بررسی وارزیابی تکنیک های داده کاوی در رتبه بندی اعتباری مشتریان بانک ها

محمد رضا صفری
دانشجوی کارشناسی علوم اقتصاد دانشگاه کردستان
safari.mohammadreza2017@gmail.com

زهره صفری
دانشجوی ارشد فناوری اطلاعات گرایش شبکه های کامپیوتری موسسه آموزش عالی صائب ابهر
safari.z.602@gmail.com

چکیده

بانک ها به عنوان بخش اصلی نظام مالی ، نقش مهمی را در تامین مالی بخش های مختلف اقتصادی بر عهده دارند. در راستای ایفای این نقش ، بانک ها با ریسک های متفاوتی روبرو هستند که یکی از عمده ترین آنها ریسک اعتباری است. ارزیابی ریسک اعتباری، یکی از مسائل مهم و پرچالش در زمینه تحلیل های مالی به شمار می آید. زیرا از این طریق می توان از وارد آمدن خسارات کلان که پیامد تصمیمات نادرست واگذاری اعتباروام به متقاضیان است، تا اندازه زیادی اجتناب کرد. علیرغم ابداعات و نوآوری های موجود در نظام بانکی هنوز هم ریسک اعتباری به عنوان دلیل عمده عدم موفقیت بانک ها محسوب میشود. یکی از مهمترین روشهای مدیریت ریسک اعتباری استفاده از سیستم های طبقه بندی برای کنترل ریسک انواع وام ها است.یکی از روش های اعتبارسنجی مشتریان استفاده از الگوریتم های داده کاوی است.در این مقاله از الگوریتم های پر کاربرد داده کاوی نظیر ماشین های بردار پشتیبان ،درختان تصمیم،شبکه های عصبی ، شبکه های بیزین، رگرسیون لجستیک وK- نزدیکترین همسایه ، برای اعتبارسنجی مشتریان بانک استفاده شده است.این الگوریتم ها برروی داده های 4117مشتری اعمال گردید و نتایج بدست آمده از این تحقیق نشان داد که مدل ایجاد شده توسط الگوریتم SVM نسبت به دیگر مدل های ایجاد شده ضریب اطمینان بالاتری (۷۵,۱۳٪)بوده است.همچنین مدل ایجاد شده توسط الگوریتم C&R Tree نیز بعد ازمدل SVMدارای ضریب اطمینان بالاتری (۷۴,۸۲٪)در مقایسه با مدل های ایجاد شده توسط دیگر الگوریتم های استفاده شده در این مقاله است.

واژگان کلیدی: ریسک اعتباری،داده کاوی، رتبه بندی مشتریان، ارزیابی مدل

Review and evaluation of data mining techniques in the credit rating of the bank's customer

Mohammad Reza Safari
Zohre Safari

Abstract

Bank as the main part of the financial system, an important role in financing various economic sectors are responsible. In order to play this role, banks are faced with different risk Credit risk is one of the major ones. Credit risk assessment, one of the most important and challenging issues in the field of financial analysis is considered. Because of the massive damage that could be the result of poor decisions Credit loan transfer to the applicants, to a large extent be avoided. Despite innovations in the banking system still credit risk as the main reason is the failure of banks. One of the most important credit risk management methods of classification systems for controlling the risk of the loans one way to authenticate customer using data mining algorithms. In this paper, the most commonly used data mining algorithms such as support vector machines, decision trees, neural networks, Bayesian networks, logistic regression and K-nearest neighbor, bank customers used to validate the algorithm is applied on customer data, 4117 . the results of this study showed that the model created by the algorithm SVM than other models created highest confidence level (% 75.13). also the model created by the algorithm C & R Tree after model SVM has a highest confidence level (74.82%) compared to models created by other algorithms used In this article

Key words: credit risk, data mining, customer ratings, evaluation model

مقدمه

فرایند جهانی شدن و به تبع آن تشدید رقابت موسسات مالی و اقتصادی در بازارهای محلی و جهانی، روز به روز ضرورت تقویت و اصلاحات سیستمی را در بنگاه های مالی و سرمایه ای در سرتاسر دنیا افزایش می دهد. بانک ها نیز از این امر مستثنا نیستند(فرد حریری، ۱۳۸۷).ارزیابی ریسک اعتباری[1]، یکی از مسائل مهم و پرچالش در زمینه تحلیل های مالی به شمار می آید. زیرا از این طریق می توان از وارد آمدن خسارات کلان که پیامد تصمیمات نادرست واگذاری اعتباروام به متقاضیان است، تا اندازه زیادی اجتناب کرد(Yu et al,2008) .علیرغم ابداعات و نوآوری های موجود در نظام بانکی هنوز هم ریسک اعتباری به عنوان دلیل عمده عدم موفقیت بانک ها محسوب می شود که بنا به گفته کارشناسان علت آن این است که معمولاً 80درصد ترازنامه یک بانک تسهیلات اعطایی به مشتریان است (Greuning et al,2009 . با مدیریت این ریسک، بانک قادر خواهد بود تا تعهدات خود را انجام دهد و برای سهامداران ارزش ایجاد کند. یکی از ابزارهای مدیریت ریسک اعتباری، سیستم رتبه بندی و امتیازدهی اعتباری می باشد که یک سیستم تحلیلی مبتنی بر آمار است و به منظور تعیین درجه ریسک متقاضیان تسهیلات اعتباری به کار می رود و به وام دهندگان در شناسایی مشتریان خوش حساب و بد حساب کمک می کند و از این طریق می توان ریسک اعتباری را مدیریت کرد.

نظام امتیازدهی اعتباری، برای اولین بار در دهه 1950تدوین شد. در واقع، ایده ایجاد تمایز بین گروه های موجود در یک جامعه بر اساس مشخصه های ذکر شده برای اعضای آن، از مقاله فیشر[2] ۱۹۳۶ نشأت میگیرد. دانهام[3] در سال ،1938اولین کسی بود که سیستمی را برای ارزیابی متقاضیان تسهیلات ارائه کرد. دوراند[4] در سال ،1941به شناسایی عوامل مهم از دید وام دهندگان پرداخت. او اولین کسی بود که دیدگاه آماری را مد نظرقرار داد و از مدل تحلیل تمایزی که بر نتایج فیشرمتمرکز بود استفاده کرد. بدین وسیله عملاً وی انگیزه توسعه چارچوب تئوریک که بدان وسیله بتوان اهمیت هرمعیار مشخص را تعیین کرد به وجود آورد. بنابراین میتوان دوراند را پایه گذار سیستم اعتباردهی امروزی دانست. حجم زیاد تقاضای اعتباری، منجر به استفاده ازمدلهای اعتبارسنجی در مؤسسات مالی شد (Eskandari,2006). مدل های مختلفی برای رتبه بندی و امتیازدهی اعتباری ارائه شده است که مدل هایی همچون مدلهای رگرسیون لجستیک، تحلیل ممیزی، شبکه های عصبی و غیره نمونه هایی از این روشها هستند (Singh et al,2013).به دلیل اهمیت اعتبار سنجی مشتریان بانک ها و موسسات اعتباری پژوهش های گسترده ای در این زمینه صورت گرفته است که برخی از این تحقیقات انجام شده در جداول (۲و۱) قابل مشاهده است.

