

مقایسه مدل بهینه رگرسیون لجستیک چندگانه و باینری برای رتبه‌بندی اعتباری مشتریان حقیقی بانک رفاه کارگران

ناهید بهارلو*، علی اکبر امین‌بیدختی** و محمدجواد محقق‌نیا***

تاریخ پذیرش: ۱۳۹۵/۰۸/۲۳

تاریخ دریافت: ۱۳۹۴/۱۱/۲۸

چکیده

یکی از وظایف اصلی موسسات مالی در حمایت از فعالان اقتصادی، اعطای تسهیلات و انجام تعهدات آنان است. پیش‌بینی ریسک اعتباری ناشی از اعطای تسهیلات و در نتیجه مدیریت این ریسک از مهم‌ترین چالش‌های پیش‌روی بانک‌ها و موسسات مالی است. هدف اصلی در این کار تهیه مدل بهینه لجستیک جهت تعیین اعتبار مشتریان حقیقی با بررسی نقش تعداد رده‌های اعتبار مشتری از دو رده خوش حساب و بدحساب تا چهار رده خوش حساب، سررسیدشده، معوق و مشکوک‌الوصول و نیز شناسایی مهم‌ترین متغیرهای مستقل تاثیرگذار بر اعتبار مشتریان حقیقی بانک رفاه کارگران است. نتایج هر دو مدل برازش لجستیک چندگانه و باینری تهیه شده در این کار نشان‌دهنده اهمیت نسبی متغیرهای سطح تحصیلات و سن مشتری حقیقی نسبت به سایر متغیرهای مستقل است. نتایج یکسان هر دو مدل می‌تواند نشان‌دهنده تاثیر کم افزایش تعداد رده‌بندی متغیر وابسته (اعتبار مشتری) و یا توزیع نامناسب تعداد مشتریان برای رده‌های مختلف اعتبار مشتری باشد.

طبقه‌بندی JEL: C45, C35, G32, G24

کلیدواژه‌ها: اعتبارسنجی، مشتریان حقیقی، بانک، برازش لجستیک، باینری، چندگانه، لجیت، آماره‌های آزمون.

* دانشجوی کارشناسی ارشد، گروه مدیریت، واحد سمنان، دانشگاه آزاد اسلامی، سمنان، پست الکترونیکی:

nahidbgh@yahoo.com

** دانشیار دانشکده مدیریت دانشگاه سمنان، سمنان، ایران - نویسنده مسئول، پست الکترونیکی:

aliaminbeidokhti@yahoo.com

*** استادیار دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبایی، تهران، پست الکترونیکی:

mgmohagh@yahoo.com

۱- مقدمه

اعتبارسنجی مشتریان بانک‌ها برای دریافت و اخذ تسهیلات بانکی امروزه یک ضرورت تردیدناپذیر در سیستم بانکی بیشتر کشورهای دنیا است. ایجاد تعادل میان عرضه و تقاضا در منابع و تسهیلات بانکی، مدیریت کاهش معوقات بانکی و رهایی از سیستم وثیقه‌محوری از جمله مواردی است که موجب شده است مشتریان بانک‌ها اعتبارسنجی شوند. در واقع بانک‌ها از طرق مختلف مشتریان خوش حساب یا بدحساب را شناسایی می‌کنند (محتشمی، ۱۳۸۹).

اعتبارسنجی مشتریان در ارائه تصویری شفاف از وضعیت و توانایی مشتری در ایفای به موقع تعهدات، بازداشتن آنان از استفاده بیش از حد منابع و سقوط به ورطه بحران مالی نقش اساسی دارد. همچنین اعتبارسنجی مشتریان بانک‌ها باعث سهولت در امور بانکی، کاهش ریسک سرمایه‌گذاری، تسهیل در رتبه‌بندی مشتریان و ایجاد بستری برای مبارزه با پولشویی خواهد شد («اعتبارسنجی مشتریان ضرورت نظام بانکی»، ۱۳۸۹).

رتبه‌بندی اعتباری یکی از ابزارهای ارزشیابی اصلی برای موسسات گوناگون در دهه‌های اخیر بوده که در مباحث مختلفی همانند مباحث مالی و حسابداری مورد استفاده قرار گرفته است. روش‌های رتبه‌بندی متفاوتی برای مباحث طبقه‌بندی و پیش‌بینی رده اعتباری مشتریان استفاده شده که مهم‌ترین آن‌ها روش‌های آماری هستند.

به طور کلی روش‌های رتبه‌بندی اعتباری را می‌توان به دو گروه عمده روش‌های پارامتریک (مانند پروبیت، لجیت، تحلیل تمایزی و رگرسیون لجستیک) و روش ناپارامتریک و داده‌کاوی (مانند درختان تصمیم‌گیری، شبکه‌های عصبی و سیستم‌های خبره) تقسیم‌بندی کرد (البرزی و همکاران ۱۳۸۹). برخی از مدل‌های ترکیبی نیز در اعتبارسنجی مورد استفاده قرار گرفته‌اند، مثل مدل عصبی فازی، مدل فازی ماشین بردار پشتیبان و مدل ترکیبی شبکه‌های عصبی.

۲- پیشینه و سابقه موضوع

عبدو^۱ و همکارانش (۲۰۰۷) با استفاده از مدل‌های رتبه‌بندی اعتباری، مشتریان بانک‌های مصر را اعتبارسنجی کردند. تکنیک‌های بکار برده شده در این کار شامل تحلیل دودویی

1- Abdou

(باینری)، تحلیل پروبیت و برازش لجستیک است. در این کار وابسته باینری درنظر گرفته شده و ۲۰ متغیر مستقل شامل: مقدار وام، تعداد اقساط، شرکت، جنسیت، وضعیت تاهل، سن، درآمد ماهیانه، ملکیت خانه، درآمدهای اضافه، تلفن خانه، موقعیت شخصی، سطح تحصیلات، وام از بانک‌های دیگر، وضعیت کارت اعتباری و ... هستند.

