



اعتبار سنجی مشتریان کارت‌های اعتباری بانکی با استفاده از الگوریتم ژنتیک

افشین شهلاei مقدم

کارشناسی ارشد مهندسی صنایع- دانشگاه صنعتی اصفهان

Afshin_shahlayi@in.iut.ac.ir

علی شالباف زاده

کارشناسی ارشد مهندسی کامپیوتر- دانشگاه صنعتی اصفهان

Ali.shalbafzadeh@gmail.com

طاهر احمدی

کارشناسی مهندسی صنایع- دانشگاه صنعتی اصفهان

Tahir_Ahmadi@yahoo.com

تقی رضوان

کارشناسی ارشد مهندسی صنایع- دانشگاه صنعتی اصفهان

Taghi_rezvan@in.iut.ac.ir

چکیده

پیشرفت در بکار گیری و جمع‌آوری داده با به خدمت گرفتن کامپیوتر و همچنین استفاده همگانی از وب و اینترنت به عنوان یک سیستم اطلاع رسانی جهانی، داده‌کاوی را به عنوان یک ابزار مفید جهت کمک به حل مسائل پیچیده مطرح ساخته است. کاهش ریسک‌ها و مخاطرات، و همچنین استفاده از فرصت‌ها در حوزه بانکداری از مسائل مورد توجه محققین می‌باشد که تحت عنوان اعتبارسنجی اشخاص حقیقی و حقوقی از آن می‌توان یاد کرد. در گذشته جهت بررسی مسائل مالی و بهویژه اعتبارسنجی از روش‌های کلاسیک داده‌کاوی نظری رگرسیون خطی، تجزیه و تحلیل ممیزی خطی و غیره استفاده می‌شده است. اما با ظهور هوش مصنوعی و هوش محاسباتی، روش‌های آنها از قبیل شبکه‌های عصبی، سیستم‌های خبره و الگوریتم‌های ژنتیک، نیز در اعتبارسنجی در بانکداری با استقبال محققین این زمینه مواجه شده است.

در این مقاله با مطالعه و بررسی عوامل مؤثر بر اعتبار فرد، با استفاده از روش تصمیم‌گیری گروهی مبتنی بر بونیسون فازی تعیین شده و سپس با استفاده از اطلاعات جمع‌آوری شده توسط پایگاه داده ایجاد شده بوسیله اکسس، الگوریتم ژنتیکی جهت ایجاد پایگاه دانش اعتبارسنجی فرد توسعه داده می‌شود.

واژه‌های کلیدی: مدل اعتبار سنجی، بونیسون فازی، الگوریتم ژنتیک

۱- معرفی

اعتبار به معنی "اعتماد داشتن و باور کردن" مترادف کلمه credit می‌باشد. اعتبار در مفهوم عام آن، حسن شهرتی است که شخص نسبت به قابلیت ایفای تعهداتش دارد، یعنی اعتماد به انجام تعهدات یک شخص در طول انجام معامله یا معاملاتش. بر این اساس، واژه اعتبار در عرف بنگاه‌های اقتصادی، مترادف با ایجاد تسهیلات مالی برای مشتریان، اعتبارات بانکی، تسهیلات اعتباری و تسهیلات اعطایی می‌باشد. هر یک از طرق مختلف تأمین مالی یا تضمین تعهدات واحدهای اقتصادی، گونه‌ای از تسهیلات اعطایی محسوب می‌گردد. بنابراین، تسهیلات اعطایی را می‌توان یک یا ترکیبی از موارد ذیل دانست:

- (1) تأمین تمام یا قسمتی از هزینه‌های مالی (جاری و ثابت) یک فعالیت اقتصادی
- (2) قبول تعهد یا تضمین آثار مالی تمام یا قسمتی از تعهدات مشتریان [1].

