Journal of Financial Management Strategy
Vol. 8, No. 29
Summer 2020

Alzahra University- Faculty of Social Sciences and Economics
Received: 2018/05/27
Accepted: 2019/05/26

Hybrid Model Binary ant Colony Algorithm and Support Vector Machine (BACO-SVM) for Feature Selection and Classification of Bank Customers with Case Study

Ali Husseinzadeh Kashan[1], Fatemeh Garousi[2]

Abstract

One of the most important issues faced by banks and financial institutions is the issue of credit risk. The significant amount of deferred bank claims around the world indicates the importance of this issue and the need to pay attention to it. So far, many efforts have been made to provide an effective model for evaluating and classification credit applicants as accurately as possible. In this regard, the present study attempts to provide a new approach for assessing the credit risk of bank customers. The support vector machine(SVM) method is combined as the main classifier of banking customers, with a feature selection method called the Binary Ant Colony Optimization Algorithm(BACO-SVM). In order to demonstrate the effectiveness of the proposed method, we used data from 85 companies from legal recipients of facilities of an Iranian bank in a 5 year interval (1393-1893) along with 16 characteristics related to each of them. The results of the BACO-SVM method have been compared with the PSO-SVM, GA-SVM, and SVM method. The results of the research indicated that BACO-SVM model has better performance in assessing credit risk rather than other methods. As the result, using the BACO-SVM method, we classify customers into two groups of good and bad account customers. Finally, in order to increase the flexibility in decision making, we will rank our good account customers with the VIKOR method. This rating will lead to a more accurate assessment of the credit risk situation of good account applicants.

Keywords: credit risk, credit rating, support vector machine, feature selection, Binary ACO, VIKOR
JEL: C53, G32, G21

---

1 . Associate Professor of Industrial Engineering, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran, Corresponding Author, Email:a.kashan@modares.ac.ir
2 . MSc Graduate of Financial Engineering, Tarbiat Modares University, Iran. Email:F_garosi@yahoo.com

راهبرد مدیریت مالی
سال هشتم، شماره بیست و نهم
تابستان ۱۳۹۹
صص ۹۲-۷۱

دانشگاه الزهرا (س)
دانشکده علوم اجتماعی و اقتصادی
تاریخ دریافت:۱۳۹۷/۰۳/۰۶
تاریخ تصویب: ۱۳۹۸/۰۳/۰۵

ارائه مدل ترکیبی الگوریتم مورچگان باینری و ماشین بردار پشتیبان (BACO-SVM) برای انتخاب ویژگی و طبقه‌بندی مشتریان بانکی به همراه مطالعه موردی[1]

علی حسین‌زاده کاشان[2]، فاطمه گروسی[3]

چکیده

یکی از مهم‌ترین مسائلی که همواره بانک‌ها و مؤسسات مالی با آن مواجه هستند، مسئله ریسک اعتباری می‌باشد. رقم قابل‌توجه مطالبات معوق بانک‌ها در سراسر جهان نشان‌دهنده اهمیت این موضوع و لزوم توجه به آن می‌باشد. ازاین‌رو تاکنون تلاش‌های بسیاری به‌منظور ارائه مدلی کارا جهت ارزیابی و طبقه‌بندی هر چه دقیق‌تر متقاضیان تسهیلات اعتباری صورت گرفته است. در این راستا، پژوهش حاضر سعی در ارائه رویکردی نو برای ارزیابی ریسک اعتباری مشتریان بانکی دارد. روش ماشین بردار پشتیبان (SVM) به‌عنوان طبقه‌بندی کننده‌ی اصلی با یک روش انتخاب ویژگی به نام الگوریتم مورچگان باینری (BACO-SVM) ترکیب می‌گردد. به‌منظور نشان دادن اثربخشی روش پیشنهادی از داده‌های مربوط به ۸۵ شرکت از تسهیلات گیرندگان حقوقی یک بانک ایرانی در یک بازه‌ی ۵ ساله (۱۳۸۹-۱۳۹۳) به همراه ۱۶ ویژگی مربوط به هر یک از آن‌ها استفاده نموده‌ایم. نتایج روش BACO-SVM با روش PSO-SVM، GA-SVM و روش SVM به‌تنهایی مقایسه گردیده است. یافته‌های پژوهش دلالت بر آن داشته که در ارزیابی ریسک اعتباری، مدل BACO-SVM نسبت به روش‌های دیگر از عملکرد خوبی برخوردار است. درنتیجه با استفاده از روش BACO-SVM به طبقه‌بندی مشتریان به دو گروه مشتریان خوش‌حساب و بدحساب می پردازیم؛ و درنهایت جهت افزایش انعطاف‌پذیری در تصمیم‌گیری، مشتریان خوش‌حساب را با استفاده از روش VIKOR رتبه‌بندی می‌کنیم. این رتبه‌بندی منجر به آن می شود که قضاوت دقیق تری درباره‌ی وضعیت ریسک اعتباری متقاضیان خوش‌حساب صورت گیرد.

واژه های کلیدی: ریسک اعتباری، رتبه‌بندی اعتباری، ماشین بردار پشتیبان، انتخاب ویژگی، الگوریتم بهینه‌سازی مورچگان باینری، VIKOR
طبقه‌بندی موضوعی: C53, G32, G21

---

۱. کد DOI مقاله: 10.22051/jfm.2020.20531.1678
۲. دانشیار دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران، نویسنده مسئول، Email: a.kashan@modares.ac.ir
۳. فارغ‌التحصیل رشته کارشناسی ارشد مهندسی مالی، دانشگاه تربیت مدرس، تهران، ایران Email:F_garosi@yahoo.com

مقدمه

هدف اصلی کلیه بانک‌ها جمع‌آوری پس‌اندازهای اشخاص حقیقی و حقوقی و اعطای تسهیلات به سازمان‌ها و اشخاص حقیقی می‌باشد. آنچه برای بانک‌ها اهمیت زیادی دارد این است که تسهیلات را به افراد یا سازمان‌هایی اعطاء کنند که بتوانند تسهیلات دریافتی را به موقع بازپرداخت نمایند. در این حالت بانک‌ها با ریسک روبه‌رو هستند که از مهم‌ترین ریسک‌هایی که بانک‌ها با آن مواجه می‌باشند ریسک اعتباری و نکول می‌باشد. کاهش و کنترل ریسک اعتباری به‌عنوان یکی از عوامل مؤثر در بهبود فرآیند اعطای اعتبار و عملکرد بانک‌ها مطرح گردیده و نقش اساسی در تداوم ارائه تسهیلات، سودآوری و بقای بانک‌ها و مؤسسات مالی ایفا می‌نمایند (فلاح‌پور و همکاران، ۱۳۹۳). ریسک اعتباری در سال‌های اخیر در سراسر دنیا بسیار موردتوجه قرارگرفته است. در کشور ما نیز با توجه به افزایش مطالبات معوق، مدیریت صحیح ریسک اعتباری یکی از مهم‌ترین چالش‌های پیش روی بانک‌هاست.

تاکنون پژوهش‌های گوناگونی (در پیشینه‌ی پژوهش مطرح‌شده‌اند) برای ارائه مدلی کارا، جهت کمک به اتخاذ تصمیمات صحیح اعطای اعتبار از مؤسسات مالی ارائه‌شده است که افزایش دقت پیش‌بینی احتمال نکول متقاضیان تسهیلات، از مهم‌ترین اهداف آن‌ها بوده است. در مدل‌های قدیم و جدید عموماً به مدل‌های رگرسیونی و هوش مصنوعی پرداخته‌شده است که هر یک دارای نقاط ضعف فراوانی است. همچنین در پژوهش‌های داخل کشور توجه بسیار کمی به مدل‌های مبتنی بر بهینه‌سازی شده است که ضرورت کار با چنین مدل‌هایی را ایجاد می‌کرد. چراکه مدل‌های بهینه‌سازی با پیدا کردن بهترین جواب همواره به دنبال کاهش خطاها هستند و به‌مانند عمده مدل‌های هوش مصنوعی به‌صورت جعبه سیاه نیستند.

در دهه اخیر روش یادگیری ماشین، به‌طور خودکار الگوهای پیچیده را در مرحله یادگیری تشخیص می دهد و تصمیم‌گیری هوشمندی بر اساس داده‌ها انجام می‌دهد از الگوریتم SVM، در هرجایی که نیاز به تشخیص الگو یا دسته‌بندی در کلاس‌های خاص باشد می‌توان استفاده کرد. آموزش نسبتاً ساده است و در بعضی از پژوهش‌ها کارایی آن نسبت به سایر طبقه‌بندی‌ها بیشتر است، ولی در مجموعه داده‌های با ابعاد بالا به دلیل پیچیدگی‌های محاسباتی کارایی آن پایین می‌آید (چن و لی، ۲۰۱۰)؛ بنابراین در این پژوهش برای بالا بردن دقت SVM از الگوریتم‌های فرا ابتکاری به‌منظور انتخاب زیرمجموعه‌ی بهینه ویژگی‌ها استفاده می‌شود.

در این پژوهش سعی گردیده است تا با به‌کارگیری رویکردهای ترکیبی روش ماشین بردار پشتیبان بر پایه الگوریتم مورچگان باینری (BACO-SVM)، روش ماشین بردار پشتیبان بر پایه الگوریتم ژنتیک (GA-SVM)، روش ماشین بردار پشتیبان بر پایه الگوریتم بهینه‌سازی ازدحام ذرات (PSO-SVM) و روش ماشین بردار پشتیبان (SVM) به‌تنهایی اقدام به طبقه‌بندی مشتریان بانکی به دودسته مشتریان خوش‌حساب و بدحساب نموده و دقت طبقه‌بندی بیان‌شده، استخراج گردد. به‌طور خلاصه هدف اصلی این مقاله مقایسه عملکرد روش‌های بیان‌شده و همچنین رتبه‌بندی مشتریان خوش‌حساب با استفاده از روش VIKOR می‌باشد که این رتبه‌بندی منجر به آن می‌شود که قضاوت دقیق‌تری درباره‌ی وضعیت ریسک اعتباری متقاضیان خوش‌حساب صورت گیرد.

داده‌های به‌کاررفته در این پژوهش اطلاعات مربوط به ۸۵ شرکت از تسهیلات گیرندگان حقوقی یک بانک ایرانی در یک بازه‌ی ۵ ساله (۱۳۸۹-۱۳۹۳) که ۶۰ نفر از آن‌ها مشتریان خوش‌حساب و ۲۵ نفر از آن ها مشتریان بدحساب محسوب می‌گردند. همچنین پس از استخراج اقلام ذکرشده از ترازنامه و صورت‌حساب سود و زیان، باید نسبت‌های مالی ورودی مدل را معرفی نمود. اهمیت بسیاری از نسبت‌های

مالی که در ادامه ذکرشده است، در پژوهش‌های نیلی و سبزواری (۱۳۸۷)، سپهردوست و برجیسیان (۱۳۹۲)، صفری، ابراهیمی و شیخ (۱۳۸۹) و... مورداستفاده قرارگرفته است و به آن‌ها توصیه‌شده است.

در این پژوهش، پس از مقدمه، توضیحاتی پیرامون مبانی نظری و پیشینه پژوهش‌ها صورت گرفته در زمینه روش‌های ارزیابی ریسک اعتباری مشتریان بانک‌ها ارائه می‌گردد. در بخش بعدی مطالبی در رابطه با روش‌های به‌کاررفته و متغیرهای پژوهش تقدیم گردیده و در پایان نیز بعد از بررسی نتایج و یافته‌های پژوهش، نتیجه‌گیری و پیشنهادهایی برای پژوهش‌های آتی بیان می‌گردد.

