

داده کاوی و پیش بینی ریسک اعتباری در بانک ها

مریم کمالیان برازجانی

کارشناس ارشد کامپیوتر نرم افزار

maryamkamalian@ymail.com

چکیده

ارزیابی اعتباری مشتریان می تواند توسط کارشناسان خبره و ارزیاب ها انجام پذیرد، لیکن این امر اغلب به علت کمبود وقت، هزینه بالا، کمبود تعداد افراد خبره و تعداد موارد ارزیابی، مقرون به صرفه نیست. با استفاده از فن آوری اطلاعات و ارتباطات که تحول عظیمی در سیستم بانکداری بوجود آورده ، می توان مدل های ارزیابی اعتباری را طراحی کرد که با استفاده از روش های علمی به جای قضاوت های ذهنی در زمان کم و با هزینه مناسب، حساب های خوب (مشتریان خوش حساب) و حساب های بد (مشتریان بد حساب) را از هم تفکیک کرد.

کلمات کلیدی: ریسک اعتباری، بانک، مدیریت، امتیازدهی

۱-مقدمه

همچنانکه دانش بیشتر و بیشتر مترادف ایجاد ثروت می شود. سازمان ها و افراد که به اطلاعات درست در لحظه درست دسترسی دارند، شانس بیشتری برای موفقیت در عصر جهانی شدن دارند.

در حال حاضر، مخازن بزرگ داده الکترونیک توسط بانکها و دیگر موسسات مالی سراسر دنیا نگه داشته می شوند. بخش های با ارزش اطلاعات در این مخازن داده تعبیه می شود [2].

کسب و کار اطلاعاتی¹ روی کشف دانش از مخازن داده الکترونیکی مختلف متمرکز است، تا تصمیم گیری های بهتر را حمایت کند. در سالهای اخیر ،

¹ - business intelligence

سیستم های کسب و کار اطلاعاتی نقشهای محوری در کمک کردن به سازمانها برای همخوان بودن با اهداف تجاری مثل نگهداری مشتری، نفوذ بازار، بهره وری و سودمندی، بازی کرده است. در اغلب موارد، این بینش ها توسط آنالیز داده های مبتنی بر تاریخ بدست می آید [1,2]. ریسک یا خطر امکان ایجاد یک ضرر مالی و یا غیرمالی به سبب محقق نشدن پیش بینی ها در مورد آینده گویند. ریسک اعتباری عبارت است از ریسکی که یک بنگاه یا نهاد مالی به علت اعطای تسهیلات به اشخاص حقیقی یا حقوقی ممکن است با آن روبه رو شود. ریسک را می توان در حوزه های مختلف مطرح نمود بایستی توجه داشت که در بنگاههای اقتصادی مانند بانک ها و بیمه ها ریسک ها بسیار پرهزینه هستند. از طرفی برای سازمان های مالی همچون بانک ها افزایش سودآوری نیازمند ریسک است. بر این اساس ارزیابی صحیح و مدیریت درست ریسک به صورتی که بتوان به صورت مناسبی در مقابل وقایع پیش بینی نشده عکس العمل نشان داد می تواند از الزامات اساسی این سازمان ها باشد [3]. چطور یک بانک ریسک اعتباری را انتخاب و مدیریت می کند بطور حیاتی برای عملکرد آن در طول زمان مهم است، در حقیقت کاهش سرمایه بواسطه تلفات وام، دلیل تقریبی اغلب شکست های شرکت ها هستند. شناسایی و رتبه بندی اعتبار ریسک اولین قدم حیاتی در مدیریت آن به طور موثر است. در اعطای تسهیلات که یکی از عمده ترین فعالیت های بانک ها و موسسات اعتباری است برای تصمیم گیری صحیح، باید درجه اعتبار و قدرت بازپرداخت اصل و سود تسهیلات دریافت کننده را تعیین نمود تا احتمال عدم برگشت اصل و سود تسهیلات اعطایی، یعنی ریسک درجه اعتبار، کاهش یابد. یکی از روش های کاهش این ریسک، طراحی نظام تعیین درجه اعتباری برای دریافت کنندگان تسهیلات است، و کانون این نظام، مدل رتبه بندی یا ارزیابی اعتباری است [3].