جدول(۱) روش های امتیاز دهی اعتباری مورد استفاده در پژوهش های انجام یافته در خارج از کشور

ردیف	مدل امتیاز دهی اعتباری	پژوهشگر ،تاریخ
۱	رگرسیون لجستیک	(Abdou et al ,2008)
۲	طبقه بندی ورگرسیون	(Zekic et al 2004)
۳	K -نزدیکترین همسایه	(Bellotti and Crook,2009)
۴	شبکه عصبی مصنوعی	(Chuang and Lin 2009)
۵	ترکیب تحلیل ممیزی وشبکه های عصبی	(Lee et al., 2002)
۶	ترکیب رگرسیون لجستیک وشبکه های عصبی	(Mileris, 2011);(Lin, 2009)
۷	درخت تصمیم	(Yanping et al., 2012); (Lee et al., 2006)
۸	ترکیب در خت تصمیم وشبکه عصبی	(Kabari and Nwachukwu, 2013)
۹	الگوریتم ژنتیک	(Dehuri et al 2008)
۱۰	ماشین بردار پشتیبان	(Huang et al 2007)
۱۱	روش های ترکیبی ماشین بردار پشتیبان	(Yu et al., 2008)

[1] Credit Risk
[2] Fisher
[3] Dunham
[4] Durand

۱۲	تحلیل پوششی داده ها	(Min and Lee, 2008);(Emel et al, 2003)

جدول (۲)روشهای امتیازدهی اعتباری مورد استفاده در پژوهشهای انجام یافته در داخل کشور

ردیف	مدل امتیاز دهی اعتباری	پژوهشگر ،تاریخ
۱	رگرسیون لجستیگ	(میرزایی و همکاران، ۱۳۹۰)؛(تهرانی و فلاح شمس، ۱۳۸۴)
۲	شبکه های عصبی	(قدسی پور و همکاران، ۱۳۹۱)
۳	تحلیل پوششی داده ها	(صفری و همکاران، ۱۳۸۹)
۴	تحلیل ممیزی	(کرمی، ۱۳۸۹)؛ (شریعت پناهی و هاشمی، ۱۳۸۸)
۵	ماشین بردار	(راعی و فلاح پور، ۱۳۸۷)

تکنیک های داده کاوی[5] همچون طبقه بندی می توانند با ارائه یک الگو یا مدل به کشف دانش پنهان در حجم زیادی از داده های تراکنش های اعتباری مشتریان بانک ها کمک کنند.در این پژوهش با توجه به ضرورت رتبه بندی مشتریان بانک ها وموسسات مالی واعتباری سعی بر آن است که با استفاده از الگوریتم های پر کاربرد داده کاوی[6] مشتریان مورد مطالعه را اعتبارسنجی نموده و مهم ترین وموثرترین ویژگی ها را در تعیین خوش حساب وبد حساب بودن مشتریان شناسایی نماییم. سپس مدل های ایجاد شده توسط الگوریتم های مختلف داده کاوی مورد ارزیابی قرار گرفته و مدل یا مدلی هایی که دارای بالاترین ضریب اطمینان می باشند به عنوان بهترین مدل ها برای رتبه بندی مشتریان بانک ها وموسسات مالی واعتباری انتخاب وارائه شود.

ریسک های مقابل صنعت بانکداری

در واقع، بانک ها با ریسک های متعددی روبرو می باشند که بر سودآوری آنها در طول فعالیت تجاری شان اثر می گذارند. مدیریت این ریسک ها همیشه یکی از بخش های عمده و اصلی مدیریت بانکی را تشکیل می داده است. بنابراین بانک ها را اینگونه نیز می توان تعریف نمود: مؤسساتی با ویژگی متعادل نمودن سود و زیان فرصت های مختلف با هدف به حداکثر رساندن سود دهی. بانک ها با ارایه قراردادهای مالی با ویژگی های مناسب و مطلوب سود و بازپرداخت به سپرده گذاران، در واقع مشوق پس انداز کردن پول در بانک ها می گردند، بانک ها از طریق معین نمودن میزان تقاضای اعتبار، سرمایه را به سمت کاربردهای سود بخش وتولیدی اجتماعی سوق می دهند.محققان منابع اصلی ریسک های بانکی را به انواع مختلف تقسیم بندی کرده اند.پایل[7] این منا بع را به ۴ دسته تقسیم نمود:

۱-ریسک اعتباری ۲- ریسک بازاری ۳-ریسک اعتباری ۴- ریسک اجرایی یا عملکردی

ریسک اعتباری ،تغییر در ارزش سرمایه به دلیل تغییرات در توانایی شرکا ء برای انجام تعهدات قرار داد هایشان است.(این تعریف بر تئوری مدل ارزش سرمایه که توسط مرتون [8] ارائه شده، اثر می گذارد).ریسک بازاری، تغییر در ارزش سرمایه به دلیل تغییر در فاکتورهای زیربنایی و اصلی اقتصاد نظیر نرخ سود، نرخ ارز، قیمت محصولات و سهام است. ریسک کاربردی ناشی از هزینه هایی است که بانک به خاطر اشتباه در انجام معاملاتی مثل شکست در پرداخت و تسویه، متحمل می شو د. ریسک اجرایی، زیان های ناشی از عدم نظارت صحیح بر کارمندان و یا زیان های حاصل از عدم استفاده از روش های مناسب را در بر می گیرد. (عریانی وعیسی زاده ،۱۳۸۴)

[6] Data Mining
[7] Pyle
[8] Merton

ریسک اعتباری

ریسک اعتباری از آن جهت در نهاد های پولی واعتباری حائز اهمیت است که منابع بکار گرفته شده برای تسهیلات در رخصت بدهی نهاد پولی (بانک) به سهامداران آن نهاد می باشد.که درصورت عدم جریان هم توان اعتبار دهی قدرت برگشت سرمایه ،سهامدارن را تضعیف می کند.از مهمترین رویکرد جهت کاهش خسارت های ریسک اعتباری ،شناسایی سنجش درجه وطبقه بندی اعتباری مشتریان می باشد که اصطلاحا به آن اعتبارسنجی گویند.بر آن اساس هرکدام از متقاضیان تسهیلات دارای یک درجه اعتبار می باشد که آن نوعی برآورد از عملکرد وباز پرداخت تسهیلات درخواستی متقاضی می باشد(دهمرد وهمکاران،۱۳۹۱).به طور کلی فنون اندازه گیری ریسک اعتباری را می توان به دو گروه عمده تقسیم بندی نمود:(عرب مازار وروبین تن ،۱۳۸۵)

۱ - الگوهای امتیاز دهی اعتباری غیر پارامتری
- برنامه ریزی ریاضی
- درخت طبقه بندی
- الگوی نزدیکترین همسایه
- فرآیند تحلیل سلسه مراتبی
- سیستم خبره
- شبکه های عصبی مصنوعی والگوریتم ژنتیک

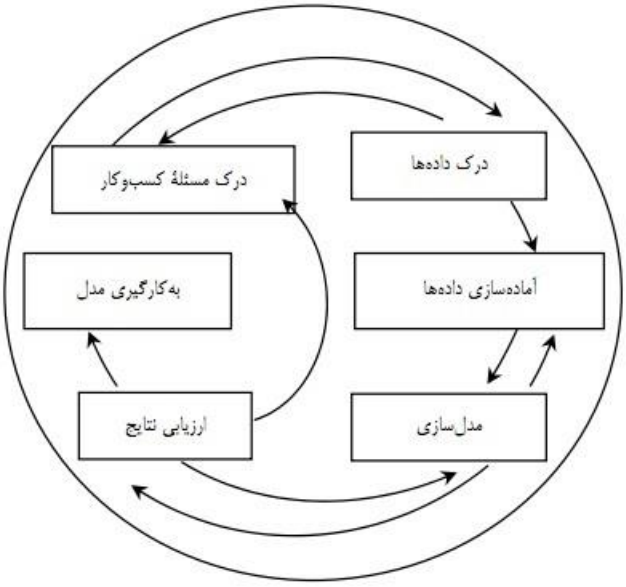
۲ - الگوهای امتیاز دهی اعتباری پارامتری
- مدل احتمال خطی
- مدل تحلیل ممیزی
- مدل لوجیت
- مدل پروبیت

موسسات اعتباری وبانکها می بایست با توجه به پیچیدگی فعالیت ها ومحیط اقتصادی پیرامونشان مدل های مناسب جهت ارزیابی امتیاز دهی مشتریان را انتخاب نمایند.

داده کاوی

داده کاوی فرآیند به خدمت گرفتن یک روش شناسی رایانه ای است که با استفاده ازتکنیک هاو الگوریتم های مختلف درجستجوی دانش نهفته درداده هاست.این فرآیند مشارکتی میان انسان و رایانه درنهایت به دنبال کشف الگوها و قواعد معناداردرمیان داده می باشد. داده کاوی ،پایگاه های داده ای بزرگ را به عنوان منبع دانش درنظرمی گیرد.با توجه به این امرکه داده کاوی فرآیند اکتشاف مدل های گوناگون ،خلاصه ها و ارزش های نشات گرفته ازمجموعه خاصی ازداده هاست، برای پیاده سازی چنین فرآیندی باید ازروش نظام یافته استفاده کرد.در این راستا متدولوژی « فرآیند استاندارد میان صنعتی داده کاوی»[9] با تجزیه و تحلیل نمایندگی های دایملر کرایسلر ایجاد شد. (رادفرد وهمکاران ،۱۳۹۳) این روش ،یک مدل فرآیندی برای داده کاوی ارائه می دهد که مروری بر چرخه ی عمرهرپروژه داده کاوی است، با توجه به روش داده کاوی CRISP یک پروژه چرخه عمری متشکل از ۶ گام دارد.این گام ها به صورت مستمر و تکراری در تمام فرآیند داده کاوی به کار گرفته می شوند(Chapman et al ,2000)شکل (۱)

[9] CRoss Industry Standard Process for Data Mining



شکل(۱) متدولوژی CRISP

دنبال کردن چنین متدولوژی است که به داده کاوی توان عمیق تر نگاه کردن به مساله را می دهد.در ادامه مقاله سعی برآن است که این متدولوژی به کار گرفته شود.

