فصل‌نامه مطالعاتی در مدیریت بانکی و بانکداری اسلامی

مؤسسه عالی آموزش بانکداری ایران
تابستان ۱۴۰۲، دوره ۹، شماره ۲۳، صفحه ۹۹- ۱۲۲

ارزیابی اثر شبکه اجتماعی افراد بر امتیاز اعتباری بانک‌ها و مؤسسه‌های اعتباری با روش‌های یادگیری عمیق ماشین و گرادیان

رضا حیدری*[۱]
سید موسی خادمی[۲]

چکیده

این پژوهش به‌دنبال بررسی اثر متغیرها و داده‌های مرتبط با شبکه اجتماعی افراد، بر امتیاز اعتباری آن‌هاست. در این پژوهش، دو هدف اصلی دنبال می‌شود: اول، کاهش عدم تقارن اطلاعاتی و دوم، افزایش شمول مالی. دستیابی به اهداف یادشده با یافتن اطلاعات معنادار در زمینه داده‌های اجتماعی افراد انجام می‌شود تا چگونگی اثر چنین داده‌هایی بر امتیاز اعتباری آن‌ها اندازه‌گیری شود. فرضیه اساسی پژوهش این است که افراد با امتیاز اعتباری بالا با افراد مشابه و هم سنخ خود ارتباط اجتماعی دارند. در این پژوهش از یک مجموعه داده، برای تأیید و توضیح اثر متغیرهای شبکه اجتماعی بر امتیاز اعتباری استفاده شد که مشتمل بر وام‌هایی بود (بیش از ۳۰۰ هزار) که یکی از بانک‌های ایرانی به افراد حقیقی پرداخت کرده بود. برای تعیین متغیرهای پژوهش با تعدادی از متخصصان بانکی و افراد خبره در حوزه اعتبارسنجی، مصاحبه عمیق صورت گرفت و در انتها، متغیرها در سه طبقه مالی، رفتاری و اجتماعی تعیین و طبقه‌بندی شد. برای آزمایش فرضیه پژوهش، ابتدا از روش‌های آمار توصیفی و در ادامه، از روش رگرسیون لجستیک و در نهایت، از انواع مدل‌های رگرسیون مبتنی بر یادگیری عمیق ماشین، از جمله گرادیان استفاده شد. نتیجه بررسی‌های روش رگرسیون لجستیک نشان داد که از نظر آماری، متغیرهای اجتماعی افراد، قابلیت پیش‌بینی احتمال نکول وام آنان را دارد. نتایج الگوریتم‌های یادگیری ماشین نیز نشان داد که اطلاعات شبکه اجتماعی، می‌تواند عملکرد پیش‌بینی نکول وام را به‌طور چشمگیری بهبود بخشد.

واژه‌های کلیدی: امتیازدهی اعتباری، ریسک نکول، شمول مالی، عدم تقارن اطلاعاتی، شبکه اجتماعی، پیوند اجتماعی بد.

طبقه‌بندی JEL: G۴۱، G۲۰، D۸۱، C۵۸، C۵۱ و C۱۱.

---

۱. کارشناس ارشد، گروه اقتصاد، دانشکده اقتصاد، دانشگاه تهران، تهران، ایران (نویسندهٔ مسئول)؛ rheydari2017@gmail.com
۲. استادیار، دانشگاه پیام نور، تهران، ایران؛ sm.khademi@pnu.ac.ir

مقدمه

در سال‌های اخیر، تغییرات و پیشرفت‌های فناوری 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 فناوری به دو طریق بر حوزه اعتبارسنجی تأثیر داشته است؛ نخست، از طریق ایجاد منابع داده‌ای جدید در خصوص مشتری و دوم، از طریق بهبود روش‌ها و قدرت محاسباتی ابزارهای اعتبارسنجی با استفاده از فناوری‌های جدید. منظور از منابع داده‌ای جدید، کلیه داده‌هایی است که از رفتارهای مشتری در استفاده از بسترهای فناوری حاصل می‌شود، مانند رفتار خرید اینترنتی وی و یا فعالیت وی در یک شبکه اجتماعی (تعداد پست‌ها، لایک‌ها، نظردهی و بازدیدها و...).

در این میان، مؤسسه‌های اعتبارسنجی با تحلیل داده‌های متنوع و جدید حاصل از رفتارهای افراد از یک سو و استفاده از مدل‌هایی با قدرت پیش‌بینی‌کنندگی بیشتر از سوی دیگر توانستند در خصوص اعتبار متقاضیان، بینش عمیق‌تری پیدا کرده و مدل‌های دقیق‌تری برای اعتبارسنجی ایجاد کنند. استفاده از روش‌های فنی جدید نیز در اعتبارسنجی به‌طور شایان توجهی در سال‌های اخیر افزایش داشته است. در همین راستا، استفاده روش‌های پیچیده و نوین در سنجش اعتبار مانند هوش مصنوعی و روش‌های پیشرفته یادگیری ماشین مانند جنگل‌های تصادفی و تقویت‌گرادیان و شبکه‌های عصبی عمیق جایگزین موارد سنتی مانند تحلیل تمایزی خطی و رگرسیون لجستیک شده‌اند. مدل‌های اعتبارسنجی جدید مبتنی بر ماشین بر خلاف روش‌های سنتی این امکان را فراهم می‌کنند تا روابط بین متغیرها بدون محدودیت در تعداد متغیرها کشف و استفاده می‌شود. برای مثال، با این روش‌ها می‌توان در دریاچه 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.

بیان مسئله

ذی‌نفعان حوزه اعتبارسنجی را می‌توان به سه دسته کلی شامل مشتریان یا وام‌گیرندگان، تولیدکنندگان گزارش‌های اعتبارسنجی یا آژانس‌های مرجع اعتبار و مؤسسه‌های تأمین مالی و اعتباری تقسیم کرد. با توجه به نیازهای این ذی‌نفعان، اهمیت استفاده از داده‌های اجتماعی به‌منظور بهبود نظام اعتبارسنجی در قالب موارد ذیل مطرح می‌شود:

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

۲. برای یک وام‌دهنده هوشمند، داده‌هایی که از شبکه‌های اجتماعی افراد به دست می‌آیند، در واقع نوعی تأیید برای این هستند که آنان چه کسی هستند، نخست، اینکه واقعی هستند و دوم اینکه، اگر فرد خوبی هستند، باید با اطرافیان خود ارتباط خوبی داشته باشند، برای مثال، باید تعداد معینی دوست داشته باشند، شماره تلفن همراه را برای مدت معینی مالک باشند و... .

۳. با یک 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 خصوص استفاده مردم از تلفن همراه و برنامه‌هایی که از این طریق استفاده می‌شوند، دست یافت.

۴. مسلم است که تولید گزارش‌های اعتبارسنجی در لحظه نیاز مشتری و مؤسسه‌های 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 با شناسایی شاخصه‌ها و رفتارهای بد اعتباری، موجب کنترل و کاهش هزینه‌های ناشی از نکول یا پیگیری و وصول مطالبات شود.

