ISSN: 2588-2791
http://www.miej.ir

فصلنامه علمی پژوهش در مدیریت و مهندسی صنایع (سال پنجم)
شماره ۱۸ / تابستان ۱۳۹۸ / ص ۱۱– ۱

ارائه یک مدل پیش بینی پویا برای امتیاز اعتباری مشتریان بانک با استفاده از داده‌های تابلویی

یاسر صمیمی[۱]، سجاد ارتشیدار[۲]
[۱] دانشیار، خواجه نصیرالدین طوسی تهران.
[۲] دانشجوی کارشناسی ارشد، خواجه نصیرالدین طوسی تهران.

نام نویسنده مسئول:
سجاد ارتشیدار

چکیده

بانک‌ها نقش اصلی را در تأمین مالی بخش‌های مختلف اقتصادی بر عهده دارند و در راستای ایفای این نقش با ریسک های متفاوتی روبرو هستند که از جمله مهمترین انواع آن، ریسک اعتباری است. برای اعمال مدیریت و کنترل این نوع از ریسک لازم است مدل مناسب برای ارزیابی کمّی متقاضیان تسهیلات وجود داشته باشد. در سنجش ریسک اعتباری و پیش بینی احتمال نکول، مدل های مختلفی از قبیل رگرسیون لجستیک، شبکه های عصبی مصنوعی، رگرسیون تحلیل بقا و تحلیل ممیزی به کار گرفته شده است. نقطه مشترک تحقیقات گذشته، عمدتاً عدم توجه به رفتار پویای مشتری و تاکید بر استفاده از داده های مقطعی است. در این تحقیق، با در نظر گرفتن پویایی رفتار مشتری در طول زمان و نامشخص در نظر گرفتن متغیر حالت با استفاده ار یک مدل مارکف به مدلسازی رفتار متقاضی و پیش بینی ریسک اعتباری پرداخته شده است. به منظور ارزیابی عملکرد روش ارائه شده، مدل رگرسیون لجستیک و مدل رگرسیون تحلیل بقا نیز از طریق تولید داده شبیه سازی مورد بررسی واقع می شود. نتایج نشان می دهد در صورت بکارگیری اطلاعات رفتار مشتری در طول زمان توانایی پیش بینی مدل به میزان قابل توجهی بهبود می یابد.

واژگان کلیدی: ریسک اعتباری، داده های تابلویی، مدل مارکوف با حالات پنهان، مدل رگرسیون.

ISSN: 2588-2791
http://www.miej.ir

مقدمه

در سال های اخیر بانکها نقش اصلی را در تأمین مالی بخشهای مختلف اقتصادی بر عهده داشته اند و در راستای ایفای این نقش با ریسکهای متفاوتی روبرو هستند که یکی از عمده ترین آنان ریسک اعتباری است. ریسک اعتباری عبارت است از احتمال اینکه بعضی از دارایی های بانک، بویژه تسهیلات اعطایی از نظر ارزش کاهش یابد و یا بی ارزش شود(مدرس ،۱۳۸۶). با توجه به اینکه سرمایه بانکها نسبت به کل ارزش داراییهای آن ها کم است، حتی اگر درصد کمی از وامها قابل وصول نباشند، بانک با خطر ورشکستگی رو بهرو خواهد شد .برای اعمال مدیریت و کنترل این ریسک، بانکها می بایست آگاهی و شناخت کافی از متقاضیان اعتبار داشته باشند. یکی از مهم‌ترین ابزارها برای انجام این امر، برخورداری از سیستم اعتبار سنجی و امتیازدهی اعتباری مشتریان است. در این پژوهش با بررسی رفتار متقاضی در طول زمان و تعیین حالت متقاضی هنگام دریافت تسهیلات و بهره گیری از مدل زنجیره مارکف به پیش بینی احتمال انتقال از یک حالت به حالت دیگر و تعیین احتمال نکول متقاضی پرداخته می‌شود.

۱- پیشینه تحقیق

ریسک اعتباری عبارت است از احتمال اینکه بعضی از دارایی های بانک، بویژه تسهیلات اعطایی از نظر ارزش کاهش یابد و یا بی ارزش شود (مدرس، ۱۳۸۶). با توجه به اینکه سرمایه بانک نسبت به کل ارزش دارایی آن کم است، حتی اگر درصد کمی از تسهیلات ارائه شده قابل وصول نباشد، بانک‌ها با خطر ورشکستگی رو به رو خواهند شد. برای اعمال مدیریت و کنترل این ریسک، بانکها می بایست ابزار مناسبی برای ارزیابی کمّی متقاضیان تسهیلات در اختیار داشته باشند. یکی از مهم‌ترین ابزار برای نیل به این هدف، سیستم اعتبارسنجی متقاضیان تسهیلات مالی است که در آن از روش‌های امتیازدهی اعتباری مشتریان بهره گرفته می‌شود. طراحی مدلی برای اندازه‌گیری و درجه‌بندی ریسک اعتباری برای نخستین بار در سال ۱۹۰۹ به وسیله جان موری[1] بر روی اوراق قرضه انجام شد (کیس[2]، ۲۰۰۳). تحقیقات انجام شده در زمینه برآورد ریسک اعتباری در ابتدا بوسیله روش‌های تجزیه و تحلیل آماری مانند مدل Z آلتمن(۱۹۸۶) صورت می‌گرفت. با گذشت زمان و با پیشرفت‌های بدست‌آمده در زمینه شبکه‌های عصبی مصنوعی در دهه۱۹۸۰، تحقیقات انجام شده در این حوزه نیز به سوی استفاده از شبکه‌های عصبی برای برآورد ریسک اعتباری اوراق قرضه سوق یافت (المرو بروسکی[3]، ۱۹۸۸). بویژه زمانیکه رابطه بین متغیرهای وابسته و مستقل مشخص نباشد، شبکه‌های عصبی مصنوعی به عنوان یک روش مناسب در بین روشهای ارزیابی ریسک اعتباری شناخته می‌شود (یو[4] و همکاران، ۲۰۰۸). از پژوهش های اخیر در این زمینه می‌توان به مطالعه خاشمن[5] (۲۰۱۰) اشاره نمود. او به تشریح یک سیستم ارزیابی ریسک اعتباری با استفاده از مدل‌های شبکه عصبی مبتنی بر الگوریتم یادگیری پس انتشار خطا می‌پردازد. در تحقیق مذکور، شبکه عصبی مصنوعی برای تصمیم گیری در مورد اعطا یا عدم اعطای وام پیاده‌سازی و آموزش دیده شده است.