مؤسسات اعتباری که بانک‌ها رکن اساسی آن شناخته می‌شوند برای در اختیار قرار دادن انواع تسهیلات اعطایی به مشتریان خود، نیاز به انجام بررسی‌های کاملی بهمنظور شناخت متقاضیان از ابعاد کیفی و کمی دارند و به طور عام، این بررسی‌ها را اعتبارسنجی می‌نامند[1]. مؤسسه‌های مختلفی را جهت سنجش اعتبار و ریسک مشتریان خود به کار می‌برند. این گستردگی به جهت استفاده از روش‌های کمی مبتنی بر اطلاعات می‌باشد که عموماً با تغییر میزان دسترسی به اطلاعات روش‌های مورد استفاده نیز تغییر می‌کند[1]. مؤسسه‌های اعتباری در گذشته به علت محدودیت تنوع وامها و همین‌طور دسترسی به اطلاعات مشتریان، از روش‌های ساده کیفی جهت سنجش اعتبار مشتریان خود استفاده می‌نمودند[2]. روش‌های کمی پس از مدت زمان نسبتاً طولانی وارد عرصه تصمیم‌گیری شدند، دلیل اصلی این تأخیر به جهت عدم دسترسی به اطلاعات مناسب در زمان مناسب می‌باشد. در بسیاری از کشورها به جهت منع قانونی در مورد حریم شخصی افراد، دشواری‌های بسیاری جهت ایجاد پایگاه داده یکپارچه و قابل دسترسی وجود دارد. بسیاری از مؤسسه‌های اطلاعاتی مربوط به مشتریان را محدود به پایگاه‌های داده خود نمودند. با این حال امروزه در بسیاری از کشورهای توسعه یافته، پایگاه‌های اطلاعاتی عظیمی جهت استفاده محققین و کارشناسان تسهیلات شکل گرفته است که با استفاده از آن بسیاری از روش‌های اعتبارسنجی کمی مورد توجه قرار گرفته‌اند. در گذشته روش‌های کلاسیک مختلفی نظریه تجزیه و تحلیل ممیزی خطی^۱، رگرسیون خطی، درخت باینری و روش حداقل برای اعتبارسنجی استفاده می‌شده که روش‌های تجزیه و تحلیل خطی و رگرسیون لجستیک دو روش عمده مورد استفاده بوده است[3 و 4]. در بین این دو روش، تجزیه و تحلیل خطی از مقبولیت بیشتری برخوردار می‌باشد. اما نقطه ضعف روش تجزیه و تحلیل خطی، فرض خطی بودن رابطه بین متغیرها و نرمال بودن آنها است، در حالی که این رابطه معمولاً غیرخطی و غیرنرمال می‌باشد. در هر دو روش تجزیه و تحلیل خطی و رگرسیون، رابطه خطی بین متغیرها وجود دارد، لذا دقت این دو روش بسیار پایین است[5]. از سوی دیگر بعضی از پژوهش‌ها نشان داده است که به جهت رابطه غیرخطی ضعیف بین متغیرها دو روش تجزیه و تحلیل خطی و رگرسیون لجستیک عملکرد خوبی دارند[6].

علاوه بر این، چند روش بر جسته در حوزه هوش مصنوعی همچون سیستم‌های خبره، شبکه‌های عصبی و الگوریتم ژنتیک نیز برای اعتبارسنجی مورد استفاده قرار گرفته‌اند. در بین این روش‌ها شبکه‌های عصبی به علت توانایی تخمین روابط پیچیده غیرخطی بیشتر مورد توجه بوده‌اند[7]. در ادبیات تحقیق اعتبارسنجی، در اکثر مسائل اعتبارسنجی شبکه‌های عصبی عملکرد دقیق‌تری از رگرسیون و تجزیه و تحلیل ممیزی خطی دارند[9 و 10]. از سوی دیگر شبکه‌های عصبی ضعف عمدہ‌ای در فرآیندهای طولانی مدت یادگیری دارند و پس از شکل‌گیری معماری بهینه، به مثابه یک جعبه سیاه^۲ عمل می‌کنند، و به همین دلیل رابطه بین متغیرهای ورودی و خروجی مشخص نمی‌باشد. علاوه بر شبکه‌های عصبی پژوهش‌هایی نیز راجع به الگوریتم ژنتیک وجود دارد [11 و 12]، ولی سیستم‌های ترکیبی^۳ بیشتر مورد توجه قرار گرفته است[13].

