طبقهٔ عوامل خارجی: امتیازات اعتباری خارجی ارائه‌شده توسط سایر دفاتر اعتباری و ردپای مالی (تعداد دفعاتی که متقاضی برای وام از سایر وام‌دهندگان درخواست داده است) به تفکیک دوره (در آخرین روز، هفتهٔ گذشته، ماه گذشته، سه‌ماههٔ آخر، و سال گذشته)؛
- محل سکونت (مالکی، اجاره‌ای، زندگی با والدین، و غیره)؛
- طبقهٔ ثبات زندگی: مدت زمان اشتغال نزد کارفرمای فعلی، همچنین، مدت زندگی در آدرس فعلی، و آدرس‌های قبلی، مدت زمان داشتن یک شماره تلفن؛
- طبقهٔ مرتبط با شغل: نوع شغل و نوع کارفرما؛
- طبقهٔ اجتماعی: تعداد دوستانی که برای بازپرداخت وام خود با مشکل مواجه بودند و در واقع در بازپرداخت وام تأخیر، بی‌نظمی، و نکول کردند؛
- طبقهٔ ترکیب اعتباری: انواع مختلف وام‌هایی که متقاضی دارد، شامل ارزی، مصرفی یا شرکتی، مسکن و خودرو، و...؛
- سوابق عملکرد اعتباری: تعداد تأخیرها و بازپرداخت‌های انجام‌نشدهٔ گذشته و مبالغ آن‌ها، همچنین تعداد تسویهٔ زودهنگام اقساط یا تسویهٔ کامل وام؛ و
- طبقهٔ عملکرد حساب: گردش حساب، الگوی هزینه، یا تعداد و مبلغ برداشت‌ها با استفاده از کارت‌های نقدی؛ و نیز محل جغرافیایی استفاده از کارت.

در این تحقیق، به‌منظور دستیابی به داده‌های اجتماعی افراد به‌منظور کشف شبکهٔ ارتباطی آنان، از سه متغیر اجتماعی جایگزین (نماینده) شامل متغیرهای زیر استفاده شد:

۱. متغیر وجود رابطهٔ پیوند اجتماعی نشان‌دهندهٔ تعداد افرادی است که با فرد موردنظر رابطه دارند یا در شبکهٔ اجتماعی وی قرار دارند؛
۲. متغیر تکرار یا ثبات پیوند اجتماعی نشان‌دهندهٔ تعداد دفعات است که فرد موردنظر در بازه زمانی موردمطالعه با افراد درون شبکهٔ اجتماعی خود ارتباط برقرار کرده است؛ و
۳. متغیر عمق رابطه یا کیفیت پیوندها نشان‌دهندهٔ مدت ارتباط در هربار برقراری ارتباط یا حجم دادوستد مالی صورت‌گرفته بین آن‌هاست.

در ادامه، به‌دلیل محدودیت های فنی و عدم امکان پذیری جمع‌آوری این قبیل داده‌ها به دو روش جایگزین اقدام شد اول از طریق بررسی شبکهٔ ارتباطی افراد با توجه به لیست

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]
[DOR: 20.1001.1.26453355.1402.16.57.2.0]

شماره‌های مندرج در دفترچه تلفن آنان و دوم از طریق بررسی خزشی[1] تراکنش‌های مالی افراد با دیگران.

در ادامه، از آزمون‌ها و روش‌های آماری استفاده شد تا اطلاعاتی در مورد گروه‌بندی و رفتار انواع وام‌گیرندگان استخراج شود. در این بخش، هدف کشف ارتباط بین داده‌های اجتماعی و امتیازات اعتباری از طریق استنباط آماری است. این کار در سه مرحله صورت پذیرفت که در ادامهٔ تحقیق تشریح خواهد شد: اول، آزمون این فرضیه که وام‌گیرندگانی که در نهایت به نکول می‌پردازند، شبکهٔ اجتماعی بد بزرگ‌تری دارند؛ دوم، آزمون اینکه آیا متغیرهای اجتماعی اطلاعات بهتری را برای ارزیابی‌کنندگان ریسک اعتباری در مورد ریسک وام‌ها دارند یا خیر؟ این کار از نظر آماری با استفاده از روش بیزی و با تحلیل شانس اطلاعات مبتنی بر وزن شواهد انجام شده است؛ و سوم، از یک مدل رگرسیون لجستیک و چند مدل پیشرفتهٔ یادگیری ماشین به‌منظور برآورد ضریب متغیرهای شبکهٔ اجتماعی در توضیح امتیازات اعتباری استفاده شد و در نهایت، نتایج حاصل از هر دو روش با یکدیگر مقایسه و تحلیل شد. در ادامه، برای نشان‌دادن عملکرد مدل‌ها از آمارهای محاسبه‌شدهٔ مربوط به ماتریس آشفتگی[2] ارائه‌شده توسط ونگ[3] و همکاران (۲۰۱۵) طبق جدول ۳ استفاده شد.

---

[1] crawling

[2] confusion matrix

[3] Wang

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]

جدول ۳
ماتریس آشفتگی

آماره‌های اندازه‌گیری	نتایج مشاهده‌های آماری		دقت[1] = (TP + TN) / N	
	منفی‌های واقعی	مثبت‌های واقعی		
صحت[2]	مثبت‌های کاذب (FP)	مثبت‌های درست (TP)	مثبت‌های پیش‌بینی‌شده	پیش‌بینی مدل
	منفی‌های درست (TN)	منفی‌های کاذب (FN)	منفی‌های پیش‌بینی‌شده	
تشخیص[3]		حساسیت[4]	آماره‌های اندازه‌گیری	

همچنین به‌منظور ارزیابی قدرت برازش مدل رگرسیون لجستیک، از شبه R^2 [5] استفاده شد.

۲-۲ داده

در بخش داده، جامعهٔ آماری وام‌گیرندگان پس از پاک‌سازی داده‌ها، ۳۰۰,۴۲۷ نفر در نظر گرفته است. شایان ذکر است برای اجرای مدل‌های مبتنی بر ماشین، لازم است از تعداد کافی و ترجیحاً زیاد داده در نمونه‌های موردبررسی استفاده شود تا ماشین بتواند انواع خطاهای ذکرشده در ماتریس به‌هم‌ریختگی را تجربه کند. در غیر این‌صورت، مدل با خطاهای over fit و under fit مواجه می‌شود.