از روش های نوینی که اخیرا در ارزیابی ریسک اعتباری مورد توجه قرار گرفته است روش مدلسازی رگرسیون بقا است. در زمینه تحلیل بقا، نارین[6] (۱۹۹۲) اولین فردی بود که از این متد برای توسعه مدلسازی امتیازدهی اعتباری استفاده کرد. او در این مقاله، رگرسیون لجستیک مرسوم را از نظر قدرت پیش بینی با روش تحلیل بقا مقایسه کرد و دریافت که مدل نمایی رگرسیون تعداد شکست ها را در هر زمان به خوبی تخمین می‌زند و از طرف دیگر روند امتیاز دهی اگر تخمین اعتبار به وسیله تحلیل بقا انجام بگیرد نتایج بهتری خواهد داد.

تحقیقات متعددی همچون توماس و همکاران[7] (۱۹۹۹)، استفانوا و توماس[8] (۲۰۰۱)، بلوتی و کروک[9] (۲۰۰۷) وآرگان و همکاران[10] (۲۰۱۲) دریافتند که مدل تحلیل بقا روشی است که می‌تواند با رگرسیون لجستیک رقابت کرده و از آن بهتر عمل کند. ویژگی مشترک همه این مدل‌ها استفاده از رگرسیون‌های پارامتریک، ناپارامتریک و یا نیمه پارامتریک برای مدلسازی زمان تا شکست است. باناسیک و همکاران[11] (۱۹۹۹) دریافتند که مقدار وام، تغییر در تشکیلات تجاری و وضعیت استخدام تأثیر قابل توجهی در وقوع نکول دارد. باسین و

[1] John Mory
[2] Kiss
[3] Elmer,Borowski
[4] Yu Wang Li
[5] Khashman
[6] Narain
[7] Thomas et al.
[8] Stepanova M and Thomas LC
[9] Belloti and Crook
[10] Argan et al.
[11] Banasik et al.

ISSN: 2588-2791
http://www.miej.ir

همکاران۱۲ (۲۰۰۵) سال‌های خدمت، هدف اخذ وام و حق بیمه را در پیش بینی قصور پراهمیت دانستند. نتیجه بررسی استفانوا و توماس (۲۰۰۱) نشان داد متغیرهایی نظیر مقدار وام، درآمد خالص، وضعیت تاهل و برخی متغیرهای رفتاری مشتری تأثیر زیادی در قصور دارند. این دو در مقاله‌ای دیگر در سال ۲۰۰۲ نشان دادند که نرخ قصور به مدت زمان بازپرداخت وام بستگی ندارد. بلوتی و کروک (۲۰۰۷) متغیرهای اقتصاد کلان مثل نرخ بهره، درآمد، بیکاری و تولید را در کنار متغیرهای مربوط به مشخصات انفرادی نظیر سن، نوع اشتغال و وضعیت تملک محل سکونت مورد بررسی قرار دادند.

۲-روش شناسی پژوهش

مدلهای رگرسیون لجستیک و پروبیت از جمله اولین ابزار مورد استفاده برای پیش بینی ریسک اعتباری محسوب می شود. از آنجا که در مدل رگرسیون لجستیک مرسوم از داده های مقطعی استفاده می شود امکان تحلیل رفتار در طول زمان وجود ندارد. برای توسعه مدلهای خطی تعمیم یافته به نحوی که امکان در نظر گرفتن همبستگی توزیع مشاهدات در طول زمان وجود داشته باشد، از روش معادلات برآورد تعمیم یافته۱۳ (GEE) استفاده می شود (مایرز و همکاران،۱۹۸۸). بدین ترتیب، چنانچه داده های رفتاری مربوط به متقاضیان در طول زمان در اختیار باشد، امکان مدلسازی داده های پانل با استفاده از این روش وجود خواهد داشت. استفاده از داده های رفتار فرد در طول زمان به صورت داده پانل یکی از روش‌هایی است که به کمک آن می توان به یک پیش بینی پویا از ریسک اعتباری دست یافت.

