

**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 Kashan¹, Fatemeh Garousi²

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

ارائه مدل ترکیبی الگوریتم مورچگان باینری و ماشین بردار پشتیبان (BACO-SVM) برای انتخاب ویژگی و طبقه‌بندی مشتریان بانکی به همراه مطالعه موردی^۱

علی حسین زاده کاشان^۱، فاطمه گروسی^۲

چکیده

یکی از مهم‌ترین مسائلی که همواره بانک‌ها و مؤسسات مالی با آن مواجه هستند، مسئله ریسک اعتباری می‌باشد. رقم قابل توجه مطالبات معوق بانک‌ها در سراسر جهان نشان‌دهنده اهمیت این موضوع و لزوم توجه به آن می‌باشد. از این رو تاکنون تلاش‌های بسیاری به منظور 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) به عنوان طبقه‌بندی کننده اصلی با یک روش انتخاب ویژگی به نام الگوریتم مورچگان باینری (BACO-SVM) ترکیب می‌گردد. به منظور نشان دادن اثربخشی روش پیشنهادی از داده‌های مربوط به ۸۵ شرکت از تسهیلات گیرندگان حقوقی یک بانک ایرانی در یک بازه 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 به تنهایی 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. در نتیجه با استفاده از روش BACO-SVM به طبقه‌بندی مشتریان به دو گروه مشتریان خوش حساب و بدحساب می‌پردازیم؛ و در نهایت جهت افزایش انعطاف‌پذیری در تصمیم‌گیری، مشتریان خوش حساب را با استفاده از روش VIKOR رتبه‌بندی می‌کنیم. این رتبه‌بندی منجر به آن می‌شود که قضاوت دقیق‌تری درباره‌ی وضعیت ریسک اعتباری متقاضیان خوش حساب صورت گیرد.

واژه‌های 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 می‌نماید (فلاح‌پور و همکاران، ۱۳۹۳). ریسک اعتباری در سال‌های اخیر در سراسر دنیا بسیار مورد توجه قرار گرفته 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 ارائه شده 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 پرداخته شده 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 می‌کند. 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 می‌دهد و تصمیم‌گیری هوشمندی بر اساس داده‌ها انجام می‌دهد از الگوریتم SVM، در هر جایی 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 فراابتکاری 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 خوش حساب و بدحساب نموده و دقت طبقه‌بندی بیان شده، استخراج گردد. به‌طور خلاصه هدف اصلی این مقاله مقایسه عملکرد روش‌های بیان‌شده و همچنین رتبه‌بندی مشتریان خوش حساب با استفاده از روش VIKOR می‌باشد که این رتبه‌بندی منجر به آن می‌شود که قضاوت دقیق‌تری درباره‌ی وضعیت ریسک اعتباری متقاضیان خوش حساب صورت گیرد.

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

مالی که در ادامه ذکر شده است، در پژوهش‌های نیلی و سبزواری (۱۳۸۷)، سپهر دوست و برجسیان (۱۳۹۲)، صفری، ابراهیمی و شیخ (۱۳۸۹) و... مورد استفاده قرار گرفته است و به آن‌ها توصیه شده 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 به کاررفته 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^۱، ۱۹۰۹). مطالعه فیشر^۲ (۱۹۳۶) اولین سیستم ارزیابی اعتبار است. در سال ۱۹۴۱ دوراند^۳ در مطالعه‌ای به تعیین ویژگی‌های مهم با دو روش پرسش از افراد خبره و روش آماری تحلیل تمایزی با تکیه بر نتایج فیشر پرداخت. او در مطالعه خود سیستم مبتنی بر امتیازدهی را برای طبقه‌بندی افراد متقاضی دریافت وام جهت خرید ماشین دست‌دوم، مورد استفاده قرار داد (دوراند، ۱۹۴۱). از جمله مطالعات دیگر می‌توان به مقاله بی‌ور^۴ (۱۹۶۷) که در زمینه برآورد موفقیت و شکست شرکت‌ها با استفاده از برخی شاخص‌های مالی است، اشاره نمود. آلتمن^۵ (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^۶ و همکاران (۲۰۰۴) از اولین پژوهشگرانی بودند که با این تکنیک که جزوه روش‌های یادگیری ماشینی است، اقدام به پیش‌بینی وضعیت اعتباری مشتریان بانک‌های ایالات متحده آمریکا 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^۷ (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

آنجیلا و ماتزو^۱ (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. از آنجایی که فرایند تخصیص اعتبار و تأمین مالی این شرکت‌ها، امری حیاتی است، آن‌ها در این مقاله سعی در ارائه یک مدل استوار^۲ به منظور تخصیص شرکت‌های کوچک و متوسط به کلاس‌های ریسک نمودند.

لی^۳ و همکاران (۲۰۱۶) به رتبه‌بندی اعتباری شرکت‌های با اندازه کوچک و متوسط^۴ 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^۵ (۲۰۱۵) در پژوهش خود تکنیک ماشین‌های بردار پشتیبان خوشه‌بندی شده (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، هزینه محاسبات را بالا می‌برد. بر طبق این گفته، مطالعه‌ی وی روش CSVM را با سایر ماشین‌های بردار پشتیبان غیرخطی مقایسه نموده است و نشان داده است که روش مذکور در طبقه‌بندی، درحالی‌که هزینه محاسباتی کمتری دارد، عملکرد خوبی از خود نشان داده است.