مبانی نظری و مروری بر پیشینه پژوهش

اندازه‌گیری و درجه‌بندی ریسک اعتباری برای نخستین بار در سال ۱۹۰۹ میلادی توسط جان موری بر روی اوراق قرضه انجام شد. اولین مدل‌های امتیازدهی اعتباری بر پایه روش‌های آماری بوده‌اند (موری[1]،۱۹۰۹). مطالعه فیشر[2] (۱۹۳۶) اولین سیستم ارزیابی اعتبار است. در سال ۱۹۴۱ دوراند[3] در مطالعه‌ای به تعیین ویژگی‌های مهم با دو روش پرسش از افراد خبره و روش آماری تحلیل تمایزی با تکیه‌بر نتایج فیشر پرداخت. او در مطالعه خود سیستم مبتنی بر امتیازدهی را برای طبقه‌بندی افراد متقاضی دریافت وام جهت خرید ماشین دست‌دوم، مورداستفاده قرار داد (دوراند،۱۹۴۱). ازجمله مطالعات دیگر می‌توان به مقاله بی‌ور[4] (۱۹۶۷) که در زمینه برآورد موفقیت و شکست شرکت‌ها با استفاده از برخی شاخص‌های مالی است، اشاره نمود. آلتمن[5] (۱۹۶۸) به‌عنوان یکی از پیشگامان بحث اعتبارسنجی، تلاش بسیاری برای یافتن یک رابطه معنی‌دار بین متغیرهای حسابداری یک شرکت و احتمال عدم توانایی در پرداخت دیون این شرکت در آینده انجام و رابطه‌ای معروف به نام Z-score را ارائه داد. این روش مبتنی بر تحلیل تمایزی خطی بین شرکت‌های خوب و بد بود.

یکی دیگر از روش‌های پرکاربرد در صنعت اعتباری، استفاده از ماشین بردار پشتیبان می‌باشد که هانگ[6] و همکاران (۲۰۰۴) از اولین پژوهشگرانی بودند که با این تکنیک که جزوه روش‌های یادگیری ماشینی است، اقدام به‌پیش‌بینی وضعیت اعتباری مشتریان بانک‌های ایالات‌متحده امریکا نموده و نتایج آن را با شبکه عصبی پیش‌خور مقایسه نمودند. طبق مطالعات آن‌ها، ماشین بردار پشتیبان همانند شبکه عصبی به‌دقت پیش‌بینی نزدیک به ۸۰ درصد دست یافت.

بلوتی و کروک[7] (۲۰۰۹) از روش ماشین بردار پشتیبان روی یک پایگاه داده مربوط به اطلاعات کارت‌های اعتباری ۲۵۰۰۰ مشتری استفاده کرده و عملکرد این روش را با روش‌های رگرسیون لجستیک، تحلیل تمایزی و نزدیک‌ترین همسایه مقایسه کردند. نتایج نشان می‌دهد اولاً روش ماشین بردار پشتیبان بالاترین نرخ طبقه‌بندی صحیح را ارائه می‌دهد، ثانیاً این تکنیک می‌تواند در انتخاب مهم‌ترین ویژگی‌های مؤثر بر ریسک نکول بسیار مفید واقع گردد.

---

1 . Jon Mory
2 . Fisher
3 . Durand
4 . Beaver
5 . Altman
6 . Huang
7 . Bellotti& Crook

آنجیللا و ماتزو[1] (۲۰۱۵) در پژوهش خود سعی در رتبه‌بندی اعتباری شرکت‌های کوچک و متوسط به کمک یکی از روش‌های تصمیم‌گیری با معیارهای چندگانه به نام ELECTRE-TRI نمودند. ازآنجایی‌که فرایند تخصیص اعتبار و تأمین مالی این شرکت‌ها، امری حیاتی است، آن‌ها در این مقاله سعی در ارائه یک مدل استوار[2] به‌منظور تخصیص شرکت‌های کوچک و متوسط به‌کلاس های ریسک نمودند.

لی[3] و همکاران (۲۰۱۶) به رتبه‌بندی اعتباری شرکت‌های با اندازه کوچک و متوسط[4] پرداختند. آن‌ها یک مدل ترکیبی از رگرسیون لجستیک و شبکه عصبی مصنوعی ارائه دادند؛ و از اطلاعات مالی شرکت‌های کوچک و متوسط کشور فنلاند در سال‌های ۲۰۰۴-۲۰۱۲ استفاده کردند؛ که این مجموعه داده شامل ۲۶۸۱ متقاضی وام که ۱۳۶۴ نفر در طبقه مشتریان بدحساب و ۱۳۱۷ در طبقه مشتریان بدحساب قرار گرفتند. نتایج اجرای مدل نشان داد که دقت مدل پیشنهادی بالاتر از زمانی است که از روش‌های رگرسیون لجستیک و شبکه عصبی مصنوعی به‌تنهایی استفاده شود، است. همین‌طور با افزایش دقت این مدل پیشنهادی به این نتیجه رسیدند. اول دقت بالای رتبه‌بندی باعث می‌شود پرتفوی وام به این شرکت‌ها عملکرد بهتری داشته باشد. دوم بانک‌ها سرمایه کمتری برای مقابله با ریسک اعتباری، ذخیره می‌کنند و می‌توان نرخ بهره کمتری از این شرکت‌ها گرفت.

هریس[5] (۲۰۱۵) در پژوهش خود تکنیک ماشین‌های بردار پشتیبان خوشه‌بندی‌شده (CSVM) را برای توسعه کارت‌های امتیازی اعتباری معرفی و استفاده نمود. او در مقاله خود به این موضوع اشاره داشته است که ایراد مدل‌های سابق مبتنی بر ماشین‌های بردار پشتیبان غیرخطی این بوده است که با افزایش حجم داده‌ها، هزینهٔ محاسبات را بالا می‌برد. بر طبق این گفته، مطالعهٔ وی روش CSVM را با سایر ماشین‌های بردار پشتیان غیرخطی مقایسه نموده است و نشان داده است که روش مذکور در طبقه‌بندی، درحالی‌که هزینهٔ محاسباتی کمتری دارد، عملکرد خوبی از خود نشان داده است.

اوزترک و همکاران[6] (۲۰۱۶) در مقاله خود با به کار گرفتن درخت‌های طبقه‌بندی و رگرسیونی، پرسپترون چندلایه‌ای، ماشین‌های بردار پشتیبان، شبکه‌های بیزی و طبقه‌بندی کننده نایو بیز، یک مدل پیش‌بینی رتبه اعتباری از چندین تکنیک هوش مصنوعی ارائه کرده‌اند که بر روی یک نمونه ناهمگن اجرا کرده‌اند. نتایج این پژوهش نشان داده است که طبقه‌بندی کننده‌های هوش مصنوعی، ازنظر دقت پیش‌بینی بهتر از مدل‌های سنتی آماری عمل کرده‌اند. بر طبق معیارهای در نظر گرفته‌شده برای دقت پیش‌بینی، طبقه‌بندی کننده‌های مبتنی بر هوش مصنوعی، از دقت ۹۰٪ فراتر رفته‌اند؛ درحالی‌که عملکرد روش‌های آماری سنتی در حدود ۷۰٪ بوده است.

افزایش درجه اهمیت موضوع ریسک اعتباری در ایران موجب شد که پژوهشگران ایرانی نیز به سمت اندازه‌گیری ریسک اعتباری در بانک‌ها و مؤسسات مالی داخلی بپردازند و مطالعاتی در این زمینه صورت گیرد.

---

1 . Angilella& Mazzu
2 . Robust
3 . Li& Niskanen& Kolehmainen& Niskanen
4 . Small and medium-sized firm (SME)
5 . Harris
6 . Özurk

عرب مازار و روئین‌تن (۱۳۸۵)، باهدف شناسایی عوامل مؤثر و تدوین مدلی برای سنجش ریسک اعتباری مشتریان حقوقی بانک کشاورزی ایران، اقدام به انجام مطالعه‌ای با استفاده از روش رگرسیون لاجیت کردند و برای این منظور، اطلاعات کیفی و مالی یک نمونه‌ی تصادفی دویست تایی از شرکت‌هایی که طی سال‌های ۱۳۷۸ تا ۱۳۸۳ از شعب بانک کشاورزی استان تهران تسهیلات اعتباری دریافت کردند را بررسی کردند. در این مطالعه ابتدا پس از بررسی پرونده‌های اعتباری هر یک از نمونه‌ها و با استفاده از تحلیل رگرسیون لاجیت، از بین ۳۶ متغیر توضیح‌دهنده کیفی و مالی، در نهایت هفده متغیر که اثر معنی‌داری بر ریسک اعتباری و تفکیک بین دو گروه از مشتریان خوش‌حساب و بدحساب را انتخاب و مدل نهایی توسط آن متغیرها، برازش شد.

پویان فر و همکاران (۱۳۹۲) در پژوهشی از روش حداقل مربعات ماشین بردار پشتیبان مبتنی بر الگوریتم ژنتیک با رویکرد پوشاننده جهت ارزیابی ریسک اعتباری مشتریان استفاده کرده‌اند. جهت سنجش کارایی و دقت مدل، مجموعه داده اعتباری آلمان از پایگاه داده UCI مورداستفاده قرارگرفته و مدل پیشنهادی در این پژوهش با مدل‌های لاجیت و روش حداقل مربعات ماشین بردار پشتیبان به‌تنهایی مقایسه شده است. نتایج حاکی از آن است که مدل پیشنهادی عملکرد بهتری داشته است.

فلاح‌پور و همکاران (۱۳۹۲) مدلی ارائه داده‌اند که در آن روش شبکه عصبی به‌عنوان طبقه‌بندی کننده اصلی مشتریان تسهیلات بانکی با الگوریتم کلونی زنبورعسل جهت انتخاب ویژگی با رویکرد پوشاننده ترکیب می‌گردد. روش‌های دیگر به‌کاررفته در این پژوهش شبکه عصبی مبتنی بر آنالیز اجزای اساسی به‌عنوان یک روش فیلتر انتخاب ویژگی و روش شبکه عصبی به‌تنهایی می‌باشند. نتایج اجرای روش‌ها بر روی مجموعه داده‌ای از ۲۵۰ نفر از مشتریان بانک کارآفرین در یک بازه زمانی سه‌ساله (۱۳۹۰-۱۳۸۷) دلالت بر بهتر بودن روش شبکه عصبی مبتنی بر کلونی زنبورعسل نسبت به دو روش دیگر داشته است. همچنین استفاده از کلونی زنبورعسل به‌عنوان یک روش انتخاب ویژگی پوشش‌دهنده نقش مهمی در افزایش دقت و سرعت در اجرای تکنیک پیش‌بینی، حذف داده‌های غیر مرتبط و افزایش قابلیت فهم‌پذیری ایفا نمود و همچنین در مقایسه با عملکرد روش انتخاب ویژگی فیلتر کننده آنالیز اجزای اساسی، کلونی زنبورعسل به‌عنوان یک روش پوشش‌دهنده در ترکیب با شبکه عصبی دقت پیش‌بینی بالاتری از خود نشان داد.

فلاح‌پور و همکاران (۱۳۹۶) در سال ۱۳۹۶ در مقاله‌ای دیگر از ماشین بردار پشتیبان با چهار تابع کرنل خطی، چندجمله‌ای، شعاعی و سیگموید به‌عنوان مدل طبقه‌بندی کننده و ترکیب آن با روش‌های انتخاب ویژگی فیلتر کننده و پوشش‌دهنده استفاده کردند. همچنین از الگوریتم ژنتیک که یکی از انواع روش‌های پوشش دهندهٔ انتخاب ویژگی است و روش‌های آنالیز اجزای اساسی، زنجیرهٔ اطلاعات و رلیف که جزء روش‌های فیلترکنندهٔ انتخاب ویژگی هستند، استفاده‌شده است. نتایج به‌دست‌آمده نشان داد که روش الگوریتم ژنتیک نسبت به روش‌های فیلتر کننده، عملکرد بهتری دارد. همچنین دقت ماشین بردار پشتیبان با توابع کرنل خطی، چندجمله‌ای، شعاعی و سیگموید در ترکیب با الگوریتم ژنتیک، با سطح اطمینان ۹۵ درصد تفاوت معناداری با هم ندارند.