۲- داده کاوی و کاربرد آن در دسته بندی مشتری

داده کاوی عبارتست از استخراج الگوها و ساختارها از روی مجموعه ای داده ها. تکنیک های داده کاوی برای کشف دانش پنهان، طرح های ناشناخته و قوانین جدید از گروه داده های بزرگ، استفاده می شود، که ممکن است برای انواع فعالیت های تصمیم گیری مفید باشد. با افزایش اقتصاد، جهانی سازی و بهتر شدن در تکنولوژی اطلاعات، مقدار زیادی از داده های مالی در حال تولید شدن و ذخیره شدن هستند. اینها می توانند به تکنیک های داده کاوی برای کشف طرح های پنهان و بدست آوردن پیش بینی ها برای گرایشات در آینده و رفتار بازار مالی معطوف شود [4]. پیشرفت های ساخته شده در تکنولوژی برای خلق سریع تر و بهتر سیستم های پیش بینی توانمند شده اند. این سیستم ها بر پایه یک ترکیب تکنیک های کاوش داده و روش های هوشمند مصنوعی مثل استدلال مبتنی بر مورد^۲ (CBR) و شبکه های عصبی هستند. ترکیب چنین سیستم پیش بینی همراه با یک استراتژی تجاری خوب شانس های فوق العاده ای برای بازگشت های بزرگ پیشنهاد می کند [5].

شبکه های عصبی توسط قابلیت های یادگیری و توانایی بهتر کردن عملکرد در طول زمان مشخص می شود. همچنین شبکه های عصبی میتواند تعمیم داده شود یعنی تشخیص اشیا جدیدی که احتمالا شبیه باشند اما نه دقیقا همسان برای اشیا قبلی. شبکه های عصبی با قابلیت ان برای استخراج معنی از داده های نادرست/غیر دقیق را می توان مورد استفاده برای شناسایی الگوهای که در غیر این صورت بیش از حد پیچیده است که توسط انسان شناسایی شده است. شبکه های عصبی به صورت کارشناس خبره در ناحیه ای که آموزش دیده اند تا در آن کار کنند عمل میکنند. اینها میتوانند برای فراهم کردن پیش بینی ها برای موقعیت

^۲ - Constant-Based Resonance

سومین همایش ملی مهندسی کامپیوتر، داده کاوی و داده های حجیم

های جدید استفاده شود و در زمان واقعی کار کند. بنابراین، داده های تاریخی در دسترس در مورد بازارهای مالی و متغیرهای متنوع میتوانند برای آموزش شبکه های عصبی برای شبیه سازی بازار استفاده شوند. بر پایه ورود مقادیر متداول/رایج متغیرهای بازار، شبکه های عصبی میتواند وضعیت در روز آینده را پیش بینی کند و برای دادن توصیه خرید/فروش استفاده شود. روش CBR بر پایه استدلال از عملکرد گذشته است. ان از یک منبع داده ذخیره شده به عنوان مواردی که همه متغیرهای بازار در این مورد را در بر می گیرد استفاده می کند. موقعی که یک مورد جدید تغذیه می شود، الگوریتم CBR نتیجه این مورد را بر پایه مواردی که در مخزنش دارد پیش بینی می کند. تکنیک های کاوش داده می تواند برای اشکار سازی طرح های مخفی که در این موارد استفاده شود که احتمالاً بعداً برای تصمیم گیری های بعدی استفاده می شود. روش های CBR در زمان واقعی میتواند استفاده شود که تجزیه و تحلیل را حقیقتاً سریع میکند و در زمان واقعی تصمیم گیری که منجر به سود های سریع/فوری میشود کمک میکند [5].