اوزترک و همکاران^۶ (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، از نظر دقت پیش‌بینی بهتر از مدل‌های سنتی آماری عمل کرده‌اند. بر طبق معیارهای در نظر گرفته شده 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 روئین تن (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 خوش حساب 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 مورد استفاده قرار گرفته 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 به کاررفته 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 سیگموئید 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 پوشش‌دهنده انتخاب ویژگی است و روش‌های آنالیز اجزای اساسی، زنجیره اطلاعات و رلیف که جزء روش‌های فیلتر کننده انتخاب ویژگی هستند، استفاده شده است. نتایج به‌دست آمده 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 سیگموئید 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 می شود و الگوریتم های متعددی جهت حل آن در زمان کوتاه تر ارائه شده است (ماریناکیس و همکاران، ۲۰۰۹). روش های انتخاب ویژگی بر اساس استفاده یا عدم استفاده از یک تابع طبقه بندی، به دودسته روش های فیلتر و پوشاننده تقسیم می شوند.

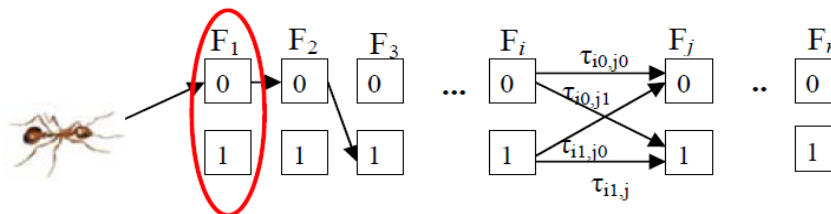
رویکرد فیلتر ویژگی ها را بدون استفاده از یک الگوریتم یادگیری ارزیابی می کند. در این رویکرد، خصوصیات ذاتی داده ها جهت ارزیابی و رتبه بندی ویژگی ها مورد استفاده قرار می گیرند (طباخی و همکاران، ۲۰۱۴).

رویکرد پوشاننده از یک الگوریتم یادگیری جهت ارزیابی هر زیرمجموعه از ویژگی ها استفاده می کند. در این رویکرد لازم است یک روش جستجو جهت یافتن زیرمجموعه بهینه به کار گرفته شود. رویکرد پوشاننده طبقه بندی را یک جعبه سیاه در نظر گرفته و زیرمجموعه ویژگی ها را بر اساس توان پیش بینی آن ها رتبه بندی می کند (حاتمی خواه ۱۳۹۲).

در سال های اخیر، الگوریتم های فرا ابتکاری مانند الگوریتم ژنتیک، بهینه سازی کلونی مورچگان و بهینه سازی ازدحام ذرات به شکل گسترده ای در زمینه انتخاب ویژگی مورد استفاده قرار گرفته اند. بیشتر الگوریتم های فرا ابتکاری انتخاب ویژگی، به علت برخی خصوصیات ذاتی خود از رویکرد پوشاننده استفاده می کنند (بوز^۱، ۲۰۰۲). به کارگیری رویکرد پوشاننده در مقایسه با رویکرد فیلتر نیازمند صرف زمان و هزینه بیشتری است اما دقت و کارایی این روش به مراتب بالاتر از روش فیلتر است. از این رو، در این پژوهش، انتخاب ویژگی بر اساس رویکرد پوشاننده و با استفاده از الگوریتم بهینه سازی کلونی مورچگان، الگوریتم جمعیت ذرات و الگوریتم ژنتیک، انجام می گیرد.

الگوریتم جمعیت مورچه باینری در انتخاب ویژگی

هیروياسو^۲ (۲۰۰۰) روشی برای حل مسائل فضای باینری با نام TACO ارائه و بعدها از آن برای حل مسائل مختلف بهینه سازی استفاده شد (اوزترک^۳ و همکاران، ۲۰۰۹). شکل (۱) چگونگی روش مذکور را به تصویر کشیده و همان طور که مشهود است، برای هر ویژگی یک گره صفر در رشته بالایی و یک گره یک در رشته پایینی در نظر گرفته شده است. هر مورچه در هر حرکت دو انتخاب (صفر یا یک) دارد که با توجه به تابع احتمالی که برای حرکت تعریف شده است، یکی را به عنوان موقعیت بعدی خود انتخاب می کند.



شکل ۱. رشته باینری به وسیله ی مسیری که مورچه از آن گذر کرده است

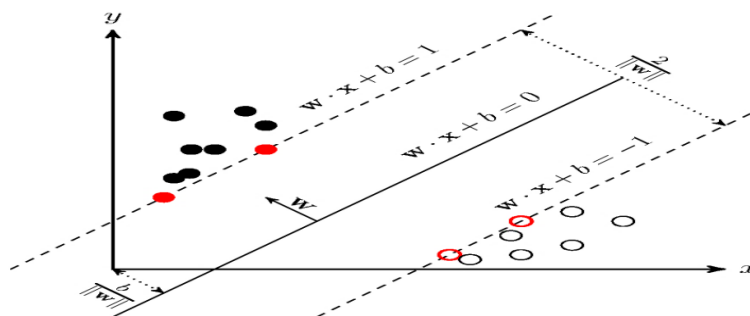
1 . Boz
2 . Hiroyasu
3 . Ozturk

هر مورچه بعد از اتمام سفرش، در مسیرهای عبوری خود مقداری فرمون به جا می گذارد. میزان فرمون به جامانده از هر مورچه با توجه به هزینه ای که مسیر طی شده به همراه داشته است، محاسبه می گردد. در قوانین حرکت نسخه باینری، مورچه ها فقط از اطلاعات فرمون مسیرها استفاده می کنند و وقتی یک مورچه مرحله تصمیم گیری را برای همه گره های رشته کامل کند به این معنی است که یک راه حل برای مسئله تولید شده است.

در روش TACO، تمام مورچه ها با توجه به شایستگی شان بر مسیرهایی که از آن ها گذشته اند، فرمون ترشح می کنند. با اقتباس از روش های موجود الگوریتم فضای گسسته ACO، سه روش به نام های 'ETACO'، 'RTACO' و 'BACO' برای بهبود TACO در (توحیدی و همکاران، ۱۳۸۶) پیشنهاد شد. این سه روش در دو مسئله طبقه بندی معنایی تصویر و بازشناسی ارقام دست نویس فارسی آزموده شدند. نتایج آزمایش ها نشان دادند روش BACO عملکرد بهتری نسبت به سایر روش های پیشنهادی و روش های مورد مقایسه دارد.

ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان در واقع یک طبقه بندی کننده دودویی است که دو کلاس را با استفاده از یک مرز خطی از هم جدا می کند. در تقسیم خطی داده ها هدف دستیابی به تابعی است که تعیین کننده ابر صفحه ای^۴ با بیشترین حاشیه می باشد. با حداکثر شدن حاشیه این ابر صفحه، تفکیک بین طبقات حداکثر می گردد. فرض کنید که $S = \{x_i, y_i\}$ یک ونه ی آموزشی است که از دو کلاس $y_i = \pm 1$ و هر کلاس از $i = 1, \dots, m$ ویژگی تشکیل شده. همان گونه که در شکل (۲) نشان داده شده است، خط $\langle wx_i + b \rangle = 0$ داده های موجود را در دو کلاس ± 1 طبقه بندی می کند. به این خط، ابر صفحه جداکننده گفته می شود. دو خط $\langle wx_i + b \rangle + 1$ و $\langle wx_i + b \rangle - 1$ ترتیب بیانگر مرز ناحیه ی دسته های $y = +1$ و $y = -1$ می باشند. به نزدیک تر داده های آموزشی به ابر صفحه های جداکننده، بردار پشتیبان^۵ نامیده می شوند.



شکل ۲. ابر صفحه ی 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 غیر خطی می توان مجموعه نقاط را از یکدیگر جدا نمود (آوسی انجین^۱، ۲۰۰۹). در ادامه به شرح هر یک از روش های فوق پرداخته شده است.

تفکیک خطی

در حالتی که داده ها را بتوان به صورت خطی از هم جدا کرد، ماشین بردار پشتیبان با در نظر گرفتن مجموعه داده های آموزشی، با استفاده از حل مسئله بهینه سازی زیر ابر صفحه بهینه با حاشیه حداکثر را پیدا می نماید:

$$\text{Min } \frac{1}{2\|W\|^2} \quad (1)$$

$$\text{St : } (wx_i + b) \geq -1 \quad i = 1, 2, \dots, m$$

که در آن $\|W\|$ نرم اقلیدسی^۲ است. بر اساس مدل فوق، کمینه شدن مقدار $\|W\|$ با توجه به محدودیت آن، منجر به حداکثر شدن پهنای صفحه می گردد.

حل این مسئله بهینه سازی دشوار است، لذا به منظور ساده تر نمودن حل آن، از ضرایب لاگراژ استفاده می گردد. سپس دوگان آن با توجه به شرط کروش کان-تاکر^۳، به صورت زیر در خواهد آمد:

$$\max W(\alpha) = \sum \alpha_i - \frac{1}{2} \sum \sum \alpha_i \alpha_j y_i y_j \langle x_i x_j \rangle \quad (2)$$

$$\text{Subject to: } \sum \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, \dots, N$$

با توجه به عدم وجود b در معادله فوق، از محدودیت های ابتدایی برای به دست آوردن آن استفاده می کنیم. پس از آنکه مقادیر α_i و b به دست آمد، می توان SVM را برای دسته بندی نمونه های جدید بکار برد. اگر x یک نمونه جدید باشد، دسته بندی آن به صورت زیر مشخص می گردد.

$$F(x) = \text{sign}[f(x, \alpha, b)] \quad (3)$$

که در آن $f(x, \alpha, b)$ به صورت زیر می باشد:

$$\begin{aligned} f(x, \alpha, b) &= wx + b \\ &= \sum \alpha_i y_i x_i x + b \\ &= \sum_{i \in SV} \alpha_i y_i x_i x + b \end{aligned} \quad (4)$$

همان گونه که قبلاً عنوان شد، SV بردارهای پشتیبان می باشند.

تفکیک غیر خطی

یک فرض بسیار مهم در SVM این است که داده ها به صورت خطی جدا پذیر باشند. در حالی که در عمل در بیشتر مواقع این فرض صحیح نیست. برای اولین بار کورتس و وپنیک در سال ۱۹۹۵ به منظور اینکه بتوانند در این حالت یک ابر صفحه ی بهینه را برای جداسازی ۲ کلاس به دست بیاورند، متغیرهای نامنفی $\varepsilon_i \geq 0$ را به عنوان مقادیر خطا برای هر بردار تعریف نمودند (کورتس^۴، ۱۹۹۵). مطابق روش ارائه شده صورت مسئله بهینه سازی تبدیل می شود به یافتن w به نحوی که معادله زیر حداقل گردد:

1 . Avci.Engine
2 . Euclidean norm
3 . Karush Kuhn-Tucker (KKT)
4 . Cortes

$$\text{Min } \frac{1}{2} \|w\|^2 + C \sum_i \varepsilon_i \quad (5)$$

$$\text{St: } y_i (< wx_i > + b) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0 \forall i$$

که در آن C پارامتر تنظیم کننده حاشیه می باشد که وظیفه ی آن برقراری تعادل بین حداکثر کردن حاشیه و حداقل کردن خطای دسته بندی بوده و همواره بزرگ تر از صفر است (کریستیانینی و تیلور^۱، ۲۰۰۰). اگر C عددی بزرگ انتخاب شود، توجه بیشتری به خطا معطوف می گردد. ماشین بردار پشتیبانی که به این صورت تعریف شده باشد را ماشین بردار پشتیبان حاشیه نرم^۲ می نامند (آبه^۳، ۲۰۰۵).