متغیرهای کمکی وابسته به زمان نیز نقش مهمی در تحلیل رفتار به صورت پویا ایفا می‌کند. در این زمینه بویژه در مدل‌های تحلیل بقا امکان استفاده از متغیرهای کمکی وابسته به زمان وجود دارد. رگرسیون نیمه پارامتریک کاکس یکی از اصلی ترین روش های مدلسازی تحلیل بقا می باشد. این رگرسیون با مدلسازی زمان تا نکول، پیش بینی به مراتب بهتری نسبت به روش‌های مرسوم مانند رگرسیون لجستیک ارائه می‌دهد(توماس و همکاران،۱۹۹۹). در مدل نیمه پارامتریک کاکس می توان از متغیرهای وابسته به زمان۱۴ استفاده نمود. بدین ترتیب، با توجه به آنکه مقدار متغیرهای کمکی مدل در طول زمان ثابت نیست، می توان از این طریق ویژگی پویایی را در پیش بینی حاصل از مدل رگرسیون کاکس به کمک متغیرهای وابسته به زمان۱۵ لحاظ نمود (آرگان ،۲۰۱۲).

هدف اصلی در مقاله پیش رو، توسعه مدلی برای پیش بینی ریسک اعتباری با تاکید بر پویایی رفتار مشتری در طول زمان است. برای این منظور، ابتدا به بررسی مدل‌های پویای مرسوم در ارزیابی ریسک اعتباری می پردازیم. در ادامه با توسعه مدل مارکف با حالات پنهان، رفتار متقاضی را به کمک فرایند مارکف سه وضعیتی مدلسازی می کنیم. در انتها با استفاده از نمودار ROC به مقایسه توانایی مدل مارکف پنهان با سایر مدل‌های پویای معرفی شده می پردازیم.

۳- مدل سازی مارکوف با حالات پنهان

مدل مارکوف در سطح مشتری با در نظر گرفتن سه وضعیت غیر قابل مشاهده تحت عنوان وضعیت با ریسک پایین، وضعیت با ریسک بالا و وضعیت نکول تعریف می شود. متغیر حالت از زنجیره‌ی مارکف مرتبه‌ی اول پیروی می‌کند و ماتریس احتمال انتقال وضعیت بین حالات سیستم به صورت زیر تعریف می شود:

$$s = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ 0 & 0 & 1 \end{bmatrix}$$ رابطه (۱)

این ماتریس نشان می دهد اگر در دوره $t-1$ مشتری در وضعیت با ریسک پایین باشد با احتمال p_{11} در دوره t نیز در همین وضعیت باقی خواهد ماند و اگر در دوره $t-1$ در وضعیت با ریسک بالا باشد با احتمال p_{22} در دوره t نیز در وضعیت با ریسک بالا باقی خواهد ماند. وضعیت سوم اشاره به حالت نکول مشتری دارد و ماتریس انتقال وضعیت به ترتیبی تعریف شده است که در سطح سه دارای حالت جاذب است. به منظور آنکه مدل ارائه شده تا حد ممکن منطبق بر شرایط دنیای واقعی باشد، فرض می شود درایه های ماتریس انتقال نسبت به مشتریان دارای ناهمگونی است. در این راستا، درایه های ماتریس انتقال به شکل زیر برای هر مشتری به صورت جداگانه قابل محاسبه است.

---

[12] Baesens et al.
[13] GEE
[14] Time-varying covariate
[15] Cox Time-varying Covariate

ISSN: 2588-2791
http://www.miej.ir

$$p_{11} = \frac{1}{e^{(d_{011}+d_{111}*X_i)}}$$
$$p_{12} = \frac{1}{e^{(d_{012}+d_{112}*X_i)}}$$
$$p_{21} = \frac{1}{e^{(d_{021}+d_{121}*X_i)}}$$
$$p_{22} = \frac{1}{e^{(d_{022}+d_{122}*X_i)}}$$
$$p_{13} = 1 - p_{11} - p_{12}$$
$$p_{23} = 1 - p_{21} - p_{22}$$

رابطه (۲)

در روابط بالا، X_i متغیری است که مجموعه اطلاعات اولیه مشتری iام را به طور خلاصه نشان می دهد. رفتار مشتری به صورت زمان های بین هر دو پرداخت متوالی اقساط رصد می شود. بعبارت دیگر، آنچه به عنوان داده قابل مشاهده در فرایند تصادفی رفتار مشتری منظور می شود مقدار متغیر پیوسته زمان های بین پرداخت ها است که مطابق فرض به صورت زیر تعریف می شود:

$$T_{it} \sim Gamma(a; b_{it})$$
$$b_{it} = e^{\alpha_{is_t}+\beta_{is_t}X_i}$$
$$\alpha_{is_t} = \begin{cases} \alpha_{i1} & if \ \ s_t = 1 \\ \alpha_{i2} & if \ \ s_t = 2 \end{cases}$$
$$\beta_{is_t} = \begin{cases} \beta_{i1} & if \ \ s_t = 1 \\ \beta_{i2} & if \ \ s_t = 2 \end{cases}$$

رابطه (۳)

به طوری که، s_t متغیر مشاهده نشدنی حالت (وضعیت) می‌باشد ($s_t = 1$ نشان دهنده وضعیت با ریسک نکول کم و $s_t = 2$ نشان دهنده وضعیت با ریسک نکول بالا می باشد و$s_t = 3$ نشان دهنده حالت نکول می‌باشد). T_{it}زمان تراکنش مشتری iام در لحظه t را بیان می کند که مطابق فرض از توزیع گاما با پارامترهای a و b_{it} پیروی می کند. پارامتر دوم متغیر تصادفی گاما در قالب معادله رگرسیون (۳-۱) با متغیر توضیحی X_i مرتبط است. مقادیر α_{i1}، β_{i1}، α_{i2}، و β_{i2} بیانگر پارامترهای عرض از مبدا و شیب در رگرسیون گاما برای مشتری iام به ترتیب در دو حالت ریسک کم و ریسک زیاد می باشد. امید ریاضی زمان تا پرداخت بعدی برای مشتری iام به شکل زیر محاسبه می شود:

$$E(T_{it}) = ab_{it}$$

رابطه (۴)