در اعتبارسنجی صورت گرفته توسط عبدو و همکاران، میزان وام و میزان درآمد ماهیانه از پارامترهای مهم تحلیل باینری هستند. متوسط طبقه‌بندی صحیح ۸۶/۷۵ درصد برای مدل تحلیل باینری است. برای تحلیل پروبیت همه متغیرها بجز سه متغیر درآمد اضافه، جنسیت و شرکت متغیرهای مهم و تاثیرگذار هستند. متوسط طبقه‌بندی صحیح ۸۹/۳۳ درصد برای مدل تحلیل پروبیت است. برای برازش لجستیک بجز سه متغیر درآمد اضافه، جنسیت و شرکت، متغیرهای دیگر مهم بوده و مدل برای حالت بدون این سه متغیر مجدد اجرا شده که متوسط طبقه‌بندی صحیح ۸۷/۹۵ درصد است.

ترایان و همکاران (۲۰۱۳) اعتبارسنجی مشتریان حقیقی بانک سپه را با مدل‌های لجستیک انجام دادند. در کار ایشان، متغیر وابسته اعتبار به صورت باینری (خوش حساب و بدحساب) در نظر گرفته شده بود. در این مرجع، ذکر شده است که محدودیت منابع مالی یکی از دلایل عمده وجود چالش تشخیص و ارزیابی توانایی مالی مشتریان برای بازپرداخت تسهیلات دریافتی آنها است.

مشخصه‌هایی همانند سن، میزان تحصیلات، جنسیت، مقدار تسهیلات و ارزش وثیقه^۱ مهم‌ترین پارامترهای لحاظ شده در کار اعتبارسنجی صورت گرفته توسط ترایان و همکاران است. نتایج اقدام ایشان نشان می‌دهد که متغیرهای سن و میزان تحصیلات مهم‌ترین پارامترهای تاثیرگذار بر اعتبار مشتریان و رتبه‌بندی اعتباری آنها است و بقیه متغیرها، ارتباط معناداری با وضعیت اعتبار مشتریان دارند.

کریمی (۲۰۱۴) برای بانک‌های تجاری ایران ریسک اعتباری را با استفاده از رگرسیون لجستیک و شبکه عصبی مصنوعی مدل‌سازی کرده است. این کار بر روی ۳۱۶ نمونه و ۵ متغیر مستقل انجام شده است. نتایج مدل‌سازی کریمی نشان می‌دهد که شبکه عصبی مصنوعی نسبت به برازش لجستیک توانایی بهتری برای شناسایی مشتریان بدحساب دارد. اعتبار به صورت یک متغیر باینری در نظر گرفته شده است. نتایج مدل‌سازی کریمی

نشان داده است که افزایش نرخ بهره و تاخیر در بازپرداخت موجب افزایش ریسک اعتباری می‌شود. مدل شبکه عصبی در کار کریمی حدود ۹۳ درصد مشتریان بدحساب و ۶۳ درصد مشتریان خوش حساب را به درستی تشخیص داده است.

گروه اسمارت دریل^۱ مطالعه موردی روی ۸۵۰ مشتری یک بانک را برای تعیین رتبه اعتباری آن‌ها با استفاده از آنالیز رگرسیون لجستیک انجام داده است. جامعه آماری شامل ۷۱۷ مشتری‌ای است که پیشتر تسهیلات دریافت کرده‌اند. در کار گروه اسمارت دریل، یک نمونه تصادفی ۵۱۳ نفری از ۷۱۷ مشتری برای ساخت مدل مورد استفاده قرار گرفته است. ۲۰۴ مشتری به عنوان نگه‌داشته^۲ شده یا نمونه ارزیابی هستند که برای تست مدل ریسک اعتباری مورد استفاده قرار می‌گیرند. سپس مدل برای طبقه‌بندی ۱۳۳ مشتری آینده به عنوان ریسک اعتباری خوب و بد بکار برده می‌شود.

در مطالعه گروه اسمارت دریل، رگرسیون لجستیک باینری برای برآزش مورد استفاده قرار گرفته است. متغیرهای پیش‌بینی کننده^۳ (مستقل) شامل سن، میزان تحصیلات، سابقه در شغل کنونی (به سال)، سابقه در آدرس کنونی (به سال)، شغل همسر (میزان درآمد همسر)، میزان وام (بدهی) به درآمد^۴ و میزان دیون اعتباری است. در این مطالعه از روش برآزش خودکار گام‌به‌گام استفاده شده است و نتایج نشان داد که متغیرهای سابقه در آدرس کنونی، میزان وام (بدهی) به درآمد و میزان دیون اعتباری مهمترین پارامترهای تعیین اعتبار مشتری هستند.

تباگری^۵ در سال ۲۰۱۵ با استفاده از مدل برآزش لجستیک به اعتبارسنجی مشتریان بانک پرداخته است. در مطالعه ایشان، متغیر وابسته حالت باینری خوش حساب و بدحساب هستند. اطلاعات ۲۸۵ مشتری خوش حساب و ۲۱۴ مشتری بدحساب برای ساخت مدل مورد استفاده قرار گرفته است. در مطالعه تباگری گفته شده که حداقل ۳۰ رکورد برای ساخت مدل لازم است و برای هرمتغیر مستقل، حدود ۳۰ تا ۵۰ رکورد باید داشته باشیم. علاوه بر این، این روش باید حداقل متغیرها را داشت (متغیرهای اصلی). مقادیر پارامترها

-
- 1- SmartDrill
 - 2- Holdout
 - 3- Predictors
 - 4- Debt-to-income Ratio
 - 5- Tabagari

در این مدل مطلق نبوده و معمولاً به صورت نسبی سنجیده می‌شوند. در این کار ۱۶ متغیر مستقل در نظر گرفته شده است. متغیرهای مهم عبارتند از: سن، بازپرداخت بدهی به ماه، مدت زمان زندگی در محل کنونی، نوع شغل، مقدار اعتبار، بدهی‌های دیگر.

مهرآرا و همکاران در سال ۱۳۸۸ کار مدل‌سازی سنجش ریسک اعتباری و اعتبارسنجی مشتریان در بانک پارسیان به روش رگرسیون لاجیت و پروبیت و مدل شبکه‌های عصبی هوشمند *GMDH* را انجام داده‌اند. برای این منظور اطلاعات و داده‌های مالی و کیفی یک نمونه تصادفی ۴۰۰ تایی از مشتریان که تسهیلات دریافت کرده‌اند مورد بررسی قرار می‌گیرد. این حجم نمونه از مشتریان دارای حساب منتهی به سال ۱۳۸۸ انتخاب شده‌اند.