روش‌شناسی و مدل‌های پژوهش

رویکرد انتخاب ویژگی

مسئله انتخاب ویژگی که به تعیین ویژگی‌های مناسب جهت بهبود عملکرد مدل می‌پردازد، یک مسئله بهینه‌سازی به‌صورت جستجو در فضای ویژگی‌ها جهت تعیین زیرمجموعه بهینه یا

نزدیک به بهینه بر اساس معیار عملکرد است این مسئله جزو مسائل NP-hard محسوب می‌شود و الگوریتم‌های متعددی جهت حل آن در زمان کوتاه‌تر ارائه‌شده است (مارینا‌کیس و همکاران،۲۰۰۹). روش‌های انتخاب ویژگی بر اساس استفاده یا عدم استفاده از یک تابع طبقه‌بندی، به دودسته روش‌های فیلتر و پوشاننده تقسیم می‌شوند.

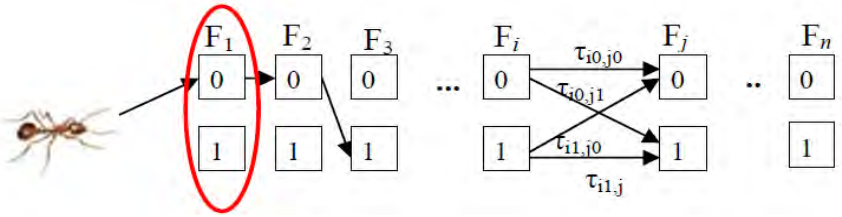
رویکرد فیلتر ویژگی‌ها را بدون استفاده از یک الگوریتم یادگیری ارزیابی می‌کند. در این رویکرد، خصوصیات ذاتی داده‌ها جهت ارزیابی و رتبه‌بندی ویژگی‌ها مورداستفاده قرار می‌گیرند (طباخی و همکاران،۲۰۱۴).

رویکرد پوشاننده از یک الگوریتم یادگیری جهت ارزیابی هر زیرمجموعه از ویژگی‌ها استفاده می‌کند. در این رویکرد لازم است یک روش جستجو جهت یافتن زیرمجموعه بهینه به کار گرفته شود. رویکرد پوشاننده طبقه‌بندی را یک جعبه سیاه در نظر گرفته و زیرمجموعه ویژگی‌ها را بر اساس توان پیش‌بینی آن‌ها رتبه‌بندی می‌کند (حاتمی‌خواه ۱۳۹۲).

در سال‌های اخیر، الگوریتم‌های فرا ابتکاری مانند الگوریتم ژنتیک، بهینه‌سازی کلونی مورچگان و بهینه‌سازی ازدحام ذرات به شکل گسترده‌ای در زمینه انتخاب ویژگی مورداستفاده قرارگرفته‌اند. بیشتر الگوریتم‌های فرا ابتکاری انتخاب ویژگی، به علت برخی خصوصیات ذاتی خود از رویکرد پوشاننده استفاده می‌کنند (بوز[1]،۲۰۰۲). به‌کارگیری رویکرد پوشاننده در مقایسه با رویکرد فیلتر نیازمند صرف زمان و هزینه بیشتری است اما دقت و کارایی این روش به‌مراتب بالاتر از روش فیلتر است. ازاین‌رو، در این پژوهش، انتخاب ویژگی بر اساس رویکرد پوشاننده و با استفاده از الگوریتم بهینه‌سازی کلونی مورچگان، الگوریتم جمعیت ذرات و الگوریتم ژنتیک، انجام می‌گیرد.

الگوریتم جمعیت مورچه باینری در انتخاب ویژگی

هیرویاسو[2] (۲۰۰۰) روشی برای حل مسائل فضای باینری با نام TACO ارائه و بعدها از آن برای حل مسائل مختلف بهینه‌سازی استفاده شد (اوزترک[3] و همکاران،۲۰۰۹). شکل (۱) چگونگی روش مذکور را به تصویر کشیده و همان‌طور که مشهود است، برای هر ویژگی یک گره صفر در رشته بالایی و یک گره یک در رشته پایینی در نظر گرفته‌شده است. هر مورچه در هر حرکت دو انتخاب (صفر یا یک) دارد که با توجه به تابع احتمالی که برای حرکت تعریف‌شده است، یکی را به‌عنوان موقعیت بعدی خود انتخاب می‌کند.



شکل ۱. رشته باینری به‌وسیله‌ی مسیری که مورچه از آن گذر کرده است

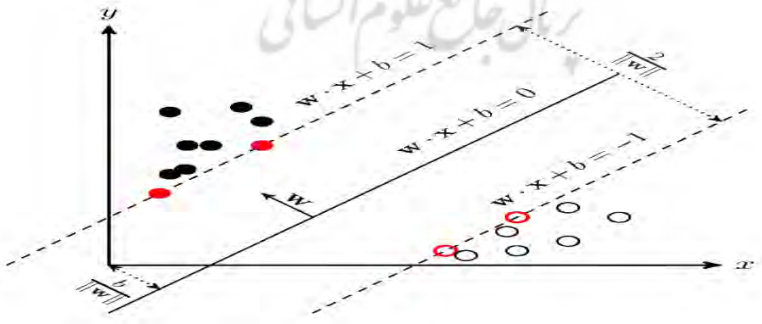
1 . Boz
2 . Hiroyasu
3 . Ozturk

هر مورچه بعد از اتمام سفرش، در مسیرهای عبوری خود مقداری فرمون به‌جا می‌گذارد. میزان فرومون به‌جامانده از هر مورچه با توجه به هزینه‌ای که مسیر طی شده به همراه داشته است، محاسبه می‌گردد. در قوانین حرکت نسخه باینری، مورچه‌ها فقط از اطلاعات فرومون مسیرها استفاده می‌کنند و وقتی یک مورچه مرحله تصمیم‌گیری را برای همه گره‌های رشته کامل کند به این معنی است که یک‌راه حل برای مسئله تولیدشده است.

در روش TACO، تمام مورچه‌ها با توجه به شایستگی‌شان بر مسیرهایی که از آن‌ها گذشته‌اند، فرمون ترشح می‌کنند. با اقتباس از روش‌های موجود الگوریتم فضای گسسته ACO، سه روش به نام‌های ETACO[1]، RTACO[2] و BACO[3] برای بهبود TACO در (توحیدی و همکاران، ۱۳۸۶) پیشنهاد شد. این سه روش در دو مسئله طبقه‌بندی معنایی تصویر و بازشناسی ارقام دست‌نویس فارسی آزموده شدند. نتایج آزمایش‌ها نشان دادند روش BACO عملکرد بهتری نسبت به سایر روش‌های پیشنهادی و روش‌های مورد مقایسه دارد.

ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان درواقع یک طبقه‌بندی کننده دودویی است که دو کلاس را با استفاده از یک مرز خطی از هم جدا می‌کند. در تقسیم خطی داده‌ها هدف دست‌یابی به تابعی است که تعیین‌کننده ابر صفحه‌ای[4] با بیشترین حاشیه می‌باشد. با حداکثر شدن حاشیه این ابر صفحه، تفکیک بین طبقات حداکثر می‌گردد. فرض کنید که $S = \{x_i . y_i\}$ یک ونه‌ی آموزشی است که از دو کلاس $y_i = \pm 1$ و هر کلاس از $i = 1. \ldots . m$ ویژگی تشکیل‌شده. همان‌گونه که در شکل (۲) نشان داده‌شده است، خط $(< wx_i b >) = 0$ داده‌های موجود را در دو کلاس ± 1 طبقه‌بندی می‌کند. به این خط، ابر صفحه جداکننده گفته می‌شود. دو خط $(< wx_i + b >) + 1$ و $(< wx_i + b >) = -1$ ه ترتیب بیانگر مرز ناحیه‌ی دسته‌های $y = -1$ و $y = +1$ می‌باشند. به نزدیک‌تر داده‌های آموزشی به ابر صفحه‌های جداکننده، بردار پشتیبان[5] نامیده می‌شوند.



شکل۲. ابر صفحه‌ی جداکننده دو گروه ۱+ و ۱-

1 . Elitism TACO
2 . Ranked - Based TACO
3 . Binary ACO
4 . Hyper plane
5 . Support Vector (SV)

در ماشین بردار پشتیبان به دو طریق خطی و غیرخطی می‌توان مجموعه نقاط را از یکدیگر جدا نمود (آوسی انجین[1]، ۲۰۰۹). در ادامه به شرح هر یک از روش‌های فوق پرداخته‌شده است.

تفکیک خطی

در حالتی که داده‌ها را بتوان به‌صورت خطی از هم جدا کرد، ماشین بردار پشتیبان با در نظر گرفتن مجموعه داده‌های آموزشی، با استفاده از حل مسئله بهینه‌سازی زیر ابر صفحه بهینه با حاشیه حداکثر را پیدا می‌نماید:

$$\begin{aligned} &\text{Min}\frac{1}{2\|W\|^2} \\ &St:\ (< wx_i + b >) \geq -1 \quad i = 1.2. \ldots . m \end{aligned} \qquad (۱)$$

که در آن $\|W\|$ نرم اقلیدسی[2] است. بر اساس مدل فوق، کمینه شدن مقدار $\|W\|$ با توجه به محدودیت آن، منجر به حداکثر شدن پهنای صفحه می‌گردد.

حل این مسئله بهینه‌سازی دشوار است، لذا به‌منظور ساده‌تر نمودن حل آن، از ضرایب لاگرانژ استفاده می‌گردد. سپس دوگان آن با توجه به‌شرط کروش کان-تاکر[3]، به‌صورت زیر در خواهد آمد:

$$\begin{aligned} &\max W(\alpha) = \sum \alpha_i - \frac{1}{2}\sum\sum \alpha_i\alpha_j y_i y_j < x_i x_j > \\ &Subject\ \ to: \sum \alpha_i y_i = 0. \alpha_i \geq 0. i = 1. \ldots . N \end{aligned} \qquad (۲)$$

با توجه به عدم وجود b در معادله فوق، از محدودیت‌های ابتدایی برای به دست آوردن آن استفاده می‌کنیم. پس‌ازآنکه مقادیر α_i و b به دست آمد، می‌توان SVM را برای دسته‌بندی نمونه-های جدید بکار برد. اگر x یک نمونه جدید باشد، دسته‌بندی آن به‌صورت زیر مشخص می‌گردد.

$$F(x) = sign[f(x. \alpha. b)] \qquad (۳)$$

که در آن $f(x. \alpha. b)$ به‌صورت زیر می‌باشد:

$$\begin{aligned} f(x. \alpha. b) &= wx + b \\ &= \sum \alpha_i y_i x_i x + b \\ &= \sum_{i \epsilon SV} \alpha_i y_i x_i x + b \end{aligned} \qquad (۴)$$

همان‌گونه که قبلاً عنوان شد، SV بردارهای پشتیبان می‌باشند.