در نظر بگیرید برای یک مشتری خاص، مقادیر y_t، $\boldsymbol{x}_t$ و $\boldsymbol{\phi}_{s_t}$ به صورت زیر تعریف می شود:

$$y_t = T_{it}$$
$$\boldsymbol{x}_t = (1; X_i)'$$
$$\boldsymbol{\phi}_{s_t} = (\alpha_{s_t}; \beta_{s_t})'$$

رابطه (۵)

به عبارت دیگر می توان نوشت: $E(y_t|\boldsymbol{x}_t) = e^{x_t'\boldsymbol{\phi}_{s_t}}$. حال در نظر بگیرید $\boldsymbol{X}_t$ و $\boldsymbol{Y}_t$ نشان دهنده ی کلیه اطلاعات قابل مشاهده تا زمان t هستند که به صورت $\boldsymbol{X}_t = (X_1; X_2; \dots; X_t)'$ و $\boldsymbol{Y}_t = (y_1; y_2; \dots; y_t)'$ تعریف می شوند. اگر فرض شود وضعیت سیستم مشخص است، تابع چگالی مشاهدات، به شرط معلوم بودن متغیر s_t، به صورت زیر محاسبه می شود.

$$\mathcal{L} = f(y_t|s_t; X_t; \theta) = \frac{{y_t}^{\alpha-1} e^{-\frac{y_t}{b_{it}}}}{{b_{it}}^{\alpha}\Gamma(\alpha)}$$

رابطه (۶)

می توان نشان داد با استفاده از اطلاعات تا دوره $t-1$، ($\boldsymbol{Y}_{t-1}$)، توزیع احتمال توام s_t و s_{t-1} به صورت زیر بدست می‌آید:

ISSN: 2588-2791
http://www.miej.ir

رابطه(۷)

$$P(s_t.s_{t-1}|\boldsymbol{Y}_{t-1};\boldsymbol{X}_t) = P(s_t|s_{t-1};\boldsymbol{Y}_{t-1};\boldsymbol{X}_t)P(s_{t-1}|\boldsymbol{Y}_{t-1};\boldsymbol{X}_{t-1}) \\ = P(s_t|s_{t-1})P(s_{t-1}|\boldsymbol{Y}_{t-1};\boldsymbol{X}_{t-1})$$

برای حصول رابطه (۷) از تئوری بیز[16] و اصل استقلال زنجیره‌های مارکف استفاده شده است. به دلیل اینکه احتمال انتقال، $P(s_t|s_{t-1})$ و احتمال فیلتر در دوره $t-1$، یعنی $P(s_{t-1}|\boldsymbol{Y}_{t-1};\boldsymbol{X}_{t-1})$، هر دو در دوره t مشخص هستند، می‌توان مقدار رابطه (۷) را بدست آورد. با جمع بستن بر روی s_{t-1} می توان توزیع شرطی حاشیه‌ای s_t را بدست آورد.

رابطه (۸)

$$P(s_t|\boldsymbol{Y}_{t-1};\boldsymbol{X}_t) = \sum_{s_{t-1}=1}^{2} p(s_t;s_{t-1}|\boldsymbol{Y}_{t-1};\boldsymbol{X}_t)$$

حال می توان توزیع توام y_t و s_t را به صورت زیر بدست آورد:

رابطه(۹)

$$f(y_t;s_t|\boldsymbol{Y}_{t-1};\boldsymbol{X}_t) = f(y_t|s_t;\boldsymbol{Y}_{t-1};\boldsymbol{X}_t)p(s_t|\boldsymbol{Y}_{t-1};\boldsymbol{X}_{t-1})$$

نتیجه فیلتر در مورد وضعیت احتمالی در دوره t به صورت زیر محاسبه می شود.

رابطه(۱۰)

$$P(s_t|\boldsymbol{Y}_t;\boldsymbol{X}_t) = \frac{f(y_t;s_t|\boldsymbol{Y}_{t-1};\boldsymbol{X}_t)}{f(y_t|\boldsymbol{Y}_{t-1};\boldsymbol{X}_t)} = \frac{f(y_t|s_t;\boldsymbol{Y}_{t-1};\boldsymbol{X}_t)p(s_t|\boldsymbol{Y}_{t-1};\boldsymbol{X}_{t-1})}{\sum_{s_t=1}^{2} f(y_t|s_t;\boldsymbol{Y}_{t-1};\boldsymbol{X}_t)\,p(s_t|\boldsymbol{Y}_{t-1};\boldsymbol{X}_{t-1})}$$

در این پژوهش فرض بر این است که هنگام ورود مشتری، فرد می‌تواند در یکی از دو حالت کم خطر یا پرخطر قرار داشته باشد. در ابتدا، احتمال قرار گرفتن در هر یک از دو وضعیت مذکور، برابر در نظر گرفته می‌شود. تابع درست نمایی مشاهدات مربوط به پرداخت های مشتری iام به صورت زیر محاسبه می‌شود:

رابطه(۱۱)

$$\mathcal{L} = log(p \times \frac{{\beta_1}^a \boldsymbol{Y}_T e^{-\boldsymbol{Y}_T\beta_1}}{\Gamma(a)} + (1-p)\frac{{\beta_2}^a \boldsymbol{Y}_T e^{-\boldsymbol{Y}_T\beta_2}}{\Gamma(a)})$$

$$\beta_1 = exp(g_{0l} + g_{1l} \times x(i))$$

$$\beta_2 = exp(g_{0h} + g_{1h} \times x(i))$$

در ادامه احتمال‌های P_1 ، P_2 و P_3 به صورت زیر محاسبه می‌شود:

$$P_1 = p \times P_{11} + (1-p) \times P_{21}$$

$$P_2 = p \times P_{12} + (1-p) \times P_{22}$$

$$P_3 = p \times P_{13} + (1-p) \times P_{23}$$

و برای پایان دوره یعنی از آخرین مشاهده تا پایان دوره که در اینجا یکساله در نظر گرفته شده است، تابع درستنمایی به صورت زیر محاسبه می‌شود:

16 Bayes

ISSN: 2588-2791
http://www.miej.ir

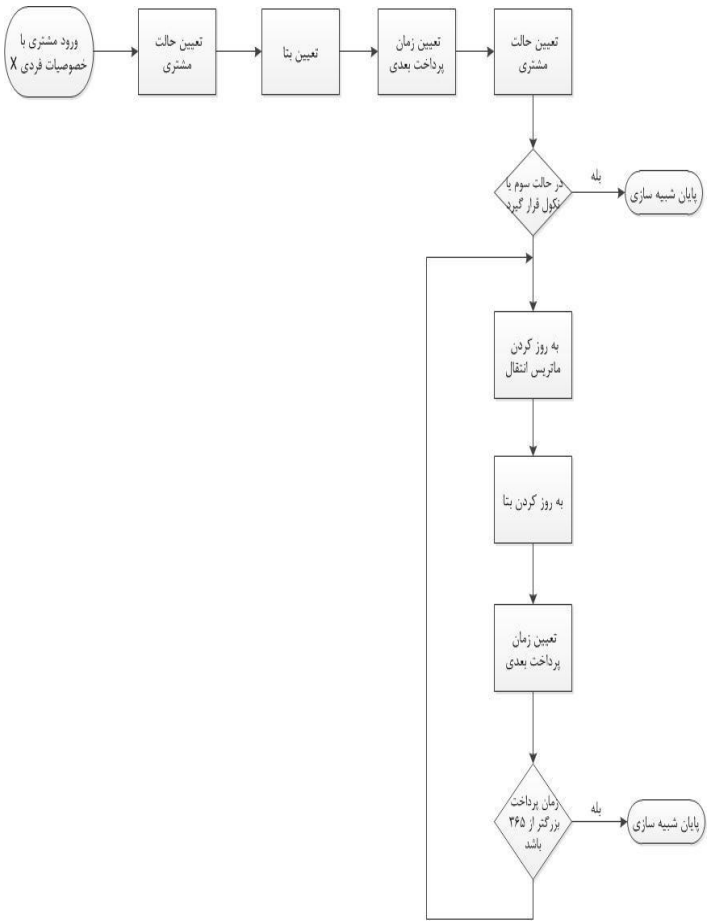
رابطه(۱۲)

$$\mathcal{L} = log[P_3 + P_1 \times (\sum_{i=0}^{a-1} \frac{e^{-(\beta_1(365-\boldsymbol{Y}_T))}(\beta_1(365-\boldsymbol{Y}_T))^i}{i!}) + P_2 \times (\sum_{i=0}^{a-1} \frac{e^{-(\beta_2(365-\boldsymbol{Y}_T))}(\beta_2(365-\boldsymbol{Y}_T))^i}{i!})]$$

تابع برآورد حداکثر درستنمایی از جمع دو رابطه ۱۱ و ۱۲ حاصل می شود. بنابراین با استفاده از آن می‌توان برای تمام پرداخت‌ها در طول سال تابع حداکثر درستنمایی را محاسبه کرد.

۴-شبیه سازی

در این بخش برای آنکه بتوان خروجی مدل ارائه شده را با سایر روشهای موجود مقایسه نمود، داده های تعدادی از متقاضیان تسهیلات مالی و رفتار آنها برای بازه زمانی یک ساله شبیه سازی می شود. در این شبیه سازی، وضعیت اولیه هر فرد در آغاز به صورت تصادفی ازیکی از دو حالت با ریسک پایین و یا با ریسک بالا انتخاب می شود. در دنیای واقعی فردی که تسهیلات دریافت می کند دارای خصوصیات و ویزگی های فردی می باشد که برای موسسه اعتباری دارای اهمیت است. خصوصیاتی از قبیل میزان درآمد، سطح تحصیلات، شغل، وضعیت تملک مسکن و .... می تواند به موسسات در تصمیم گیری در مورد ارزیابی اعتباری متقاضیان تسهیلات کمک شایانی کند. طبیعتا در شبیه سازی نیز خصوصیات اولیه شخص مد نظر قرار گرفت. برای این منظور، متغیر X_i بعنوان برایند خصوصیات متقاضی iام در نظر گرفته شده و داده آن با استفاده از توزیع نرمال تولید می شود. سه حالت برای دریافت کننده تسهیلات در نظر گرفته می شود. حالت یک یا با ریسک نکول کم، حالت دو یا با ریسک نکول بالا و حالت سه که نشان‌دهنده نکول است. شکل زیر روند شبیه‌سازی را به صورت خلاصه نشان می‌دهد:



شکل۱ :نمودار روند شبیه‌سازی

ISSN: 2588-2791
http://www.miej.ir

برای تحلیل داده شبیه سازی، دو موضوع مورد توجه قرار گرفته است: تعداد پرداخت ها طی یک ماه و زمان سپری شده از آخرین پرداخت. اگر تعداد پرداخت ها طی یک ماه مد نظر باشد، با اطلاع از اینکه فرد در کدام حالت قرار داشته باشد احتمال رخداد تعداد پرداخت با استفاده از توزیع پواسون تحلیل می شود. تعداد پرداخت هر فرد در هر ماه با np_{ij}، برای $i = 1. \dots . N$ و $j = 1. \dots . 12$ مشخص می شود. رابطه زیر احتمال رخداد np_{ij} پرداخت را برای یک دوره ۳۱ روزه محاسبه می‌کند.

رابطه(۱۳)

$$p\_np1_{ij} = P_1 \frac{e^{\left(\frac{-1}{exp(g_{0l}+g_{1l}\times x_i)31}\right)} \times \left(\frac{1}{exp(g_{0l}+g_{1l}\times x_i)\,31}\right)^{np_{ij}\times a}}{(np_{ij} * a)!}$$

رابطه(۱۴)

$$p\_np2_{ij} = P_2 \frac{e^{\left(\frac{-1}{exp(g_{0h}+g_{1h}\times x_i)31}\right)} \times \left(\frac{1}{exp(g_{0h}+g_{1h}\times x_i)\,31}\right)^{np_{ij}\times a}}{(np_{ij} \times a)!}$$

در رابطه بالا p_np1_{ij} احتمال وقوع تعداد پرداخت برای زمانی است که مشتری در حالت یک قرار داشته باشد. همچنین، p_np2_{ij} نیز احتمال رخداد تعداد پرداخت در یک ماه مربوط به زمانی است که مشتری در حالت دو باشد. بازه زمانی در نظر گرفته شده ماه های ۳۱روزه است. پس از به دست آوردن احتمال رخ دادن np_{ij} پرداخت، احتمال قرار گرفتن در حالت های مختلف در انتهای ماه به روز می شود. فرمول های زیر نشان دهنده احتمال های به روز شده در انتهای ماه می باشد:

رابطه (۱۵)

$$P_1 = \frac{p\_np1_{ij}}{p\_np1_{ij} + p\_np2_{ij}}$$
$$P_2 = \frac{p\_np2_{ij}}{p\_np1_{ij} + p\_np2_{ij}}$$
$$P_3 = 1 - P_1 - P_2$$

P_1احتمال قرار گرفتن در حالت اول و P_2 احتمال قرار گرفتن در حالت دوم است.

حالت دیگر محاسبه احتمال ها در انتهای ماه با در نظر گرفتن زمان تا آخرین پرداخت است. در این حالت با توجه به این که مشتری در چه وضعیتی قرار دارد و چنانچه تعداد پرداخت ها حداقل یکی باشد خواهیم داشت:

رابطه(۱۶)

$$p\_np1_{ij} = P_1 \times \prod_{i=1}^{np_{ij}} \frac{{\beta_1}^a \boldsymbol{Y}_i e^{-\boldsymbol{Y}_i\beta_1}}{\Gamma(a)} \times \sum_{i=0}^{a-1} \frac{\left(\beta_1(\boldsymbol{Y}_T")\right)^i}{i!}$$

رابطه(۱۷)

$$p\_np2_{ij} = P_2 \times \prod_{i=1}^{np_{ij}} \frac{{\beta_2}^a \boldsymbol{Y}_i e^{-\boldsymbol{Y}_i\beta_2}}{\Gamma(a)} \times \sum_{i=0}^{a-1} \frac{\left(\beta_2(\boldsymbol{Y}_T")\right)^i}{i!})$$

رابطه(۱۸)

$$P_1 = \frac{p\_np1_{ij}}{(p\_np1_{ij} + p\_np2_{ij})}$$
$$P_2 = 1 - P_1$$
$$P_3 = 0$$

ISSN: 2588-2791
http://www.miej.ir

اگر تعداد پرداخت برابر صفر باشد احتمال قرار گرفتن در حالت های مختلف در انتهای ماه به صورت زیر محاسبه می‌شود:

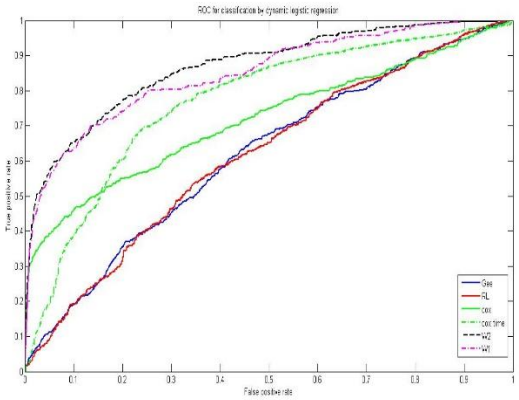
$$p\_np1_{ij} = P_1 \sum_{i=0}^{a-1} \frac{e^{\beta_1(\boldsymbol{Y}_T'')}\left(\beta_1(\boldsymbol{Y}_T'')\right)^i}{i!}$$ رابطه (۱۹)

$$p\_np2_{ij} = P_2 \sum_{i=0}^{a-1} \frac{e^{\beta_2(\boldsymbol{Y}_T'')}\left(\beta_2(\boldsymbol{Y}_T'')\right)^i}{i!}$$ رابطه(۲۰)

$$P_1 = \frac{p\_np1_{ij}}{(p\_np1_{ij} + p\_np2_{ij} + P_3)}$$
$$P_2 = \frac{p\_np2_{ij}}{(p\_np1_{ij} + p\_np2_{ij} + P_3)}$$
$$P_3 = 1 - P_1 - P_2$$ رابطه(۲۱)