کاربردهای داده کاوی

داده کاوی در بسیاری از شاخه ها همچون بازاریابی،امور مالی،بانکداری،تولید،پزشکی،مدیریت ارتباط با مشتری،ردیابی ، ینی خرابی ها و آموزش های سازمانی کاربرد دارد.برخی از کاربردهای داده کاوی به طور خلاصه به شرح زیر است:

- کاربرد های معمول تجاری: ازقبیل تحلیل ومدیریت بازار ،تحلیل سبد بازار ،بازاریابی هدف،فهم رفتار مشتری،تحلیل ومدیریت ریسک
- مدیریت وکشف فریب:کشف فریب تلفنی ،کشف فریب های بیمه ای واتومبیل ،کشف حقه های کارت اعتباری،کشف تراکنش های مشکوک مالی(پول شویی)
- متن کاوی:پالایش متن (نامه های الکترونیکی ،گروه های خبری،جستجوی مقالات وموضوعات خاص وغیره)
- پزشکی :کشف ارتباط علامت وبیماری ،تحلیل آرایه های DNA، وتصاویر پزشکی
- ورزش: آمارهای ورزشی
- وب کاوی:پیشنهاد صفحات مرتبط ،بهبود ماشین های جستجوگر یا شخصی سازی حرکت در وب سایت

جدول(۳) بیانگر میزان نفوذ کاربرد داده کاوی در صنایع مختلف است که آمار موجود نشانگر استفاده از داده کاوی به عنوان یکی از ابزاردر مدیریت ارتباط با مشتری وبانکداری می باشد.(جماعت و عسگری،۱۳۸۹)

جدول (۳) میزان نفوذ کاربرد داده کاوی در صنایع مختلف

مدیریت ارتباط با مشتری	٪۲۶,۱
بانک داری	٪۲۳,۹
بازاریابی مستقیم	٪۲۰,۳
شناسایی جرم	٪۱۸,۸
وب کاوی	٪۱۰,۱
خرده فروشی	٪۱۰,۱
بیمه	٪۸,۷
مالی ولیزینگ	٪۷,۲
تجارت الکترونیک	٪۵,۸
امنیت وضد تروریسم	٪۶,۳
سرمایه گذاری وبورس	٪۲,۹

روند پیش برد تحقیق

تعریف مسئله

با توجه به اهمیت موضوع ریسک اعتباری در بانک ها در این پژوهش سعی بر آن است تا با استفاده از علم داده کاوی، اطلاعات مشتریان بانک مورد مطالعه را به دو گروه مشتریان خوش حساب وبد حساب دسته بندی کنیم. در این پژوهش الگوریتم های پر کاربرد داده کاوی نظیر درختان تصمیم،شبکه های عصبی،ماشین بردار،شبکه های بیزین،لجستیک و... هرکدام به طور مجزا بر روی داده های مورد مطالعه اعمال می گردد و مدلی هایی ایجاد خواهد شد وسپس برای اعتبارسنجی مدل ایجاد شده از داده های آزمون استفاده می گردد.درپایان مدل های ایجاد شده با الگوریتم های گفته شده با هم مقایسه ومدل هایی که دارای ضریب اطمینان بالا تر هستند انتخاب ومورد بررسی ارزیابی قرار می گیرند.

شناخت داده ها

کارایی داده کاوی مستقیماً ارتباط با داده های مورد استفاده دارد. هر اندازه داده ها دقیق تر وجامع تر و باکیفیت بهتری باشند خروجی داده کاوی کارتر خواهد بود .بنابراین انتخاب و جمع آوری داده های درست ، توصیف آنها ، یکپارچه سازی قالب آنها به منظور استفاده در داده کاوی اهمیت بسیار بالایی برخوردار می باشد.در این مقاله از مجموعه داده های اعتباری (UCI) http://paginas.fe.up.pt/~ec استفاده شده است.

آماده سازی داده ها

آماده سازی داده ها ،حدود ۶۰تا ۹۰ در صد زمان مورد نیاز کاوش را صرف کرده و ۷۵ تا ۹۰ درصد موفقیت پروژه های داده کاوی به آن مربوط می شود(قنبری وهمکاران ،۱۳۹۳).عدم آماده سازی داده یا آماده سازی ضعیف پروژه سبب شکست کامل پروژه می شود.دو وظیفه اساسی برای آماده سازی دادها وجود دارد:

۱ .سازمان دهی داده ها به شکل استاندارد به طوری که برای پردازش با داده کاوی ودیگر ابزارهای با مبنای رایانه ای آماده باشند.

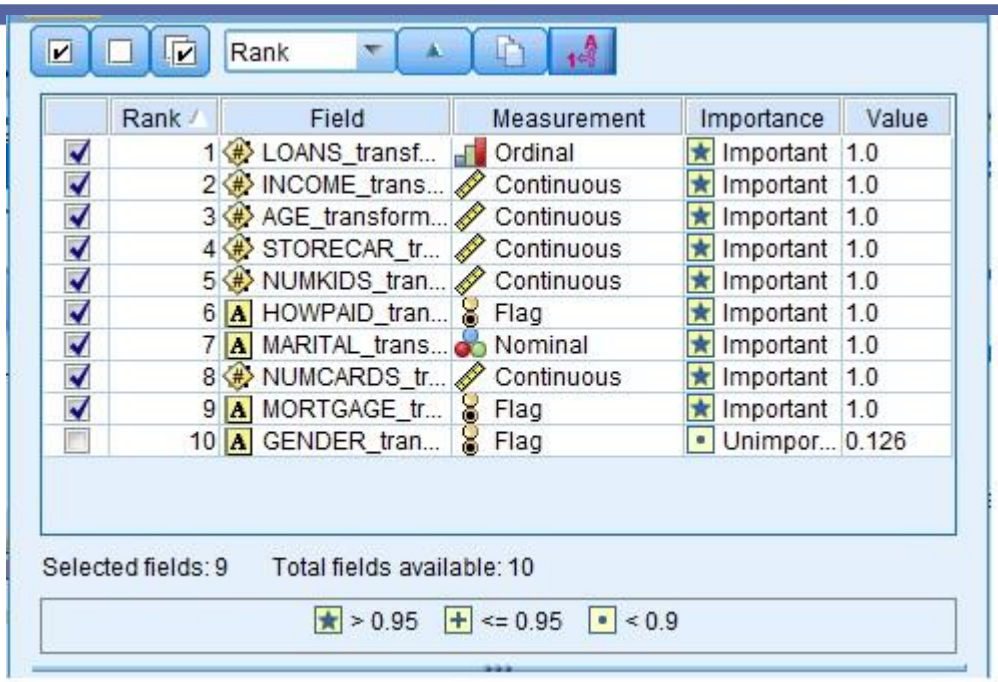
۲.مجموعه داده ها طوری آماده شوند که به بهترین عملکرد روش های داده کاوی منتهی شوند(مشکانی وناظمی ،۱۳۸۸)

داده های انتخاب شده در این تحقیق اطلاعات مربوط۴۱۱۷ نفر از مشتریان مورد مطالعه می باشد.داده های مورد مطالعه که به صورت یک فایل اکسل بوده است، به نرم افزار داده کاوی به کار رفته در این پروژه یعنی IBM SPSS Modeler 18 منتقل گردید . این فایل دارای ۱۲ ویژگی از ویژگی های مشتریان بانک مورد نظر بوده که این ویژگی هادر جدول (۴) دیده می شوند.