۵. ایجاد یک سیستم اعتبارسنجی مناسب، موجب بهبود رشد و توسعه اقتصادی می‌شود. این موضوع برای مثال با اعطای تسهیلات برای راه‌اندازی یک تجارت ساده یا تأمین سرمایه در گردش یک مزرعه یا آموزش یک مهارت خاص امکان‌پذیر می‌شود. دقیقاً به همین دلیل است که نیاز به استفاده از دیتاهای غیرمالی در ارزیابی‌های اعتباری مهم است.

۶. ایجاد فرهنگ دفاع از اعتبار خود در افراد یک جامعه به‌معنای رشد ثروت پنهان آن جامعه است. اینکه افراد چه کسی را می‌شناسند، مهم است. شبکه اجتماعی افراد اهمیت دارد. این ایده که

افرادی که در یک شبکه اجتماعی در ارتباط هستند، ویژگی‌ها و رفتارهای بسیار مشابهی با هم دارند، مفهوم و فرهنگ بسیار قوی و بسیار قدرتمندی است که جامعه را به‌سمت خودکنترلی هدایت می‌کند.

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

با توجه به موارد گفته‌شده، مسئله که در پژوهش به آن توجه می‌شود، آن است که در وضع موجود متغیرهای استفاده‌شده توسط نظام اعتبارسنجی کشور که به‌طور عمده بر داده‌های مالی و رفتار بانکی مشتریان تکیه دارد، برای امتیازدهی اعتباری افراد کافی نیستند و به‌دنبال پاسخ به این پرسش است که آیا با افزودن داده‌های حاصل از فعالیت‌های ارتباطی و روابط اجتماعی افراد می‌توان این نظام اعتبارسنجی را بهبود بخشید؟

اهداف پژوهش

به‌طور خلاصه، اهداف اصلی این پژوهش 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 یا یکسان نداشته باشند. به بیان دیگر، افراد برای وام گرفتن همه اطلاعات خود را ابراز نمی‌کنند، بنابراین، بانک‌ها اطلاعات محدود دارند و در نهایت، به استعلام‌ها از بانک مرکزی و مرکز ملی اعتبارسنجی (مؤسسه رتبه‌بندی ایران) اکتفا می‌کنند. حال، اگر برای دسترسی به اطلاعات افراد، روش‌های دیگری وجود داشته باشد، این تقارن اطلاعاتی برای مؤسسه‌های اعتباری کم می‌شود.

- بانک‌ها با بررسی داده‌های اجتماعی به‌موازات داده‌های مالی، شانس بیشتری برای شناخت مشتریان خواهند داشت و فقط به اطلاعات اظهاری یا سابقه مالی و اعتباری آنان اکتفا نمی‌کنند.
- افزایش در شمول مالی[2]: به این معنا که بانک‌ها با استفاده از داده‌های اجتماعی در کنار داده‌های مالی، می‌توانند در خصوص مشتریان موجود و افراد باسابقه مالی کم، اطلاعات دقیق‌تری به‌دست آورند و در نتیجه، دامنه ارائه خدمات اعتباری خود را گسترش داده و همگانی کنند.

---

1. Information asymmetry
2. Financial inclusion

فرضیه پژوهش

فرضیه این پژوهش براساس گزاره خبری زیر تدوین و پیشنهاد شده است: «در فرایند تشکیل یک شبکه اجتماعی، افراد ترجیح می‌دهند با افرادی دارای خلق و خو و شخصیت شبیه به خود روابط اجتماعی برقرار کنند».

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

فرضیه صفر عبارت است از: کسانی که وام‌ها را نکول می‌کنند (وام‌گیرندگان بد) روابط اجتماعی بد کمتر یا مساوی با کسانی دارند که وام‌های خود را به‌موقع بازپرداخت می‌کنند.

توضیح اینکه در این پژوهش، فارغ از تعاریف مرسوم در نظام بانکی کشور برای طبقه‌بندی مطالبات، وام‌گیرنده در صورت گذشت زمان بیش از ۹۰ روز از سررسید بدهی، به‌عنوان نکول‌کننده تعریف می‌شود. این زمان برای افراد عادی و خوب تا ۳۰ روز و برای افراد تأخیرکننده ۳۰ تا ۹۰ روز در نظر گرفته شده است.

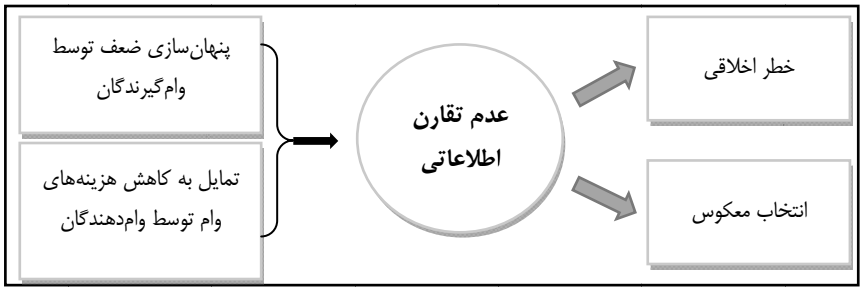
مرور ادبیات و پیشینه پژوهش

ویک و ان تویگا[2] (۲۰۱۶) روی سه مزیت هنگام استفاده از داده‌های مربوط به رسانه‌های اجتماعی در ارزیابی ریسک های اعتباری تأکید کردند: نخست، غلبه بر اثرهای نامطلوب انتخاب با جذب وام‌گیرندگان ارزشمند که سابقه مالی محدودی دارند. دوم، درک نیازهای واقعی مشتری و در نتیجه، کاهش اثرهای مخاطره اخلاقی. 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 در خصوص شرایط و ضوابط وام با هدف کسب سود بیشتر، اطلاعات کاملی ارائه نمی‌دهند. در طرف مقابل، این شواهد وجود دارد که وام‌گیرنده نیز اطلاعاتی را که نشان‌دهنده توانایی ضعیف بازپرداخت آن‌هاست، اعلام نمی‌کند. این دو موضوع، می‌تواند به یک وضعیت نامتقارن اطلاعاتی منجر شود که در نهایت، باعث انتخاب نامطلوب یا یک مورد خطر اخلاقی می‌شود. یان، یو و ژوآ[3] (۲۰۱۵) این وضعیت را در شکل ۱ به تصویر کشیده‌اند.