تفکیک غیرخطی

یک فرض بسیار مهم در SVM این است که داده‌ها به‌صورت خطی جدا پذیر باشند. درحالی‌که در عمل در بیشتر مواقع این فرض صحیح نیست. برای اولین بار کورتس و وپنیک در سال ۱۹۹۵ به‌منظور اینکه بتوانند در این حالت یک ابر صفحه‌ی بهینه را برای جداسازی ۲ کلاس به دست بیاورند، متغیرهای نامنفی $\varepsilon_i \geq 0$ را به‌عنوان مقادیر خطا برای هر بردار تعریف نمودند(کورتس[4]، ۱۹۹۵). مطابق روش ارائه‌شده صورت‌مسئله بهینه‌سازی تبدیل می‌شود به یافتن w به‌نحوی‌که معادله زیر حداقل گردد:

---

1 . Avci.Engin
2 . Euclidean norm
3 . Karush Kuhn–Tucker (KKT)
4 . Cortes

$$Min\ \frac{1}{2}\|w\|^2 + C\sum_i \varepsilon_i \qquad (۵)$$
$$St: \quad y_i(< wx_i > +b) \geq 1 - \varepsilon_i . \varepsilon_i \geq 0\ \forall i$$

که در آن C پارامتر تنظیم‌کننده حاشیه می‌باشد که وظیفه‌ی آن برقراری تعادل بین حداکثر کردن حاشیه و حداقل کردن خطای دسته‌بندی بوده و همواره بزرگ‌تر از صفر است (کریستیانینی و تیلور[1]، ۲۰۰۰). اگر C عددی بزرگ انتخاب شود، توجه بیشتری به خطا معطوف می‌گردد. ماشین بردار پشتیبانی که به این صورت تعریف‌شده باشد را ماشین بردار پشتیبان حاشیه نرم[2] می‌نامند (آبه[3]، ۲۰۰۵).

همچنین در حالتی که داده‌ها جداناپذیر بوده و همچنین کلاس‌ها دارای همپوشانی هستند، جدا کردن کلاسها توسط مرز خطی همواره با بروز خطا همراه می‌باشد. به‌منظور حل مشکل مزبور می‌توان ابتدا داده‌ها را با استفاده از یک تبدیل غیرخطی φ، از فضای اولیه به فضایی با بعد بالاتر منتقل کرد با این هدف که در فضای جدید، کلاس‌ها تداخل کمتری با یکدیگر داشته باشند (بورگس[4]، ۱۹۹۸). پس از نگاشت داده‌ها به فضای بالاتر، با استفاده از معادلات قبل و جایگزین کردن x_i با $\varphi(x_i)$، ابر صفحه بهینه به دست خواهد آمد.

انتقال از بعد پایین به بعدی بالاتر را می‌توان توسط توابع کرنل[5] انجام داد. تعدادی از توابع کرنل موجود به شرح ذیل می‌باشند:

$$K(x_i . x_j) = x_i^T x_j + c \qquad \text{کرنل خطی} \qquad (۶)$$
$$K(x_i . x_j) = (\alpha x_i^T x_j + c)^d \qquad \text{کرنل چندجمله‌ای} \qquad (۷)$$
$$K(x_i . x_j) = exp\left(-\|x_i - x_j\|^2/\sigma^2\right) \qquad \text{کرنل تابع پایه شعاعی} \qquad (۸)$$

در معادلات فوق، c عبارت ثابت، d درجه چندجمله‌ای و سیگما (σ) پارامتر قابل تنظیم می‌باشد. کرنل تابع پایه شعاعی در دو مقوله نسبت به کرنل‌های دیگر دارای نقطه قوت بوده و درنتیجه عملکرد مطلوب‌تری در مسائل طبقه‌بندی اعتباری از خود نشان داده است. این کرنل به‌صورت غیرخطی نمونه‌ها را به ابعاد فضایی بالاتری نگاشت می‌نماید، بنابراین برخلاف کرنل خطی، در حالت‌هایی که روابط بین کلاس‌ها و ویژگی‌ها به‌صورت غیرخطی است، کاربرد خواهد داشت. دومین نقطه قوت، تعداد پارامترهایی است که بر پیچیدگی انتخاب مدل تأثیرگذار می‌باشند. کرنل چندجمله‌ای تعداد پارامترهای بیشتری نسبت به RBF دارا می‌باشد (لین[6] و همکاران، ۲۰۱۱). ازاین‌رو در این پژوهش از کرنل تابع شعاعی استفاده‌شده است.

معیارهای ارزیابی مدل

مسلماً انجام هر کاری بدون ارزیابی آن ارزش چندانی ندارد و راهی به سمت بهبود و انتخاب بهترین راه پیدا نمی‌کند. در این پژوهش نیز قصد داریم تا پس از تغییر دادن برخی پارامترها، شرایط و...، آن‌ها را با یکدیگر مقایسه نماییم تا بهترین شرایط جهت پیش‌بینی طبقهٔ اعتباری مشتریان انتخاب گردد. برای مقایسهٔ این حالات، باید ابتدا معیارهای ارزیابی مناسبی را تعریف نمود تا نتیجه مقایسه، برای تصمیم‌گیری قابل‌اتکا باشد. در پژوهش ژانگ، گائو و شی(۲۰۱۴)[7]، چندین معیار ارزیابی برای

1 . Cristianini, Shawe-Taylor
2 . Soft-Margin
3 . Abe
4 . Burges
5 . Kernel
6 . Lean
7 . Zhang& gao& shi

ارزیابی طبقه‌بندی اعتباری استفاده گردیده که در این پژوهش قصد استفاده از آن‌ها را داریم. پیش از معرفی این معیارها، به برخی از تعاریف در ارزیابی طبقه‌بندی کننده‌های دودویی می‌پردازیم.

نتیجه‌ای که مدل، اعتبار یک مشتری را پیش‌بینی می‌کند، ممکن است اعتبار بد (Positive)، یا اعتبار خوب (Negative) باشد. حال در این راه، چهار حالت زیر ممکن است اتفاق بیفتد:

الف) نتیجه پیش‌بینی بدحساب باشد و مشتری بانکی هم طبق طبقه‌بندی تجربی بانک بدحساب باشد که به آن True Positive می‌گویند و به اختصار با TP نشان می‌دهند.

ب) نتیجه پیش‌بینی بدحساب باشد؛ اما مشتری بانکی طبق طبقه‌بندی تجربی، واقعاً خوش‌حساب باشد که آن را با False Positive و به اختصار با FP نشان می‌دهند. طبق تعاریف در علم آمار، در اینجا خطای نوع یک صورت گرفته است.

ج) نتیجه پیش‌بینی خوش‌حساب باشد و مشتری بانکی هم طبق طبقه‌بندی تجربی بانک، واقعاً خوش‌حساب باشد که به آن True Negative می‌گویند و به اختصار با TN نشان می‌دهند.

د) نتیجه پیش‌بینی خوش‌حساب باشد؛ حال‌آنکه طبق طبقه‌بندی تجربی بانک، مشتری واقعاً بدحساب باشد که به آن False Negative می‌گویند و آن را با FN نشان می‌دهند... حال در ادامه به تعریف و بیان رابطهٔ معیارهای مورداستفاده خواهیم پرداخت.

صحت ($Accuracy$): این نسبت عبارت است از نسبت پیش‌بینی‌های درست به کل پیش‌بینی‌ها.

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (۹)$$

حساسیت (Recall): بیان‌گر نسبت پیش‌بینی درست اعتبارهای بد به کل اعتبارهای واقعاً بد می‌باشد.

$$Recall = \frac{TP}{TP + FN} \quad (۱۰)$$

تشخیص (Specificity): برابر نسبت پیش‌بینی درست اعتبارهای خوب به کل اعتبارهای واقعاً خوب می‌باشد.

$$Specificity = \frac{TN}{TN + FP} \quad (۱۱)$$

دقت (Precision): بیانگر نسبت پیش‌بینی‌های درست اعتبارهای بد به کل اعتبارهای بدی که پیش‌بینی شده اند.

$$Precision = \frac{TP}{TP + FP} \quad (۱۲)$$

برای بررسی تعداد ویژگی‌های بعد از انتخاب ویژگی، نسبت به تعداد ویژگی‌های اصلی، پارامتری به نام F_r تعریف می‌شود که با رابطه زیر به دست می‌آید:

$$F_r = \frac{p - q}{p} \quad (۱۳)$$

که p تعداد ویژگی‌های اصلی و q تعداد ویژگی‌ها بعد از انتخاب ویژگی است. طبق این رابطه هر چه مقدار F_r به ۱ نزدیک‌تر باشد، کاهش تعداد ویژگی‌ها بیشتر بوده و مطلوب‌تر است.

توصیف داده‌ها

در این پژوهش، عمل طبقه‌بندی اعتباری را بر روی مشتریان یکی از بانک‌های ایرانی انجام خواهیم داد. مشتریانی که قصد طبقه‌بندی آن‌ها را داریم، داده‌های مربوط به ۸۵ شرکت از تسهیلات گیرندگان حقوقی یک بانک ایرانی در یک بازه‌ی ۵ ساله (۱۳۸۹-۱۳۹۳) به همراه ۱۶ ویژگی مربوط به هر یک از آن‌ها می‌باشد.۱۶ آیتم موجود در ترازنامه و صورت‌حساب سود و زیان این شرکت‌ها می‌باشد. اهمیت بسیاری از نسبت‌های مالی که در ادامه ذکرشده است، در پژوهش‌های نیلی و سبزواری (۱۳۸۷)، سپهردوست و برجیسیان

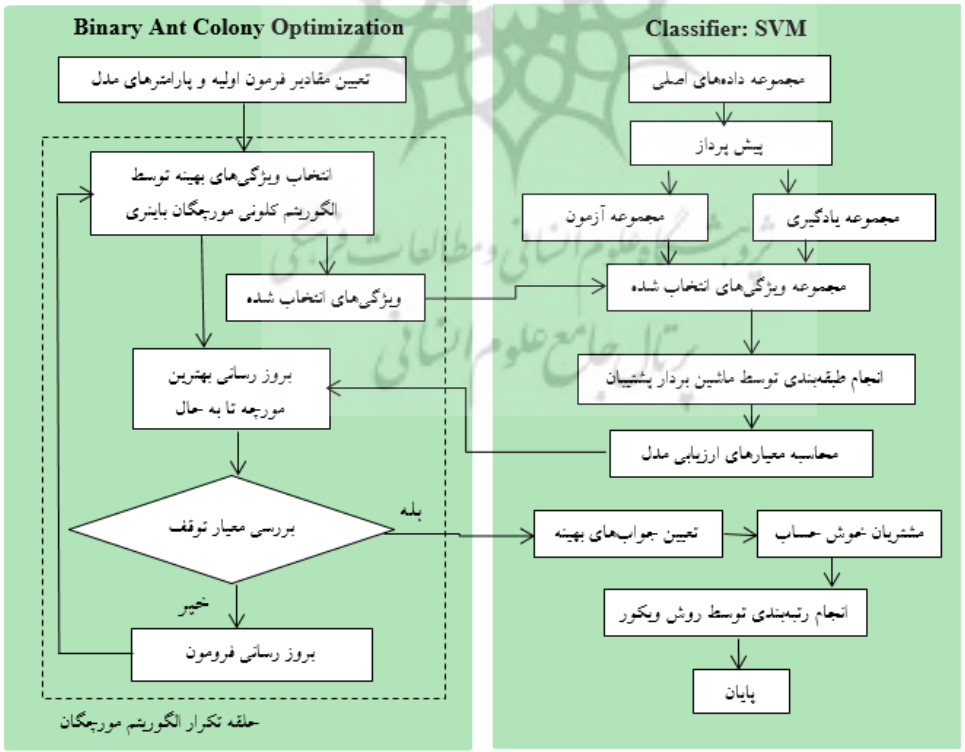
(۱۳۹۲)، صفری، ابراهیمی و شیخ (۱۳۸۹) و... مورداستفاده قرارگرفته است و به آن‌ها توصیه‌شده است. این ۱۶ آیتم عبارت‌اند از: حساب‌های دریافتنی، موجودی مواد و کالا، دارایی‌های جاری، دارایی‌های ثابت، کل دارایی‌ها، بدهی‌های جاری، کل بدهی‌ها، حقوق صاحبان سهام و سود انباشته که از ترازنامه این شرکت‌ها جمع‌آوری‌شده است و سرمایه، فروش، سود ناخالص، سود عملیاتی، هزینه‌های مالی، سود قبل از مالیات و سود خالص پس از مالیات که از صورت‌حساب سود و زیان آن‌ها اخذ گردیده است. این داده‌های خام جمع‌آوری‌شده به‌منظور محاسبهٔ نسبت‌های مالی مورداستفاده قرار می‌گیرد. پس از محاسبهٔ مقادیر میانگین نسبت‌های مالی، باید این نسبت‌ها را به‌صورت مقادیر بین صفر و یک نرمال‌سازی نمود. رابطهٔ مورداستفاده برای نرمال‌سازی داده‌ها در قالب رابطهٔ (۱۸) ارائه‌شده است که در آن x_iها، مقادیر مربوط به یک نسبت مالی خاص، برای ۸۵ شرکت می‌باشند از این رابطه در پژوهش لیو و موتودا (۱۹۹۸)[1] استفاده گردیده است.

$$Normalized\ value\ of\ x_i = \frac{x_i - \min_i x_i}{\max_i x_i - \min_i x_i} \qquad (۱۴)$$

لازم به ذکر است برای جمع‌آوری این داده‌ها از دو پایگاه اینترنتی کدال[2] و پردازش اطلاعات مالی ایران[3] بهره گرفته‌شده است.