۵- یافته های پژوهش

در این بخش نمودار *ROC* مربوط به مدل های رگرسیون لجستیک، معادلات برآورد تعمیم یافته، رگرسیون کاکس، رگرسیون کاکس با متغیرهای وابسته به زمان و دو مدل ارائه شده در این مقاله مبتنی بر فرایند مارکوف با حالات پنهان برای داده های شبیه سازی شده ترسیم می شود. شبیه سازی برای جامعه ای شامل ۱۰۰۰نفر متقاضی اجرا شده است و خروجی نمودار آن به صورت زیر می باشد:



شکل ۲ نمودار *ROC* مدل‌های مختلف

همانطور که در نمودار مشخص شده است، مدل های رگرسیون لجستیک و معادلات تعمیم یافته نسبت به دیگر مدل ها در پیش بینی احتمال نکول با دقت کمتری عمل نموده اند. خط آبی نشان دهنده رگرسیون حاصل از معادلات تعمیم یافته است و خط قرمز رنگ رگرسیون لجستیک را مشخص می کند.

مدل بعدی که مورد بررسی قرار گرفت رگرسیون نیمه پارامتریک کاکس است. رگرسیون کاکس با در نظر گرفتن زمان تا پرداخت، پیش بینی از احتمال نکول ارائه می دهد که به مراتب بهتر از دو مدل اولیه عمل می کند. با وابسته کردن متغیر کمکی به زمان در رگرسیون کاکس، نتایج پیش بینی نسبت به حالت اولیه کاکس بهتر است.

مدلسازی اولیه ای پژوهش براساس تعداد پرداخت در ماه بوده است که برتری آن نسیت به مدل های قبلی مشهود است. شبیه سازی بر اساس زمان های بین پرداخت نتایجی مشابه به تعداد پرداخت در ماه را ارائه می کند.

همچنین توانایی مدل در تمایز بین پیش بینی نکول را می توان با استفاده از سطح زیر منحنی *ROC* اندازه گیری کرد. سطح زیر منحنی *ROC* بدین‌گونه تفسیر می شود: احتمال اینکه یک مدل پیش بینی، احتمال نکول پیش بینی شده بیشتری نسبت به یک مدل دیگرداشته باشد. هرچقدر این مقدار بیشتر باشد، مدل در تمایز بین دو رخداد تواناتر است.

ISSN: 2588-2791
http://www.miej.ir

در جدول زیر مقادیر سطح زیر منحی *ROC* برای مدل های مختلف نمایش داده شده‌است. آزمون فرضی به صورت زیر بر روی *AUC* انجام گرفته است:

رابطه(۲۲)

$$\begin{cases} H_0 : AUC = 0.5 \\ H_1 : AUC \neq 0.5 \end{cases}$$

جدول۱ :جدول مقایسه سطح زیر نمودار *ROC*

	AUC	SE	Z	P	%95 CI	
					Loewr	Upper
LR	۰,۶۰۳	۰,۰۳۱	۳,۲۹۷	۰,۰۰۱	۰,۵۴۲	۰,۶۶۴
GEE	۰,۶۰۳	۰,۰۳۱	۳,۲۹۷	۰,۰۰۱	۰,۵۴۲	۰,۶۶۴
Cox	۰,۷۵۰	۰,۰۲۹	۸,۵۶۶	۰,۰۰۰۱	۰,۶۹۲	۰,۸۰۷
Cox TV	۰,۷۶۳	۰,۰۲۷	۸,۴۰۹	<۰,۰۰۰۱	۰,۶۷۱	۰,۸۱۷
Model1	۰,۸۴۱	۰,۰۲۶	۱۳,۲۷۸	<۰,۰۰۰۱	۰,۷۹۱	۰,۸۹۱
Model2	۰,۸۲۴	۰,۰۲۵	۱۲,۷۷۹	<۰,۰۰۰۱	۰,۷۷۵	۰,۸۷۴

با توجه به اینکه مقادیر *P-value* محاسبه شده برای مدل های مختلف کمتر از۰,۰۵ است ، بنابراین فرض صفر رد می‌شود و مدل ها توانایی پیش بینی را دارا می باشند. با توجه به اینکه نمودار *ROC* نسبت به همه حد آستانه ها نمودار را رسم می کند در یک حد آستانه خاص مقایسه نمودار ها شاید با حد آستانه دیگری متفاوت باشد. سطح زیر نمودار معیار مناسبی برای مقایسه قدرت پیش بینی است. با توجه به مقادیر بدست آمده و اینکه هر سطح ریز منحنی به یک نزدیکتر باشد شرایط بهتری در پیش بینی دارد، مشاهده می شود که مقایسه انجام گرفته در نمودار *ROC* به درستی تفاوت بین مدل ها را نمایش می دهد.