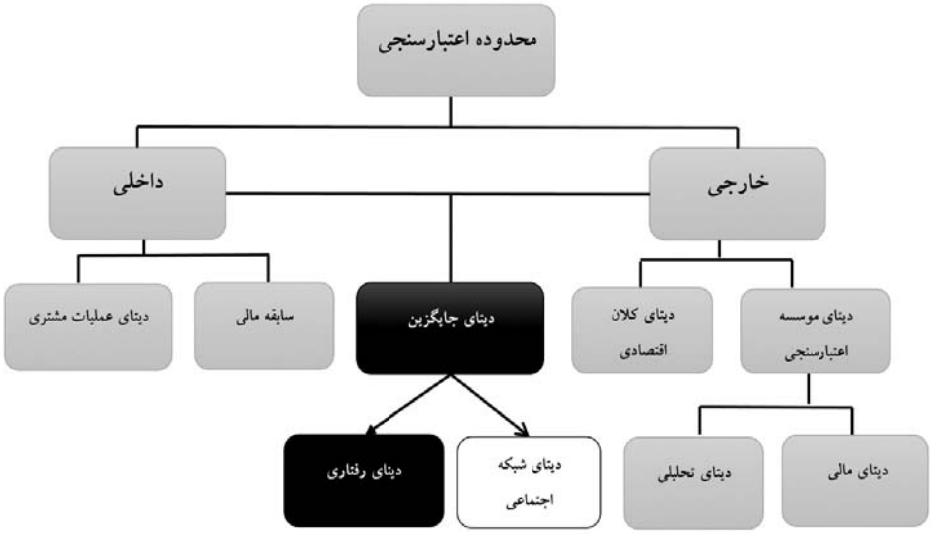
---

1. Default Probablity
2. Weke & Ntwiga
3. Yan, Yu & Zhao



شکل ۱. علل و آثار عدم تقارن اطلاعاتی در اعتبار

یکی از راه‌های تقویت قدرت پیش‌بینی رفتار اعتباری مشتریان، یافتن و استفاده از منابع جدیـدی از داده‌ها و اطلاعات مشتری است. داده‌های رفتاری و اجتماعی در ادبیات مدیریت ریسک، به‌عنـوان «داده‌های جایگزین»[1] شناخته می‌شوند که می‌تواند از منابع داخلی و خـارجی سـازمان تـأمین شـود. محدوده اعتبارسنجی در استفاده از انواع داده‌ها را می‌توان در قالب شکل ۲ نمایش داد.



شکل ۲. انواع داده‌ها در حوزه اعتبارسنجی

1. Alternative data

در بررسی ادبیات مرتبط با موضوع اعتبارسنجی، متغیرهای مؤثر بر امتیاز اعتباری را می‌توان به دو دسته سنتی (مالی) و جایگزین (رفتاری و اجتماعی) طبقه‌بندی کرد. در جدول‌های ۱ و ۲ به برخی از این پژوهش‌ها اشاره می‌شود.

جدول ۱. پژوهش‌هایی در خصوص متغیرهای سنتی

پژوهشگر (پژوهشگران)	متغیرهای سنتی
لئو و کروک [1] (۲۰۱۶)	میانگین ارزش تراکنش‌های بانکی، تعداد برداشت وجه نقد، تغییرات محدودیت اعتباری، نسبت ماه‌های معوقه، مبلغ بازپرداخت‌شده و مانده باقی‌مانده
روسلی [2] (۲۰۱۳)	تحصیلات و اشتغال
اسکاردوتیر و همکاران [3] (۲۰۱۹)	سن و مبلغ مخارج در ماه منتهی به درخواست وام
استینیکر و گوارتز [4] (۱۹۸۹)	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
کروپا، شوارز، آرمینگر و زیگلر [5] (۲۰۱۳)	سایر بدهی‌ها یا تعهدها در زمان درخواست وام و عملکرد تاریخی اعتبارگیرنده و رفتار پرداخت وی در گذشته
وست [6] (۲۰۰۰)	طول عمر حساب، سابقه اعتباری، طبقه‌بندی اشتغال، عملکرد (گردش) حساب، دارایی تحت مالکیت، سال اقامت، سایر وام‌های موجود، نوع سکونت، میزان وام، هدف از دریافت وام، سال استخدام و وضعیت حساب پس‌انداز

جدول ۲. پژوهش‌هایی در خصوص متغیرهای جایگزین

پژوهشگر (پژوهشگران)	متغیرهای جایگزین
ماسیوتین [7] (۲۰۱۵)	داده‌های موجود در رسانه‌های اجتماعی (مثل وضعیت تأهل: شامل متأهل، در عشق، پیچیده است و غیره) فعالیت (تعداد زیاد ویدئوهای ارسال‌شده) متوسط (۳-۷) تعداد اشتراک صفحه و غیره
بیورکگرن و گریسن [8] (۲۰۱۵)	حجم و توزیع تماس‌ها طی زمان، میزان شارژ نقدی و اعتباری موبایل

1. Leow & Crook
2. Rusli
3. Óskarsdóttir et al.
4. Steenackers & Goovaerts
5. Kruppa, Schwarz, Arminger & Ziegler
6. West
7. Masyutin
8. Bjorkegren & Grissen

پژوهشگر (پژوهشگران)	متغیرهای جایگزین
سوزا و همکاران[1] (۲۰۱۶)	طلاق، بیکاری و بیماری
اسکاردوتیر[2] (۲۰۱۹)	مدت زمان تماس تلفنی و تعداد تماس‌های تلفنی دریافتی از افراد بدحساب و دارای تأخیر پرداخت
بروکت و گلدن[3] (۲۰۰۷)	داده‌های مربوط به قبض‌های تلفن از ارائه‌دهندگان خدمات مخابراتی زمان مکالمه به‌ازای هر تماس و زمان استفاده از تلفن همراه
هان، پی و کامبر[4] (۲۰۱۱)	متن‌ها و فعالیت‌ها، تصاویر، ویدئوها، وبلاگ‌ها، علایق و پیوندهای شبکه‌های اجتماعی افراد
روسلی (۲۰۱۳)	تعداد دنبال‌کنندگان، دوستان و نوع (های) آن‌ها

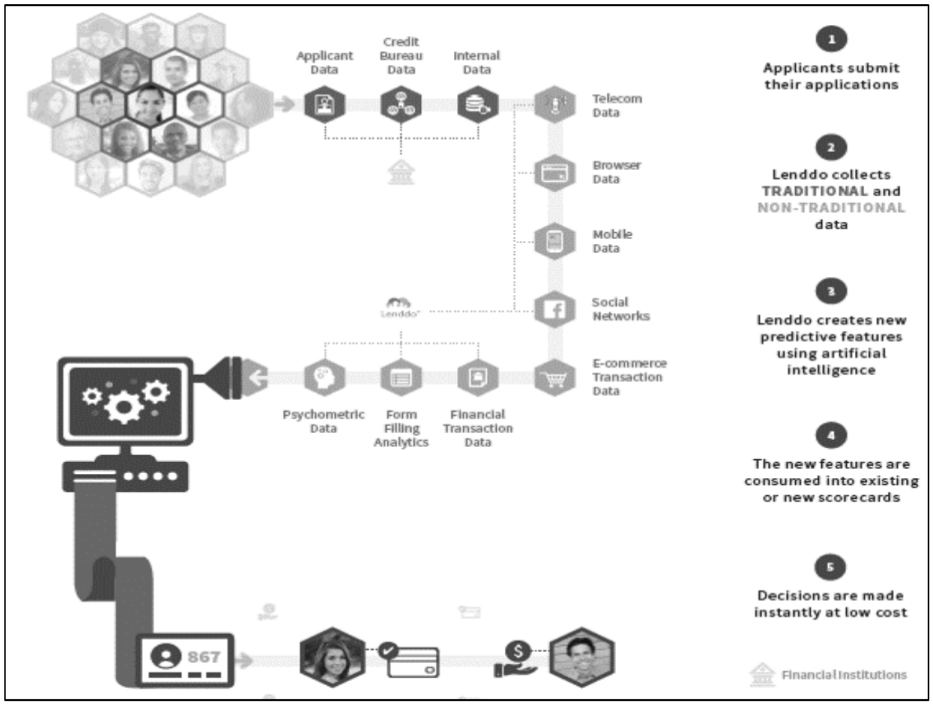
افرم [5] یک فین تک است که مدل خود برای ارزیابی ریسک اعتباری افراد را از طریق شماره تلفن آن‌ها طراحی کرده و از این طریق، به داده‌های مربوط به رسانه‌های اجتماعی و اطلاعات بازاریابی در تلفن‌های هوشمند افراد دست یافته است. گزارش شده است که افرم، در خصوص افراد، بیش از ۷۰۰۰۰ ویژگی اطلاعاتی به دست آورده که می‌تواند نتیجه اعتبار آنان را پیش‌بینی کند (ردراپ[6]، ۲۰۱۷).