استفاده از رویکردهای داده کاوی می تواند به بانکها در شناسایی مشتریان پرریسک بسیار کمک نماید. رویکردهای داده کاوی اغلب رویکردهای آماری سنتی که در بخش های مختلف تجارت بکار گرفته می شدند را بهبود داده اند. رویکردهای پردازش داده با بهبود مدل های مختلف یادگیری با پیدا کردن متغیرهای مهم و همچنین اضافه کردن متغیرهای جدیدتر و با شناسایی روابط مختلف و غیر خطی بین این متغیرها سعی در بهبود روشهای سنتی در زمینه های مختلف تجاری بخصوص صنعت بانکی دارد. در نتیجه با مدل های جدید تر روابط بین داده ها به صورت مناسبی کشف شده و در نتیجه هزینه های سازمان کاهش و سود سازمان افزایش می یابد. روشهای مختلف داده کاوی می تواند در موارد زیر به بانکها کمک شایانی را ارایه دهد.

- جذب مشتریان جدی
- حفظ مشتریان
- توسعه خطوط جدید تولیدی
- تشخیص ادعاهای جعلی
- هماهنگی گروه های آماری و بازاریابی [5] و

۳- تحقیقات پیشین

پیرا موتو از سیستم های عصبی فازی بهره می برد و از سه دسته از داده های واقعی ارزیابی ریسک اعتباری استفاده می کند. او بیان می کند که نتایج آموخته شده که از طریق بکار گیری سیستم های عصبی فازی بدست آمده، از طریق هر کاربری قابل درک است. در حالیکه به صورت قوانین اگر-آنگاه می باشد و هر تصمیمی که توسط این سیستم عصبی فازی گرفته می شود را می توان از طریق این قوانین تحلیل کرد [3].

وست در کار خود از شبکه های عصبی برای امتیاز دهی اعتباری استفاده می کند و عملکرد پنج شبکه MLP، MOE، RBF، LVQ و FAR را مورد بررسی قرار می دهد. او هم چنین نتایج شبکه های عصبی را با روش های آماری قدیمی هم چون تحلیل متمایز کننده خطی، رگرسیون لجستیک، نزدیک ترین همسایگی، تخمین تراکم مرکزی و درخت تصمیم مقایسه می کند. نتایج نشان داد با وجود اینکه MLP رایج ترین توپولوژی شبکه عصبی است که مورد استفاده قرار گرفته است، اما RBF و MOE بهترین عملکرد را داشته اند. MOE تا حدودی از RBF صحت بیشتری را داراست. اما بطور کلی نتایج بیانگر این امر بود که شبکه های عصبی می تواند تا حدودی صحت پیش بینی را بالا ببرد. در میان روش های قدیمی نیز روش رگرسیون لجستیک بهترین عملکرد را داشته است. بر این اساس رگرسیون لجستیک می تواند جایگزین خوبی برای شبکه های عصبی

باشد[4].

تیوالا از مدل های دسته بندی استفاده کرده است که استفاده از آن ها در چند سال اخیر در بسیاری از شاخه ها بسیار رایج شده است. او صحت پنج دسته بندی کننده که برای ریسک اعتباری بکار گرفته شده است را مورد بررسی قرار داده است و چهار مجموعه داده واقعی را برای این منظور بکار گرفته است. این بررسی در شرایطی انجام گرفت که داده هایی موجود نبودند یا به اصطلاح داده ها کامل نبودند. در این حالت نزدیک ترین همسایگی بدترین روش معین شد. بیزین ها موثرترین در مقایسه با رگرسیون لجستیک و درخت تصمیم بود. مشکل انتخاب توپولوژی منجر به عملکرد ضعیف شبکه عصبی شد. بطور کلی بهترین ها درخت تصمیم و بیزین ها بودند که بیشترین درجه صحت را داشتند[5].