در مقاله مهرآرا و همکاران پس از بررسی پرونده‌های اعتباری هر یک از مشتریان، ۱۱ متغیر توضیح‌دهنده مورد ارزیابی قرار می‌گیرد. نتایج مقاله ضمن دلالت بر تایید نظریه‌های اقتصادی و مالی نشان می‌دهد که عملکرد پیش‌بینی الگوی شبکه عصبی (درصد پیش‌بینی‌های صحیح آن) به مراتب بهتر از الگوهای اقتصادسنجی متعارف لاجیت و پروبیت است و در زمینه عوامل موثر بر ریسک اعتباری نشان می‌دهد که از بین این متغیرها، نوع وثیقه و نسبت بدهی دارای بیشترین اثر بر متغیر احتمال نکول هستند. همچنین سابقه همکاری، نسبت جاری، نسبت آتی و نسبت مالکانه دارای اثر معمولی و سایر متغیرها کم اثر هستند. نوع وثیقه، نوع مالکیت محل فعالیت، سابقه همکاری مشتریان با بانک، سرمایه مشتری، نسبت جاری، نسبت آتی، دارایی جاری به دارایی کل، گردش دارایی کل، گردش سرمایه جاری، نسبت بدهی و نسبت مالکانه متغیرهای مدل‌سازی در این کار بوده‌اند.

در بیشتر موارد، نویسندگان مراجع مختلف بیان کرده‌اند که یک سری خاص از داده‌ها برای موسسات خاصی تهیه شده و در دسترس است، بنابراین متغیرهای بکار برده شده برای ساخت مدل رتبه‌بندی بستگی به داده‌های تهیه شده و در دسترس دارد (هند^۱ و همکارانش (۲۰۰۵، ۲۰۱۴)؛ هوانگ و ژن^۲ (۲۰۰۵) و ساسترسیک^۳ و همکارانش (۲۰۰۹)).

1- Hand

2- Huang and Tzeng

3- Sustersic

۳- مدل برازش لجستیک

اجرای مدل‌های رگرسیون و به طور ویژه رگرسیون خطی برای مفاهیم علمی رواج بسیاری دارد، اما باید توجه داشت که استفاده از رگرسیون خطی زمانی مجاز است که متغیر وابسته، کمی باشد. اغلب گفته می‌شود که متغیرهای مستقل نیز باید کمی باشند، اما ترکیبی از متغیرهای فاصله‌ای و ترتیبی نیز مجاز شمرده شده است. البته چنانچه برخی از متغیرها اسمی باشند باید به صورت متغیرهای مجازی دو وجهی با کدهای ۰ و ۱ آورده شوند. در این صورت مدل رگرسیون مورد استفاده قدری متفاوت بوده و با نام رگرسیون لجستیک شناخته می‌شود. در حالتی که متغیر وابسته بیش از دو طبقه یا رده داشته باشد، همانند اعتبار مشتری در این کار که شامل رده‌های خوش حساب، سررسید، معوق و مشکوک‌الوصول است، مدل برازش لجستیک چند گانه^۱ استفاده می‌شود.

در روش اشاره شده، مدل توسط رویه تکراری حداکثر درست‌نمایی ساخته می‌شود. در ابتدا مقادیر اختیاری برای ضرایب رگرسیون در نظر گرفته می‌شوند و مدل اولیه ساخته می‌شود. با پیش‌بینی داده‌های مشاهده شده بر اساس این مدل، خطاها بررسی شده و ضرایب رگرسیون طوری اصلاح می‌شوند که درست‌نمایی مدل جدید بهتر از مدل قبلی باشد. این رویه تا همگرایی مدل تکرار می‌شود، یعنی تا زمانی که درست‌نمایی دو مدل متوالی تفاوت چندانی نداشته باشد (مدیرپلاس، ۱۳۹۲).

مدل برازش لجستیک را می‌توان به عنوان مدل خطی تعمیم‌یافته‌ای که از تابع لجیت به عنوان تابع پیوند استفاده کرد. معادله کلی رگرسیون لجستیک به صورت معادله (۱) است (ویکی‌پدیا، ۲۰۱۵).

$$\text{Logit}(p) = \text{Log} \left[\frac{p}{1-p} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (1)$$

در رابطه (۱)، p احتمال وقوع رویداد، β عرض از مبدا و β_k ضریب مرتبط با متغیر مستقل (پیش‌بینی کننده) x است. متغیر وابسته، y_i لگاریتم نسبت احتمال وقوع رویداد به احتمال عدم وقوع آن است.

مدل بکار گرفته شده بر اساس توزیع دو جمله‌ای، یعنی احتمال است. احتمال رخداد داده نمونه به صورت معادله (۲) است.

$$L = \prod_{i=1}^n p^{y_i} (1 - p)^{1 - y_i} \quad (2)$$

لگاریتم گیری از طرفین و ساده‌سازی، لگاریتم احتمال رخداد را نتیجه می‌دهد (معادله (۳)) که در آن y_i مقادیر مشاهده شده و p_i مقادیر محاسباتی معادل آن‌ها است.

$$\log\text{-likelihood} = \sum_{i=1}^n [y_i \ln p_i + (1 - y_i) \ln(1 - p_i)] \quad (3)$$

هدف برازش لجستیک در اینجا ماکزیمم کردن مقدار احتمال رخداد (LL) با فرض تعریف p_i به صورت معادله (۴) است که این عمل منجر به محاسبه β_i ها خواهد شد. به عبارتی بهینه‌سازی تابع هدف منجر به نتایج بهینه β_i ها خواهد شد. مدل مورد استفاده در این کار مدل لجستیک چندگانه^۱ و باینری^۲ موجود در نرم‌افزار SPSS 22 است.

$$p = P(y = 1) = \frac{1}{1 + e^{-\beta_0 - \sum_{i=1}^k \beta_i x_i}} \quad (4)$$