همچنین در حالتی که داده ها جداناپذیر بوده و همچنین کلاس ها دارای همپوشانی هستند، جدا کردن کلاسها توسط مرز خطی همواره با بروز خطا همراه می باشد. به منظور حل مشکل مزبور می توان ابتدا داده ها را با استفاده از یک تبدیل غیر خطی ϕ ، از فضای اولیه به فضایی با بعد بالاتر منتقل کرد با این هدف که در فضای جدید، کلاس ها تداخل کمتری با یکدیگر داشته باشند (بورگس^۴، ۱۹۹۸). پس از نگاشت داده ها به فضای بالاتر، با استفاده از معادلات قبل و جایگزین کردن x_i با $\phi(x_i)$ ، ابر صفحه بهینه به دست خواهد آمد.

انتقال از بعد پایین به بعدی بالاتر را می توان توسط توابع کرنل^۵ انجام داد. تعدادی از توابع کرنل موجود به شرح ذیل می باشند:

$$\text{کرنل خطی } K(x_i, x_j) = x_i^T x_j + c \quad (6)$$

$$\text{کرنل چندجمله ای } K(x_i, x_j) = (\alpha x_i^T x_j + c)^d \quad (7)$$

$$\text{کرنل تابع پایه شعاعی } K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / \sigma^2) \quad (8)$$

در معادلات فوق، c عبارت ثابت، d درجه چندجمله ای و سیگما (σ) پارامتر قابل تنظیم می باشد. کرنل تابع پایه شعاعی در دو مقوله نسبت به کرنل های دیگر دارای نقطه قوت بوده و در نتیجه عملکرد مطلوب تری در مسائل طبقه بندی اعتباری از خود نشان داده است. این کرنل به صورت غیر خطی نمونه ها را به ابعاد فضایی بالاتری نگاشت می نماید، بنابراین برخلاف کرنل خطی، در حالت هایی که روابط بین کلاس ها و ویژگی ها به صورت غیر خطی است، کاربرد خواهد داشت. دومین نقطه قوت، تعداد پارامترهایی است که بر پیچیدگی انتخاب مدل تأثیر گذار می باشند. کرنل چندجمله ای تعداد پارامترهای بیشتری نسبت به RBF دارا می باشد (لین^۶ و همکاران، ۲۰۱۱). از این رو در این پژوهش از کرنل تابع شعاعی استفاده شده است.

معیارهای ارزیابی مدل

مسلماً انجام هر کاری بدون ارزیابی آن ارزش چندانی ندارد و راهی به سمت بهبود و انتخاب بهترین راه پیدا نمی کند. در این پژوهش نیز قصد داریم تا پس از تغییر دادن برخی پارامترها، شرایط و... آن ها را با یکدیگر مقایسه نماییم تا بهترین شرایط جهت پیش بینی طبقه اعتباری مشتریان انتخاب گردد. برای مقایسه این حالات، باید ابتدا معیارهای ارزیابی مناسبی را تعریف نمود تا نتیجه مقایسه، برای تصمیم گیری قابل اتکا باشد. در پژوهش ژانگ، گائو و شی^۷ (۲۰۱۴)، 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 بد حساب باشد و مشتری بانکی هم طبق طبقه‌بندی تجربی بانک بد حساب باشد که به آن True Positive می‌گویند و به اختصار با TP نشان می‌دهند.

ب) نتیجه پیش‌بینی بد حساب باشد؛ اما مشتری بانکی طبق طبقه‌بندی تجربی، واقعاً خوش حساب 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 خوش حساب باشد و مشتری بانکی هم طبق طبقه‌بندی تجربی بانک، واقعاً خوش حساب باشد که به آن True Negative می‌گویند و به اختصار با TN نشان می‌دهند.

د) نتیجه پیش‌بینی خوش حساب باشد؛ حال آنکه طبق طبقه‌بندی تجربی بانک، مشتری واقعاً بد حساب باشد که به آن False Negative می‌گویند و آن را با FN نشان می‌دهند... حال در ادامه به تعریف و بیان رابطه معیارهای مورداستفاده خواهیم پرداخت.

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

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (9)$$

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

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

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

$$Specificity = \frac{TN}{TN + FP} \quad (11)$$

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

$$Precision = \frac{TP}{TP + FP} \quad (12)$$

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:

$$F_r = \frac{p - q}{p} \quad (13)$$

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 بازه ۵ ساله (۱۳۸۹-۱۳۹۳) به همراه ۱۶ ویژگی مربوط به هر یک از آن‌ها می‌باشد. ۱۶ آیتم موجود در ترازنامه و صورت حساب سود و زیان این شرکت‌ها می‌باشد. اهمیت بسیاری از نسبت‌های مالی که در ادامه ذکر شده است، در پژوهش‌های نیلی و سبزواری (۱۳۸۷)، سپهر دوست و برجیسین