شاخص دیگری که می توان برای مقایسه مدل‌ها به کار برد، مجموع *sensitivity* و*specificity* است که به صورت زیر تعریف می شوند:

حساسیت۱۷: بیانگر مقادیر پیش بینی شده مثبت درست در مقابل همه خروجی های مثبت واقعی است.

$$sensitivity(\%) = \frac{TP}{TP + FN} \times 100$$

رابطه (۲۳)

اگر خروجی واقعی مثبت و مقدار پیش بینی نیز مثبت باشد این حالت را*TP* می نامیم.

اگر خروجی واقعی مثبت و مقدار پیش بینی منفی باشد این حالت را *FN* می نامیم.

اگر خروجی واقعی منفی و مقدار پیش بینی نیز منفی باشد این حالت را *TN* می نامیم.

اگر خروجی واقعی منفی و مقدار پیش بینی مثبت باشد این حالت را*FP* می نامیم.

اختصاصی بودن۱۸: بیانگر مقادیر پیش بینی شده منفی درست در مقابل همه خروجی های منفی واقعی است.

$$specificity(\%) = \frac{TN}{TN + FP} \times 100$$

رابطه (۲۴)

با توجه به اینکه شاخص حساسیت بیانگر پیش بینی از تمام خروجی های واقعی مثبت است و شاخص اختصاصی بیانگر پیش بینی از همه خروجی های منفی واقعی است، مجموع این دو شاخص توانایی مدل در پیش بینی از مقدار واقعی خروجی ها را نشان می دهد.در جدول زیر این شاخص برای مدل های مورد بحث آورده شده‌است:

[17] sensitivity
[18] specificity

ISSN: 2588-2791
http://www.miej.ir

جدول *Error! No text of specified style in document.* **-۲ مقایسه مجموع *sensitivity* و *specificity***

	sensitivity + specificity
LR	۱,۱۹۹
GEE	۱,۱۹۹
Cox	۱,۴۰۶
Cox TVC	۱,۳۹۷
Model1	۱,۵۹۲
Model2	۱,۵۶۳

نتیجه گیری

در این پژوهش ابتدا به معرفی مدل‌های مارکف با حالت پنهان پرداختیم. در روش مدل‌سازی ارائه شده در این پژوهش مشتری با خصوصیات فردی مشخص در نظر گرفته شده است. این مشتری می تواند در سه حالت با ریسک نکول پایین، ریسک نکول پایین و نکول قرار گیرد. این امر حائز اهمیت است که حالتی که مشتری در طول زمان بازپرداخت تسهیلات در آن قرار میگیرد مشخص نیست. بنابراین با به کار گیری مدل مارکف پنهان رفتار مشتری در طول زمان مدلسازی شده است. در این راستا در انتهای هر دورهٔ زمانی نیاز به تعیین حالت مشتری و به روز کردن پارامترهای تعیین کننده نظیر احتمال های ماتریس انتقال می‌باشد. برای این منظور دو رویکرد تعداد پرداخت در ماه و زمان تا آخرین پرداخت برای به روز کردن احتمال وضعیت مشتری در نظر گرفته شده است. برای مقایسه نتایج اجرای مدل با مدل‌های مرسوم نظیر رگرسیون لجستیک و رگرسیون کاکس، از داده های شبیه‌سازی استفاده شد. نتایج این مقایسه بر روی نمودار ROC بیان گردید که برتری محسوس دو مدل ارائه شده در این مقاله را نشان می‌دهد. بعنوان موضوعی در ادامه این مطالعه، توجه به این نکته لازم است که در بخش خصوصیات مشتری، لزوماً کلیه خصوصیات فردی قابل مشاهده نیست. افزودن یک متغیر تصادفی جدید با توزیع پیشین مشخص می‌تواند در واقعی تر شدن نتایج نقش مهمی را ایفا کند.

ISSN: 2588-2791
http://www.miej.ir

منابع و مراجع

[1] Lawless J. ,*Statistical Models and Methods for Lifetime Data* ,2nd Edition, A JOHN WILEY & SONS, INC., PUBLICATION.
[2] Mayers R.,Montgomery D.,Vining G.,Robison T.,*General linear Model With Applications in Engineering and the Sciences*,Second Edition, JOHN WILEY.
[3] Kimiagari A.,Haidari M.(2012). Calculating the best cut off point using logistic regression and neural network on credit scoring problem- A case study of a commercial bank. *African Journal of Business Management*, 7(16).
[4] Brooks C., (2014). *Introductory Econometrics for Finance*, 3rd Edition, Cambridge University Press, New York.
[5] Andreeva, G. 2006. European generic scoring models using survival analysis. *Journal othe Operational research Society*, 57, 1180-1187.
[6] Bellotti, T. & Crook, J. 2009. Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society*, 60, 1699-1707.
[7] Crook, J. N., Edelman, D. B. & Thomas, L. C. 2007. Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183, 1447-1465.
[8] Stepanova, M. & Thomas, L. 2001. PHAB scores: proportional hazards analysis behavioural scores. *Journal of the Operational Research Society*, 1007-1016.
[9] Stepanova, M. & Thomas, L. 2002. Survival analysis methods for personal loan data. *Operations Research*, 50, 277-289.
[10] Gujarati Damodar N., Porter Dawn C. (2009). Basic Econometrics, 5rd Edition, The McGraw−Hill Companies, New York.
[11] Hamilton James D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica, 57(2), 357-384.