شرکت برنامه‌های ارتباطی وی چت [7]، ۶۰۰ میلیون کاربر در سراسر چین دارد. این شرکت از طریق سرویسی به نام وی لی دای [8] تا سقف ۳۰۰۰۰ دلار وام ارائه می‌کند و هنگام تصمیم‌گیری اعتباری، به اطلاعات و محتوای تولیدشده توسط کاربر در برنامه متکی است. این شرکت به‌دلیل تصمیم‌های سریع خود مشهور است (وید، شانو مک تیگ[9]، ۲۰۱۶).

در مثالی دیگر، لندو[10] 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

1. Sousa et al.
2. Óskarsdóttir
3. Brockett & Golden
4. Han, Pei & Kamber
5. Affirm
6. Redrup
7. Wechat
8. Weilidai
9. Wade, Shan & McTeague
10. lenddo

رفتاری، داده‌های مرورگر، داده‌های ایمیل، داده‌های شبکه اجتماعی و داده‌های تلفن همـراه اسـتفاده می‌کند (شکل ۳).

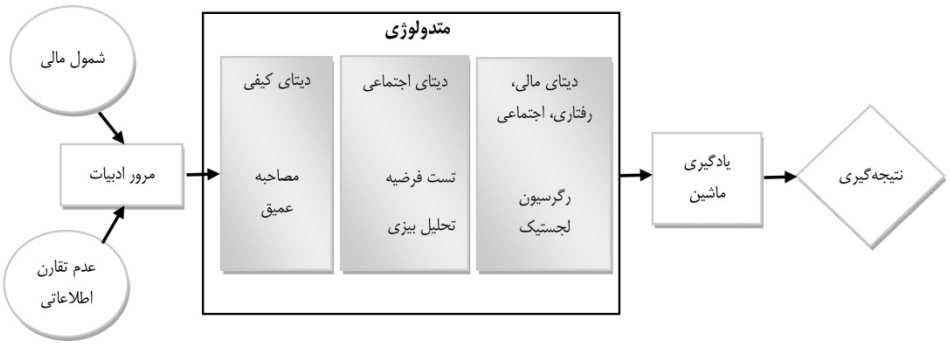


شکل ۳. مدل کسب‌وکار لندو
منبع: ردراپ (۲۰۱۷)

جدا از رسانه‌های اجتماعی، بسیاری از وام‌دهندگان از دیدگاه شبکه‌هـای اجتمـاعی، از داده‌هـای مخابراتی استفاده کرده‌اند. در ابتدا، ترکیب داده‌های مربوط بـه قـبض‌هـای تلفـن از ارائـه‌دهنـدگان خدمات مخابراتی به تجزیه و تحلیل الگوهای پرداخت و پرداخت‌ها محدود مـی‌شـد. بـا وجـود ایـن، متغیرهای دیگری مانند دقیقه به‌ازای هر تماس، زمان استفاده از تلفن همـراه و مـدت زمـان اسـاس شکل‌گیری شبکه‌های اجتماعی قابل اعتماد بودند (بروکت و گلدن، ۲۰۰۷).

روش‌شناسی پژوهش

در این پژوهش، از دو روش پژوهش کیفی و کمی استفاده شده است. در روش پـژوهش کیفـی کـه به‌منظور جمع‌آوری و تحلیل داده‌های غیرعددی برای شناخت مفاهیم، دیدگاه‌هـا و تجـارب مختلـف است از مطالعه کتابخانه‌ای و مصاحبه عمیق با خبرگان استفاده شده است. در روش پـژوهش کمـی، به بررسی سیستماتیک فرضیه‌های پژوهش با جمع‌آوری داده‌های قابل سنجش و انجام تکنیک‌هـای آماری، ریاضی پرداخته شده است. چارچوب مطالعات و نتیجه‌گیری در این پژوهش، به‌طـور خلاصـه در شکل ۴ نشان داده شده است.



شکل ٤. نقشه راه پژوهش

همان‌طور که در شکل ۴ مشاهده می‌شود، با استفاده از روش مصاحبه عمیق کیفی با متخصصان بانکداری و اعتبارسنجی، ابتدا متغیرهـای اجتمـاعی کاندیـدای بررسـی مشـخص شـدند. شناسـایی متغیرهای اجتماعی با حدس و گمانه‌زنی در خصوص انواع روابط اجتماعی افراد انجام شده و پـس از اجماع و تأیید شهودی و منطقی اثر آن‌هـا توسـط خبرگـان، مجموعـه داده‌ای در خصـوص وام‌هـای پرداختی توسط یک بانک خصوصی به اشخاص حقیقی در پنج سال گذشته و در بـازار ایـران تـأمین شد. تمام داده‌ها برای حفاظت از حریم خصوصی به‌صورت رمزگـذاری و هـش‌شـده دریافـت شـدند. مجموعه داده این پژوهش دارای ۳۰۳۱۱۱ وام پرداختی (مشاهده) است. در مرحلـه بعـد، فیلـدهـا یـا ویژگی‌های هر یک از مشاهدات براساس یافته‌های حاصل از مصاحبه کیفی که به شرح طبقات زیـر خلاصه‌بندی می‌شود، مشخص شد.