مدل پیشنهادی

شکل (۳) نحوه عملکرد مدل پیشنهادی را نشان می‌دهد. هم‌چنین شرح دقیق هر مرحله در ادامه بیان‌شده است.



شکل۳. نمودار مدل پیشنهادی

1 . Liu & Motoda
2 . www.codal.ir
3 . www.fipiran.com

تعیین پارامترهای بهینه مدل

برای اینکه جواب‌های بهتری از این مدل بگیریم باید پارامترهای مدل را تنظیم کنیم زیرا در الگوریتم مورچگان اگر پارامترهای الگوریتم به‌خوبی انتخاب نشوند، احتمال رکود و همگرایی به بهینه‌های محلی در آن زیاد خواهد بود. پارامترهایی که در نسخه‌ی باینری مورچگان مهم است که تعیین شوند تعداد مورچگان و نرخ تبخیر می‌باشد.

تعیین مقدار فرمون هر ویژگی

مقدار فرمون به‌صورت یک ماتریس وارد می‌گردد که تعداد ستون‌های آن برابر با تعداد کل ویژگی‌های موجود در مجموعه داده و تعداد سطرهای آن برابر با ۲ است. درواقع برای هر ویژگی دو حالت انتخاب شدن و نشدن (سطر اول) و یا انتخاب شدن در یک‌راه حل (سطر دوم) در نظر گرفته می‌شود. در ابتدا برای هر ویژگی مقدار فرمون سطر اول و دوم یکسان است و پس از انجام هر تکرار به‌روزرسانی می‌شود. در این پژوهش مقدار اولیه فرمون را یک در نظر گرفتیم.

فاز اصلی: پس از تخصیص مقادیر بهینه به پارامترها، وارد فاز اصلی مدل پیشنهادی می‌شویم. این فاز با انتخاب ویژگی توسط الگوریتم مورچگان باینری آغاز می‌گردد، پس از انتخاب ویژگی‌ها، طبقه‌بندی با SVM صورت گرفته و تابع برازش محاسبه می‌گردد. معیارهای توقف الگوریتم بررسی‌شده و بر این اساس اجرای الگوریتم ادامه یافته و یا متوقف می‌گردد. شرح کامل هر گام از فاز اصلی در ادامه بیان‌شده.

انتخاب ویژگی با استفاده از الگوریتم مورچگان باینری

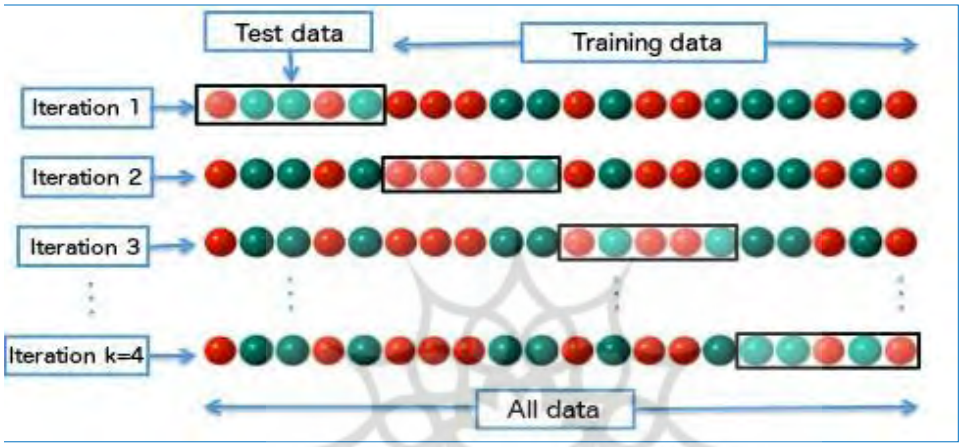
در حل مسئله انتخاب ویژگی با استفاده از الگوریتم جمعیت مورچگان باینری به ازای هر ویژگی، یک گره (بیت) صفر و یک گره (بیت) یک با جایگاه مشخص و یکسان در رشته باینری در نظر گرفته می‌شود. ازاین‌رو طول رشته بیت‌های یک و صفر در رشته‌های شکل (۱) برای حل یک مسئله خاص با تعداد ویژگی‌های آن مسئله برابر است. در این مسئله چنانچه یک مورچه برای یک ویژگی خاص، بیت حاوی مقدار صفر را برگزیند به‌منزله عدم مشارکت ویژگی و چنانچه بیت یک را برگزیند به‌منزله انتخاب ویژگی و مشارکت آن تلقی می‌شود. در این روش، تمام مورچه‌ها در آغاز در ابتدای رشته بیت‌ها قرارگرفته و با شروع کار الگوریتم بر اساس تابع احتمال انتخاب (۱۵) گره بعدی خود را انتخاب می‌کنند:

$$P^{k}_{ix.jy}(t) = \frac{\tau_{ix.jy}}{\tau_{ix.j0} + \tau_{ix.j1}} \qquad (۱۵)$$

این رابطه، احتمال انتخاب بیت با مقدار $y\epsilon\{0.1\}$ در گره‌ی بعدی (گره j ام) برای مورچه k ام که در زمان t در موقعیت $x\epsilon\{0.1\}$ در گره i ام قرار داشته را مشخص می‌کند. همچنین $\tau_{i0.j0}$، $\tau_{i0.j1}$، $\tau_{i1.j0}$، $\tau_{i1.j1}$ به ترتیب بیانگر فرومون بین مسیرهای متصل‌کننده گره‌های i ام و j ام به ترتیب روی یال‌های (۰به۰)، (۱به۰)، (۰به۱) و (۱به۱) است. لازم به یادآوری است که بنا به محدودیت‌های الگوریتم مورچه باینری همواره محدودیت j=i+1 برقرار است. یعنی این‌که هر مورچه در هر گره صرفاً می‌تواند به گره بعدی خود که از قبل مکان آن تعیین‌شده حرکت کرده و یکی از زیر گره‌های صفر و یک را برگزیند.

طراحی سیستم اعتبارسنجی

به‌منظور پیشگیری از وقوع انطباق بیش‌ازحد و اینکه مدل ریاضی، به‌جای یادگیری داده‌ها، آن‌ها را حفظ نکند و همچنین انتخاب بهترین پارامترهای مدل، عمل اعتبارسنجی را به شیوه اعتبارسنجی متقاطع k لایه‌ای انجام می‌دهیم. این عملگر ابتدا داده‌ها را به k زیرمجموعه مجزای تقریباً مساوی تقسیم می‌کند، پس‌ازآن هر بار (k-1) زیرمجموعه از داده‌ها را به‌عنوان داده‌های آموزشی و یک زیرمجموعه را به‌عنوان داده‌های تست در نظر می‌گیرد و این کار را k بار انجام می‌دهد به‌طوری‌که هر زیرمجموعه یک‌بار به‌عنوان داده تست قرار می‌گیرد. دقت پیش‌بینی برای آن زیرمجموعه محاسبه‌شده و دقت کلی مدل از میانگین آن‌ها به دست می‌آید. در این پژوهش مقدار k را ۵ در نظر گرفتم. شکل (۴) نحوه ی کار اعتبار سنجی متقاطع k لایه ای را نشان می دهد.



شکل ٤. روش اعتبار سنجی متقاطع k لایه‌ای

انجام طبقه‌بندی توسط SVM

کرنل تابع پایه شعاعی در دو مقوله نسبت به کرنل‌های دیگر دارای نقطه قوت بوده و درنتیجه عملکرد مطلوب‌تری در مسائل طبقه‌بندی اعتباری از خود نشان داده است. این کرنل به‌صورت غیرخطی نمونه‌ها را به ابعاد فضایی بالاتری نگاشت می‌نماید، بنابراین برخلاف کرنل خطی، در حالت‌هایی که روابط بین کلاس‌ها و ویژگی‌ها به‌صورت غیرخطی است، کاربرد خواهد داشت. دومین نقطه قوت، تعداد پارامترهایی است که بر پیچیدگی انتخاب مدل تأثیرگذار می‌باشند. کرنل چندجمله‌ای تعداد پارامترهای بیشتری نسبت به RBF دارا می‌باشد (لین[1] و همکاران،۲۰۱۱). ازاین‌رو در این پژوهش از کرنل تابع شعاعی استفاده‌شده است. مقدار پارامتر C و σ را به ترتیب ۱۰ و ۰,۹ در نظر می‌گیریم. به‌وسیله‌ی الگوریتم SVM مشتریان به دودسته‌ی خوش‌حساب و بدحساب طبقه‌بندی می‌شوند.

ویژگی‌هایی که توسط الگوریتم مورچگان انتخاب می‌شوند به‌عنوان ورودی‌های SVM هستند. مشتریان با توجه به ویژگی‌های انتخاب‌شده توسط الگوریتم SVM طبقه‌بندی می‌شوند.

محاسبه تابع برازش

تابع برازش کیفیت یک‌راه حل را اندازه‌گیری می‌کند. پس از انجام طبقه‌بندی، به محاسبه تابع برازش می‌پردازیم. در این مطالعه، از معیار حساسیت (Recall) به‌عنوان تابع برازش استفاده کرده‌ایم. معیار حساسیت

1 . Lin

نشان‌دهنده نسبتی از مشتریان بدحساب است که به‌درستی طبقه‌بندی‌شده‌اند؛ زیرا این معیار بسیار حائز اهمیت می‌باشد چراکه هزینه‌ای که از عدم‌تشخیص صحیح مشتریان بدحساب مؤسسات اعتباری را تهدید می‌نماید، بسیار بیشتر از هزینه عدم‌تشخیص صحیح مشتریان خوش‌حساب می‌باشد. این هزینه شامل از دست دادن اصل‌وفرع تسهیلات به همراه هزینه پیگیری مطالبات معوق می‌باشد. با توجه به مفهوم معیار مذکور، اختلاف آن از عدد یک را می‌توان به‌عنوان ریسک اعتباری روش مورداستفاده در نظر گرفت. منظور از ریسک اعتباری اعطای اعتباری به مشتریان بدحساب می‌باشد که مدل توانایی تشخیص درست آن‌ها را نداشته است. معیار دیگری که برای ما مهم است معیار F_1 می‌باشد که برای بررسی تعداد ویژگی‌های بعد از انتخاب ویژگی استفاده می‌شود. ر چه مقدار آن به ۱ نزدیک‌تر باشد، کاهش تعداد ویژگی‌ها بیشتر بوده و مطلوب‌تر است. این معیار طبق رابطه (۱۷) محاسبه می‌شود.