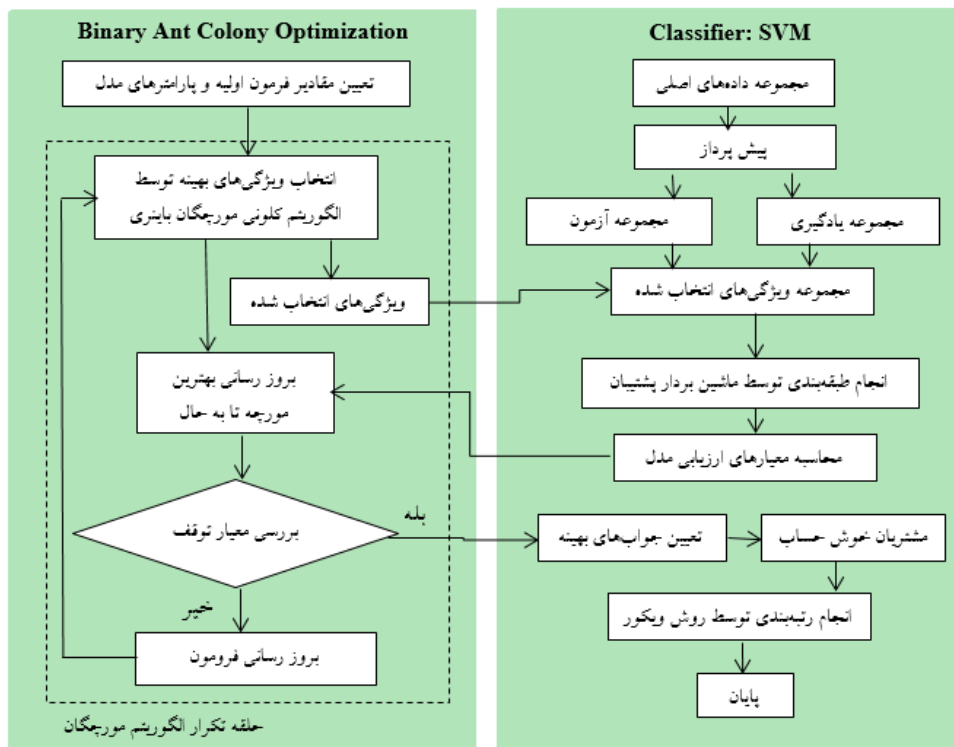
(۱۳۹۲)، صفری، ابراهیمی و شیخ (۱۳۸۹) و... مورد استفاده قرار گرفته است و به آن‌ها توصیه شده است. این ۱۶ آیتم عبارت‌اند از: حساب‌های دریافتی، موجودی مواد و کالا، دارایی‌های جاری، دارایی‌های ثابت، کل دارایی‌ها، بدهی‌های جاری، کل بدهی‌ها، حقوق صاحبان سهام و سود انباشته که از ترازنامه این شرکت‌ها جمع‌آوری شده 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 جمع‌آوری شده به منظور محاسبه نسبت‌های مالی مورد استفاده قرار می‌گیرد. پس از محاسبه مقادیر میانگین نسبت‌های مالی، باید این نسبت‌ها را به صورت مقادیر بین صفر و یک نرمال‌سازی نمود. رابطه مورد استفاده برای نرمال‌سازی داده‌ها در قالب رابطه (۱۸) ارائه شده است که در آن x_i ها، مقادیر مربوط به یک نسبت مالی خاص، برای ۸۵ شرکت می‌باشند از این رابطه در پژوهش لیو و موتودا (۱۹۹۸)^۱ استفاده گردیده است.

$$\text{Normalized value of } x_i = \frac{x_i - \min_i x_i}{\max_i x_i - \min_i x_i} \quad (14)$$

لازم به ذکر است برای جمع‌آوری این داده‌ها از دو پایگاه اینترنتی کدال^۲ و پردازش اطلاعات مالی ایران^۳ بهره گرفته شده است.

مدل پیشنهادی

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



شکل ۳. 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 به خوبی 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 به صورت 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. در واقع 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 به روزرسانی 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 بررسی شده و بر این اساس اجرای الگوریتم ادامه یافته و یا متوقف می‌گردد. شرح کامل هر گام از فاز اصلی در ادامه بیان شده.

انتخاب ویژگی با استفاده از الگوریتم مورچگان باینری

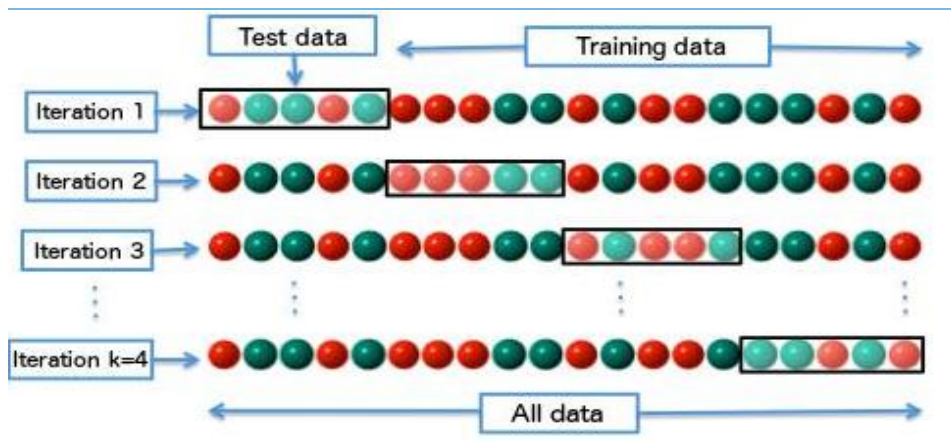
در حل مسئله انتخاب ویژگی با استفاده از الگوریتم جمعیت مورچگان باینری به ازای هر ویژگی، یک گره (بیت) صفر و یک گره (بیت) یک با جایگاه مشخص و یکسان در رشته باینری در نظر گرفته می‌شود. از این رو طول رشته بیت‌های یک و صفر در رشته‌های شکل (۱) برای حل یک مسئله خاص با تعداد ویژگی‌های آن مسئله برابر است. در این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

$$P_{ix.jy}^k = \frac{\tau_{ix.jy}}{\tau_{ix.j0} + \tau_{ix.j1}} \quad (15)$$

این رابطه، احتمال انتخاب بیت با مقدار $y \in \{0,1\}$ در گره‌ی بعدی (گره j ام) برای مورچه k ام که در زمان t در موقعیت $x \in \{0,1\}$ 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 فرمون بین مسیرهای متصل‌کننده گره‌های i ام و j ام به ترتیب روی یال‌های (۰ به ۰)، (۰ به ۱)، (۱ به ۰) و (۱ به ۱) است. لازم به یادآوری است که بنا به محدودیت‌های الگوریتم مورچه باینری همواره محدودیت $j=i+1$ برقرار است. یعنی این که هر مورچه در هر گره صرفاً می‌تواند به گره بعدی خود که از قبل مکان آن تعیین شده حرکت کرده و یکی از زیر گره‌های صفر و یک را برگزیند.