جدول (۴) ویژگی های مورد انتخاب

نام ویژگی	نام ویژگی به لاتین
شماره	ID
سن	AGE
جنسیت	GENDER
متاهل/مجرد بودن	MARITAL
میزان در آمد	INCOME
تعداد فرزندان	NUMKIDS
تعداد کارت های اعتباری	NUMCARDS
نوع پرداخت اقساط	HOWPAID
وام مسکن	MORTGAGE
تعداد ماشین	STORECAR
تعداد وام های دیگر	LOANS
ریسک پرداخت اقساط	RISK

با توجه به ماهیت تحقیق که رتبه بندی می باشد تصمیم گرفته شد تا بر روی ویژگی ریسک (خوش حساب یا بد حساب بودن مشتری) تمرکز کینم.در ادامه تحقیق با استفاده از ابزار های پیش پردازش داده این نرم افزار ،کارهای فیلترکردن ،حذف ویژگی های اضافی و در صورت نیاز تبدیل و ویژگی اسمی به عددی،تعداد ۹ ویژگی مهم وموثر بر روی فیلد هدف (ریسک) توسط گره Feature Selection شناسایی و برای ادامه کاوش انتخاب شدند. در شکل (۲) فیلدهای موثر وغیر موثر نمایش داده شده است.که فیلد GENDER,ID به دلیل موثر نبودن در فیلد هدف حذف شدند.



شکل(۲) انتخاب مهمترین ویژگی ها

در این مرحله قبل از انجام مدل سازی، داده هارا با استفاده از گره Partition به دو دسته آزمایش وآزمون تقسیم نمودیم. در این تحقیق(۳۱۴۸) ۷۷٪برای ازمایش مدل استفاده شده است و(۹۶۹)٪۲۳ برای تست وآزمون مدل در نظر گرفته شد.

مدل سازی

پس از انجام مراحل قبلی ، حال می توان به مدل سازی پرداخت.در این مرحله باید یکی از تکنیک های داده کاوی را با متناسب با صورت مسئله اعمال کرد. به طور خلاصه می توان گفت که ماموریت اصلی کاوش داده ها به عهده این گام است .انتخاب تکنیک و الگوریتم مناسب بسیار تعیین کننده است. پس از انتخاب مدل و تعیین پارامتر ها ،بخش های کوچکی از پروژه تعریف شده و پس از اجرا شدن ،در هر مرحله تست می شوند تا کیفیت مدل ایجاد شده تضمین شود.در این مرحله اگر مدل مورد نظر دقت لازم را نداشت کیفیت مطلوب را حاصل نکرد،باید به مرحله قبل برگشت و به تغییر پارامترهای مدل بپردازیم و مجدداً مدل را تست کنیم.در این مقاله اگوریتم های مختلف داده کاوی برروی داده های مشتریان اعمال گردید که این الگوریتم عبارتند از:درخت تصمیم[۱۰]،ماشین بردار پشتیبان[۱۱]، روش دسته بندی بیزین[۱۲] ،شبکه های عصبی[۱۳]،لجستیک[۱۴]،K –نزدیکترین همسایه[۱۵]

الگوریتم درخت تصمیم

درخت تصمیم یکی از روش های ناپارامتری رده بندی کردن است که با توجه به نوع متغیر وابسته به دو دسته رده بندی درختی[۱۶] برای متغیر رسته ای و رگرسیون درختی [۱۷]برای متغیر پیوسته تقسیم می شود. رده بندی درختی در راستای روش

[10] Decision tree algorithm
[11] Support vector machine
[12] Bayesian classifier
[13] Neural network algorithm
[14] Logistic regression algorithm
[15] K-Nearest Neighbor Algorithm

[16] Classification Tree
[17] Regression Tree

هایی نظیر تحلیل ممیزی و رگرسیون لجستیک است. در این روش مجموعه ای از شرط های منطقی به صورت یک الگوریتم با ساختار درختی برای رده بندی یا پیش بینی یک پیامد به کار می رود .(Lee,2006) با توجه به این که شاخص های متنوع و روش های گوناگونی برای تعیین درخت تصمیم معرفی شده است، الگوریتم های متنوع و گوناگونی نیز ارائه شده است که مهم ترین و شناخته شده ترین الگوریتم ها به ترتیب عبارتند از:CHAID,CART,ID3,C4.5,QUEST,CRUISE

در این تحقیق از انواع الگوریتم های درخت تصمیم برای اعتبارسنجی مشتریان بانک استفاده شده است که هر کدام از این الگوریتم نتایج ای متفاوت را ارائه داده اند. این نتایج والگورتیم ها در جداول (۱۱-۵) نشان داده شده است.

جدول (۵)الگوریتمC5.0

'Partition'	1_Training		2_Testing	
Correct	۲۳٦۹	۷۵,۲۵%	۷۲۲	۷٤,۵۱%
Wrong	۷۷۹	۲٤,۷۵%	۲٤۷	۲۵,٤۹%
Total	۳۱٤۸		۹٦۹	

جدول (۶)الگوریتمDescrimiant

'Partition'	1_Training		2_Testing	
Correct	2235	۷۱%	٦۸٤	۷۰,۵۹%
Wrong	913	۲۹%	۲۸۵	۲۹,٤۱%
Total	۳۱٤۸		۹٦۹	

جدول (۷) الگوریتمCHAID

'Partition'	1_Training		2_Testing	
Correct	2382	۷۵/۶۷%	۷۱۵	۷۳,۷۹%
Wrong	766	۲٤,۳۳%	۲۵٤	۲٦,۲۱%
Total	۳۱٤۸		۹٦۹	

جدول (۸)الگوریتمRandom -tree

'Partition'	1_Training		2_Testing	
Correct	2116	٦۷,۲۲%	٦٤٦	٦٦,٦۷%
Wrong	1032	۳۲,۷۸%	۳۲۳	۳۳,۳۳%
Total	۳۱٤۸		۹٦۹	

جدول (۹)الگوریتمC&R Tree

'Partition'	1_Training		2_Testing	
Correct	2383	۷۵/۷%	۷۲۵	۷٤,۸۲%
Wrong	765	۲٤,۳%	۲٤٤	۲۵,۱۸%
Total	۳۱٤۸		۹٦۹	

جدول (۱۰)الگوریتمTree-AS

'Partition'	1_Training		2_Testing	
Correct	2372	۷۵/۳۵%	۷۲۲	۷٤,٥۱%
Wrong	776	۲٤,٦٥%	۲٤۷	۲٥,٤۹%
Total	۳۱٤۸		۹٦۹	

جدول (۱۱)الگوریتمQUEST

'Partition'	**1_Training**		**2_Testing**	
Correct	2313	**۷۳/۴۸%**	703	**۷۲,٥٥%**
Wrong	835	**۲٦,٥۲%**	266	**۲۷,٤٥%**
Total	۳۱٤۸		۹٦۹	

ماشین بردارپشتیبان

در کاربردهای امروزی یادگیری ماشین، ماشین بردار پشتیبان به عنوان یکی از قدیمی ترین و دقیق ترین متدها در میان الگوریتم های معروف شناخته می شود. همچنین ماشین بردار پشتیبان یکی از روش های یادگیری با ناظر ست که از آن برای طبقه بندی و رگرسیون استفاده می کنند. مبنای کاری دسته بندی کننده این مدل، دسته بندی خطی داده بوده و درتقسیم خطی داده ها سعی بر آن است خطی انتخاب شود که حاشیه اطمینان بیشتری را داشته باشد. البته ماشین بردار پشتیبان در دسته بندی غیرخطی هم کاربرد دارد. به طورکلی این الگوریتم از یک نگاشت غیرخطی برای تبدیل داده های اصلی به ابعاد بالاتر استفاده می کند. (Fazli and momeni ,2013)

جدول (۱۲) الگوریتم LSVM

'Partition'	1_Training		2_Testing	
Correct	۲۳۲٦	۷۳,۸۹%	۷۱۲	۷۳,٤۸%
Wrong	۸۲۲	۲٦,۱۱%	۲۸۷	۲٦,٥۲%
Total	۳۱٤۸		۹٦۹	

جدول (۱۳) الگوریتم SVM

'Partition'	1_Training		2_Testing	
Correct	2383	۷٥,۷%	۷۲۸	۷٥,۱۳%
Wrong	765	۲٤,۳%	۲٤۱	۲٤,۸۷%
Total	۳۱٤۸		۹٦۹	