پس از بررسی ویژگی‌های اطلاعاتی یا متغیرهای مرتبط با هریک از افراد یا همان مشاهده‌های جامعهٔ آماری وام‌گیرندگان، مجموعاً ۴۲ ویژگی انتخاب شد که به‌شرح جدول ۴ است.

[1] accuracy
[2] precision
[3] specificity
[4] recall
[5] pseudo R^2

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]
[DOR: 20.1001.1.26453355.1402.16.57.2.0]

جدول ۴
ماتریس متغیرها یا ویژگی‌های افراد

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]

R	متغیرها	نوع متغیر	توضیحات
۱	نوع حساب	مالی	قرض‌الحسنه، کوتاه‌مدت، و بلندمدت
۲	مانده حساب	مالی	مانده حساب در روز بررسی
۳	گردش بستانکار حساب	مالی	واریزی به حساب در شش ماه گذشته
۴	گردش بدهکار حساب	مالی	برداشت از حساب در شش ماه گذشته
۵	سن	مالی	۲۱ تا ۶۹ سال
۶	جنسیت	مالی	مرد (۶۳)، زن (۳۷)
۷	وضعیت تأهل	مالی	متأهل (۷۳)، مجرد (۱۵)، مطلقه (۷)، بیوه (۵)
۸	تعداد فرزندان	رفتاری	دارای فرزند (۳۱)، بدون فرزند (۶۹)
۹	تحت تکفل	رفتاری	تعداد تحت تکفل: ۱ (۲۲)، ۲ (۵۰)، ۳ (۱۸)، ۴ (۰۹)، ۵ یا بالاتر (۰۱)
۱۰	تحصیلات	رفتاری	راهنمایی یا کمتر (۷۶)، دبیرستان (۰۳)، دانشگاهی(۲۱)
۱۱	مدت ارتباط با بانک	رفتاری	تاریخ افتتاح اولین حساب
۱۲	شغل	رفتاری	۱۷ نوع شغل
۱۳	مدت اشتغال	مالی	از تاریخ اولین استخدام
۱۴	نوع کارفرما	مالی	۱ـ خویش‌فرما، ۲ـ دولتی، ۳ـ خصوصی، ۴ـ هیچ‌کدام
۱۵	مبلغ درآمد	مالی	درآمد اعلامی ماهانهٔ ۴ طبقه
۱۶	نوع درآمد	مالی	بیکار، دانشجو، مستمری‌بگیر، مرخصی،... ۸ طبقهٔ اصلی
۱۷	وضعیت محل زندگی	رفتاری	مالک (۸۸)، خانهٔ والدین (۶)، اجاره (۶)
۱۸	مدت سکونت در آدرس فعلی	رفتاری	۳ سال یا کمتر (۴۲)، ۳ تا ۷ سال (۳۱)، > ۷ سال (۲۷)
۱۹	داشتن ماشین	رفتاری	دارای ماشین (۶۷)، بدون ماشین (۳۳)
۲۰	خوش‌حسابی	مالی	سه طبقهٔ خوش‌حساب، معوق، و نکول
۲۱	مشخصات اعتبار (۴ متغیر)	مالی	مبلغ بدهی، نرخ، مدت تسهیلات، و دلیل وام
۲۲	مبلغ اقساط ماهانه	مالی	جمع مبلغ اقساط
۲۳	معرفی ضامن	رفتاری	وضعیت دودویی
۲۴	ارائهٔ چک تضمین	رفتاری	وضعیت دودویی
۲۵	نمرهٔ اعتبارسنجی داخلی	مالی	مقداری
۲۶	نمرهٔ اعتبارسنجی ملی	مالی	مقداری
۲۷	معرفی موبایل معتبر	رفتاری	داشتن موبایل به نام خود
۲۸	معرفی تلفن ثابت منزل	رفتاری	داشتن تلفن ثابت منزل معتبر
۲۹	معرفی تلفن ثابت محل کار	رفتاری	داشتن تلفن ثابت برای آدرس شرکت

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]

۳۰	معرفی املاک و مستغلات	مالی	ارائهٔ مستندات املاک و مستغلات
۳۱	ارزش املاک و مستغلات	مالی	اظهاری
۳۲	سابقهٔ چک برگشتی	رفتاری	تاریخ، مبلغ، و تعداد چک‌های برگشتی
۳۳	الگوی استفاده کارت اعتباری	رفتاری	تسویه سر دوره یا تقسیط
۳۴	تاریخ ورود به بانک	رفتاری	مدت زمانی که از ورود به بانک سپری شده است
۳۵	داشتن موبایل‌بانک	رفتاری	مدت زمانی که از آخرین ورود می‌گذرد
۳۶	داشتن اینترنت‌بانک	رفتاری	مدت زمانی که از آخرین ورود می‌گذرد
۳۷	وجود پیوند اجتماعی بد (باینری)	اجتماعی	۰ (۱۳)، ۱ (۸۷)
۳۸	تعداد پیوند اجتماعی بد	اجتماعی	۱-۲ (۱۲)، ۳-۹ (۵۶)، ۱۰-۱۴ (۲۱)، ۱۵ و بالاتر (۱۱)
۳۹	عمق پیوند اجتماعی بد (متوسط زمان‌های ارتباط)	اجتماعی	متوسط زمانی ارتباط با افراد بدحساب: ۱ سال یا کمتر (۳۲)، ۱ تا ۳ سال (۳۱)، > ۳ سال (۳۷)

در جامعهٔ آماری تحقیق، داده‌های مرتبط با بسیاری از متغیرها یا ویژگی‌های فوق وجود داشته ویا قابل‌محاسبه است. به‌منظور دستیابی به داده‌های اجتماعی مدنظر تحقیق، از چند روش به‌صورت تکی یا ترکیبی استفاده شد:

اول، با استفاده از شمارهٔ موبایل مشتری و شمارهٔ موبایل مخاطبین وی. این روش در صورتی قابل‌استفاده است که مشتری در اپلیکیشن موبایل‌بانک خود دسترسی اپلیکیشن را به دفترچه تلفن خود داده باشد. همچنین، در مواردی که مشتری با استفاده از امکان دعوت دوستان، کسانی را به بانک دعوت کرده باشد و/یا خود توسط افرادی دیگر به بانک دعوت شده باشد، این ارتباط قابل‌شناسایی خواهد بود.