طراحی سیستم اعتبارسنجی

به منظور پیشگیری از وقوع انطباق بیش از حد و اینکه مدل ریاضی، به جای 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، پس از آن هر بار $(k-1)$ زیرمجموعه از داده‌ها را به عنوان داده‌های آموزشی و یک زیرمجموعه را به عنوان داده‌های تست در نظر می‌گیرد و این کار را k بار انجام می‌دهد به طوری که هر زیرمجموعه یک بار به عنوان داده تست قرار می‌گیرد. دقت پیش‌بینی برای آن زیرمجموعه محاسبه شده و دقت کلی مدل از میانگین آن‌ها به دست می‌آید. در این پژوهش مقدار k را ۵ در نظر گرفتیم. شکل (۴) نحوه‌ی کار اعتبارسنجی متقاطع k لایه‌ای را نشان می‌دهد.



شکل ۴. روش اعتبارسنجی متقاطع k لایه‌ای

انجام طبقه‌بندی توسط SVM

کرل تابع پایه شعاعی در دو مقوله نسبت به کرل‌های دیگر دارای نقطه قوت بوده و در نتیجه عملکرد مطلوب‌تری در مسائل طبقه‌بندی اعتباری از خود نشان داده است. این کرل به صورت غیرخطی نمونه‌ها را به ابعاد فضایی بالاتری نگاشت می‌نماید، بنابراین برخلاف کرل خطی، در حالت‌هایی که روابط بین کلاس‌ها ویژگی‌ها به صورت غیرخطی است، کاربرد خواهد داشت. دومین نقطه قوت، تعداد پارامترهایی است که بر پیچیدگی انتخاب مدل تأثیرگذار می‌باشند. کرل چندجمله‌ای تعداد پارامترهای بیشتری نسبت به RBF دارد (لین^۱ و همکاران، ۲۰۱۱). از این رو در این پژوهش از کرل تابع شعاعی استفاده شده است. مقدار پارامتر C و σ را به ترتیب ۱۰ و ۰٫۹ در نظر می‌گیریم. به وسیله‌ی الگوریتم SVM مشتریان به دودسته‌ی خوش حساب و بد حساب طبقه‌بندی می‌شوند.

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

محاسبه تابع برازش

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

نشان دهنده نسبتی از مشتریان بدحساب است که به درستی طبقه بندی شده اند؛ زیرا این معیار بسیار حائز اهمیت می باشد چرا که هزینه ای که از عدم تشخیص صحیح مشتریان بدحساب مؤسسات اعتباری را تهدید می نماید، بسیار بیشتر از هزینه عدم تشخیص صحیح مشتریان خوش حساب می باشد. این هزینه شامل از دست دادن اصل و فرع تسهیلات به همراه هزینه پیگیری مطالبات معوق می باشد. با توجه به مفهوم معیار مذکور، اختلاف آن از عدد یک را می توان به عنوان ریسک اعتباری روش مورد استفاده در نظر گرفت. منظور از ریسک اعتباری اعطای اعتباری به مشتریان بدحساب می باشد که مدل توانایی تشخیص درست آن ها را نداشته است. معیار دیگری که برای ما مهم است معیار F_1 می باشد که برای بررسی تعداد ویژگی های بعد از انتخاب ویژگی استفاده می شود. رچه مقدار آن به ۱ نزدیک تر باشد، کاهش تعداد ویژگی ها بیشتر بوده و مطلوب تر است. این معیار طبق رابطه (۱۷) محاسبه می شود.

بررسی معیار توقف

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

به روزرسانی فرمون

در پایان هر تکرار از الگوریتم، فرمون ابتدا تبخیر و سپس به روزرسانی می شود. رابطه مورد استفاده جهت تبخیر فرمون به صورت زیر است:

با توجه به اینکه برای محاسبه ی دقت مدل از اعتبار سنجی چندلایه ای استفاده کرده بودیم که مجموعه داده ها به K زیرمجموعه تقسیم شده بودند؛ بنابراین $CVREC$ میانگین حساسیت (درصد مشتریان بدحسابی که به تی دسته بندی شده اند) K زیرمجموعه است.

$$CVREC = \frac{\sum_i Recall_i}{K} \quad i = 1.2. \dots .K \quad (16)$$

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

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

$$\tau_{ij} = (1 - \rho)\tau_{ij} + \sum_{k=1}^{Number\ of\ Ants} \Delta\tau_{ij}^k + e\Delta\tau_{ij}^{Best} \quad (17)$$

$$\Delta\tau_{ij}^k = \begin{cases} CVREC^K & \text{if ant } k \text{ use edge } (i, j) \text{ in tour} \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

این به روزرسانی که با افزایش مقدار فرمون ویژگی های منتخب همراه است منجر به آن می شود که در تکرارهای بعدی نیز احتمال انتخاب این ویژگی ها افزایش یابد. به این ترتیب الگوریتم پس از چندین تکرار به یک راه حل واحد همگرا می گردد.

تعیین جواب بهینه

در هر تکرار از الگوریتم فوق از بین راه حل های ساخته شده بهترین راه حل بر اساس معیار حساسیت تعیین و ذخیره می گردد. پس از پایان تمام تکرارها مقایسه و جواب بهینه تعیین می شود.

نتایج حاصل از مدل BACO-SVM

پارامترهای مهم در نسخه ی باینری مورچگان، تعداد مورچگان و نرخ تبخیر می باشد. انتخاب مقادیر ۲۰، ۳۰، ۴۰ برای تعداد مورچگان و انتخاب مقادیرهای ۰،۲، ۰،۳، ۰،۵ برای نرخ تبخیر، پارامترهای SVM، $C=10$ و $\sigma = 0.9$ قرار می دهیم. مدل BACO-SVM باید ۹ بار ران شود. ما این مدل را ۹ بار با ۱۰۰ تکرار ران می کنیم. همان طور که در جدول (۲) نشان داده شده است ما با تغییر تعداد مورچه ها و نرخ تبخیر، مدل را ۹ بار اجرا می کنیم که در جدول (۲) شماره مدل بیانگر این است که هر مدل با چه پارامترهایی اجرا شده است که مقدار پارامترهای هر مدل در جدول (۱) آمده است.