- طبقه وام: مقدار و نوع وام‌های دریافتی و در نهایت هدف از وام.
- طبقه دارایی متقاضی: مالکیت دارایی، مالکیت خودرو، درآمد و نوع درآمد.
- طبقه شرایط زندگی: تحصیلات، وضعیت تأهل، تعداد فرزندان و تعداد افراد تحت تکفل دیگر.
- جمعیت‌شناسی: سن و جنسیت.
- طبقه عوامل خارجی: امتیازهای 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، هفته گذشته، ماه گذشته، سه ماهه آخر و سال گذشته).
- محل سکونت (مالکی، اجاره‌ای، زندگی با والدین و غیره).
- طبقه ثبات زندگی: مدت زمان اشتغال نزد کارفرمای فعلی، همچنین، مدت زندگی در آدرس فعلی و آدرس‌های قبلی، مدت زمان داشتن یک شماره تلفن.
- طبقه مرتبط با شغل: نوع شغل و نوع کارفرما.
- طبقه اجتماعی: تعداد دوستانی که برای بازپرداخت وام خود با مشکل مواجه بودند و در واقع در بازپرداخت وام تأخیر، بی نظمی و نکول کردند.
- طبقه ترکیب اعتباری: انواع مختلف وام‌هایی که متقاضی دارد شامل ارزی، مصرفی یا شرکتی، مسکن و خودرو و...
- سوابق عملکرد اعتباری: تعداد تأخیرها و عدم بازپرداخت‌های گذشته و مبالغ آن‌ها، همچنین تعداد تسویه زودهنگام اقساط یا تسویه کامل وام.
- طبقه عملکرد حساب: گردش حساب و الگوی هزینه یا تعداد و مبلغ برداشت‌ها با استفاده از کارت‌های نقدی و نیز محل جغرافیایی استفاده از کارت.

در این پژوهش، به‌منظور دستیابی به داده‌های اجتماعی افراد و کشف شبکه ارتباطی آنان از سه متغیر اجتماعی جایگزین (نماینده)، شامل متغیرهای زیر استفاده شد.

۱. متغیر وجود رابطه پیوند اجتماعی نشان‌دهنده تعداد افرادی است که با فرد مدنظر رابطه دارند یا در شبکه اجتماعی وی قرار دارند.
۲. متغیر تکرار یا ثبات پیوند اجتماعی نشان‌دهنده تعداد دفعات است که فرد مدنظر در بازه زمانی مورد مطالعه با افراد درون شبکه اجتماعی خود ارتباط برقرار کرده است.

۳. متغیر عمق رابطه یا کیفیت پیوندها نشان‌دهنده مدت ارتباط در هر بار برقراری ارتباط یا حجم دادوستد مالی انجام‌شده بین آن‌ها است.

از سوی دیگر، به‌دلیل نبود زیرساخت‌های جمع‌آوری این قبیل داده‌ها به دو روش اقدام شد. ابتدا از طریق بررسی شبکه ارتباطی افراد با توجه به فهرست شماره‌های مندرج در دفترچه تلفن آنان و دوم از طریق بررسی خزشی[1] تراکنش‌های مالی افراد با دیگران.

در ادامه، برای استخراج اطلاعاتی در خصوص گروه‌بندی و رفتار انواع وام‌گیرندگان، از آزمون‌ها و روش‌های آماری استفاده شد. در این بخش، هدف کشف ارتباط بین داده‌های اجتماعی و امتیازهای اعتباری از طریق استنباط آماری است. این کار در سه مرحله انجام شد که در ادامه مقاله تشریح خواهد شد. نخست، آزمون این فرضیه که وام‌گیرندگانی که در نهایت به نکول می‌پردازند، شبکه اجتماعی بد بزرگ‌تری دارند. دوم، آزمون اینکه آیا متغیرهای اجتماعی، برای ارزیابی‌کنندگان ریسک اعتباری در خصوص ریسک وام‌ها، اطلاعات بهتری 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. سوم، از یک مدل رگرسیون لجستیک و چند مدل پیشرفته یادگیری ماشین به‌منظور برآورد ضریب متغیرهای شبکه اجتماعی در توضیح امتیازهای اعتباری استفاده شد و در نهایت، نتایج حاصل از هر دو روش تحلیل شده و با یکدیگر مقایسه شدند.

جدول ۳. ماتریس به‌هم‌ریختگی

آماره‌های اندازه‌گیری	نتایج مشاهده‌های آماری		**دقت[2] =(TP + TN) / N**	
	منفی‌های واقعی	**مثبت‌های واقعی**		
صحت[3]	مثبت‌های کاذب (FP)	مثبت‌های درست (TP)	مثبت‌های پیش‌بینی‌شده	پیش‌بینی مدل
	منفی‌های درست (TN)	منفی‌های کاذب (FN)	منفی‌های پیش‌بینی‌شده	
	تشخیص[4]	حساسیت[5]	آماره‌های اندازه‌گیری	

1. Crawling
2. Accuracy
3. Precision
4. Specificity
5. Recall

برای نشان دادن عملکرد مدل‌ها از آمارهای محاسبه‌شده مربوط به ماتریس به‌هم‌ریختگی[1] ارائه‌شده توسط ونگ، ما، هوآنگ و شو[2] (۲۰۱۵) طبق جدول ۳ استفاده شد. همچنین، به‌منظور ارزیابی قدرت برازش مدل رگرسیون لجستیک، از شبه R^2 استفاده شد[3].

یافته‌های پژوهش

در بخش تجزیه و تحلیل پژوهش، جامعه آماری وام‌گیرندگان پس از پاک‌سازی داده‌ها ۳۰۰۴۲۷ نفر در نظر گرفته است. شایان ذکر است که برای اجرای مدل‌های مبتنی بر ماشین، لازم است از تعداد کافی و زیاد داده در نمونه‌های بررسی‌شده استفاده شود تا ماشین بتواند انواع خطاهای ذکرشده در ماتریس به‌هم‌ریختکی را تجربه کند. در غیر این صورت، مدل با خطاهای over fit و under fit مواجه می‌شود.

پس از بررسی ویژگی‌های اطلاعاتی یا متغیرهای مرتبط با هر یک از افراد یا همان مشاهده‌های جامعه آماری وام‌گیرندگان، در مجموع ۴۲ ویژگی انتخاب شد که به شرح جدول ۴ است.

جدول ٤. ویژگی‌های استفاده‌شده در تجزیه و تحلیل

R	متغیرها	نوع متغیر	توضیحات
۱.	نوع حساب	مالی	شامل 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 (۰/۶۳)، زن (۰/۳۷)
۷.	وضعیت تأهل	مالی	متأهل (۰/۷۳)، مجرد (۰/۱۵)، مطلقه (۰/۷) و بیوه (۰/۵)
۸.	تعداد فرزندان	رفتاری	دارای فرزند (۰/۳۱) و بدون فرزند (۰/۶۹)
۹.	تحت تکفل	رفتاری	تعداد تحت تکفل: ۱ (۰/۲۲)، ۲ (۰/۵۰)، ۳ (۰/۱۸)، ۴ (۰/۰۹) و ۵ یا بالاتر (۰/۰۱)
۱۰.	تحصیلات	رفتاری	راهنمایی یا کمتر (۰/۷۶)، دبیرستان (۰/۰۳) و دانشگاهی (۰/۲۱)
۱۱.	مدت ارتباط با بانک	رفتاری	تاریخ افتتاح نخستین حساب