دوم، با استفاده از شماره حساب یا کارت یا شبای مشتری و اینکه این شماره‌ها به‌عنوان مبدأ یا مقصد تراکنش‌های سایر افراد مورداستفاده قرار گرفته شده باشند. پس از استخراج اطلاعات تراکنش‌های مالی فی مابین مشتریان و با توجه به تعداد یا فراوانی تکرار تراکنش‌های بین افراد، متوسط دورهٔ زمانی تراکنش با افراد بدحساب (مدتی که از اولین تراکنش فرد با افراد بدحساب تا آخرین تراکنش فرد با افراد بدحساب گذشته است)، داده‌های مربوط به متغیرهای وجود، ثبات، و کیفیت ارتباط اجتماعی بین افراد استخراج و درکنار سایر داده‌های مالی و رفتاری افراد قرار گرفت.

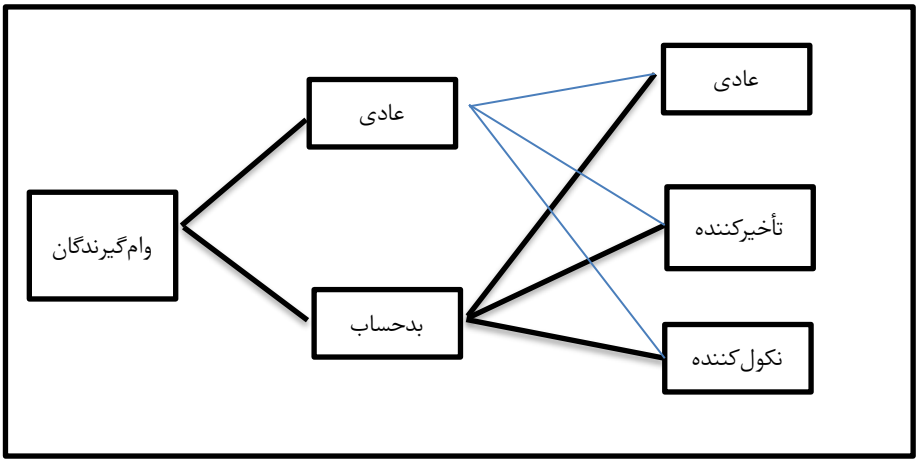
[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]

۳ بحث

۳-۱ تجزیه‌وتحلیل آماری

در ادامه، نتایج تجزیه‌وتحلیل اثر متغیرهای وجود ارتباط اجتماعی، ثبات اجتماعی، و کیفیت اجتماعی در رفتار خوب و بد وام‌گیرندگان با استفاده از میانگین و انحراف معیار برای هر دو گروه وام‌گیرندگان خوب و بد متناسب با درخت شکل ۵ به‌شرح جداول ۵ تا ۷ و جداول مرتبط با تحلیل بیزی (جداول ۸ و ۹) و روش‌های یادگیری ماشین (جداول ۱۰ و ۱۱) ارائه خواهد شد.



شکل ۵ انواع وام‌گیرندگان

جدول ۵

آمار توصیفی برای اثر وجود پیوند اجتماعی با نکول‌کنندگان

اثر متغیر وجود پیوند اجتماعی با افراد نکول‌کننده		
	عادی	نکول
تعداد نمونه	۲۷۵۷۶۷	۲۴۶۶۰
میانگین (MEAN)	۰/۱۴۱	۰/۱۹
انحراف از معیار (STD. D)	۰/۴۴	۰/۵۱۱

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]

جدول ۶

آمار توصیفی برای اثر وجود پیوند اجتماعی با افراد دارای تأخیر و معوقات

اثر متغیر وجود پیوند اجتماعی با افراد دارای تأخیر و معوقات		
	عادی	نکول
تعداد نمونه	۲۷۵,۷۶۷	۲۴,۶۶۰
میانگین (MEAN)	۱/۴۷۱	۱/۶۷۹
انحراف از معیار (STD. D)	۰/۶۵۲	۰/۷۲۱

جدول ۷

آمار توصیفی برای اثر وجود پیوند اجتماعی با افراد عادی

اثر متغیر وجود پیوند اجتماعی با افراد عادی یا خوش‌حساب		
	عادی	نکول
تعداد نمونه	۲۷۵,۷۶۷	۲۴,۶۶۰
میانگین (MEAN)	۲/۳۶۶	۲/۲۷۸
انحراف از معیار (STD. D)	۰/۸۰۳	۰/۸۳۲

همان‌طور که در جداول بالا مشاهده می‌شود، بین میانگین‌های دو گروه از وام‌گیرندگان در هنگام بررسی اثر پیوندهای اجتماعی تفاوت وجود دارد که می‌تواند به‌طور تصادفی یا به‌دلیل اثر متغیرهای اجتماعی وام‌گیرندگان رخ داده باشد. همان‌طور که مشاهده می‌شود، در مشتریان بدحساب میانگین تعداد پیوندهای آن‌ها با سایر مشتریان بدحساب (دارای معوق یا نکول‌کننده) از همین معیار در مورد مشتریان خوش‌حساب بالاتر است و درعین‌حال، با مقایسهٔ میانگین پیوندهای دو گروه از مشتریان بد و خوش‌حساب با افراد عادی و خوش‌حساب مشاهده می‌شود که مشتریان بدحساب پیوند اجتماعی کمتری با مشتریان عادی دارند. این نتایج به این معنی است که اثر داده‌های اجتماعی در ارزیابی ریسک اعتباری و پیش‌بینی امتیازات اعتباری توجیه‌پذیر است. بررسی نتایج آزمون Mann- Whitney U نیز نشان داد که می‌توان فرضیهٔ اثر متغیرهای اجتماعی در احتمال تأخیر و نکول‌کردن را تاحدی موردپذیرش قرار داد.