جدول ۱. تنظیم پارامترهای الگوریتم مورچگان

نام مدل/پارامترها	تعداد مورچه ها	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
۴	۳۰	۰،۲	۱۰	۰،۹
۵	۳۰	۰،۳	۱۰	۰،۹
۶	۳۰	۰،۵	۱۰	۰،۹
۷	۴۰	۰،۲	۱۰	۰،۹
۸	۴۰	۰،۳	۱۰	۰،۹
۹	۴۰	۰،۵	۱۰	۰،۹

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

مدل - معیارها	Precision		Specificity		Recall		Accuracy		Fr
	Training	Test	Training	Test	Training	Test	Training	Test	
۱	۰،۹۴	۰،۸۶	۰،۹۸	۰،۹۰	۱،۰۰	۰،۷۶	۰،۹۸	۰،۸۹	۰،۲۵
۲	۰،۸۵	۰،۸۴	۰،۹۴	۰،۹۳	۰،۹۸	۰،۸۴	۰،۹۵	۰،۹۱	۰،۳۱
۳	۰،۸۱	۰،۸۴	۰،۹۰	۰،۹۳	۰،۹۷	۰،۸۴	۰،۹۲	۰،۹۱	۰،۵۶
۴	۰،۸۶	۰،۸۵	۰،۹۴	۰،۹۳	۰،۹۷	۰،۹۲	۰،۹۵	۰،۹۳	۰،۲۵
۵	۰،۸۱	۰،۸۱	۰،۹۰	۰،۹۲	۰،۹۶	۰،۹۲	۰،۹۲	۰،۹۲	۰،۳۱
۶	۰،۷۹	۰،۷۹	۰،۸۹	۰،۹۰	۰،۹۹	۰،۹۲	۰،۹۲	۰،۹۱	۰،۳۸
۷	۰،۷۸	۰،۸۵	۰،۸۸	۰،۹۳	۰،۹۸	۰،۸۸	۰،۹۱	۰،۹۲	۰،۵۶
۸	۰،۷۹	۰،۷۷	۰،۸۹	۰،۸۸	۰،۹۷	۰،۹۶	۰،۹۱	۰،۹۰	۰،۳۱
۹	۰،۷۹	۰،۸۹	۰،۸۹	۰،۹۵	۰،۹۸	۰،۹۶	۰،۹۲	۰،۹۵	۰،۵۶

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

این ۹ مدل را با پارامترهای مختلفی که برای هر کدام 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. در نتیجه 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. در واقع مدل ویکور از طریق ارزیابی گزینه‌ها بر اساس معیارها، گزینه‌ها را اولویت‌بندی 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، ارزیابی شده 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 تصمیم‌گیری، 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:

$$\begin{aligned} f^+ &= \text{Max } f_{ij} \\ f^- &= \text{Min } f_{ij} \end{aligned} \quad (19)$$

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 ایده آل و مقدار تأسف (R) بیانگر حداکثر ضرر گزینه i ام از دوری نقطه ایده آل است.

$$S_j = \sum_{i=1}^n w_i \cdot \frac{f_i^* - f_{ij}}{f_i^* - f_i^-} \quad (20)$$

$$R_j = \max_i w_i \cdot \frac{f_i^* - f_{ij}}{f_i^* - f_i^-} \quad (21)$$

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

$$Q_i = v \left[\frac{S_i - S^*}{S^- - S^*} \right] + (1 - v) \left[\frac{R_i - R^*}{R^- - R^*} \right] \quad (22)$$

$$S^- = \text{Max } S_i, S^* = \text{Min } S_i \quad (23)$$

$$R^- = \text{Max } R_i, R^* = \text{Min } R_i$$

v مقداری بین صفر و یک است و وزن برای استراتژی ماکسیمم مطلوبیت گروهی است. در اینجا v با مقدار ۰٫۵ در نظر گرفته شده 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 به شرط 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.

$$Q(A_2) - Q(A_1) \geq \frac{1}{m-1} \quad (24)$$

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، پرداخته شده است. جهت نشان دادن تأثیر الگوریتم انتخاب ویژگی در افزایش دقت مدل، SVM بدون الگوریتم انتخاب ویژگی اجرا شده است. میانگین نتایج در جدول (۳) ارائه شده 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)	۰,۸۸۸(۱)	۰,۵۶۲۵(۲)
GA-SVM	۰,۸۲۱۳(۲)	۰,۸۷۱۳(۳)	۰,۸۰۰۸(۳)	۰,۶۴۴۶(۲)	۰,۶۲۵(۱)
PSO-SVM	۰,۸۱۱۷(۳)	۰,۹۵(۲)	۰,۷۵(۴)	۰,۶۱۵۳(۴)	۰,۴۳۷(۳)
SVM	۰,۷۶۴۶(۴)	۰,۵۴۰۶(۴)	۰,۸۵۶۹(۲)	۰,۶۲۱۶(۳)	۰(۴)