۴- داده‌ها و اطلاعات مورد استفاده در مدل

جامعه آماری در این تحقیق، مشتریان حقیقی حدود ۱۲ شعبه بانک رفاه کارگران در استان تهران مشتمل بر پرونده‌های حدود ۳۶۰ مشتری حقیقی است. برای دستیابی به اطلاعات و داده‌های مورد نیاز در این پژوهش، پرسشنامه‌ای طراحی و تعریف شد که پس از بررسی و اظهار نظر توسط همکاران مدیریت ریسک و پژوهش بانک رفاه به شکل نهایی تدوین شد. در این مطالعه، متغیر وابسته اعتبار مشتریان حقیقی بود که دارای چهار رده مختلف شامل: خوش حساب، سررسید گذشته، معوق و مشکوک الوصول است. متغیرهای مستقل مهم مورد استفاده در این کار نیز طیف وسیعی از اطلاعات شخصی و مالی مشتری است که اکثر آن‌ها متغیرهای پیوسته و عددی بوده و شامل: جنسیت، سطح تحصیلات، سن، متوسط درآمد ماهیانه، وضعیت تاهل، شغل، وضعیت ملکیت منزل، شغل همسر، مبلغ تسهیلات (درخواست)، نرخ بهره، ارزش وثیقه و تعداد اقساط است.

1- Multinomial

2- Binary

قابلیت پایایی پرسشنامه با استفاده از ضریب آلفای کرونباخ سنجیده شد. با لحاظ کردن همه متغیرها مقدار این ضریب ۰/۶۳۱ و با حذف دو متغیر شغل همسر و وضعیت ملکیت، ۰/۶۹۹ به دست آمد (جدول (۱)).

جدول (۱) - پایایی نهایی پرسشنامه

آماره پایایی		
تعداد ایت‌ها	آلفای کرونباخ بر اساس استاندارد	آلفای کرونباخ
۱۰	۰/۶۲۵	۰/۶۹۸

خلاصه آماری متغیرهای مورد استفاده در این کار در جدول (۲) آورده شده است. داده‌ها در این مطالعه به دو دسته آموزش و آزمایش تقسیم شدند. حدود ۷۵ درصد داده‌ها برای ساخت مدل برازش (آموزش) و مابقی برای ارزیابی و تست مدل (آزمایش) مورد استفاده قرار گرفته‌اند.

جدول (۲) - خلاصه آماری متغیرهای مورد استفاده در مدل

Case Processing Summary						
Cases						
Total		Excluded		Included		
Percent	N	Percent	N	Percent	N	
۱۰۰/۰	۳۵۹	۰/۰	۰	۱۰۰/۰	۳۵۹	Y (اعتبار)
۱۰۰/۰	۳۵۹	۲/۸	۱۰	۹۷/۲	۳۴۹	X1 (جنسیت)
۱۰۰/۰	۳۵۹	۵/۸	۲۱	۹۴/۲	۳۳۸	X2 (مدری)
۱۰۰/۰	۳۵۹	۱/۷	۶	۹۸/۳	۳۵۳	X3 (سن)
۱۰۰/۰	۳۵۹	۷/۲	۲۶	۹۲/۸	۳۳۳	X4 (درآمد)
۱۰۰/۰	۳۵۹	۳/۹	۱۴	۹۶/۱	۳۴۵	X5 (تاهل)
۱۰۰/۰	۳۵۹	۳/۱	۱۱	۹۶/۹	۳۴۸	X6 (شغل)
۱۰۰/۰	۳۵۹	۸/۴	۳۰	۹۱/۶	۳۲۹	X7 (ملکیت)
۱۰۰/۰	۳۵۹	۱۵/۶	۵۶	۸۴/۴	۳۰۳	X8 (همسر شغل)
۱۰۰/۰	۳۵۹	۴/۲	۱۵	۹۵/۸	۳۴۴	X9 (تسهیلات میزان)
۱۰۰/۰	۳۵۹	۴/۷	۱۷	۹۵/۳	۳۴۲	X10 (بهره)
۱۰۰/۰	۳۵۹	۹/۵	۳۴	۹۰/۵	۳۲۵	X11 (وثیقه ارزش)
۱۰۰/۰	۳۵۹	۰/۰	۰	۱۰۰/۰	۳۵۹	X12 (اقساط)

۱-۴- برآزش مدل لجستیک چندگانه

با لحاظ اثرات متقابل متغیرهای مستقل در کنار اثرات اصلی این متغیرها، مدل برآزش به صورت گام به گام پیشرو اجرا شد. در این روش، در ابتدا هیچ متغیر مستقلی وارد مدل نمی شود، سپس متغیرهای مستقل به ترتیب اهمیت وارد مدل می شوند و بعد از ورود هر متغیر، مدل نهایی چک می شود. اگر متغیر وارد شده تاثیر زیادی در مدل نداشته باشد، حذف می شود و در غیر این صورت متغیر در مدل می ماند. این کار تا بررسی تاثیر تمام متغیرهای مستقل در مدل انجام می گیرد. مدل نهایی باقی مانده، مدل بهینه ای هم از نظر تعداد متغیر مستقل بهینه شده و هم از نظر کفایت مدل.

در مقابل روش گام به گام پیشرو، روش گام به گام پس رو قرار دارد که در آن ابتدا همه متغیرهای مستقل در مدل قرار گرفته، سپس یک به یک متغیرها کنترل شده و متغیری که تاثیر چندانی نداشته باشد از مدل خارج شده و در غیر این صورت در مدل باقی می ماند. در این مطالعه از روش گام به گام پیشرو استفاده شده است.

۱-۱-۴- نتایج مدل بهینه لجستیک چندگانه

نیکیویی برآزش در جدول (۳) ارائه شده است. با توجه به مقدار احتمال در هر دو آزمون که نزدیک به ۱ است، این نتایج نشان می دهد مدل نهایی به حد کافی به داده ها برآزش یافته است و داده ها با فرضیات مدل سازگار هستند.