الگوریتم شبکه عصبی

اساس روش شبکه عصبی برپایه ساختار مغز انسان است. پس از آنکه ایده شبکه عصبی به ذهن خطور کرد مبانی ریاضی آن پایه ریزی می شود. برمبنای کارکرد آماری، شبکه های عصبی در رگرسیون و سریهای زمانی برای مدلسازی و پیش بینی به کار می رود، و از آنجا که شبکه ای از اجزای به هم مرتبط است آن را شبکه عصبی نامیده اند. این اجزاء از مطالعات سیستم های عصبی زیستی الهام می گیرد. به عبارت دیگر، شبکه عصبی کوششی است تا با استفاده از اجزایی که شبیه سلول های عصبی زیستی رفتار می کنند ماشین هایی ایجاد گردد که مانند مغز انسان کار کند. شبکه عصبی دارای قابلیت های طبقه بندی الگوها، قابلیت

یادگیری و تعمیم بوده و بنابراین، از این ساختارها می توان برای پیش بینی شرایط آینده براساس تجارب گذشته استفاده کرد. شبکه عصبی یک پردازنده توزیع شده موازی بزرگی است که از واحدهای پردازشی ساخته شده و دارای یک تمایل طبیعی برای ذخیره دانش تجربی و ارائه پیشنهادات مناسب است. شبکه ها باید طوری طراحی شوند که بتوانند براساس مجموعه های ورودی؛ خروجی های مناسب و مدنظر را ایجاد کنند. برای آموزش شبکه های عصبی می توان از الگوریتم هایی چون یادگیری رو به جلو یا مدل پرسپترون چند لایه[18] بهره جست.(Breuel and Shafait,2016)(Mahdavi et al,2013) (Kayvanjoo et al 2014)

جدول (۱۴)الگوریتم Neural Net (MLP)

'Partition'	1_Training		2_Testing	
Correct	2372	۷۵/۳۵%	۷۲۲	۷۰,۱۸%
Wrong	776	۲٤,٦٥%	۲٤۷	۲۹,۸۲%
Total	۳۱٤۸		۹٦۹	

جدول (۱۵)الگوریتم Neural Net (RBF)

'Partition'	1_Training		2_Testing	
Correct	2322	۷۳/۷۶%	۷۰٤	۷۲,٦٥%
Wrong	826	۲٦,۲٤%	۲٦٥	۲۷,۳٥%
Total	۳۱٤۸		۹٦۹	

رگرسیون لجستیک

رگرسیون لجستیک یکی از ابزارهای مهم داده کاوی است و اغلب در مواردی کاربرد می یابد که متغیر پاسخ دو بخشی (باینری) باشد (Oded and Rokach,2010)در این روش ضرورت دارد متغیر پاسخ عدد صحیح یا نسبی باشد.

جدول (۱٦)الگوریتم Logistic

'Partition'	1_Training		2_Testing	
Correct	2320	۷۳,۷%	۷۱۱	۷۳,۳۷%
Wrong	828	۲٦,۳%	۲۸٥	۲٦,٦۳%
Total	۳۱٤۸		۹٦۹	

جدول (۱۷) الگوریتم GLM

'Partition'	1_Training		2_Testing	
Correct	۲۳۲۷	۷۳,۹۲%	۷۱۱	۷۳,۳۷%
Wrong	۸۲۱	۲٦,۰۸%	۲٥۸	۲٦,٦۳%
Total	۳۱٤۸		۹٦۹	

[18] Multi Perceptron Layer(MPL)

جدول (۱۸)الگوریتمGLMM

'Partition'	1_Training		2_Testing	
Correct	1084	٥٨,٤٥%	٥٦٧	٥٨,٥١%
Wrong	1303	٤١,٥٥%	٤٠٢	٤١,٤٩%
Total	٣١٤٨		٩٦٩	

شبکه های بیزین

یک روش بسیار مهم، روش بیز ساده است که بیز سطحی و بیز مستقل نیز نامیده می شود. ساخت این روش بسیار ساده است و نیازی به برنامه های تخمین پارامتر تکرارشونده پیچیده ندارد؛ یعنی می توان از آن برای مجموعه داده های بسیار وسیع استفاده کرد که درنهایت، این روش معمولاً فوق العاده عمل می کند. بیز ساده یکی از قدیمی ترین الگوریتم های دسته بندی رسمی است و هنوز حتی در ساده ترین شکل بسیار مؤثر است. از این مدل در دسته بندی متون و جداسازی اسپم ها به طور گسترده استفاده می شود.(salari and adib,2010)

جدول (۱۹)الگوریتمBayes Net

'Partition'	1_Training		2_Testing	
Correct	2352	٧٤/٧١%	٧٢٣	٧٤,٦١%
Wrong	796	٢٥,٢٩%	٢٤٦	٢٥,٣٩%
Total	٣١٤٨		٩٦٩	

k نزدیکترین همسایگی

به طورکلی K-NNروشی برای طبقه بندی یک عنصر در یک مجموعه است که این کار را براساس نزدیکترین خصوصیات سایر اعضای موجود در مجموعه)نمونههای آموزشدهنده(انجام میدهد. K-NNدارای یکی از ساده ترین الگوریتم های آموزش برای پیش بینی داده ها است. الگوریتم K-NN به گروه موسوم به الگوریتم های یادگیرنده تنبل[19] تعلق دارد و داده ها را به گونه ای ذخیره می نماید که تا زمانی که دستوری برای برآورد جدید نباشد، فرایند آموزش روی این داده ها انجام نخواهد شد (Nemes et al,2008)

جدول (۲۰)الگوریتمK-NN

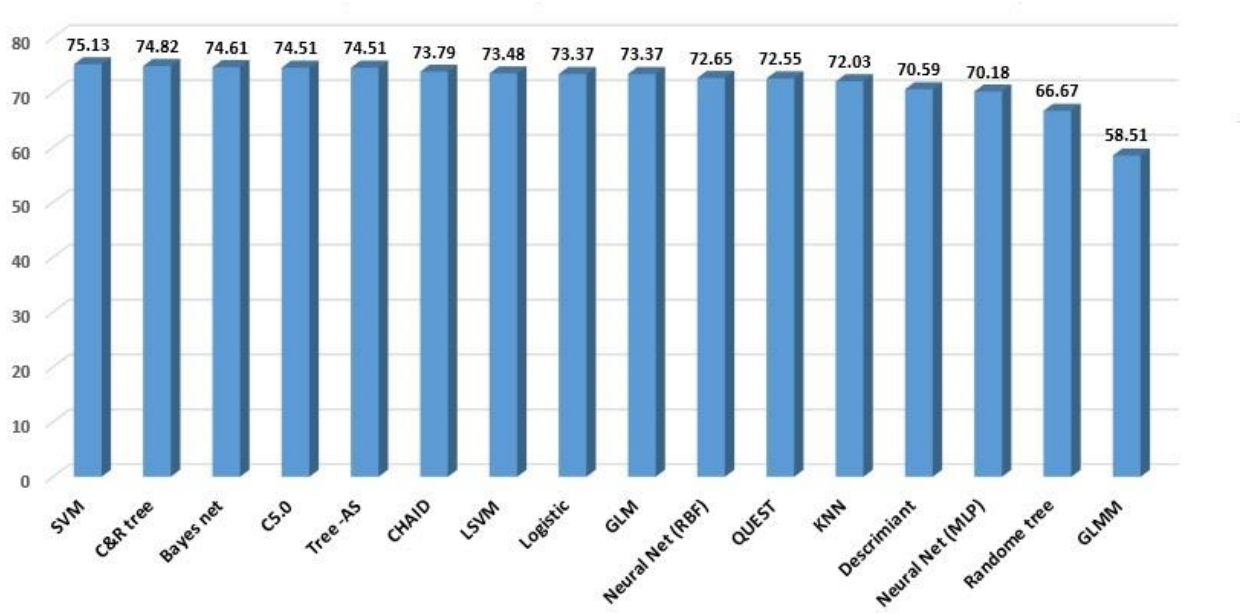
'Partition'	1_Training		2_Testing	
Correct	٢٤٨٠	٧٨/٧٨%	٦٩٨	٧٢,٠٣%
Wrong	٦٦٨	٢١,٢٢%	٢٧١	٢٧,٩٧%
Total	٣١٤٨		٩٦٩	