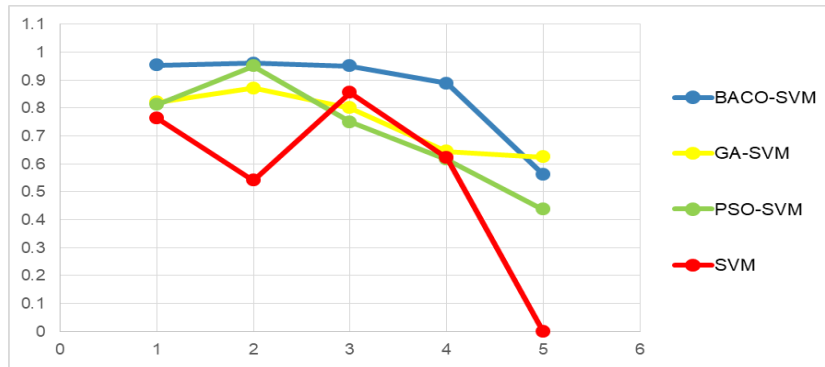
در هر ستون عدد اول نشان‌دهنده دقت روش موردنظر بر اساس معیارهای پژوهش می‌باشد و عدد دوم که در پرانتز قرار داده شده است نشانگر رتبه روش مزبور در میان دیگر روش‌ها می‌باشد. همان‌طور که در جدول (۳) ملاحظه می‌شود روش BACO-SVM در چهار معیار اول رتبه‌ی یک را کسب کرده و در معیار Fr رتبه‌ی دوم را کسب کرده. همین‌طور 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 انجام شده 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 کرده‌اند.

جدول ۳. مجموع رتبه‌های به‌دست‌آمده برای هر مدل

SVM	PSO-SVM	GA-SVM	BACO-SVM
۱۷(۴)	۱۶(۳)	۱۱(۲)	۶(۱)

بنابراین مدل BACO-SVM به‌عنوان بهترین مدل انتخاب می‌شود؛ زیرا به دلیل ایجاد زبان‌های 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 خوش حساب 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، همه مشتریان بدحساب را با دقت ۹۶ درصد پیش‌بینی می‌نماید و کلیه 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 ۴. مقایسه نتایج حاصل از مدل‌ها

بعد از طبقه‌بندی مشتریان به وسیله‌ی مدل BACO-SVM، مشتریان خوش حساب را با استفاده از روش VIKOR رتبه‌بندی می‌کنیم. برای رتبه‌بندی با روش VIKOR، از معیارهایی که توسط مدل BACO-SVM انتخاب شده‌اند استفاده می‌کنیم. در نتیجه در جدول (۵) نماد مشتریان خوش حساب و معیارهای انتخابی توسط مدل BACO-SVM ارائه شده‌اند. ۸ معیار توسط مدل BACO-SVM انتخاب شدند وزن معیارها با روش آنترپی به دست آمده، 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 هرچه قدر کمتر باشد بهتر است.

جدول ۴. وزن معیارها با استفاده از روش آنترپی

۰,۰۱۹	گردش دارایی ثابت	۰,۰۲۵	نسبت آنی
-۰,۰۰۳۵	نسبت بدهی	-۰,۰۰۵۳	دوره متوسط وصول مطالبات
۰,۰۱۳۷	سرمایه در گردش به کل دارایی‌ها	۱,۰۱۷	ROE
۰,۰۰۷	نسبت پوشش بهره	۰,۰۰۰۸	نسبت حقوق صاحبان سهام

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

Specificity	Recall	Accuracy	تعداد ویژگی	منبع اطلاعات	روش
0.95 (1)	0.96 (1)	0.9529 (1)	۱۶	مدل پیشنهادی	BACO-SVM
0.906 (4)	0.3333 (5)	0.739 (5)	۲۴	هوانگ و همکاران (۲۰۰۷)	GA-SVM
0.9233 (2)	0.4863 (4)	0.7761 (3)	۲۴	ژو و همکاران (۲۰۰۹)	GS-LSSVM
0.9067 (3)	0.5245 (3)	0.7693 (4)	۲۴	یو و همکاران (۲۰۱۱)	DS-LSSVM
0.8923 (5)	0.5555 (2)	0.7846 (2)	۲۴	یو و همکاران (۲۰۱۱)	DOE-LSSVM

نتیجه گیری و بحث

طراحی و اجرای سیستم مدیریت امور اعتباری نهادهای مالی با توجه به سیاست‌های کلان اعتباری و استراتژی‌های بلندمدت این مؤسسات می‌تواند نقش به‌سزایی 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 به دقت خروجی حاصل از تقسیم نمودن مشتریان به دودسته خوش حساب 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 SVM بر 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 GA, PCA, BACO 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 پیشنهادی‌های آتی استفاده از سایر روش‌های انتخاب ویژگی همچون کرم شب‌تاب، الگوریتم استعماری و... پیشنهاد می‌شود. همچنین انتخاب ویژگی‌هایی همچون نسبت آبی، 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 کلیه 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 تصمیم‌گیرندگان 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 اکسپوزر 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. فلاح پور، 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 داده کاوی و منطق فازی. فصلنامه علمی پژوهشی مطالعات مدیریت صنعتی، شماره ۲۵، صفحات. ۸۵-۱۰۷.
- توحیدی، ح. نظام‌آبادی پور، ح. سرزیدی، 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
- سپهر دوست، ح.؛ برجسیان، 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 بانک‌های 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.
- فلاح پور، س. راعی، ر.؛ هندیجانی زاده، م. ۱۳۹۳. رویکرد شبکه عصبی مصنوعی مبتنی بر کلونی زنبور عسل مصنوعی جهت تخمین رتبه اعتباری مشتریان بانک‌ها.
- فلاح پور، س. نوروزیان لکوان، ع.؛ هندیجانی زاده، ۱۳۹۶. کاربرد روش ترکیبی ماشین بردار پشتیبان و انتخاب ویژگی به منظور پیش‌بینی درماندگی مالی شرکت‌های پذیرفته شده 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 تجزیه و تحلیل 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 Accounting 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.
- BURGESS, C. (1998). tutorial on support vector machines for pattern recognition. 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.
- FALLAHOPOUR, 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)
- FALLAHOPOUR, 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. FALLAHOPOUR, 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.