بررسی معیار توقف

پس از محاسبه تابع برازش، معیارهای توقف الگوریتم بررسی می گردند. در اینجا ما معیار توقف را تعداد تکرارها در نظر گرفته‌ایم. وقتی الگوریتم به تعداد تکرارهای معینی رسید متوقف می‌شود.

به‌روزرسانی فرومون

در پایان هر تکرار از الگوریتم، فرومون ابتدا تبخیر و سپس به‌روزرسانی می‌شود. رابطه مورداستفاده جهت تبخیر فرومون به صورت زیر است:

با توجه به اینکه برای محاسبه‌ی دقت مدل از اعتبار سنجی چندلایه‌ای استفاده کرده بودیم که مجموعه داده‌ها به K زیرمجموعه تقسیم‌شده بودند؛ بنابراین $CVREC$ میانگین حساسیت (درصد مشتریان بدحسابی که به تی دسته‌بندی‌شده‌اند) K زیرمجموعه است.

$$CVREC = \frac{\sum_i Recall_i}{K} \quad i = 1.2. \ldots . K \qquad (۱۶)$$

فرومون مسیرهایی که مورچه طی کرده به‌وسیله‌ی یک نرخ تبخیر ثابت ρ آپدیت می‌شود. پارامتر نرخ تبخیر برای اجتناب از افزایش پیوسته در میزان فرومون هر ویژگی است. این پارامتر به الگوریتم کمک می کند تا تصمیمات اشتباهی که گرفته است را فراموش کند.

و آنگاه فرومون روی مسیری که مورچه عبور کرده ترشح می‌شود. $\Delta\tau_{ij}^{k}$ مقدار فرومونی است که مورچه k ام ترشح می کند. رویکردهای مختلفی برای به‌روزرسانی فرمون طراحی‌شده است. در این الگوریتم استراتژی مورچه نخبه به کار گرفته می‌شود که $\Delta\tau_{ij}^{Best}$ فرومونی است که توسط بهترین مورچه که تاکنون بوده ترشح می‌شود. e وزن فرمونی است که توسط بهترین مورچه ترشح می‌شود. بنابراین ؛ رابطه زیر برای بروز رسانی فرمون مورداستفاده قرار می گیرد:

$$\tau_{ij} = (1-\rho)\tau_{ij} + \sum_{k=1}^{Numberof Ants} \Delta\tau_{ij}^{k} + e\Delta\tau_{ij}^{Best} \qquad (۱۷)$$

$$\Delta\tau_{ij}^{k} = \begin{cases} CVREC^{K} & \text{if ant k use edge } (i.j) \text{ in tour} \\ 0 & otherwise \end{cases} \qquad (۱۸)$$

این به‌روزرسانی که با افزایش مقدار فرمون ویژگی‌های منتخب همراه است منجر به آن می‌شود که در تکرارهای بعدی نیز احتمال انتخاب این ویژگی‌ها افزایش یابد. به‌این‌ترتیب الگوریتم پس از چندین تکرار به یک‌راه حل واحد همگرا می‌گردد.

تعیین جواب بهینه

در هر تکرار از الگوریتم فوق از بین راه‌حل‌های ساخته‌شده بهترین راه‌حل بر اساس معیار حساسیت تعیین و ذخیره می‌گردد. پس از پایان تمام تکرارها مقایسه و جواب بهینه تعیین می‌شود.

نتایج حاصل از مدل BACO-SVM

پارامترهای مهم در نسخه‌ی باینری مورچگان، تعداد مورچگان و نرخ تبخیر می‌باشد. انتخاب مقادیر ۲۰، ۳۰، ۴۰ برای تعداد مورچگان و انتخاب مقدارهای ۰,۲، ۰,۳، ۰,۵ برای نرخ تبخیر، پارامترهای SVM، C=10 و $\sigma = 0.9$ قرار می‌دهیم. مدل BACO-SVM باید ۹ بار ران شود. ما این مدل را ۹ بار با ۱۰۰ تکرار ران می‌کنیم. همان‌طور که در جدول (۲) نشان داده‌شده است ما با تغییر تعداد مورچه‌ها و نرخ تبخیر، مدل را ۹ بار اجرا می‌کنیم که در جدول (۲) شماره مدل بیانگر این است که هر مدل با چه پارامترهایی اجراشده است که مقدار پارامترهای هر مدل در جدول (۱) آمده است.

جدول ۱. تنظیم پارامترهای الگوریتم مورچگان

نام مدل/پارامترها	تعداد مورچه‌ها	نرخ تبخیر	C	σ
۱	۲۰	۰,۲	۱۰	۰,۹
۲	۲۰	۰,۳	۱۰	۰,۹
۳	۲۰	۰,۵	۱۰	۰,۹
٤	۳۰	۰,۲	۱۰	۰,۹
٥	۳۰	۰,۳	۱۰	۰,۹
٦	۳۰	۰,۵	۱۰	۰,۹
۷	٤۰	۰,۲	۱۰	۰,۹
۸	٤۰	۰,۳	۱۰	۰,۹
۹	٤۰	۰,۵	۱۰	۰,۹

جدول ۲. مقادیر معیارها در مدل BACO-SVM با تعداد تکرار ۱۰۰

مدل- معیارها	Accuracy		Recall		Specificity		Precision		
	Test	Training	Test	Training	Test	Training	Test	Training	Fr
۱	۰,۸۹	۰,۹۸	۰,۷٦	۱,۰۰	۰,۹۰	۰,۹۸	۰,۸٦	۰,۹٤	۰,۲٥
۲	۰,۹۱	۰,۹٥	۰,۸٤	۰,۹۸	۰,۹۳	۰,۹٤	۰,۸٤	۰,۸٥	۰,۳۱
۳	۰,۹۱	۰,۹۲	۰,۸٤	۰,۹۷	۰,۹۳	۰,۹۰	۰,۸٤	۰,۸۱	۰,٥٦
٤	۰,۹۳	۰,۹٥	۰,۹۲	۰,۹۷	۰,۹۳	۰,۹٤	۰,۸٥	۰,۸٦	۰,۲٥
٥	۰,۹۲	۰,۹۲	۰,۹۲	۰,۹٦	۰,۹۲	۰,۹۰	۰,۸۱	۰,۸۱	۰,۳۱
٦	۰,۹۱	۰,۹۲	۰,۹۲	۰,۹۹	۰,۹۰	۰,۸۹	۰,۷۹	۰,۷۹	۰,۳۸
۷	۰,۹۲	۰,۹۱	۰,۸۸	۰,۹۸	۰,۹۳	۰,۸۸	۰,۸٥	۰,۷۸	۰,٥٦
۸	۰,۹۰	۰,۹۱	۰,۹٦	۰,۹۷	۰,۸۸	۰,۸۹	۰,۷۷	۰,۷۹	۰,۳۱
۹	۰,۹٥	۰,۹۲	۰,۹٦	۰,۹۸	۰,۹٥	۰,۸۹	۰,۸۹	۰,۷۹	۰,٥٦

راهنما: مقادیر با قلم ضخیم بیانگر بیشترین مقدار میانگین هر معیار می‌باشند

این ۹ مدل را با پارامترهای مختلفی که برای هرکدام در نظر گرفتیم با ۱۰۰ تکرار اجرا می‌کنیم که نتایج را در جدول (۲) ملاحظه می‌فرمایید. برای مقایسه‌ی مدل BACO-SVM با مدل PSO-SVM، GA-SVM و SVM به‌تنهایی در اینجا از بین ۹ بار اجرای مدل، مدلی که بهترین نتیجه را داشته انتخاب می‌کنیم برای فرآیند مقایسه مدل شماره ۹، در معیارهای Accuracy، Recall، Specificity، Precision و Fr از همه‌ی مدل‌ها بهتر عمل کرده. درنتیجه برای مدل BACO-SVM، ما پارامترهای مدل شماره‌ی ۹ (تعداد مورچه‌ها ۴۰ و نرخ تبخیر ۰,۵) را در نظر می‌گیریم. برای مقایسه‌ی معیارها فقط نتایج معیارهای تست را در همه‌ی مدل‌ها در نظر می‌گیریم.

رتبه‌بندی به‌وسیله‌ی روش VIKOR

یک روش تصمیم‌گیری چند شاخصه توافقی است که توسط آپریکویچ و تی زنگ بر اساس روش ال پی متریک توسعه یافت. شاخص رتبه‌بندی گزینه‌ها در این روش بر اساس میزان نزدیکی آن به جواب ایده ال می‌باشد. دلیل استفاده از مفهوم روش ویکور در روش ترکیبی پیشنهادی آن است که ویژگی و قابلیت‌های این روش، در عمل می‌تواند به انتخاب فردی که نسبت به سایرین به جواب ایده آل نزدیک‌تر بوده و از بیشینه مطلوبیت گروهی و کمینه تأثر انفرادی نیز برخوردار باشد، کمک می‌نماید (توکلی مقدم و همکاران، ۱۳۹۱).

هدف روش VIKOR تمرکز بر رتبه‌بندی و انتخاب از بین یک مجموعه از آلترناتیوها در مسئله‌ای با معیارهای متعارض است که یک لیست رتبه‌بندی توافقی به‌علاوه یک یا چند راه‌حل توافقی برای آن مشخص می‌کند. راه‌حل توافقی یک راه‌حل شدنی است که نزدیک‌ترین راه‌حل به راه‌حل ایده آل است. در اینجا جواب سازشی نزدیک‌ترین جواب موجه به جواب ایده آل است که کلمه سازش به یک توافق متقابل اطلاق می‌گردد. درواقع مدل ویکور از طریق ارزیابی گزینه‌ها بر اساس معیارها، گزینه‌ها را اولویت-بندی یا رتبه‌بندی می‌کند. در این مدل معیارها وزن‌دهی نمی‌شوند بلکه معیارها از طریق روش‌های دیگر ارزیابی می‌شود و سپس گزینه‌ها بر اساس معیارها و با ترکیب در ارزش معیارها، ارزیابی‌شده و رتبه‌بندی می‌شوند. در این مدل همواره چند گزینه مختلف وجود دارد که این گزینه‌ها بر اساس چند معیار به‌صورت مستقل ارزیابی می‌شوند و در نهایت گزینه‌ها بر اساس ارزش، رتبه‌بندی می‌گردند.

منظور از توافق یا سازش جوابی است که بر اساس توافق متقابل بین معیارها حاصل می‌شود.

رویکردهای اصلی برای تصمیم‌گیری، شامل تئوری مطلوبیت هستند.

مبنای روش حداقل نمودن بردار ارزیابی آلترناتیوها از نقطه ایده آل مثبت است:

محاسبه‌ی یک فاصله ترکیبی با فرمول خاص، برای هر آلترناتیو

راه‌حل ارجح: آلترناتیوی که کوچک‌ترین فاصله ترکیبی را دارد

مراحل روش VIKOR به ترتیب زیر است:

۱. تشکیل ماتریس تصمیم: ماتریس تصمیم با توجه به ارزیابی همه گزینه‌ها برای معیارهای مختلف تشکیل می‌شود.

۲. تعیین وزن معیارها: در این مرحله با توجه به ضریب اهمیت معیارهای مختلف در تصمیم-گیری، با استفاده از روش‌هایی مانند آنتروپی یا AHP و ... بردار وزن تعریف می‌شود؛ که ما در اینجا برای تعیین وزن معیارها از روش آنتروپی استفاده کردیم.