جدول (۳) - جدول نیکیویی برآزش مدل بهینه لجستیک چندگانه

کای دو	درجه آزادی	سطح معناداری
۴۲۲/۸۱۵	۴۵۰	۰/۸۱۷
۲۳۸/۱۷۳	۴۵۰	۱/۰۰

در جدول (۴) نتایج آزمون نسبت درستنمایی مدل نهایی برآزش یافته در مقابل مدل صفر، ارائه شده است. مقدار آماره کای دو ۳۱/۲۰۱ با ۱۲ درجه آزادی است که با توجه به مقدار احتمال به دست آمده در سطح معنی داری ۰/۰۵ فرض صفر رد می شود و می توان گفت عملکرد مدل نهایی بهتر از مدل صفر است.

جدول (۴) - نتایج اطلاعات برازش مدل بهینه لجستیک چندگانه

مدل	معیار برازش مدل			آزمون‌های نسبت بخت	
	<i>-2 Log Likelihood</i>			کای دو	درجه آزادی
مدل صفر	۲۸۳/۸۱۲				
مدل نهایی	۲۵۲/۶۱۲			۳۱/۲۰۱	۱۲

جدول (۵) که آزمون نسبت درست‌نمایی است نیز سهم هر متغیر مستقل در مدل را بررسی می‌کند. پیشتر گفته شد فرض صفری که در این آزمون مورد بررسی قرار می‌گیرد، صفر بودن ضریب متغیر مستقل در مدل است. با توجه به مقدار احتمال‌های معنی‌داری ارائه شده در جدول که همگی کمتر از ۰/۰۵ است، می‌توان نتیجه گرفت که متغیرهای مستقل مدرک، سن و نوع مالکیت در سطح معنی‌داری ۰/۰۵ سهم مهمی در مدل نهایی دارند.

جدول (۵) - آزمون نسبت درست‌نمایی مدل بهینه لجستیک چندگانه

اثر	معیار برازش مدل			آزمون‌های نسبت بخت	
	<i>-2 Log Likelihood of Reduced Model</i>			کای دو	درجه آزادی
مدل صفر ^۱	۲۵۲/۶۱۲			۰/۰۰۰	۰
<i>Madrak</i> ^۲	۲۷۱/۲۱۳			۱۸/۶۰۱	۶
<i>AgeFG</i> ^۳	۲۶۶/۸۳۲			۱۴/۲۲۰	۶

ضریب‌های تقریبی R^2 کاکس و اسنل، ناگلرکز و مک فادنز برای مدل نهایی برازش یافته در جدول (۶) محاسبه شده‌اند. نتایج نشان می‌دهند که قدرت مدل در تبیین واریانس متغیر وابسته توسط متغیرهای مستقل متوسط است. البته با توجه به اینکه تمام متغیرهای مستقل باقی مانده در مدل رسته‌ای هستند، مقدار این ضرایب به طور نسبی خوب است.

۱- مقدار ثابت مدل - عرض از مبدا

۲- سطح تحصیلات مشتری

۳- سن مشتری

مقایسه مدل بهینه رگرسیون لجستیک چندگانه و باینری برای ... ۱۵۷

جدول (۶) - مقادیر R^2 مدل بهینه لجستیک چندگانه

۰/۱۵۷	<i>Cox and Snell</i>
۰/۱۹۵	<i>Nagelkerke</i>
۰/۱۰۴	<i>McFadden</i>

جدول (۷) رده‌بندی نتایج عملی مدل رگرسیون لجستیک چندگانه نهایی برازش یافته را نشان می‌دهد. براساس این جدول از ۱۳۶ مشتری خوش حساب، پیش‌بینی شده است که تعداد ۱۳۶ مشتری در دسته خوش حساب قرار دارند (۱۰۰ درصد این مشتریان درست رده‌بندی شده‌اند). از ۲۶ مشتری سررسید شده، ۱۴ مشتری معوق و ۷ مشتری مشکوک الوصول، همه خوش حساب تشخیص داده شده‌اند (صفر درصد این مشتریان درست تشخیص داده شده‌اند). به صورت کلی حدود ۷۵ درصد مشتریان را مدل به درستی رده‌بندی کرده است.

جدول (۷) - رده‌بندی نتایج عملی مدل بهینه برازش یافته

<i>Percent Correct</i>	پیش‌بینی شده				مشاهده شده
	<i>Mashkok</i>	<i>Moavagh</i>	<i>Sarresid</i>	<i>Khoshesab</i>	
٪۱۰۰	۰	۰	۰	۱۳۶	<i>Khoshesab</i>
۰٪	۰	۰	۰	۲۶	<i>Sarresid</i>
۰٪	۰	۰	۰	۱۴	<i>Moavagh</i>
۰٪	۰	۰	۰	۷	<i>Mashkok</i>
۷۴/۳٪	۰٪	۰٪	۰٪	۱۰۰٪	<i>Overall Percentage</i>

جدول برآورد پارامترهای مدل (جدول ۸)، تاثیر هر متغیر مستقل در مدل را نشان می‌دهد. مجذور نسبت ضریب هر متغیر (B) به خطای استاندارد آن ($Std. Error$) برابر آماره والد^۱ است. اگر سطح معنی‌داری آماره والد کوچک‌تر از ۰/۰۵ باشد، پارامتر از صفر فاصله دارد یا عبارتی پارامتر بر نتایج تاثیر گذار است. ضریب پارامتر مرتبط با رده

^۱ $Wald = (B/Std.Error)^2$ به عنوان مثال $6/0.85 = (7.063/0.85)^2$ مقدار آماره والد برای رده اول سطح تحصیلات (مدرک) خواهد بود.

آخر هر متغیر مستقل رسته‌ای (گسسته) به شرط ثابت بودن، مقداری اضافی است و در مدل وارد نمی‌شود (مثلاً برای رده سوم سطح تحصیلات یا مدرک مشتری، ضریب مقدار ثابت صفر می‌باشد و این رده در معادله پیش‌بینی وارد نمی‌شود).