انتخاب آزمون Mann Whitney U به‌دلایل زیر بود:

الف- متغیرهای اندازه‌گیری‌شده بر اساس یک مقیاس (تعداد روابط اجتماعی بدحساب و نکول) قابل‌رتبه‌بندی‌اند؛

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]

ب- نمونه‌ها مستقل‌اند؛ و
ج- نمونه‌ها به‌اندازهٔ کافی بزرگ‌اند، زیرا n2 ≥۳۰ و n1.

به‌منظور انجام‌دادن تجزیه‌وتحلیل بیزی بر مبنای طبقه‌بندی انجام‌شده طبق شکل ۵، ابتدا شانس خوش‌حسابی کل نمونه با تقسیم احتمال خوب‌بودن مشتری در نمونهٔ انتخابی، یعنی P(G)، بر احتمال بدبودن مشتری در نمونهٔ انتخابی، یعنی $P(B)$، محاسبه می‌شود. احتمال خوب‌بودن یک مشتری در جمعیت از تقسیم تعداد مشتریان خوب موجود در جمعیت بر تعداد کل جمعیت و احتمال بدبودن یک مشتری در جمعیت از تقسیم تعداد مشتریان بد موجود در جمعیت بر تعداد کل جمعیت به‌دست می‌آید. در این تقسیم، به‌دلیل یکسان‌بودن مخرج هر دو احتمال و حذف آن هنگام تقسیم، شانس خوش‌حساب جمعیت از فرمول زیر به‌دست می‌آید:

شانس خوش‌حسابی جمعیت $= Opop = P(G)/\,P(B) =$ ۲۷۵۷۶۷ ÷ ۲۴۶۶۰ = ۱۸۳/۱۱

برای شروع کار، ابتدا افراد موجود در نمونهٔ انتخابی با شمارش متغیر وجود ارتباط با افراد بد، به ۲۰ طبقهٔ دلخواه قسمت شد. در ادامه، شانس خوش‌حسابی و شانس اطلاعات هر طبقه با توجه به فرمول فوق محاسبه شد تا به‌عنوان وزن آن طبقه (وزن شاهد) مورداستفاده قرار گیرد. وزن شاهد برای همهٔ طبقات محاسبه شد تا طبقاتی که قدرت تمایز بیشتری دارند و به پیش‌بینی بهتر احتمال نکول کمک می‌کنند، مشخص شود. ضمناً وزن‌های شاهد پرت از محاسبات حذف شدند. جدول ۸ محاسبات آماری هریک از طبقات را خلاصه می‌کند.

جدول ۸

احتمالات، شانس خوش‌حسابی، شانس اطلاعات، و وزن شواهد برای طبقات بد

طبقهٔ بد	تعداد خوب	تعداد نکول	ODDS	Px	P(x|G)	P(x|B)	I(x)	W.o.E
۰	۱۴۷۸۲۷	۱۲۸۸۹	۱۱/۴۶۹۲	۰/۵۳۵۰	۰/۵۳۶۱	۰/۵۲۲۷	۱/۰۲۵۶	۰/۰۲۵۳
۱	۴۳۹۷۶	۳۸۷۳	۱۱/۳۵۴۵	۰/۱۵۹۳	۰/۱۵۹۵	۰/۱۵۷۱	۱/۰۱۵۴	۰/۰۱۵۲
۲	۲۶۷۹۱	۲۴۴۸	۱۰/۹۴۴۰	۰/۰۹۷۳	۰/۰۹۷۲	۰/۰۹۹۳	۰/۹۷۸۷	-۰/۰۲۱۶
۳	۱۸۲۵۵	۱۶۶۶	۱۰/۹۵۷۴	۰/۰۶۶۳	۰/۰۶۶۲	۰/۰۶۷۶	۰/۹۷۹۸	-۰/۰۲۰۴
۴	۱۲۶۱۱	۱۲۴۴	۱۰/۱۳۷۵	۰/۰۴۶۱	۰/۰۴۵۷	۰/۰۵۰۴	۰/۹۰۶۵	-۰/۰۹۸۱
۵	۸۵۸۵	۷۹۱	۱۰/۸۵۳۴	۰/۰۳۱۲	۰/۰۳۱۱	۰/۰۳۲۱	۰/۹۷۰۵	-۰/۰۲۹۹

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]
[DOR: 20.1001.1.26453355.1402.16.57.2.0]