۳. تعیین نقطه ایده آل مثبت و منفی: برای هر معیار، بهترین و بدترین هر یک را در میان همه گزینه‌ها تعیین کرده و به ترتیب همانند زیر نشان می‌دهیم:

$$f+= Max\ f_{ij} \qquad (۱۹)$$
$$f-= Min\ f_{ij}$$

۴. محاسبه شاخص مطلوبیت (S) و شاخص نارضایتی (R) برای هر گزینه:

دو مفهوم اساسی در محاسبات VIKOR، سودمندی (S) و تأسف (R) است. مقدار سودمندی (S) بیانگر فاصله نسبی گزینه i ام از نقطه ایده‌آل و مقدار تأسف (R) بیانگر حداکثر ضرر گزینه i ام از دوری نقطه ایده‌آل است.

$$S_j = \sum_{i=1}^{n} w_i . \frac{f_i^* - f_{ij}}{f_i^* - f_i^-} \qquad (۲۰)$$

$$R_j = \max_i w_i . \frac{f_i^* - f_{ij}}{f_i^* - f_i^-} \qquad (۲۱)$$

محاسبه شاخص VIKOR: برای هر گزینه شاخص ویکور با رابطه زیر محاسبه می‌شود:

$$Q_i = v\left[\frac{S_i - S^*}{S^- - S^*}\right] + (1-v)\left[\frac{R_i - R^*}{R^- - R^*}\right] \qquad (۲۲)$$

$$S^- = Max\ S_i\ , S^* = Min\ S_i \qquad (۲۳)$$
$$R^- = Max\ R_i\ , R^* = Min\ R_i$$

v مقداری بین صفر و یک است و وزن برای استراتژی ماکسیمم مطلوبیت گروهی است. در اینجا v با مقدار ۰٫۵ در نظر گرفته‌شده است. در تصمیم‌گیری با تکنیک VIKOR دو شرط نهایی داریم: در گام پایانی از این تکنیک، گزینه‌های بر اساس مقادیر S، R، Q در سه گروه از کوچک به بزرگ مرتب می‌شوند. بهترین گزینه آن است که کوچک‌ترین مقدار Q را داشته باشد به‌شرط آنکه دو شرط زیر برقرار باشد:

شرط یک: اگر A_1 و A_2 در میان m گزینه رتبه اول و دوم را داشته باشند، باید رابطه زیر برقرار باشد.

$$Q(A_2) - Q(A_1) \geq \frac{1}{m-1} \qquad (۲۴)$$

شرط دو: گزینه A_1 باید حداقل در یکی از گروه‌های R و S به‌عنوان رتبه برتر شناخته شود. اگر شرط نخست برقرار نباشد هر دو گزینه بهترین خواهند بود. اگر شرط دوم برقرار نباشد گزینه A_1 و A_2 هر دو به‌عنوان گزینه برتر انتخاب می‌شوند. مزیت مدل VIKOR در این است که الزاماً در این مدل جهت ارزیابی گزینه‌ها بر اساس معیارها، نیازی به استفاده از نظرات کارشناسان نیست بلکه می‌توان از داده‌های خام استفاده کرد.

نتایج پژوهش

در این قسمت به مقایسه‌ی مدل‌ها، پرداخته‌شده است. جهت نشان دادن تأثیر الگوریتم انتخاب ویژگی در افزایش دقت مدل، SVM بدون الگوریتم انتخاب ویژگی اجراشده است. میانگین نتایج در جدول (۳) ارائه‌شده است.

جدول ۲. مقایسه نتایج حاصل از مدل‌ها

مدل-معیار	Accuracy	Recall	Specificity	Precision	Fr
BACO-SVM	۰,۹۵۲۹(۱)	۰,۹۶(۱)	۰,۹۵(۱)	۰,۸۸۸۸(۱)	۰,۵۶۲۵(۲)
GA-SVM	۰,۸۲۱۳(۲)	۰,۸۷۱۳(۳)	۰,۸۰۰۸(۳)	۰,٦٤٤٦(۲)	۰,٦۲۵(۱)
PSO-SVM	۰,۸۱۱۷(۳)	۰,۹۵(۲)	۰,۷۵(٤)	۰,٦۱۵۳(٤)	۰,٤۳۷(۳)
SVM	۰,۷٦٤٦(٤)	۰,۵٤۰٦(٤)	۰,۸۵٦۹(۲)	۰,٦۲۱٦(۳)	۰(٤)

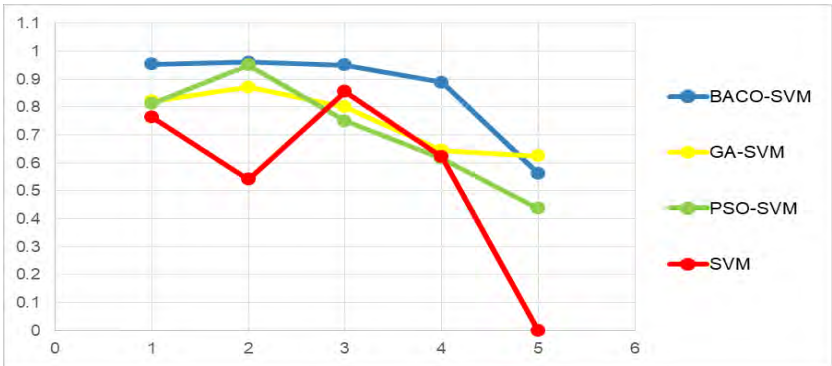
در هر ستون عدد اول نشان‌دهنده دقت روش موردنظر بر اساس معیارهای پژوهش می‌باشد و عدد دوم که در پرانتز قرار داده‌شده است نشانگر رتبه روش مزبور در میان دیگر روش‌ها می‌باشد. همان‌طور که در جدول (۳) ملاحظه می‌شود روش BACO-SVM در چهار معیار اول رتبه‌ی یک را کسب کرده و در معیار Fr رتبه‌ی دوم را کسب کرده. همین‌طور روش SVM هم در ۳ معیار رتبه‌ی آخر را کسب کرده و این نشان‌دهنده‌ی تأثیر انتخاب ویژگی می‌باشد.

در جدول (۴) مقایسه بر مبنای مجموع رتبه‌های موجود در جدول (۳) برای هر مدل انجام‌شده است. مقدار کمتر مجموع رتبه‌ها برای یک مدل، نشان‌دهنده نتایج میانگین بهتر در میان سایر مدل‌ها می‌باشد. اگرچه در بعضی موارد این کمیت دقت پایینی دارد اما استفاده از آن در آمار غیر پارامتری مرسوم است (کاشف و نظام‌آبادی‌پور،۱۳۹۳). با توجه به جدول (۴)، مدل BACO-SVM در رتبه اول، GA-SVM رتبه دوم، PSO-SVM رتبه سوم و مدل SVM هم‌رتبه‌ی چهارم را کسب کرده اند.

جدول ۳. مجموع رتبه‌های به‌دست‌آمده برای هر مدل

BACO-SVM	GA-SVM	PSO-SVM	SVM
٦ (۱)	۱۱(۲)	۱٦(۳)	۱۷(٤)

بنابراین مدل BACO-SVM به‌عنوان بهترین مدل انتخاب می‌شود؛ زیرا به دلیل ایجاد زیان‌های غیرمنتظره آتی و مطالبات مشکوک الوصول و معوق بانکی که از پیش‌بینی اشتباه مشتریان بدحساب ناشی می‌شود، پیش‌بینی درست این مشتریان به‌مراتب مهم‌تر از مشتریان خوش‌حساب است؛ لذا ما به دنبال مدلی هستیم که مشتریان بدحساب را به‌خوبی پیش‌بینی نماید. با توجه به اینکه دو معیار Accuracy و Recall این خواسته را برای ما پژوهشگر می‌سازند، لذا در بین مدل‌ها، مدل BACO-SVM دارای بالاترین مقادیر Accuracy و Recall می‌باشد که مقادیر آن به ترتیب ۹۵,۲۹٪ و ۹۶٪ می‌باشد. به‌عبارت‌دیگر این مدل، همهٔ مشتریان بدحساب را با دقت ۹۶ درصد پیش‌بینی می‌نماید و کلیهٔ مشتریان را با صحت ۹۵,۲۹ درصد طبقه‌بندی می‌کند. البته در معیار Fr رتبه‌ی دوم را کسب کرده. این معیار که مربوط به انتخاب تعداد ویژگی‌ها می‌باشد هر چه این معیار به یک نزدیک باشد بهتر است. با توجه به اینکه هر چه معیارها به یک نزدیک‌تر باشند نشان‌دهنده‌ی عملکرد بهتر مدل است بنابراین در شکل ۵ هم مشاهده می‌کنید که مقدار معیارها در مدل BACO-SVM نسبت به مدل‌های دیگر به یک نزدیک‌تر بوده و مدل BACO-SVM نسبت به مدل‌های دیگر بهتر عمل کرده است.



شکل ٤. مقایسه نتایج حاصل از مدل‌ها

بعد از طبقه‌بندی مشتریان به‌وسیله‌ی مدل BACO-SVM، مشتریان خوش‌حساب را با استفاده از روش VIKOR رتبه‌بندی می‌کنیم. برای رتبه‌بندی با روش VIKOR، از معیارهایی که توسط مدل BACO-SVM انتخاب‌شده‌اند استفاده می‌کنیم. درنتیجه در جدول (۵) نماد مشتریان خوش‌حساب و معیارهای انتخابی توسط مدل BACO-SVM ارائه‌شده‌اند. ۸ معیار توسط مدل BACO-SVM انتخاب شدند وزن معیارها با روش آنتروپی به‌دست‌آمده، بعد از محاسبه‌ی وزن‌ها، معیارهای منفی (مانند هزینه) که با افزایش آن مطلوبیت کاهش می‌یابد. معیارهای مثبت (مانند سود) که با افزایش آن مطلوبیت افزایش می‌یابد؛ که در اینجا معیارهای دوره متوسط وصول مطالبات و نسبت بدهی معیارهای منفی ما هستند یعنی مقدار آن‌ها هرچقدر کمتر باشد بهتر است.

جدول ٤. وزن معیارها با استفاده از روش آنتروپی

نسبت آنی	۰,۰۰۲۵	گردش دارایی ثابت	۰,۰۱۹
دوره متوسط وصول مطالبات	-۰,۰۰۵۳	نسبت بدهی	-۰,۰۰۳۵
ROE	۱,۰۱۷	سرمایه در گردش به‌کل دارایی‌ها	۰,۰۱۳۷
نسبت حقوق صاحبان سهام	۰,۰۰۰۸	نسبت پوشش بهره	۰,۰۰۷

در ادامه جهت قیاس عملکرد مدل پیشنهادی، نتایج حاصل با نتایج پژوهش‌های قبلی مقایسه شده‌اند. همان‌گونه که در جدول (۶) مشاهده می‌کنید مدل پیشنهادی پژوهش در هر سه معیار بهترین عملکرد را داشته.

جدول ۵. مقایسه نتایج پژوهش با پژوهش‌های مشابه

روش	منبع اطلاعات	تعداد ویژگی	Accuracy	Recall	Specificity
BACO-SVM	مدل پیشنهادی	١٦	0.9529 (1)	0.96 (1)	0.95 (1)
GA-SVM	هوانگ و همکاران (۲۰۰۷)	٢٤	0.739 (5)	0.3333 (5)	0.906 (4)
GS-LSSVM	ژو و همکاران (۲۰۰۹)	٢٤	0.7761 (3)	0.4863 (4)	0.9233 (2)
DS-LSSVM	یو و همکاران (۲۰۱۱)	٢٤	0.7693 (4)	0.5245 (3)	0.9067 (3)
DOE-LSSVM	یو و همکاران (۲۰۱۱)	٢٤	0.7846 (2)	0.5555 (2)	0.8923 (5)

نتیجه‌گیری و بحث

طراحی و اجرای سیستم مدیریت امور اعتباری نهادهای مالی با توجه به سیاست‌های کلان اعتباری و استراتژی‌های بلندمدت این مؤسسات می‌تواند نقش به سزایی در اتخاذ تصمیمات مناسب و اثربخش ایفا نماید. در این راستا، هدف محوری این پژوهش ارائه نمودن یک طبقه‌بندی کننده کارا برای ارزیابی ریسک اعتباری مشتریان بانکی با توجه به‌دقت خروجی حاصل از تقسیم نمودن مشتریان به دودسته خوش‌حساب و بدحساب، تعریف گردیده بود. با توجه به نتایج خروجی حاصل از دقت پیش‌بینی کلی و معیار Recall و مقایسه‌های صورت پذیرفته، می‌توان از رویکرد SVMبر پایه کلونی بهینه‌سازی مورچگان باینری که در مقایسه با روش‌های رقیب خود عملکرد بهتری از خود نشان داد، به‌عنوان یک طبقه‌بندی کننده ترکیبی پیشنهادی برای بررسی ریسک اعتباری متقاضیان تسهیلات از مؤسسات مالی استفاده نمود.