جدول (۸) - برآورد پارامترهای مدل

	Y (عتبار) ^a	B	Std. Error	Wald	df	Sig.	Exp(B)
Sarresid	Intercept	-۲۰/۵۱۱	۱/۰۳۳	۴۰۱/۸۳۷	۱	۰/۰۰۰	
	[Madrak=1]	۲/۶۰۳	۱/۰۵۵	۶/۰۸۵	۱	۰/۰۱۴	۱۳/۵۰۱
	[Madrak=2]	۱/۹۸۲	۱/۰۸۸	۳/۳۱۹	۱	۰/۰۳۸	۷/۲۵۶
	[Madrak=3]
	[AgeFG=1]	۱۶/۲۹۲	۰/۴۹۱	۱۱۰۰/۴۴۹	۱	۰/۰۰۰	۱۱۸۹۸۸۹۵/۶۹۰
	[AgeFG=2]	۱۷/۱۹۰	۰/۰۰۰	.	۱	.	۲۹۲۱۷۰۰۳/۴۳۷
	[AgeFG=3]
Moavagh	Intercept	-۲/۲۴۳	۱/۱۴۰	۳/۸۷۱	۱	۰/۰۴۹	
	[Madrak=1]	۰/۹۳۵	۰/۸۴۶	۱/۲۲۳	۱	۰/۳۶۹	۲/۵۴۸
	[Madrak=2]	۰/۵۰۰	۰/۹۰۵	۰/۳۰۵	۱	۰/۵۸۱	۱/۶۴۹
	[Madrak=3]
	[AgeFG=1]	-۱/۷۵۰	۱/۰۸۴	۲/۶۰۷	۱	۰/۱۰۶	۰/۱۷۴
	[AgeFG=2]	-۰/۲۶۳	۰/۸۸۹	۰/۰۸۷	۱	۰/۸۶۸	۰/۸۶۹
	[AgeFG=3]
Mashkok	Intercept	-۳۶/۱۹۴	۲۴۵۰/۱۰۷	۰/۰۰۰	۱	۰/۹۸۸	
	[Madrak=1]	۱۷/۱۶۹	۲۴۵۰/۱۰۷	۰/۰۰۰	۱	۰/۹۹۴	۲۸۵۸۸۷۶۵/۴۸۱
	[Madrak=2]	۱۵/۵۱۱	۲۴۵۰/۱۰۷	۰/۰۰۰	۱	۰/۹۹۵	۵۴۴۷۷۵۱/۸۳۰
	[Madrak=3]
	[AgeFG=1]	۱۶/۳۷۸	۰/۸۷۲	۳۵۲/۵۴۳	۱	۰/۰۰۰	۱۲۹۷۴۲۷۴/۹۴۲
	[AgeFG=2]	۱۷/۲۶۱	۰/۰۰۰	.	۱	.	۳۱۳۷۱۵۶۷/۶۰۹
	[AgeFG=3]

با توجه به ضرایب متغیرها در این برازش، مدل نهایی مطابق معادلات (۱) تا (۴) به صورت پیش‌بینی می‌شود که در آن $X_۲$ سطح تحصیلات، $X_۳$ سن مشتری، $\pi_۱$ احتمال

خوش حسابی، π_2 احتمال سررسید شدن، π_3 معوق شدن و π_4 احتمال مشکوک الوصول شدن مشتری است.

$$\text{logit}(\pi_2) = \ln\left(\frac{\pi_2}{\pi_1}\right) = -20.511 + 2.603X_1' + 1.982X_2' + 16.292X_3' + 17.190X_4' \quad (1)$$

$$\text{logit}(\pi_3) = \ln\left(\frac{\pi_3}{\pi_1}\right) = -2.243 + 0.935X_1' + 0.5X_2' - 1.75X_3' - 0.263X_4' \quad (2)$$

$$\text{logit}(\pi_4) = \ln\left(\frac{\pi_4}{\pi_1}\right) = -36.194 + 17.169X_1' + 15.511X_2' + 16.378X_3' + 17.261X_4' \quad (3)$$

$$\pi_1 = P(Y = 1 | X_2, X_3, X_4) = 1 - \sum_{j=2}^4 \pi_j \quad (4)$$

۲-۱-۴- نتایج مدل بهینه لجستیک باینری

با توجه به اینکه تعداد مشاهدات رده‌های سوم و چهارم متغیر وابسته کم است، می‌توان این رده‌ها را در رده دوم ادغام کرده و به این ترتیب یک متغیر وابسته جدید با دو رده خوش حساب و بد حساب ایجاد می‌شود. این کار موجب بهبود برازش نخواهد شد، چراکه پیشتر محدوده رده‌ها (خوش حسابی، بد حسابی و ...) مشخص شده و نمی‌توان این محدوده‌ها را در این مرحله با هم ادغام کرد.

در این بخش مدل رگرسیون لجستیک را به داده‌های موجود با این متغیر وابسته جدید برازش می‌دهیم. در اینجا نیز نمونه ۷۵ درصدی از جامعه آماری به عنوان نمونه‌های ساخت مدل استفاده می‌شوند. مدل رگرسیون لجستیک بهینه را با استفاده از روش پیشرو گام به گام به دست می‌آوریم. این روش در دو گام انجام شده و در هر مرحله متغیری که بیشترین تاثیر را در مدل داشته، وارد مدل می‌شود و متغیری که وجودش در مدل معنی‌دار نباشد از مدل حذف می‌شود. مقدار آماره نیکویی برازش کای دو هوسمر-لمشو در این روش برابر ۱/۳۳۰ با ۵ درجه آزادی و مقدار معنی‌داری ۰/۸۸۵ است که نشان دهنده کفایت مدل برای توصیف داده‌های موجود است (جدول (۹)).

جدول (۹) - آماره هوسمر-لمشو برای مدل بهینه لجستیک باینری

آزمون Hosmer and Lemeshow			
مرحله	کای دو	درجه آزادی	سطح معناداری
۲	۱/۷۳۰	۵	۰/۸۸۵

خلاصه اطلاعات مدل برازش یافته در جدول (۱۰) ارائه شده است. مقدار آماره منفی دو برابر لگاریتم درستنمایی (۱۸۶/۸۵۷) میزان واریانس تبیین نشده متغیر وابسته را نشان می‌دهد و بالطبع هر چقدر مقدار آن کمتر باشد، بهتر است. در مقابل دو ضریب R^2 کاکس و اسنل (۰/۱۱۲) و R^2 ناگلکرکز (۰/۱۶۴) میزان واریانس تبیین شده متغیر وابسته توسط متغیرهای مستقل در مدل را نشان می‌دهد که هر چقدر به ۱ نزدیک‌تر باشد، بهتر است.