۶	۵۷۷۰	۵۶۳	۱۰/۲۴۸۷	۰/۰۲۱۱	۰/۰۲۰۹	۰/۰۲۲۸	۰/۹۱۶۵	-۰/۰۸۷۲
۷	۳۸۹۲	۴۰۹	۹/۵۱۵۹	۰/۰۱۴۳	۰/۰۱۴۱	۰/۰۱۶۶	۰/۸۵۰۹	-۰/۱۶۱۴
۸	۲۶۵۶	۲۴۴	۱۰/۸۸۵۲	۰/۰۰۹۷	۰/۰۰۹۶	۰/۰۰۹۹	۰/۹۷۳۴	-۰/۰۲۷۰
۹	۱۷۹۵	۱۶۸	۱۰/۶۸۴۵	۰/۰۰۶۵	۰/۰۰۶۵	۰/۰۰۶۸	۰/۹۵۵۴	-۰/۰۴۵۶
۱۰	۱۲۳۱	۱۲۰	۱۰/۲۵۸۳	۰/۰۰۴۵	۰/۰۰۴۵	۰/۰۰۴۹	۰/۹۱۷۳	-۰/۰۸۶۳
۱۱	۷۵۴	۸۱	۹/۳۰۸۶	۰/۰۰۲۸	۰/۰۰۲۷	۰/۰۰۳۳	۰/۸۳۲۴	-۰/۱۸۳۴
۱۲	۵۷۶	۶۰	۹/۶۰۰۰	۰/۰۰۲۱	۰/۰۰/۰۰۲ ۱	۰/۰۰۲۴	۰/۸۵۸۵	-۰/۱۵۲۶
۱۳	۳۵۸	۴۴	۸/۱۳۶۴	۰/۰۰۱۳	۰/۰۰۱۳	۰/۰۰۱۸	۰/۷۲۷۶	-۰/۳۱۸۰
۱۴	۲۳۰	۲۴	۹/۵۸۳۳	۰/۰۰۰۸	۰/۰۰۰۸	۰/۰۰۱۰	۰/۸۵۷۰	-۰/۱۵۴۳
۱۵	۱۵۱	۱۱	۱۳/۷۲۷۳	۰/۰۰۰۵	۰/۰۰۰۵	۰/۰۰۰۴	۱/۲۲۷۵	۰/۲۰۵۰
۱۶	۱۲۱	۹	۱۳/۴۴۴۴	۰/۰۰۰۴	۰/۰۰۰۴	۰/۰۰۰۴	۱/۲۰۲۲	۰/۱۸۴۲
۱۷	۸۰	۶	۱۳/۳۳۳۳	۰/۰۰۰۳	۰/۰۰۰۳	۰/۰۰۰۲	۱/۱۹۲۳	۰/۱۷۵۹
۱۸	۴۳	۲	۲۱/۵۰۰۰	۰/۰۰۰۱	۰/۰۰۰۲	۰/۰۰۰۱	۱/۹۲۲۶	۰/۶۵۳۷
۱۹	۳۸	۵	۷/۶۰۰۰	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۲	۰/۶۷۹۶	-۰/۳۸۶۲
۲۰	۲۷	۳	۹/۰۰۰۰	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۸۰۴۸	-۰/۲۱۷۱

با بررسی بیشتر، تعدادی از طبقات که شانس اطلاعاتی کم، دورافتاده، یا ناهنجار در مقایسه با سایر طبقات بودند و هیچ کمکی به تحلیل نمی‌کردند، حذف و در طبقات دیگر ادغام شدند. در ادامه با استفادهٔ مجدد از فرمول فوق، مجدداً شانس خوش‌حسابی هر طبقه محاسبه شد تا به‌عنوان وزن شاهد آن طبقه مورداستفاده قرارگیرد. وزن شاهد برای هر شش طبقه محاسبه شد تا طبقاتی که قدرت تمایز بیشتری دارند و به پیش‌بینی بهتر احتمال نکول کمک می‌کنند، مشخص شوند (جدول ۹).

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]

جدول ۹

احتمالات، شانس‌ها، شانس اطلاعات، و وزن شواهد برای روابط نکول‌شده

طبقه بد	تعداد خوب	تعداد نکول	ODDS	Px	P(x|G)	P(x|B)	I(x)	W.o.E
۰	۲۴۴۹۸۲	۲۰۹۸۹	۱۱/۶۷۱۹	۰/۵۵۹۴	۰/۸۸۸۴	۰/۸۵۱۱	۱/۰۴۳۷	۰/۰۴۲۸
۱	۲۴۹۰۶	۲۸۶۲	۸/۷۰۲۳	۰/۱۶۶۶	۰/۰۹۰۳	۰/۱۱۶۱	۰/۷۷۸۲	– ۰/۲۵۰۸
۲	۴۵۹۸	۶۱۶	۷/۴۶۴۳	۰/۱۰۱۸	۰/۰۱۶۷	۰/۰۲۵۰	۰/۶۶۷۵	– ۰/۴۰۴۲
۳	۱۰۱۶	۱۵۰	۶/۷۷۳۳	۰/۰۶۹۳	۰/۰۰۳۷	۰/۰۰۶۱	۰/۶۰۵۷	– ۰/۵۰۱۴
۴	۲۱۱	۳۴	۶/۲۰۵۹	۰/۰۴۸۲	۰/۰۰۰۸	۰/۰۰۱۴	۰/۵۵۵۰	– ۰/۵۸۸۹
۵	۴۷	۸	۵/۸۷۵۰	۰/۰۳۲۶	۰/۰۰۰۲	۰/۰۰۰۳	۰/۵۲۵۴	– ۰/۶۴۳۷
۶	۷	۱	۷/۰۰۰۰	۰/۰۲۲۰	۰/۰۰۰۰	۰/۰۰۰۰	۰/۶۲۶۰	– ۰/۴۶۸۵

محاسبات جدول ۹ نشان می‌دهد که بانک‌ها در حالتی که وام‌گیرنده در طبقهٔ صفر (ردیف اول جدول بالا) قرار دارد، یعنی دارای حداقل پیوند یا ارتباط با افراد بدحساب است، می‌توانند ۱/۰۴۳۷ برابر بیشتر در خصوص ریسک اعتباری وی اطلاعات کسب کنند.

در ادامهٔ این بخش، نتایج رگرسیون لجستیک بر روی انواع متغیرهای مختلف (به تفکیک مالی F، رفتاری B، و اجتماعی S) یک‌بار به‌تنهایی و یک‌بار با ورود متغیرهای اجتماعی اجرا شد. در هر بار اجراشدن مدل، نتایج آماره‌های اندازه‌گیری محاسبه، طبقه‌بندی، و برای تفسیر با وضعیت‌های دیگر مقایسه می‌شوند. شایان ذکر است حجم نمونهٔ مورداستفاده در این بخش ۴۹,۳۲۰ وام‌گیرنده است که به‌صورت تصادفی و با توزیعی مشابه از بین دو گروه بد و خوش‌حساب از جامعهٔ تحقیق انتخاب شده است. نمونهٔ انتخابی با نسبت ۷۰ و ۳۰ به‌ترتیب برای ساختن مدل و تست نتایج مورداستفاده قرارگرفته است. در نتیجه، حجم نمونهٔ تست شامل ۱۴۷۹۶ مشاهده و حجم نمونهٔ مدل‌سازی دارای ۳۴۵۲۴ مشاهده است.