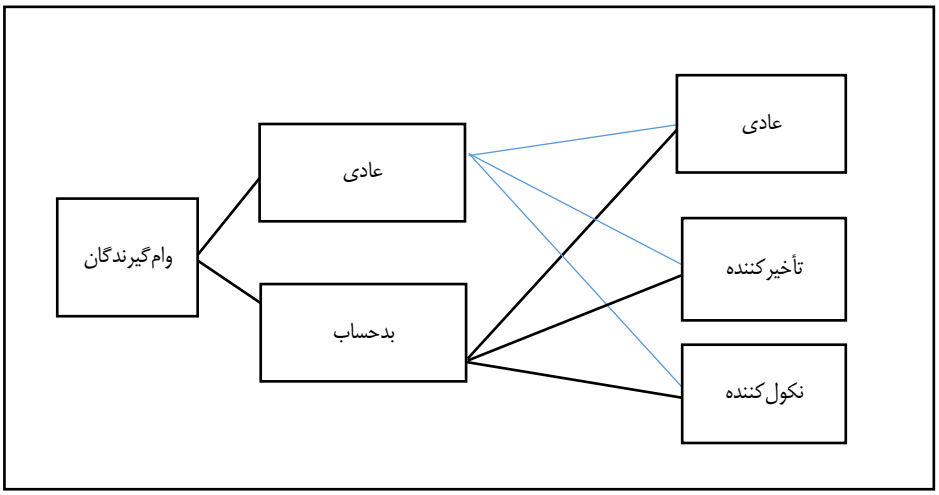
1. Confusion matrix
2. Wang, Ma, Huang & Xu
3. Pseudo R^2

R	متغیرها	نوع متغیر	توضیحات
.۱۲	شغل	رفتاری	۱۷ نوع شغل
.۱۳	مدت اشتغال	مالی	از تاریخ نخستین استخدام
.۱۴	نوع کارفرما	مالی	۱ خویش فرما، ۲ دولتی، ۳ خصوصی و ۴ هیچ‌کدام
.۱۵	مبلغ درآمد	مالی	درآمد اعلامی ماهانه ۴ طبقه
.۱۶	نوع درآمد	مالی	بیکار، دانشجو، مستمری‌بگیر، مرخصی، .... ۸ طبقه اصلی
.۱۷	وضعیت محل زندگی	رفتاری	مالک (۰/۸۸)، خانه والدین (۰/۶) و اجاره (۰/۶)
.۱۸	مدت سکونت در آدرس فعلی	رفتاری	۳ سال یا کمتر (۰/۴۲)، ۳ تا ۷ سال (۰/۳۱) و > ۷ سال (۰/۲۷)
.۱۹	داشتن ماشین	رفتاری	دارای ماشین (۰/۶۷) و بدون ماشین (۰/۳۳)
.۲۰	خوش‌حسابی	مالی	شامل سه طبقه خوش‌حساب، معوق و نکول
.۲۱	مشخصات اعتبار (۴ متغیر)	مالی	شامل مبلغ بدهی، نرخ، مدت تسهیلات و دلیل وام
.۲۲	مبلغ اقساط ماهانه	مالی	جمع مبلغ اقساط
.۲۳	معرفی ضامن	رفتاری	وضعیت دودویی
.۲۴	ارائه چک تضمین	رفتاری	وضعیت دودویی
.۲۵	نمره اعتبارسنجی داخلی	مالی	مقداری
.۲۶	نمره 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
.۳۰	معرفی املاک و مستغلات	مالی	ارائه مستندات املاک و مستغلات
.۳۱	ارزش املاک و مستغلات	مالی	اظهاری
.۳۲	سابقه چک برگشتی	رفتاری	شامل 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 اینترنت بانک	رفتاری	مدت زمانی که از آخرین ورود می‌گذرد.
.۳۷	وجود پیوند اجتماعی بد	اجتماعی	۰ (۰/۱۳) و ۱ (۰/۸۷)
.۳۸	تعداد پیوند اجتماعی بد	اجتماعی	۱-۲ (۰/۱۲)، ۳-۹ (۰/۵۶)، ۱۰-۱۴ (۰/۲۱)، ۱۵ و بالاتر (۰/۱۱)
.۳۹	عمق پیوند اجتماعی بد (زمان)	اجتماعی	متوسط زمانی از نخستین ارتباط با شبکه بد: ۱ سال یا کمتر (۰/۳۲)، ۱ تا ۳ سال (۰/۳۱) و > ۳ سال (۰/۳۷)

در جامعه آماری پژوهش، داده‌های مرتبط با بسیاری از متغیرها یا ویژگی‌های بیان‌شده وجود داشته یا محاسبه‌شدنی است. به‌منظور دستیابی به داده‌های اجتماعی مدنظر داده‌های، با توجه به کلیدهایی مثل شماره موبایل مشتری و شماره موبایل مخاطبان وی (با رجوع به دفترچه تلفن مشتری)، شماره کارت و شماره شبای مقصد تراکنش‌های مشتری و اطلاعات تراکنش‌های مالی بین مشتریان با یکدیگر مشخص شده و با توجه به شمارش فراوانی تراکنش‌های بین افراد و سوابق انجام این تراکنش‌ها، داده‌های مربوطه متغیرهای وجود، ثبات و کیفیت ارتباط اجتماعی بین افراد استخراج شد و در کنار سایر داده‌های مالی و رفتاری افراد قرار گرفت.

تجزیه‌وتحلیل آماری

در ادامه، نتایج تجزیه و تحلیل اثر متغیرهای وجود ارتباط اجتماعی، ثبات اجتماعی و کیفیت اجتماعی بر 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 به شرح جدول‌های ۵ تا ۷، جدول‌های مرتبط با تحلیل بیزی (جدول‌های ۸ و ۹) و روش‌های یادگیری ماشین (جدول‌های 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)	۰/۱۴	۰/۱۹
انحراف از معیار (STD. D)	۰/۴۴	۰/۵۱

جدول ٦. آمار توصیفی برای اثر وجود پیوند اجتماعی با افراد دارای تأخیر و معوق

اثر متغیر وجود پیوند اجتماعی با افراد دارای تأخیر و معوقات		
	عادی	**نکول**
تعداد نمونه	۲۷۵۷۶۷	۲۴۶۶۰
میانگین (MEAN)	۱/۴۷	۱/۶۷۹
انحراف از معیار (STD. D)	۰/۶۵	۰/۷۲

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	۲۷۵۷۶۷	۲۴۶۶۰
میانگین (MEAN)	۲/۳۶۶	۲/۲۷۸
انحراف از معیار (STD. D)	۰/۸۰	۰/۸۳

همان‌طور که در جدول‌های بالا مشاهده می‌شود، بین میانگین های دو گروه از وام‌گیرندگان نام‌برده 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، میانگین تعداد پیوندها در مشتریان بدحساب در مقایسه با سایر مشتریان بدحساب (دارای معوق یا نکول‌کننده) از همین معیار در مورد مشتریان خوش‌حساب بالاتر است و در عین حال، با مقایسه میانگین پیوندهای

دو گروه از مشتریان بدحساب و خوش‌حساب با افراد عادی و خوش‌حساب مشاهده می‌شود که مشتریان بدحساب، با مشتریان عادی پیوند اجتماعی کمتری دارند. این نتایج به این معنا است که اثر داده‌های اجتماعی در ارزیابی ریسک اعتباری و پیش‌بینی امتیازهای اعتباری توجیه‌پذیر است. بررسی نتایج آزمون U من ویتنی نیز نشان داد که می‌توان فرضیه اثر متغیرهای اجتماعی بر احتمال تأخیر و نکول کردن را تا حدی پذیرش کرد. انتخاب آزمون U من ویتنی به دلایل زیر بود:

الف. متغیرهای اندازه‌گیری‌شده می‌توانند براساس یک مقیاس (تعداد روابط اجتماعی بدحساب و نکول) رتبه‌بندی شوند.