جدول (۱۰) - خلاصه مدل بهینه لجستیک باینری

خلاصه مدل			
مرحله	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
۱	۱۹۴/۹۴۰	۰/۰۷۲	۰/۱۰۵
۲	۱۸۶/۸۵۷	۰/۱۱۲	۰/۱۶۴

جدول (۱۱) رده‌بندی نتایج عملی مدل رگرسیون لجستیک برازش یافته را نشان می‌دهد. با توجه به این جدول، ۱۳۶ تا از ۱۳۶ مشتری خوش حساب (صددرصد) درست رده‌بندی شده‌اند. هیچ‌یک از ۴۶ مشتری بد حساب به درستی رده‌بندی نشده‌اند (تشخیص داده نشده‌اند). در حالت کلی ۷۴/۳ درصد از موارد به درستی رده‌بندی شده‌اند. با توجه به اینکه پیچیدگی این مدل نسبت به مدل اصلی (چندگانه) کمتر است و آماره نیکویی برازش نیز نشان دهنده توصیف بهتر داده‌ها توسط این مدل است در نتیجه استفاده از این مدل توصیه می‌شود.

همانطور که جدول (۱۱) نشان می‌دهد از ۱۳۶ مشتری خوش حساب، پیش‌بینی شده است که تعداد ۱۳۶ مشتری در دسته خوش حساب قرار دارند (صددرصد این مشتریان درست رده‌بندی شده‌اند). از ۴۷ مشتری بد حساب همه خوش حساب تشخیص داده شده‌اند (صفر درصد مشتریان بد حساب درست تشخیص داده شده‌اند). به صورت کلی حدود ۷۵

درصد مشتریان را مدل به درستی رده‌بندی کرده است. برآورد پارامترهای این مدل نیز در جدول (۱۲) ارائه شده است. مدل نهایی به صورت معادله (۵) خواهد بود.

$$\text{Logit}(\pi) = -3/324 + 2/0.55X_1' + 1/349X_2' + 0/0.95X_3' + 1/133X_4' \quad (5)$$

جدول (۱۱) - جدول رده‌بندی نتایج عملی مدل بهینه لجستیک باینری

Predicted ^c			Observed	Y (اعتبار)	مرحله
Selected Cases ^b					
Percentage Correct	Y (اعتبار)				
	Badhesab	Khoshesab			
۱۰۰/۰	۰	۱۳۶	Khoshesab	Y (اعتبار)	۱
۰/۰	۰	۴۷	Badhesab		
۷۴/۳			Overall Percentage		
۱۰۰/۰	۰	۱۳۶	Khoshesab	Y (اعتبار)	۲
۰/۰	۰	۴۷	Badhesab		
۷۴/۳			Overall Percentage		

جدول (۱۲) - پارامترهای وارد شده در مدل بهینه لجستیک باینری

Variables in the Equation							
	B	S.E.	Wald	df	Sig.	Exp(B)	
Step 1 ^a	Madrak		۱۰/۰۷۱	۲	۰/۰۰۷		
	Madrak(1)	۱/۹۴۸	۰/۶۴۰	۹/۲۵۷	۱	۰/۰۰۲	۷/۰۱۲
	Madrak(2)	۱/۳۶۶	۰/۶۷۹	۴/۰۵۵	۱	۰/۰۴۴	۳/۹۲۱
	Constant	-۲/۵۳۹	۰/۶۰۰	۱۷/۹۲۴	۱	۰/۰۰۰	۰/۰۷۹
Step 2 ^b	Madrak		۱۱/۱۶۵	۲	۰/۰۰۴		
	Madrak(1)	۲/۰۵۵	۰/۶۵۱	۹/۹۷۷	۱	۰/۰۰۲	۷/۸۰۸
	Madrak(2)	۱/۳۴۹	۰/۶۸۵	۳/۸۷۲	۱	۰/۰۴۹	۳/۸۵۳
	AgeFG			۷/۴۹۵	۲	۰/۰۲۴	
	AgeFG(1)	۰/۰۹۵	۰/۸۶۶	۰/۰۱۲	۱	۰/۹۱۲	۱/۱۰۰
	AgeFG(2)	۱/۱۳۳	۰/۸۳۳	۱/۸۵۲	۱	۰/۱۷۴	۳/۱۰۵
Constant	-۳/۳۲۴	۱/۰۲۱	۱۰/۶۰۳	۱	۰/۰۰۱	۰/۰۳۶	

a. Variable(s) entered on step 1: Madrak.

b. Variable(s) entered on step 2: AgeFG.

۵- جمع‌بندی و نتیجه‌گیری

آنچه مسلم است، پیچیدگی رفتار افراد ناشی از تاثیرپذیری آن از شرایط محیط موجب عدم دسترسی به مدل‌های پیشگوی مطلق برای تبیین رتبه اعتباری مشتریان است. این موضوع باعث شده که روش‌ها و تکنیک‌های مختلف مورد استفاده در بحث اعتبارسنجی مشتریان، نتوانند نتایج قاطع و خیلی خوبی به دست دهند، بنابراین نتایج این روش‌ها به طور نسبی سنجیده می‌شود.

هدف اصلی در این کار تهیه مدل بهینه و بررسی اثر تعداد رده‌های اعتبار مشتریان حقیقی برای پیش‌بینی وضعیت اعتبار مشتریان بالقوه بانک رفاه کارگران در دریافت تسهیلات بوده تا با این ابزار، ریسک اعطای تسهیلات مبنی بر عدم بازپرداخت مشتریان به حداقل برسد. اکثر مطالعات انجام شده در این زمینه به بررسی اعتبار مشتری در دو رده مشخص خوش حساب و بدحساب پرداخته‌اند. در این مطالعه سعی شد با افزایش رده اعتبار به چهار رده خوش حساب، سررسید شده، معوق و مشکوک‌الوصول اثر این رده‌بندی در صحت و دقت نتایج نهایی بررسی شود. در مدل برازش لجستیک چندگانه، تمامی چهار رده اعتبار مشتری حقیقی در نظر گرفته شده است و در مدل برازش لجستیک باینری رده‌های سررسید شده، معوق و مشکوک‌الوصول به رده بدحساب گروه‌بندی شده‌اند. با این گروه‌بندی باینری، مشتریانی از رده سررسید شده که به رده خوش حساب خیلی نزدیک بوده‌اند هم بدحساب در نظر گرفته خواهند شد. در این مطالعه با این فرض که می‌توان این گروه‌بندی را انجام داد برازش لجستیک باینری انجام شده است.