نتایج حاصل از محاسبهٔ آماره‌های اندازه‌گیری در حالت‌های مختلف استفاده از انواع متغیرهای مالی، رفتاری، و اجتماعی به شرح جدول ۱۰ است.

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]

جدول ۱۰

عملکرد مدل رگرسیون لجستیک به تفکیک انواع مختلف متغیرهای مستقل

		مالی (F)	مالی و اجتماعی (F+S)	رفتاری (B)	رفتاری و اجتماعی (B+S)	مالی و رفتاری (F+B)	مالی و رفتاری و اجتماعی (F+B+S)
قدرت پیش‌بینی	حساسیت Recall	۰/۵۷	۰/۵۷	۰/۵۷	۰/۵۸	۰/۶۲	۰/۶۲
	صحت Precision	۰/۵۷	۰/۵۷	۰/۵۷	۰/۵۸	۰/۵۷	۰/۵۸
	F1 امتیاز	۰/۵۷	۰/۵۷	۰/۵۷	۰/۵۸	۰/۵۸	۰/۵۸
	AUC	۰/۵۷	۰/۵۷	۰/۵۷	۰/۵۸	۰/۵۸	۰/۵۸
	Log Loss	۱۴/۸۸	۱۴/۸۸	۱۴/۷۶	۱۴/۶۲	۱۴/۵۷	۱۴/۵۷
سازگاری و توضیح‌دهی	شبه R^2	٪ ۹/۵	٪ ۹/۷	٪ ۵/۰	٪ ۵/۲	٪ ۱۳/۳	٪ ۱۳/۳

با توجه به نتایج فوق، واضح است که استفاده از داده‌های مالی به‌تنهایی ۹/۵ درصد از تغییرات بین احتمالات نکول را توضیح می‌دهد. با افزودن ویژگی‌های رفتاری و اجتماعی به آن، افزایش ۳/۸ درصدی به‌دست آمد. همچنین دقت مدل از ۵۷ درصد به ۵۸ درصد افزایش یافت (مقایسهٔ ستون اول و آخر نتایج).

در بخش پایانی، الگوریتم‌های یادگیری ماشین بر روی جامعهٔ آماری تحقیق اجرا شد. دقت مدل از رگرسیون لجستیک فراتر رفت، و به‌علاوه log loss به میزان زیادی، همان‌طور که در جدول ۱۱ مشاهده می‌شود، کاهش یافت.

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]

جدول ۱۱

جدول مقایسه‌ای یادگیری ماشین

	فقط با ویژگی های سنتی		با ویژگی‌های جایگزین	
یادگیری ماشین	Accuracy	Log Loss	Accuracy	Log Loss
K-Nearest Neighbors	۰/۶۰/۶۰۷	۴/۴۰۸	۰/۵۹۱	۴/۴۵۱
Support Vector Machine	۰/۶۰/۶۰/۶۶۴	۰/۶۰/۶۱۴	۰/۶۰/۶۶۸	۰/۶۰/۶۰۹
Decision Trees	۰/۵۸۱	۱۴/۴۶	۰/۵۸۳	۱۴/۴۰۵
Random Forests	۰/۶۰/۶۳۴	۰/۹۵۲	۰/۶۰/۶۳۶	۰/۸۲۴
AdaBoost	۰/۶۰/۶۶۸	۰/۶۰/۶۸۹	۰/۶۰/۶۷۳	۰/۶۰/۶۸۹
Extreme Gradient (XGBoost)	۰/۶۰/۶۷۶	۰/۶۰/۶۰۴	۰/۶۰/۶۸۰	۰/۵۹۹
Naïve Bayes (Gaussian)	۰/۶۰/۶۰۴	۰/۹۱۴	۰/۶۰/۶۳۵	۱/۳۱۶
Linear Discriminant Analysis	۰/۶۰/۶۶۷	۰/۶۰/۶۱۳	۰/۶۰/۶۷۷	۰/۶۰/۶۰۴
Quadratic Discriminant Analysis	۰/۵۸۳	۰/۹۰۷	۰/۵۳۵	۱۰/۱۵۹

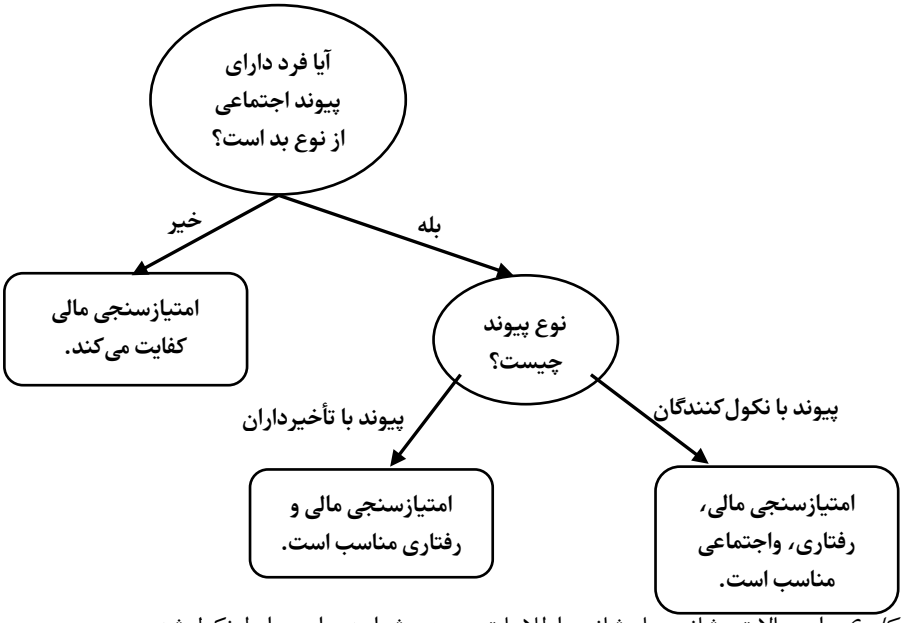
۳-۲ یافته‌ها، کاربردها، و پیشنهادها

در این تحقیق خلاصه یافته‌ها به‌شرح موارد زیر است:

- داشتن پیوندهای اجتماعی با افراد بدحساب یا دارای تأخیر در احتمال نکول تأثیر دارد، هرچند اهمیت یا وزن این اثر ضعیف است؛
- اگرچه واردکردن متغیرهای اجتماعی موجب بهبود بسیار ناچیز در توضیح‌دهندگی مدل‌های اعتبارسنجی می‌شود، با توجه به حجم کلان تسهیلات اعطایی، موجب صرفه‌جویی‌های قابل‌توجه مالی شده و لذا قابل‌صرف‌نظرکردن نیستند؛
- داشتن دوستان و خانوادهٔ بدحساب در شبکهٔ ارتباطی وام‌گیرنده با احتمال نکول وی رابطهٔ مثبت و بر این اساس با نمرهٔ اعتباری فرد رابطه منفی دارد؛
- چنانچه در شبکهٔ اجتماعی اشخاص، افراد بدحساب از جنس نکول‌کننده وجود داشته باشد، باید ریسک و امتیاز اعتباری آنان بر مبنای متغیرهای مالی، رفتاری، و اجتماعی ارزیابی شود؛ و
- چنانچه در شبکهٔ اجتماعی اشخاص، افرادی وجود داشته باشند که با تأخیر بدهی خود را بازپرداخت می‌کنند، این موضوع تأثیر مستقیمی در امتیاز اعتباری آنان ندارد. در این موارد، متغیرهای مالی سنتی و رفتاری برای محاسبهٔ امتیاز اعتباری کافی است (شکل ۶).

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]



شکل ۶ . احتمالات، شانس‌ها، شانس اطلاعات، و وزن شواهد برای روابط نکول‌شده

موارد زیر را می‌توان به‌عنوان کاربردهای نتایج تحقیق برشمرد:

- کمک به وام‌دهندگان برای شناسایی بهتر متقاضیان وام (تقارن اطلاعاتی)؛
- کمک به مشتریان برای دریافت خدمات مالی با ارائهٔ اطلاعات اجتماعی (رفع محرومیت مالی)؛
- کمک به پیش‌بینی بهتر رفتار وام‌گیرندگان و در نتیجه سودآوری بیشتر و کاهش زیان نکول؛ و
- کمک به سلامت اجتماعی و ارتقای فرهنگ مالی با القای ارزشمندی داده‌های اجتماعی.

در پایان، پیشنهادهای مبتنی بر نتایج تحقیق به‌شرح موارد زیر است:

- اقدام برای جمع‌آوری داده‌های اجتماعی با روش‌های تشویقی مثل تخفیف توسط بانک‌ها و مؤسسات مالی؛
- تأسیس اپراتور شبکهٔ اجتماعی برای محاسبهٔ امتیاز اجتماعی افراد و ایجاد API برای بانک‌ها و سایر تولیدکنندگان و استفاده‌کنندگان امتیاز اعتباری؛

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]
[DOR: 20.1001.1.26453355.1402.16.57.2.0]

- حمایت و ترغیب شرکت‌های فین‌تک وام‌دهی P2P برای پیاده‌سازی راهکاری خلاقانه برای جمع‌آوری داده‌های اجتماعی؛
- وضع قانون محافظت از داده‌های شخصی و الزامات قانونی برای استفاده از داده‌های اجتماعی؛
- توافق با تأمین‌کنندهٔ داده‌های اجتماعی مثل شرکت مخابرات در مورد داده‌های تلفن‌های ثابت و همراه مشتریان با هدف کشف پیوندهای اجتماعی؛
- تدوین و به‌روز نگهداری مدل اعتبارسنجی مبتنی بر داده‌های مالی، رفتاری، و اجتماعی توسط مراجع قانونی و نظارتی مثل بانک مرکزی؛
- ترویج و اشاعهٔ فرهنگ اهمیت و ارزشمندی پیوندهای اجتماعی؛ و
- استفاده از داده‌های اجتماعی در لحظه برای اعطای وام‌های مصرفی و کسب‌وکاری مانند موردمطالعهٔ LENDDO.

از سوی دیگر، با وجود مزایای قابل‌توجهی که استفاده از داده‌های اجتماعی برای اعتبارسنجی دارد، محدودیت‌هایی نیز می‌تواند وجود داشته باشد:

۱. داده‌های افراد، اعم از مالی، رفتاری، و اجتماعی از طرف قانون‌گذار محافظت می‌شود و دسترسی به آن‌ها می‌تواند چالش‌های زیادی را ایجاد کند و برای همین در برخی از کشورها قوانین سختی در این مورد وجود دارد.
۲. احتمال اینکه افراد خود را در شبکه‌های اجتماعی به‌نحوی نشان دهند که افراد خوبی برای بانک هستند، وجود دارد. این چالش تحت عنوان بازی با سیستم شناخته می‌شود ولی در بلندمدت به‌دلیل سازوکار نامرئی خودکنترلی به نفع بانک‌هاست.
۳. استفاده و به‌کارگیری روش‌های نوآورانه نه‌تنها پیچیدگی روش‌های اعتبارسنجی را افزایش داده است، بلکه هم‌زمان گنگ‌بودن را نیز به آن اضافه کرده است. اصولاً قابلیت تعبیرپذیری و تفسیر نتایج به‌دست‌آمده از تکنیک‌های یادگیری ماشین یکی از مهم‌ترین نگرانی‌های قانون‌گذار و صنعت اعتبارسنجی است. این نگرانی با ورود متغیرهای جدید اجتماعی، افزایش متغیرها، و پیچیده‌شدن الگوریتم‌ها بیشتر شده و باعث ایجاد نگرانی در مورد صحت و عادلانه‌بودن خروجی‌های مدل خواهد شد.