در تحقیق دیگری نیز از شبکه عصبی نظارتی مبتنی بر الگوریتم یادگیری پس انتشار برای سیستم ارزیابی ریسک اعتباری استفاده شد. در اینجا سه شبکه عصبی به منظور تصمیم گیری در مورد اینکه یک درخواست اعتباری قبول یا رد شود، پیاده سازی و بکار گرفته شده است. هم چنین از یک مجموعه از داده های واقعی برای تست استفاده شده است. در شبکه عصبی بکار گرفته شده از ۲۴ خصیصه عددی استفاده شده است. هدف محقق معرفی مدل شبکه عصبی مناسب و طرح یادگیری مناسب برای سیستم ریسک اعتباری می باشد که بهینه ترین خروجی را حاصل می کند. خروجی و نتیجه ایی که از این امر حاصل شد بیانگر این موضوع است که در صنعت اعتباری، شبکه عصبی را می توان به عنوان ابزاری دقیق و صحیح برای تحلیلات اعتباری نسبت به باقی ابزار ها بکار برد. نرخ دقت ۹۹.۲۵٪ و ۷۳.۱۷٪ با استفاده از مجموعه داده آموزشی و تست به ترتیب به دست آمد. دقت کلی برابر با ۸۳.۶٪ بود و یادگیری این مدل عصبی تقریباً در ۱۸۴ ثانیه انجام شد. بنابراین پیشنهاد شد که این سیستم عصبی می

تواند بطور کارا در بررسی درخواست های اعتباری به صورت خودکار بکار آید[7].

۳.۱-مدل های ترکیبی

در دهه اخیر تعدادی از بانک های بزرگ دنیا سیستم های پیچیده ایی برای مدلسازی ریسک اعتباری طراحی و ایجاد کرده اند. خروجی این مدل ها نقش مهمی در مدیریت ریسک بانک و فرایند ارزیابی عملکرد ایفا خواهند کرد. بنابراین محققان حوزه مالی و کاربران در مورد صحت پیش بینی ها در زمان ساخت مدل بسیار حساس هستند رویکرد عام و رایج برای ارزیابی ریسک اعتباری بکارگیری برخی تکنیک های دسته بندی بروی داده های مشابه مشتریان قدیمی است، هم متعهد و هم خطا کار، تا رابطه بین خصوصیات و ریسک بالقوه به دست آید. یک جزء اصلی و اساسی که برای این مهم مورد نیاز است یک دسته بندی کننده صحیح و مناسب است تا متقاضیان را در دسته های خوب و بد دسته بندی کند. اگرچه تقریباً همه روش های دسته بندی را می توان برای ارزیابی ریسک اعتباری استفاده کرد، برخی دسته بندی کننده های ترکیبی که دو یا چند روش دسته بندی را ترکیب می کنند صحت بیشتری را در پیش بینی نسبت به روش های فردی نشان داده اند. تحقیق در حوزه دسته بندی کننده های ترکیبی در حال حاضر در حوزه ارزیابی ریسک اعتباری در حال رشد می باشد[10].

۳.۱.۱-مدل های ترکیبی و جمعی امتیازدهی اعتباری و ریسک

بانکی

لین یک رویکرد ترکیبی دو مرحله ایی را برای ریسک اعتباری در صنعت بانکداری ارائه می کند. در این تحقیق سه مدل دومرحله ایی ترکیبی از رگرسیون لجستیک و شبکه عصبی مصنوعی پیشنهاد شده است. منطق زیر بنایی تحلیلات او این است که یک رویکرد جدید را پیشنهاد کند که سه نوع مدل دو مرحله ایی رگرسیون لجستیک و شبکه های عصبی را بکار بگیرد تا بررسی کند

که آیا این مدل ها صحت بیشتری از مدل های قدیمی دارند یا خیر. در پیش بینی، مدل ترکیبی دو مرحله ای عملکرد بهتری داشت و قدرت پیش بینی بالاتری نسبت به رگرسیون لجستیک و رویکرد های شبکه های عصبی داشت [6].

لی و چن عملکرد امتیاز دهی اعتباری را از طریق استفاده از روش مدلسازی ترکیبی با شبکه عصبی مصنوعی و رگرسیون محدود چند متغیره انطباقی بررسی می کنند. مدل ترکیبی پیشنهادی آن ها از عملکرد تحلیلات متمایز کننده، رگرسیون لجستیک، شبکه های عصبی مصنوعی و رگرسیون محدود چند متغیره انطباقی بهتر بود [8].