ب. نمونه‌ها مستقل هستند.

ج. نمونه‌ها به‌اندازه کافی بزرگ هستند؛ زیرا n۲ ≥۳۰ و n۱.

به‌منظور انجام تجزیه و تحلیل بیزی بر مبنای طبقه‌بندی انجام‌شده طبق شکل ۵، ابتدا شانس خوش‌حسابی کل نمونه با تقسیم احتمال خوب بودن مشتری در نمونه انتخابی یعنی P(G) بر احتمال بد بودن مشتری در نمونه انتخابی یعنی $P\ (B)$ محاسبه می‌شود. احتمال خوب بودن یک مشتری در جمعیت از تقسیم تعداد مشتریان خوب موجود در جمعیت بر تعداد کل جمعیت و احتمال بد بودن یک مشتری در جمعیت از تقسیم تعداد مشتریان بد موجود در جمعیت بر تعداد کل جمعیت به دست می‌آید. در این تقسیم، به‌دلیل یکسان بودن مخرج هر دو احتمال و حذف آن هنگام تقسیم، شانس خوش‌حسابی جمعیت از رابطهٔ زیر به‌دست می‌آید.

$$Opop\ =\ P\ (G)/\ P\ (B)$$ رابطهٔ ۱)

شانس خوش‌حسابی جمعیت = ۱۱/۱۸۳ = ۲۴۶۶۰ ÷ ۲۷۵۷۶۷

برای شروع کار، ابتدا افراد موجود در نمونه انتخابی با شمارش متغیر وجود ارتباط با افراد بد، به ۲۰ طبقه دل‌خواه قسمت شد. در ادامه، شانس خوش‌حسابی و شانس اطلاعات هر طبقه با توجه به فرمول بالا محاسبه شد تا به‌عنوان وزن آن طبقه (وزن شاهد) استفاده شود. وزن شاهد برای همه طبقات محاسبه شد تا طبقاتی که قدرت تمایز بیشتری دارند و به پیش‌بینی بهتر احتمال نکول کمک می‌کنند، مشخص شود. در ضمن، وزن‌های شاهد پرت از محاسبه‌ها حذف شدند. محاسبه‌های آماری هر یک از طبقات، در جدول ۸ خلاصه می‌شود.

جدول ۸. احتمالات، شانس خوش‌حسابی و شانس اطلاعات و وزن شواهد برای طبقات بد

طبقه 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
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	۰/۰۰۲۱	۰/۰۰۲۴	۰/۸۵۸۵	−۰/۱۵۲۶
۱۳	۳۵۸	۴۴	۸/۱۳۶۴	۰/۰۰۱۳	۰/۰۰۱۳	۰/۰۰۱۸	۰/۷۲۷۶	−۰/۳۱۸
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 در مقایسه با سایر طبقات، شانس اطلاعاتی کم یا دورافتاده یا ناهنجار داشتند و هیچ کمکی به تحلیل نمی‌کردند، حذف و در طبقات دیگر ادغام شدند. در ادامه با استفاده مجدد از فرمول بالا، دوباره شانس خوش‌حسابی هر طبقه محاسبه شد تا به‌عنوان

وزن شاهد آن طبقه استفاده شود. 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).

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

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

در ادامه پژوهش، نتایج رگرسیون لجستیک روی انواع متغیرهای مختلف (به تفکیک مالی F، رفتاری B و اجتماعی S) یک بار به‌تنهایی و یک بار با ورود متغیرهای اجتماعی اجرا شد. در هر بار اجرا شدن مدل، نتایج آماره‌های اندازه‌گیری محاسبه، طبقه‌بندی شده و برای تفسیر با وضعیت‌های دیگر مقایسه می‌شوند. شایان ذکر است که حجم نمونه استفاده‌شده در این بخش ۴۹۳۲۰ وام‌گیرنده است که به‌صورت تصادفی و با توزیعی مشابه از بین دو گروه بدحساب و خوش‌حساب از جامعه پژوهش انتخاب شده است. نمونه انتخابی با نسبت ۷۰ و ۳۰ به‌ترتیب برای ساختن مدل و آزمایش نتایج استفاده شده است. در نتیجه، حجم نمونه آزمایش شامل ۱۴۷۹۶ مشاهده و حجم نمونه مدل‌سازی دارای ۳۴۵۲۴ مشاهده است.

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

جدول ۱۰. عملکرد مدل رگرسیون لجستیک به‌تفکیک 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	/۶۲	/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 بالا، 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 به دست آمد. همچنین، دقت مدل از ۵۷ درصد به ۵۸ درصد افزایش یافت (مقایسه ستون اول و آخر).

جدول ۱۱. جدول مقایسه‌ای یادگیری ماشین

یادگیری ماشین	فقط با ویژگی‌های سنتی		با ویژگی‌های جایگزین	
	Accuracy	Log Loss	Accuracy	Log Loss
K-Nearest Neighbors	۰/۶۰۷	۴/۴۰۷	۰/۵۸۹	۴/۴۵
Support Vector Machine	۰/۶۶۴	۰/۶۱۴	۰/۶۷	۰/۶۰۷
Decision Trees	۰/۵۷۱	۱۴/۴۶	۰/۵۸	۱۴/۴۰۳
Random Forests	۰/۶۲۵	۰/۹۵	۰/۶۳۶	۰/۸۲۳
Extreme Gradient (XGBoost)	۰/۶۶۶	۰/۶۰۴	۰/۶۸	۰/۵۸۹
Naïve Bayes (Gaussian)	۰/۶۰۴	۰/۹۱	۰/۶۴	۱/۳۱۶
Quadratic Discriminant Analysis	۰/۵۷۳	۰/۹۰۵	۰/۵۲۵	۱۰/۱۴۹

در بخش پایانی، الگوریتم‌های یادگیری ماشین روی جامعه آماری پژوهش اجرا شد. دقت مدل از رگرسیون لجستیک فراتر رفت و log loss به‌میزان زیادی همان‌طور که در جدول زیر مشاهده می‌شود، کاهش یافت (جدول ۱۱).