ارزیابی مدل

در این مرحله مدل ها ی ایجاد شده در مرحله قبل مقایسه وارزیابی میشوند. معیارهای ارزیابی شامل دقت پیش بینی، سرعت و... می باشد. روشهای ارزیابی شامل ماتریس اغتشاش ،منحنی ROC،تابع هزینه و غیره است که در این تحقیق از ماترس اغتشاش استفاده شده است و نتایج به دست آمده از ارزیابی مدل ها در شکل (۳) به صورت نمودار

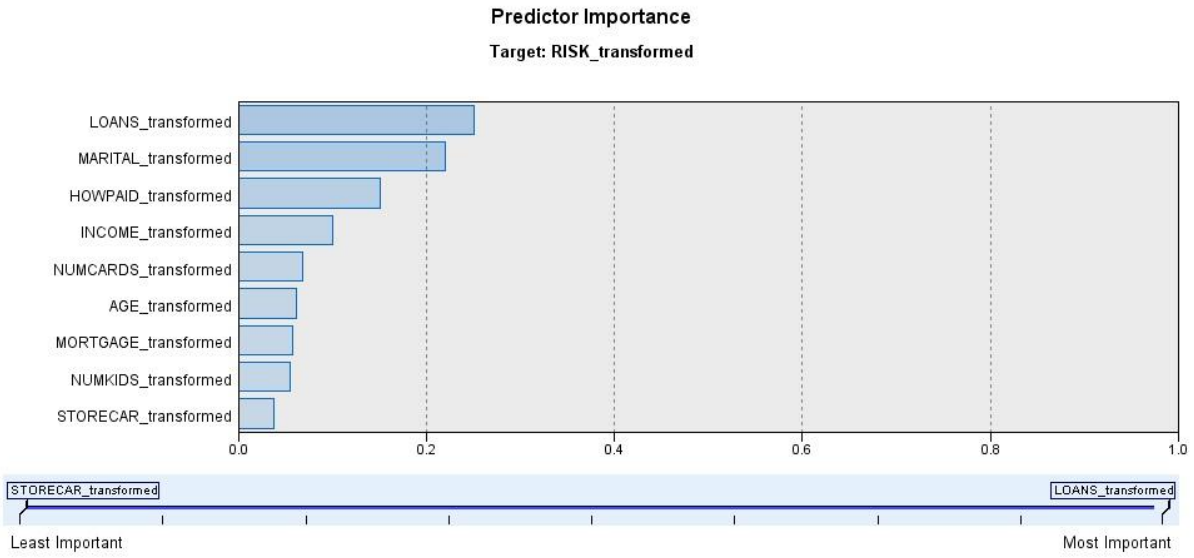
[19] lazy learning algorithms

قابل مشاهده است. هدف ارزیابی، انتخاب مدل مناسب و یا در صورت امکان، مدل بهینه در بین تعدادی از مدل هاست . پس از مدل سازی حال می بایست به ارزیابی نتایج حاصل از مدل پرداخت .ارزیابی نتایج باعث بهبود مدل شده و مدل را قابل استفاده می کند.

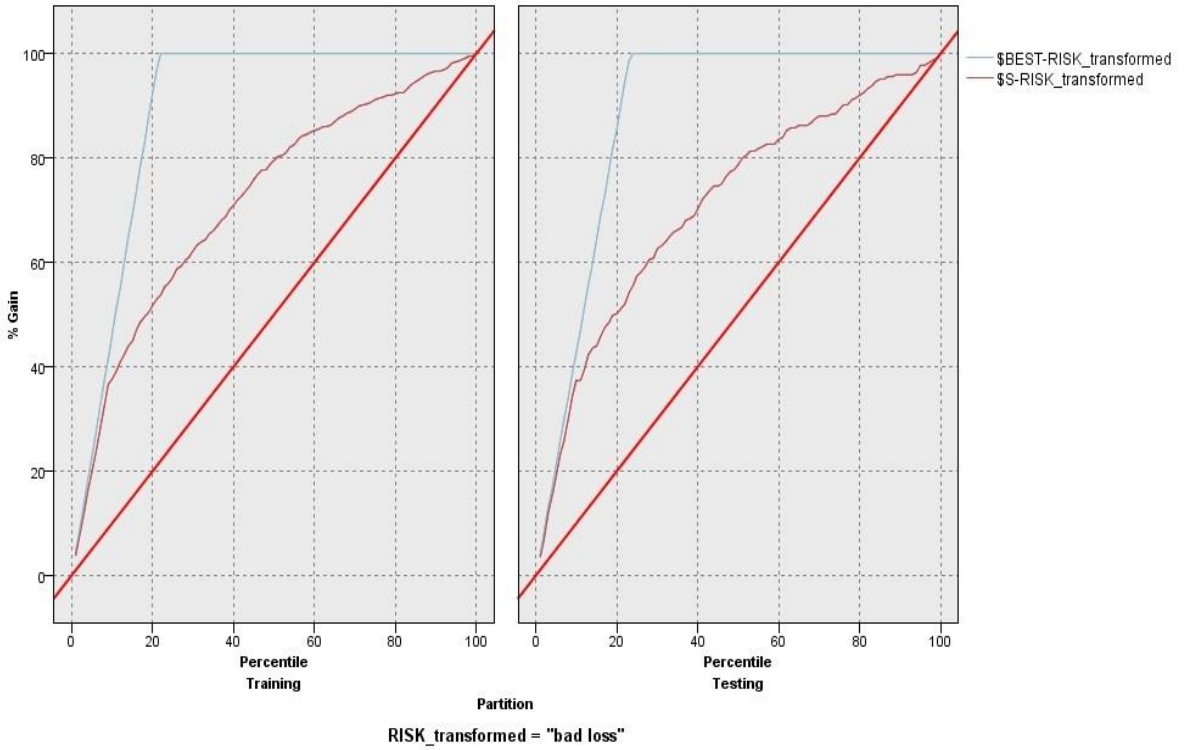


شکل (۳) نتیجه مقایسه بین الگوریتم های داده کاوی استفاده شده در رتبه بندی مشتریان بانک

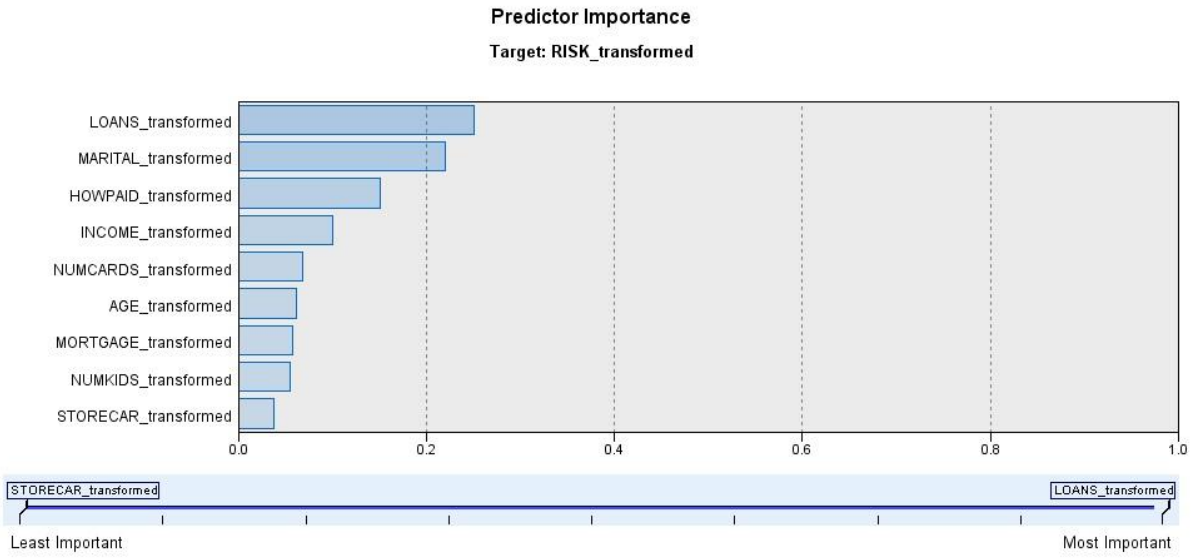
همان طور که در شکل (۳) مشاهده می شود الگوریتم SVMاز الگوریتم های ماشین بردار بالاترین ضریب اطمینان یعنی مقدار ٪۷۵,۱۳رادارا می باشد.وبه عنوان بهترین مدل از نظر ضریب اطمینان ،از میان الگوریتم های استفاده شده در این مقاله انتخاب می گردد.الگوریتم C&RTreeنیز از الگوریتم های درخت تصمیم بعد از الگوریتم SVM دارای بالاترین ضریب چاطمینان ٪۷۴,۸۲ می باشد.در شکل های (۷-۴) مهم ترین ویژگی های موثر درفیلد هدف وهم چنین نمودار ارزیابی مدل های ایجاد شده توسط دو الگوریتم SVM,C&RTreeنمایش داده شده است.نتایج این دو مدل حاکی از آن است که ویژگی های،تعداد وام های دیگر مشتری ،متاهل /مجرد بودن،نوع پرداخت اقساط وام ومیزان درآمد مشتری به ترتیب تاثیر به سزایی بروی خوش حساب /بد حساب بودن مشتری در مقایسه با ویژگی های دیگر دارند.