تسای و چن از یادگیری ماشینی ترکیبی برای امتیاز دهی استفاده کردند. آن ها چهار روش ترکیبی را با داده های واقعی بانکی در تایوان بررسی کردند. در اینجا چهار روش دسته بندی با عنوان درخت تصمیم، شبکه های عصبی، مدل دسته بندی بیزین و رگرسیون لجستیک و دو متد خوشه بندی کامینز و بیشینه سازی انتظارات بکار گرفته شده است. نتایج نشان داده که در روش های تکی و فردی رگرسیون لجستیک در روش های دسته بندی نسبت به دیگر روش ها عملکرد بهتری داشت و در خوشه بندی روش بیشینه سازی انتظارات. نتایج آزمایش نشان داد که مدل ترکیبی که مبتنی بر رگرسیون لجستیک و شبکه عصبی است می تواند بیشترین صحت را به ارمغان بیاورد [9].

۴- نتیجه گیری و جمع بندی

در این مقاله سعی بر آن بود که با نگاهی گذرا مفاهیم و کاربرد های سیستم های چند عاملی، تکنیک های یادگیری ماشینی و سیستم های ترکیبی بطور مختصر ارائه شود. هم چنین مرور ادبیات تحقیقی موضوع بیانگر این امر بود که استفاده از تکنیک های یادگیری ماشینی و به خصوص شبکه های عصبی مصنوعی می

تواند نتایج خوب و موثری در این حوزه ارائه دهد و نسبت به روش های قدیمی رتبه بندی اعتباری و دیگر تکنیک ها، نتایج صحیح تر و با دقت بیشتری ارائه دهد. همانطور که در این مقاله مرور شد، مدل های مختلفی برای امتیازدهی اعتباری ارائه شده است. بسیاری از این مدل ها و رویکردهایی که توسط محققان بکار گرفته شد منتج به ارائه عملکردی برتر نسبت به مدل های قبلی شد.

۵- مراجع

- [1] Marshall, Andrew, Leilei Tang, and Alistair Milne. "Variable reduction, sample selection bias and bank retail credit scoring." *Journal of Empirical Finance* 17, no. 3 (2010): 501-512.
- [2] Hsieh, Nan-Chen. "An integrated data mining and behavioral scoring model for analyzing bank customers." *Expert systems with applications* 27, no. 4 (2004): 623-633.
- [3] Piramuthu, Selwyn. "Financial credit-risk evaluation with neural and neurofuzzy systems." *European Journal of Operational Research* 112, no. 2 (1999): 310-321.
- [4] Jackson, Patricia, and William Perraudin. "Regulatory implications of credit risk modelling." *Journal of Banking & Finance* 24, no. 1-2 (2000): 1-14.
- [5] Pandey, T. N., D. Choudhury, A. K. Jagdev, and S. Dehuri. "Comparison of classification techniques used for credit risk assessment in financial modeling." *International Journal of Management, IT and Engineering* 3, no. 5 (2013): 180.
- [6] Yeh, I-Cheng, and Che-hui Lien. "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients." *Expert Systems with Applications* 36, no. 2 (2009): 2473-2480.
- [7] García, Vicente, A. I. Marqués, and José Salvador Sánchez. "On the use of data filtering techniques for credit risk prediction with instance-based models." *Expert Systems with Applications* 39, no. 18 (2012): 13267-13276.
- [8] Lee, Tian-Shyug, and I-Fei Chen. "A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines." *Expert Systems with Applications* 28, no.

4 (2005): 743-752.

[9] Chen, Peng, Bing Wang, Hau-San Wong, and De-Shuang Huang. "Prediction of protein B-factors using multi-class bounded SVM." *Protein and peptide letters* 14, no. 2 (2007): 185-190.

[10] Tsai, Chih-Fong, and Ming-Lun Chen. "Credit rating by hybrid machine learning techniques." *Applied soft computing* 10, no. 2 (2010): 374-380.