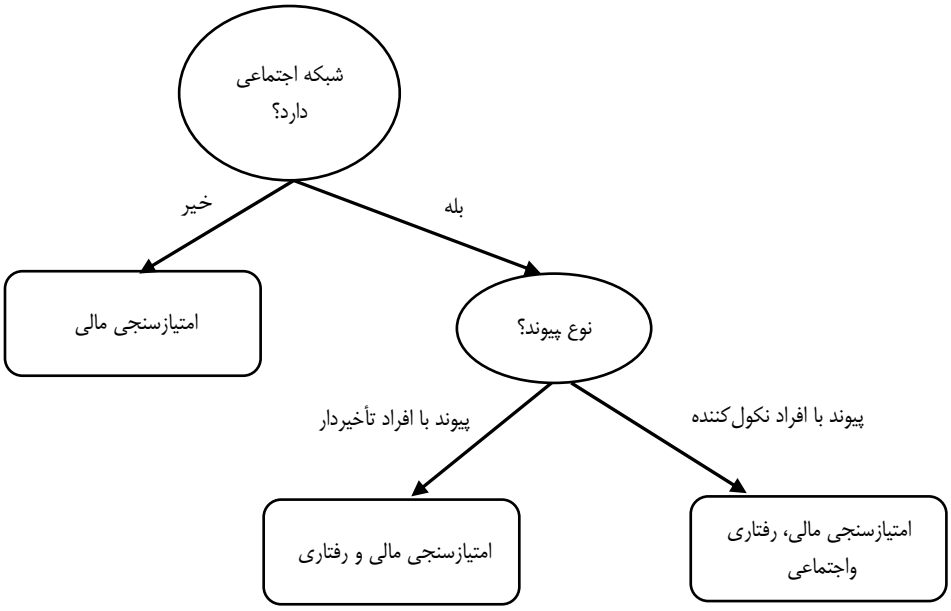
نتیجه‌ها، کاربردها و پیشنهادها

در این پژوهش، خلاصه یافته‌ها به شرح موارد زیر است:

- پیوندهای اجتماعی ضعیف یا بد بر احتمال نکول تأثیر دارد، هرچند اهمیت یا وزن این اثر ضعیف است.
- داشتن دوستان و خانواده بدحساب در شبکه ارتباطی وام‌گیرنده با احتمال نکول رابطه مثبت و بر این اساس با نمره اعتباری فرد رابطه منفی دارد.
- به‌دلیل اهمیت و ایجاد صرفه‌جویی‌های کلان مالی، نمی‌توان از اثر متغیرهای اجتماعی صرف‌نظر کرد.
- داده‌های شبکه‌های اجتماعی با داده‌های رفتاری تفاوت دارد. اولی، نتیجه انتخاب آگاهانه افراد و دومی، نتیجه شخصیت و سلیقه است، برای مثال انتخاب ماشین و انتخاب دوست.
- چنانچه در شبکه اجتماعی اشخاص، افراد نکول‌کننده وجود داشته باشند، باید ریسک و امتیاز اعتباری بر مبنای مالی، رفتاری و اجتماعی ارزیابی شود. وجود پیوندهای اجتماعی با افرادی که دارای بدهی معوق هستند یا در بازپرداخت‌های خود تأخیر دارند، تأثیر مستقیمی بر امتیاز اعتباری ندارد. در این موارد، متغیرهای مالی سنتی و رفتاری برای محاسبه امتیاز اعتباری کافی هستند (شکل ۶).

موارد زیر را می‌توان به‌عنوان کاربردهای نتایج پژوهش برشمرد:

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



شکل ٦. روش امتیازسنجی با توجه به انواع پیوندهای اجتماعی

در پایان، پیشنهادهایی مبتنی بر نتایج پژوهش به شرح موارد زیر است:

- اقدام برای جمع‌آوری داده‌های اجتماعی با روش‌های تشویقی مثل تخفیف توسط بانک‌ها و مؤسسه‌های مالی.
- تأسیس یک اپراتور شبکه اجتماعی برای محاسبه امتیاز اجتماعی افراد و ایجاد API برای بانک‌ها و سایر تولیدکنندگان و استفاده‌کنندگان امتیاز اعتباری.
- حمایت و ترغیب شرکت‌های فین تک وام‌دهی P2P برای پیاده‌سازی راهکای خلاقانه برای جمع‌آوری داده‌های اجتماعی
- وضع قانون محافظت از داده‌های شخصی و الزامات قانونی برای استفاده از داده‌های اجتماعی.
- توافق با تأمین‌کننده داده‌های اجتماعی مثل شرکت مخابرات در خصوص 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 اشاعه فرهنگ اهمیت و ارزشمندی پیوندهای اجتماعی.
- استفاده از داده‌های اجتماعی در لحظه برای اعطای وام‌های مصرفی و کسب‌وکاری مانند مورد مطالعه لندو.

از سوی دیگر، با وجود مزایای شایان توجهی 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 مشکلات زیادی ایجاد کند، به همین دلیل در برخی از کشورها در این باره قوانین سختی وجود دارد.
۲. احتمال اینکه افراد خود را در شبکه‌های اجتماعی به‌نحوی نشان دهند که افراد خوبی برای بانک هستند وجود دارد. این مشکل، با عنوان بازی با سیستم شناخته می‌شود، اما در 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 گنگ بودن را نیز به آن اضافه کرده است. اصولاً قابلیت تعبیرپذیری و تفسیر نتایج به‌دست‌آمده از تکنیک‌های یادگیری ماشین، یکی از نگرانی‌های مهم قانون‌گذار و صنعت اعتبارسنجی است. این نگرانی، با ورود متغیرهای جدید اجتماعی و افزایش متغیرها و پیچیده شدن الگوریتم‌ها بیشتر شده و باعث ایجاد نگرانی در خصوص صحت و عادلانه بودن خروجی‌های مدل خواهد شد.

منابع و مآخذ

انگلیسی

Bjorkegren, D., & Grissen, D. (2015). Behavior revealed in mobile phone usage predicts loan repayment. ***Marketing Science eJournal***, DOI:10.2139/ssrn.2611775

Brockett, P. L. & Golden, L. L. (2007). Biological and psychobehavioral 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.

Han, J., Pei, J., & Kamber, M. (2011). ***Data mining: concepts and techniques*** (3th ed.). Elsevier.

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.

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. ***Business Informatics***, 33 (3), 15–23.

Siddiqi, N. (2017). ***Intelligent Credit Scoring: Building and Implementing Better Credit Risk Scorecards*** (2th ed.). Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

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

Óskarsdóttir, M., Bravo, C., Sarraute, C., Vanthienen, J., & Baesens, B. (2019). The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics. ***Applied Soft Computing***, 74, 26-39.

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. ***The Wall Street Journal.*** Available in: https://www.wsj.com/articles/SB10001424127887324883604578396852612756398

Steenackers, A., & Goovaerts, M. (1989). A credit scoring model for personal loans. ***Insurance: Mathematics and Economics***, 8(1), 31-34.

Wade, M. R., Shan, J., & McTeague, L. (2016). ***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.

Weke, P., & Ntwiga, D. B. (2016). Consumer lending using social media data. *International journal of scientific research, 3*.

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.