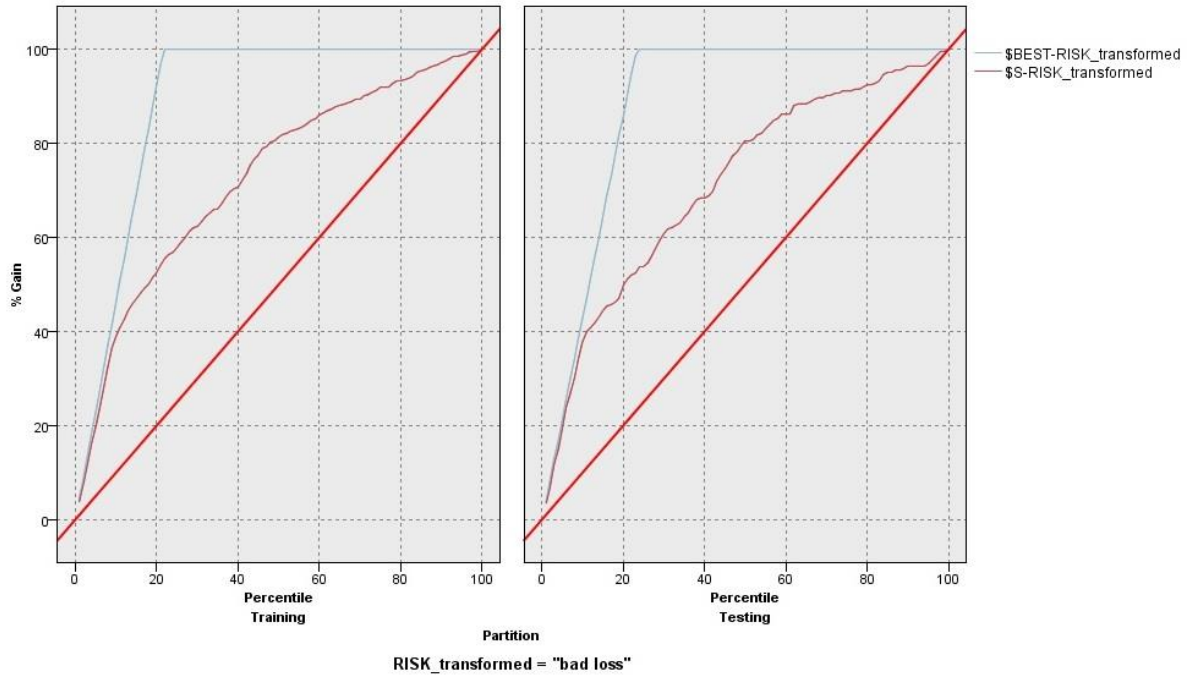
شکل(۴) مهم ترین ویژگی ها در ایجاد مدل C&R Tree



شکل(5)نمودارارزیابی مدل C&R Tree



شکل(۶) مهم ترین ویژگی ها در ایجاد مدل SVM



شکل(۷) نمودارارزیابی مدل SVM

بحث و نتیجه گیری

در این تحقیق که هدف اصلی آن بررسی وارزیابی استفاده الگوریتم های داده کاوی در طبقه بندی مشتریان بانک بوده است،نتایج مدل ماشین بردار پشتیبان(SVM) در مقایسه با مدلهای درخت تصمیم ،شبکه های عصبی ،شبکه های بیزین و رگرسیون لجستیک و... بررسی شد. یافته های تحقیق حاکی از آنست که در طبقه بندی متقاضیان، مدل ماشین بردار پشتیبان نسبت به مدلهای دیگر بطورمعناداری از دقت کلی بیشتری برخورداراست .در مقاله حاضر ابتدا سعی شد با استفاده از ابزار های پیش پردازش داده وگره Feature Selection فیلد هایی که تاثیری در فیلد هدف نداشته اند حذف

شوند.سپس داده های مورد مطالعه به دو قسمت آزمایش وآزمون تقسیم گردید.در ادامه این تحقیق مدل هایی که توسط داده های آزمایش ایجاد شده بود توسط داده های آزمون مورد ارزیابی قرار گرفت . نتایج بدست آمده از این تحقیق نشان داد که مدل ایجاد شده توسط الگوریتم SVM نسبت به دیگر مدل های ایجاد شده ضریب اطمینان بالاتری بوده است.همچنین مدل ایجاد شده توسط الگوریتم C&R Tree نیز بعد ازمدل SVMدارای ضریب اطمینان بالاتری در مقایسه با مدل های ایجاد شده توسط دیگر الگوریتم های استفاده شده در این مقاله است.

بنابراین می توان گفت که استفاده از الگوریتم های مختلف داده کاوی نظیر ماشین بردار پشتیبان ودرختان تصمیم برای ایجاد مدل های اعتبارسنجی با ضریب اطمینان بالا باعث می شود تا کارشناسان اعتبارسنجی بتوانند با صرف زمان و هزینه کم و افزایش رضایت مشتری اعتبارسنجی به طبقه بندی مشتریان خود بپردازند.هم چنین می توان برای تحقیقات آینده از ترکیب این الگوریتم های استفاده کرد ومدل هایی با سرعت وضریب اطمینان بالاتر ایجاد نمود.