نتایج حاصل به‌خوبی بر اهمیت به‌کارگیری رویکرد انتخاب ویژگی که در آن سه روش BACO، PCA، GA به‌عنوان یک روش پوشاننده نمایندگان آن بودند، دلالت داشته و هر سه روش نسبت به طبقه‌بندی کننده SVM از دقت پیش‌بینی بالاتری برخوردار بودند برای پیشنهادهای آتی استفاده از سایر روش‌های انتخاب ویژگی همچون کرم شب‌تاب، الگوریتم استعماری و... پیشنهاد می‌شود. همچنین انتخاب ویژگی‌هایی همچون نسبت آنی، دوره متوسط وصول مطالبات، ROE، نسبت حقوق صاحبان سهام، گردش دارایی ثابت، نسبت بدهی، سرمایه در گردش به‌کل دارایی‌ها، نسبت پوشش بهره توسط روش انتخاب ویژگی حاکی از اهمیت بالا این ویژگی‌ها در ارزیابی ریسک اعتباری مشتریان برای داده‌های به‌کاررفته در این پژوهش می‌باشد. در پژوهش‌های آتی می‌توان شاخص‌های اقتصاد کلان در مدل به‌عنوان بخشی از داده‌های ورودی به‌عنوان یک عامل تأثیرگذار در ریسک اعتباری مشتریان استفاده شود.

از بین کلیهٔ مدل‌های موجود، بهترین مدل از منظر معیار Recall و Accuracy، مدل BACO-SVM بود. این مدل برتر که صحت پیش‌بینی مشتریان بدحساب آن ۹۶٪ و صحت پیش‌بینی کل مشتریان را ۹۵,۲۹٪ می‌داد. البته لازم به ذکر است که برای تصمیم‌گیرندگانی که برای آن‌ها برتری در کل معیارهای ارزیابی اهمیت دارد، باید گفت که در این پژوهش، بازهم مدل BACO-SVM در ۵ معیار برتر بود. برای پژوهش‌های آتی استفاده از روش‌های بهینه‌سازی پورتفولیوی اعتباری پس از طبقه‌بندی اعتباری مشتریان، به‌منظور تخصیص نسبت‌های مناسب از بودجه در نظر گرفته‌شده برای تسهیلات اعتباری پیشنهاد می‌شود. همین‌طور دخیل کردن اکسپوژر در زمان نکول و نرخ بازیافت به‌منظور ارائه تصویر روشن‌تری از زیان اعتباری و به تبع آن رتبه‌بندی اعتباری بهتر که نیازمند بستر اطلاعاتی مناسبی در بانک‌های کشورمان است، برای پژوهش‌های آتی پیشنهاد می‌شود.

منابع

- پویانفر، ا. فلاح‌پور، س. عزیزی، م. ۱۳۹۲. رویکرد حداقل مربعات ماشین بردار پشتیبان مبتنی بر الگوریتم ژنتیک جهت تخمین رتبه اعتباری مشتریان بانک‌ها. مجله مهندسی مالی و مدیریت اوراق بهادار، شماره ۱۷.
- تقوی، م.؛ نادعلی، ا. ۱۳۹۱. طبقه‌بندی متقاضیان تسهیلات اعتباری بانکی با استفاده از داده‌کاوی و منطق فازی. فصلنامه علمی پژوهشی مطالعات مدیریت صنعتی، شماره ۲۵, صفحات. ۸۵-۱۰۷.
- توحیدی، ح. نظام‌آبادی پور، ح. سریزدی، س. ۱۳۸۶. انتخاب ویژگی با استفاده از الگوریتم جمعیت مورچگان باینری. اولین کنگره فازی و سیستم‌های هوشمند، دانشگاه فردوسی مشهد، صص.۲۶۹-۲۷۵.
- حاتمی خواه ن.۱۳۹۲. بررسی روش‌های مبتنی بر انتخاب ویژگی. دانشگاه صنعتی مالک اشتر، مجتمع ICT 5
- سپهردوست، ح.؛ برجیسیان، ع. ۱۳۹۲. برآورد احتمال نکول تسهیلات پرداختی بانک با استفاده از رگرسیون لاجیت. فصلنامه علمی پژوهشی برنامه‌ریزی و بودجه، صفحات. ۳۱-۵۲.
- صفری، س. ابراهیمی، م.؛ شیخ، م. ۱۳۸۹. مدیریت ریسک اعتباری مشتریان حقوقی در بانک های تجاری با رویکرد تحلیل پوششی داده‌ها. پژوهش‌های مدیریت در ایران.
- عرب مازار، ع.؛ رویین تن، پ. ۱۳۸۵. ریسک اعتباری مشتریان بانکی مطالعه موردی بانک کشاورزی. دو فصلنامه علمی پژوهشی جستارهای اقتصادی.
- فلاح‌پور، س. راعی، ر.؛ هندیجانی زاده، م. ۱۳۹۳. رویکرد شبکه عصبی مصنوعی مبتنی بر کلونی زنبورعسل مصنوعی جهت تخمین رتبه اعتباری مشتریان بانک‌ها.
- فلاح‌پور، س. نوروزیان لکوان، ع.؛ هندیجانی زاده، ۱۳۹۶. کاربرد روش ترکیبی ماشین بردار پشتیبان و انتخاب ویژگی به‌منظور پیش‌بینی درماندگی مالی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران. نشریه‌ی تحقیقات مالی، دوره ۱۹ شماره ۱
- نیلی، م.؛ سبزواری، ح. ۱۳۸۷. برآورد و مقایسه مدل درجه‌بندی اعتباری لاجیت با روش تجزیه‌وتحلیل سلسله مراتبی (AHP). مجله علمی پژوهشی شریف، صفحات. ۱۰۵-۱۱۷.
- ALTMAN, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The journal of finance, 23, 589-609.
- ANGILELLA, S. & MAZZU, S. 2015. The financing of innovative SMEs: A multicriteria credit rating model. European Journal of Operational Research, 244, 540-554. Applications. 37,pp.4902-4909.
- ARAB MAZAR, AS. ROINEN, P. 2006. Credit Risk of Banking Customers A Case Study of Agricultural Bank. Two Economic Research Quarterly. (in Persian)
- AVCI. ENGIN. (2009). Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithm-support vector machines. Expert systems with applications, pp.1391-1402.
- BEAVER, W. (1967). Financial ratios as Predicators of Failure. Journal of Acounting Resarch.
- BELLOTTI, T. AND CROOK, J. 2009. Support vector machines for credit scoring and discovery of significant features. Expert Systems with Applications,36(2), pp.3302-3308.
- BOZ, O, 2002. Feature Subset Selection by Using Sorted Feature Relevance. ICMLA, pp.147-153.
- BURGES, C. (1998). tutorial on support vector machines for patternrecognition. Data Mining and Knowledge Discovery, 121-167.
- CALEND,AR. NADALI, A. 2012. Classification of Bank Credit Facility Applicants Using Data Mining and Fuzzy Logic. Journal of Industrial Management Studies, No. 25, Pages. 85-107. (in Persian)

- CHEN. L, LI. C, 2010. Combination of feature selection approaches with SVM in credit scoring. Expert Systems with Applications.37,4902-4909.CORTES, C. V.(1995). Support-vector networks. Machine Learning. pp.273-297.
- FALLAHPOUR, S. NOROUZIAN LAKVAN, AS. HENDIJANI ZADEH,2017. Application of the Support Vector Machine Combined Method and Feature Selection to Predict the Financial Misery of Companies Listed in Tehran Stock Exchange. Journal of Financial Research, Volume 19 Number 1. (in Persian)
- FALLAHPOUR, S. RAEI, R. HENDIJANI ZADEH, M. 2014. Artificial Bee Colony-Based Neural Network Approach for Estimating Credit Rating of Bank Clients. (in Persian)
- FISHER, R. (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics, pp.179-188.
- HARRIS, T. 2015. Credit scoring using the clustered support vector machine. Expert Systems with Applications, 42, 741-750.
- HATAMI-EH, No. 2013. Investigation of feature selection methods. Malik Ashtar University of Technology, ICT 5. (in Persian)
- HIROYASU, T. MIKI, M. ONO, Y. AND MINAMI, Y. 2000. Ant colony for continuous functions. The Science and Engineering, Doshisha University, 20.
- HUANG, Z. CHEN, H. HSU, C. J. CHEN, W. H. & WU, S. (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. Decision Support Systems, pp.543-558.
- LEAN, Y. XIAO, Y. SHOUYANG, W. A. & K.K. L. (2011). Credit risk evaluation using least squares SVM classifier with design of experiment for parameter selection. Expert Systems with Applications, 1 pp.5392-15399.
- LI, K. NISKANEN, J. KOLEHMAINEN, M. & NISKANEN, M.2016,Financial innovation: Credit default hybrid model for SME lending. Expert Systems with Applications, 61, pp.343-355.
- NILI, M.؛ SABZEVARI, H. 2008. Estimation and Comparison of the Logit Credit Rating Model with AHP. Sharif Scientific Journal, Pages. 105-117. (in Persian)
- ÖZBAKIR, L. BAYKASOĞLU, A. KULLUK, S. AND YAPICI, H.2009. TACO-miner: An ant colony based algorithm for rule extraction from trained neural networks. Expert Systems with Applications, 36(10), pp.12295-12305.
- OZTURK, H. NAMLI, E. & ERDAL, H. I. 2016. Modelling sovereign credit ratings: The accuracy of models in a heterogeneous sample. Economic Modelling, 54, pp.469-478.
- POUYANFAR, A. FALLAHPOUR, SA. ؛AZIZI, M.2013. Genetic Algorithm Based on Genetic Algorithm for Least Squares Approach to Estimating Credit Rating of Bank Clients. Journal of Financial Engineering and Securities Management, No. 17. (in Persian)
- SAFARI, S. EBRAHIMI, M. SHEIKH, M. 2010. Credit Risk Management of Legal Clients in Commercial Banks with Data Envelopment Analysis Approach. Management Research in Iran. (in Persian)
- TABAKHI. S, MORADI. P, AKHLAGHIAN. F, 2014. An unsupervised feature selection algorithm based on ant colony optimization.Engineering Applications of Artificial Intelligence.32,112-123. pp.335-345
- TAWHIDI, H. NIZAM ABADIPOUR, H. SERIZADI, Q. 2007. Feature selection using binary ant population algorithm. First Congress on Fuzzy and Intelligent Systems, Ferdowsi University of Mashhad, pp. 269-275. (in Persian)
- WHITEHORSE, H. ؛BERJISSIAN, A. 2013. Estimation of Bank Payment Facility Failure Using Logit Regression. Journal of Planning and Budgeting, Pages. 31-52. (in Persian)
- ZHANG, Z. GAO, G. & SHI, Y. 2014. Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzification and penalty factors. European Journal of Operational Research,237, pp.335-345.