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]
[DOR: 20.1001.1.26453355.1402.16.57.2.0]

فهرست منابع

آیین‌نامهٔ نظام سنجش اعتبار، مصوب ۱۳۸۵ مجلس
الزامات، ضوابط، و فرایندهای اجرایی فعالیت تسهیلات یاران (بانک مرکزی ج.ا.ا.)
حیدری، رضا و خادمی، سید موسی . امتیازدهی اعتباری براساس رفتاربانکی و ارزیابی شبکه ارتباطی افراد با تکنیک‌های یادگیری عمیق ماشین و گرادیان، پایان‌نامهٔ کارشناسی ارشد در رشتهٔ مدیریت فنّاوری اطلاعات، گرایش کسب‌وکارهای الکترونیک، دانشگاه پیام نور تهران غرب، ۱۴۰۰.
عینی آرش. مدیریت بهینهٔ ریسک اعتباری. *اقتصاد و بانکداری اسلامی، ۱۳۹۷* ، ۷(۲۵)، ۶۷ـ۹۶.
لبافی سمیه و کیا، علی اصغر و ملکی، مصطفی. شناسایی عوامل مؤثر بر درگیرسازی و مشارکت کاربر در شبکه اجتماعی ـ علمی. *ریسرچ گیت ۱۳۹۹* .
نبی‌میبدی، ، سیروس. مفهوم‌سازی عوامل موفقیت شبکه‌های اجتماعی برخط از دیدگاه کاربری و سیاست‌گذاری: مطالعهٔ تحلیلی، ۱۳۹۶.
نبی‌میبدی، مرتضی و علیدوستی، سیروس ونظری، مریم. عوامل مؤثر بر موفقیت شبکهٔ اجتماعی حرف‌های تخصصی: یافته‌های مطالعه کیفی در ایران، ۱۳۹۷.
وهابی، فتانه و اسدی، سعید و گنجه‌فر، سهیل . مدل‌سازی رفتار اطلاعاتی کاربران پایگاه‌های اطلاعاتی با روش شبکه عصبی با تأکید بر تعاملات پیشین آن‌ها با نتایج جست‌وجو، ۱۴۰۰.
Ackert, L. & Deaves, R. (2009). Behavioral finance: Psychology, decision-making, and markets: Cengage Learning.
Banasik, J., Crook, J. N. & Thomas, L. C. (1999). Not if but when will borrowers default. *Journal of the Operational Research Society*, 50(12), 1185-1190.
Berndt, A. & Gupta, A. (2009). Moral hazard and adverse selection in the originate-to-distribute model of bank credit. *Journal of Monetary Economics*, 56(5), 725-743.
Bjorkegren, D. & Grissen, D. (2015). Behavior revealed in mobile phone usage predicts loan repayment.
Brockett, P. L. & Golden, L. L. (2007). Biological and psycho behavioral correlates of credit scores and automobile insurance losses: Toward an explication of why credit scoring works. *Journal of Risk and Insurance*, *74*(1), 23-63.
Fitzgerald, R. (2018). How Lenddo EFL Uses Data and Personality Analyses to Increase Access to Financial Services in

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]
[DOR: 20.1001.1.26453355.1402.16.57.2.0]

EmergingEconomies. Retrieved from https://www.cardrates.com/news/lenddoefl-helps-emerging-economies-access-financialservices/

General Data Protection Regulation, (2016).

Han, J., Pei, J. & Kamber, M. (2011). Data mining: Concepts and technique. *Elsevier*.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer, vol. 112.

Kosinski, M., Stillwell, D. & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences, 110(15), 5802-5805.

Kruppa, J., Schwarz, A., Arminger, G. & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. *Expert Systems with Applications*, 40(13), 5125-5131.

Kumire, J. (2019). Open Banking in the UK: what's happened so far. Retrieved from https://pulse.11fs.com/research-reports/2019/08/open-banking-in-the-uk-whatshappened-so-far.

Leow, M. & Crook, J. (2016). A new mixture model for the estimation of credit card Exposure at *Default. European Journal of Operational Research, 249(2), 487-497.*

Masyutin, A. (2015). Credit scoring based on social network data.

McEvoy, M. J. & Chakraborty, T. (2014). Enabling financial inclusion through alternative data.

Naeem Siddiqi (2017). *Intelligent credit scoring building and implementing better credit risk scorecards.* Second Edition. John Wiley & Sons, Inc., Hoboken, New Jersey. Published simultaneously in Canada.

Newman, M. (2010). *Networks: an introduction*. Oxford University Press, Oxford.

Óskarsdóttir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B. & Vanthienen, J. (2017). 175 Social network analytics for churn prediction in

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]

telco: Model building, evaluation and network architecture. Expert Systems with Applications, 85, 204-220.

Redrup, Y. (2017). How email and smartphone data could help you get a loan. Technology. Retrieved from: http://www.afr.com/technology/how-email-and-smartphone-data-couldhelp-you-get-a-loan-20171212-h02zi0#ixzz534zFfQmg

Rusli, E. M. (2013). Bad credit? Start tweeting. WALL ST. J., Apr, 1.

Schuermann, T. (2004). What do we know about loss given default?

Steenackers, A. & Goovaerts, M. (1989). A credit scoring model for personal loans.

Sousa, M. R., Gama, J., & Brandão, E. (2016). A new dynamic modeling framework for credit risk assessment. Expert Systems with Applications, 45, 341-351.

Wade, M. R., Shan, J., & McTeague, L. Strategies for responding to digital disruption. Retrieved from https://www.imd.org/research/insightsimd/strategies-for-responding-to-digital-disruption2/

Wang, G., Ma, J., Huang, L. & Xu, K. (2012). Two credit scoring models based on dual strategy ensemble trees. *Knowledge-based systems*, 26, 61-68.

Wei, Y., Yildirim, P., Van den Bulte, C. & Dellarocas, C. (2015). Credit scoring with social network data.

Weke, P., & Ntwiga, D. B. (۲۰۱۶). Consumer lending using social media data. West, D. (2000). Neural network credit scoring models. *Computers & Operations Research*, 27(11), 1131-1152.

Yan, J., Yu, W. & Zhao, J. L. (2015). How signaling and search costs affect information asymmetry in P2P lending: the economics of big data. *Financial Innovation*, 1(1), 1.

[Downloaded from jmbr.mbri.ac.ir on 2026-11-06]

[DOR: 20.1001.1.26453355.1402.16.57.2.0]