منابع

- فردحریری، علیرضا، عبدلی، قهرمان ،(۱۳۹۴). الگوسازی سنجش ریسک اعتباری مشتریان حقوقی بانک رفاه. *فصلنامه نظریه های کاربردی اقتصاد*، سال دوم ،شماره ۱،صفحات۲۴-۱.
- عریانی،بهاره،عیسی زاده ، سعید،(۱۳۸۴). رتبه بندی ریسک اعتباری مشتریان حقوقی بانک ها به روش تحلیل فراگیرداده ها(مطالعه موردی بانک کشاورزی استان تهران)
- دهمرد, نظر. سیف الدین پور, سمیرا. اسفندیاری, مرضیه. (زمستان ۱۳۹۱). اعتبار سنجی مشتریان بانک با استفاده از رویکرد امتیاز دهی اعتباری ،مطالعه موردی شعب بانک سپه در زاهدان. *پژوهش های مدیریت عمومی*، ۱۳۵-۱۵۲.
- عرب مازار، عباس، و رویین تن، پونه (۱۳۸۵). عوامل موثر بر ریسک اعتباری؛ مطالعه موردی بانک کشاورزی. *جستارهای اقتصادی*، سال سوم، ۶،ص ۴۵-۸۰.
- رادفرد،رضا؛نظافتی ،نوید؛یوسفی اصلی ،سعید (۱۳۹۳). طبقه بندی مشتریان اینترنت بانک با کمک الگوریتم های داداه کاوی ؛نشریه مدیریت فناوری اطلاعات ،دانشگاه تهران ؛دوری ۶،شماره ۱، بهار ۱۳۹۳،ص ۹۰-۷۱
- جماعت ,علی ؛عسگری ,فرید (۱۳۸۹). مدیریت ریسک اعتباری در سیستم بانکی با رویکرد داده کاوی، فصلنامه مطالعه کمی در مدیریت ،ص۱۱۵-۱۲۹.
- قنبری،صدیقه،پاشازاده،سعید،بیورانی،حسین (۱۳۹۳). ارزیابی عملکرد تکنیک های مختلف داده کاوی به منظور اعتبارسنجی مشتریان بانکی،نهمین سپوزیوم پیشرفت های علوم وتکنولوژی ،مشهد .
- مشکانی،علی؛ ناظمی،عبدالرضا (۱۳۸۸). مقدمه ای بر داده کاوی، داشنگاه فردوسی مشهد،چاپ اول .
- میرزایی، حسین. نظریان، رافیک. باقری، رعنا (۱۳۹۰). بررسی عوامل مؤثر بر ریسک اعتباری اشخاص حقوقی بانکها. *فصلنامه روند پژوهشهای اقتصادی*، شماره ۹.
- تهرانی، رضا. فلاح شمس، میرفیض.(۱۳۸۴). طراحی و تبیین مدل ریسک اعتباری در نظام بانکی کشور. *مجله علوم اجتماعی و انسانی دانشگاه شیراز*، ۶۰-۴۵، ۴۳.
- قدسی پور، حسن. سالاری، میثم. دلاوری، وحید (۱۳۹۱). ارزیابی ریسک اعتباری شرکت های وام گیرنده از بانک با استفاده از تحلیل سلسله مراتبی فازی وشبکه عصبی ترکیبی درجه بالا، *نشریه بین المللی مهندسی صنایع و مدیریت تولید*، شماره ۱،جلد.۲۳.
- صفری، سعید. ابراهیمی شقاقی، مرضیه. شیخ، محمدجواد .(۱۳۸۹) .مدیریت ریسک اعتباری مشتریان حقوقی در بانک های تجاری با رویکرد تحلیل پوششی داد ه ها)رتبه بندی اعتباری.(پژوهش های مدیریت در ایران)مدرس علوم انسانی)، ۱۴، ۱۳۷- ۱۶۴.
- کرمی، مصطفی .(۱۳۸۹) ارائه مدلی برای ارزیابی وضعیت ریسک اعتباری مشتریان بل از اعطای تسهیلات بانکی. بانک، ۱۴۲-۱۴۶.
- شریعت پناهی،سید مجید ، هاشمی برکادهی، سیما. (۱۳۸۷) ارائه مدلی برای اعتبار سنجی مشتریان در بانک صنعت و معدن. فصلنامه مطالعات حسابداری، ۸۲-۶۱، ۲۱.

–راعی، رضا. فلاح پور، سعید(۱۳۸۷) کاربرد ماشین بردار پشتیبان در پیش بینی درماندگی مالی شرکتها با استفاده از نسبتهای مالی. *بررسی های حسابداری و حسابرسی*، ۱۷- ۵۳. ۳۴.

– .Salari M, Adib F. Ten best data mining algorithms. 13th Student conference on electrical engineering. Tehran: Tarbiat Modares University; 2010. [Text in Persian]
– .Oded, M and Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook .Prof*, Springer New York Dordrecht Heidelberg London, DOI 10.1007/978-0-387-09823-4
– Abdou, H., Pointon, J. and El-Marsy,A. (2008). "Neural nets versus conventional techniques in credit scoring
in Egyptian banking." *Expert Systems with Applications*, 35, 1277-1279
– Bellotti, T. and Crook, J. (2009). "Support vector machines for credit scoring and discovery of significant
features." *Expert Systems with Applications*, 36, 3302
– Breuel TM, Shafait F. Automlp: Simple, effective, fully automated learning rate and size adjustment. University of Kaiserslautern, 67663, Germany: Kaiserslautern; 2010. Available From: http://snowbird.djvuzone.org/2010/abstracts/ 163.pdf. Accessed Apr 14, 2016
– Chapman, P. & Clinton, J. & Kerber, R. & Khabaza, T. & einartz, T. & Shearer, C. & Wirth, R. (2000). CRISP-DM Step-Data Mining Guide. SPSS Inc CRISPMWP .
– Chuang, C. and Lin, R. (2009). "Constructing a reassigning credit scoring model." *Expert Systems with*
Applications, 36,1685-1687.
– Dehuri, S., Patnaik, S., Ghosh, A., & Mall, R., (2008),"Application of elitist multi-objective genetic algorithm for classification rule generation". Applied Soft Computing ,Vol.8,pp.1,2,3,5
– Emel, AhmetBurak. Oral, Muhittin. Reisman, Arnold. Yolalan, Reha. (2003). A credit scoring approach for the commercial banking sector. *Socio*-Economic *Planning Sciences*, 37, 103–123
– Eskandari, S. (2006). "A Model for Customer Credit Scoring in Financial Market for Avoiding Ill-Credit Decision." MSc, School of Social & Economic Science, Alzahra University, Tehran.
– Fazli H, Momeni H. Comparison and evaluation of data mining algorithms, decision tree and SVM application for intrusion detection. In: Proceedings of 8th Symposium progress in science and technology 2013, Mashhad. Iran; 2013.[Text in Persian]
– Greuning, Hennie van. Bratanovic, Sonja Brajovic. (2009). Analyzing Banking Risk: A Framework for Assessing Corporate Governance and Risk Management. World Bank Publications
– Huang, C. L., Chen, M. C. and Wang, C. J. (2007). "Credit scoring with a data mining approach based on
support vector machine." *Expert Systems with Applications*, 33(2007)
– Kabari, L. G. Nwachukwu, E. O. (2013). Credit Risk Evaluating System Using Decision Tree – Neuro Based Model. *International Journal of* Engineering *Research &Technology*, 2, 2738-2745.
– KayvanJoo AH, Ebrahimi M, Haqshenas G. Prediction of hepatitis C virus interferon/ribavirin therapy outcome based on viral nucleotide attributes using machine learning algorithms. BMC Res Notes 2014;7:565.
– Lee SK. On Classifcation and Regression Trees for Multiple Responses and Its Application. J Classif. 2006 Jun 1; 23(1):123–41.
– Lee, T. S., Chiu, C. C., Chou, Y. C., Lu, C. J. (2006). Mining the customer credit using classification and regressiontree and multivariate adaptive regression splines. Computational Statistics & Data Analysis, 50, 1113-1130.
– Lee, Tian-Shyug. Chiu, Chih-Chou. Lu, Chi-Jie. Chen, I-Fei. (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems* with *Applications*, 23, 245-254.
– Lin, S. (2009). A new two-stage hybrid approach of credit risk in banking industry. *Expert Systems with Applications*, 36, 8333–8341
– Mahdavi M, Taheri M, Lotfi F. Application of data mining in neural networks. In: Proceedings of 8th Symposium on progress in science and technology 2013. Mashhad. Iran; 2013. [Text in Persian]
– Mileris, Ricardas. Boguslauskas, Vytautas. (2011). Credit Risk Estimation Model Development Process: Main Steps and Model Improvement. *InzinerineEkonomika-Engineering Economics*, 22, 126-133.
– Min, Jae H. Lee, Young-Chan .(2008). A practical approach to credit scoring. *Expert Systems with Applications*, 35, 1762–1770

– Nemes, A., Roberts, R.T., Rawls, W.J., Pachepsky, Ya.A., and Van Genuchten, M.Th. 2008. Software to estimate -33 and -1500 kPa soil water retention using the non-parametric k-Nearest Neighbor technique. Environm Modell. Softw. 23: 254-255
– Singh, S., Murthi, B. P. S., & Steffes, E. (2013). Developing a measure of risk adjusted revenue (RAR) in credit cards market: Implications for customer relationship management. *European Journal of Operational Research*, *224*(2), 425
– Yanping,Y.,Zhengming, Q., Min, Y.,Rui, G.,Liting, F.,Penghui, G. (2012) Research on the Application of Decision Tree to the Analysis of Individual Credit Risk. *Information Technology*, 25, 209-214
– Yu, L., Wang, S., Lai, K. (2008). Credit risk assessment with a multistage neural network ensemble learning approach. *Expert Systems with Applications*. 34, 1434– 1444.
– Yu, L., Wang, S., Lai, K. K. and Zhou, L. (2008). Bio-Inspired Credit Risk Analysis: Computational Intelligence with Support Vector Machines, Springer
– Zekic-Susac, M., Sarlija, N. and Bensic, M. (2004). "Small Business Credit Scoring: A Comparison of Logistic Regression, Neural Networks, and Decision Tree Models." *26th International Conference on Information Technology Interfaces*. Croatia.