به نظر می‌رسد تفکیک مشتریان فقط به دو رده خوش حساب و بدحساب نتواند رویکرد مناسبی برای تعیین اعتبار مشتریان باشد، چراکه موارد متعددی وجود خواهد داشت که تصمیم‌گیری در مورد انتساب مشتریان فقط به این دو گروه مشکل خواهد بود. با گسترش رده اعتباری به چهار رده (در این کار) نتایج مدل‌سازی به مراتب بهتر و درصد پیش‌بینی بهتر خواهد بود. این امر منوط به داشتن مقادیر زیادی داده صحیح و دقیق از تمامی رده‌ها است (توزیع مناسبی بین تعداد موارد رده‌ها وجود داشته باشد). البته در این کار بخش اعظم مشتریان در رده خوش حساب قرار دارند و در سایر رده‌ها تعداد کمتری مشتری است. این عدم توازن می‌تواند نتایج مدل‌سازی را به میزان زیادی تحت تاثیر قرار دهد.

به طور خلاصه بر اساس نتایج مدل سازی در این کار، مهم ترین یافته های به دست آمده در این تحقیق عبارتند از:

۱- با توجه به ماهیت اعتبار مشتری که به راحتی قابل اندازه گیری کمی و عددی نیست، مدل های موجود تنها ابزاری برای کاهش ریسک بوده و به هیچ عنوان مدل های صرف پیشگو نخواهند بود. به عبارت دیگر، وابستگی این مدل ها به داده های واقعی صحیح، دقیق و زیاد اجتناب ناپذیر است. بنابراین دستیابی به چنین داده هایی شرط اساسی برای تهیه مدل ها خواهد بود.

۲- نتایج مربوط به درصد پیش بینی درست در طبقه بندی رده ها برای دو مدل برازش لجستیک باینری و چندگانه نشان می دهد که هر دو مدل، نتایج یکسانی از لحاظ درصد پیش بینی درست داشته اند (هر دو ۷۴/۳ درصد). به عبارت دیگر، عملکرد هر دو مدل برای پیش بینی درست رده های اعتباری یکسان است.

۳- بر اساس مدل سازی های انجام شده در این تحقیق، یکی از مهم ترین خروجی ها، شناسایی اولیه مهم ترین پارامترهای موثر بر تعیین اعتبار مشتریان حقیقی است. رتبه بندی و تعیین اهمیت نسبی این پارامترها از دیگر خروجی های مهم این تحقیق است. با توجه به اهمیت نسبی پارامترها در مدل لجستیک بهینه باینری و چندگانه می توان گفت که به طور نسبی، سطح تحصیلات و سن مشتری از مهم ترین پارامترهای موثر در تعیین اعتبار مشتری در این کار هستند.

منابع

الف - فارسی

البرزی، محمد، محمدابراهیم محمدپورزند و محمد خان بابایی (۱۳۸۹)، «بکارگیری الگوریتم ژنتیک در بهینه‌سازی درختان تصمیم‌گیری برای اعتبارسنجی مشتریان بانک‌ها»، نشریه مدیریت فناوری اطلاعات، ۲(۴).

محتشمی، عباس (۱۳۸۹)، «اعتبارسنجی مشتریان اعتباری در بانکداری نوین»:

<http://as-mohtashami.blogfa.com/post-49.aspx>

مدیرپلاس (۱۳۹۲)، «رگرسیون لجستیک»:

http://modirplus.com/side_info.php?id=88

مهرآرا، محسن، میثم موسایی، مهسا تصویری و آیت حسن‌زاده (۱۳۸۸)، «رتبه بندی اعتباری

مشتریان حقوقی بانک پارسیان»، فصلنامه مدل‌سازی اقتصادی، ۳(۱۰)، ۱۲۱-۱۵۰.

ویکی‌پدیا، (۲۰۱۵)، «رگرسیون لجستیک»:

Retrieved from <https://fa.wikipedia.org>

ب - انگلیسی

Abdou, H., A. El-Masry and J. Pointon (2007), "On the Applicability of Credit Scoring Models in Egyptian Banks", *Banks and Bank Systems*, 2(1), 4-20.

Hand, D., S. Sohn and Y. Kim (2005), "Optimal Bipartite Scorecards", *Expert Systems with Applications*, 29(3), 684-690. doi:10.1016/j.eswa.2005.04.032.

Hand, D. J. and N. M. Adams (2014), "Selection Bias in Credit Scorecard Evaluation", *Journal of the Operational Research Society*, 65(3), 408-415. doi:10.1057/jors.2013.55.

Karimi, A. (2014), "Credit Risk Modeling for Commercial Banks", *International Journal of Academic Research in Accounting, Finance and Management Sciences*, 4(3), 187-192.

Ong, C., J. Huang and G. Tzeng (2005), "Building Credit Scoring Models Using Genetic Programming", *Expert Systems with Applications*, 29(1), 41-47.

doi:10.1016/j.eswa.2005.01.003

SmartDrill, "Data Mining | Decision Support", Retrieved from <https://smartdrill.com/>

- Sustersic, M., D. Mramor and J. Zupan (2009), "Consumer Credit Scoring Models with Limited Data", *Expert Systems with Applications*, 36(3), 4736-4744. doi:10.1016/j.eswa.2008.06.016.
- Tabagari, S. (2015), "Credit Scoring by Logistic Regression", (MS), University of Tartu.
- Torabian, A. and K. Azizi (2013), "Credit Scoring of Real Customers: A Case Study in Saderat Bank of Iran", *European Online Journal of Natural and Social Sciences*, 2(3), 2725-2735. Retrieved from www.european-science.com.