¹ Linear Discriminate Analysis

² Black Box

³ Hybrid system

دساوی و همکاران [6] در سال 1998 پژوهشی انجام دادند که طی آن با استفاده از تکنیک‌های نظریه تجزیه و تحلیل ممیزی خطی و رگرسیون لجستیک^۱ اقدام به طبقه‌بندی مشتریان وام‌های بین المللی در سه کشور آمریکا، آلمان و استرالیا به سه طبقه خوب، ضعیف و بد نمودند. البته در پایان پژوهش، محققین به این نتیجه رسیدند که طبقه‌بندی به دو دسته خوب و بد، ارجحیت دارد. در ادامه با استفاده از متغیرهای تعیین شده و مجموع داده‌های جمع آوری شده، از شبکه‌های عصبی مصنوعی به منظور طبقه‌بندی مشتریان به دو دسته خوب و بد بهره بردن. محققین در این پژوهش با استفاده از اطلاعات مربوط به اعتبارات اخذ شده در سه کشور فوق، حد فاصل سال‌های 1988 تا 1991 به فهرست متغیرهای پیش‌بینی کننده اعتبار رسیدند که این فهرست عبارت است از: تعداد کارت‌های اعتباری اصلی، مالکیت منزل، حقوق و سایر درآمدها، نمره صداقت در ارایه اطلاعات بر مبنای گزارش اعتباری دایره اعتبارات بانک، مدت زمان سابقه در شغل فعلی، تعداد اعضای تحت تکفل، تعداد استعلام در هفت ماه گذشته، پرداخت ماهانه مناسب با درآمد، بد حسابی مشتری در 12 ماه گذشته، مجموع بدھی به تناسب درآمد، سن متقاضی، مدت زمان حضور در آدرس فعلی، تعداد حساب‌های موجود بنابر گزارش دایره اعتبارات، تعداد حساب‌های فعال بنا بر گزارش دایره اعتبارات، تعداد وام‌های اخذ شده در گذشته و نمره اطلاعات بر اساس گزارش دایره اعتبارات.

پژوهشی دیگر توسط [14] در سال 2000 انجام شد که در آن با استفاده از شبکه‌های عصبی و تکنیک‌های تجزیه و تحلیل عددی، مشتریان به دو دسته خوب و بد تقسیم شدند.

در این پژوهش نیز از متغیرهای مختلفی جهت توسعه سیستم اعتبار سنجی شامل: طول عمر حساب، تاریخ اعتبار، طبقه‌بندی شغلی، بررسی وضعیت حساب، دارایی‌های شخصی، مدت زمان اقامت، سایر وام‌های اخذ شده، طبقه‌بندی سکونت، مقدار وام، مقصود از وام، مدت زمان اشتغال (سابقه کار)، وضعیت حساب پسانداز، استفاده شد.

در پژوهش دیگری، یوباس و همکاران [15] در سال 1997 انجام شد با استفاده از روش استنتاج فازی^۲ و درخت‌های تصمیم^۳، عملکرد مشتریان اعتباری را به دو دسته پرداخت کنندگان خوب و پرداخت کنندگان ضعیف تقسیم نمود. در این پژوهش از عوامل مختلفی همچون: وضعیت اشتغال درخواست کننده، مدت زمان حضور در بانک، ارزش وام مسکن، تعداد فرزندان، مدت زمان حاضر در شغل فعلی، وضعیت اقامتی، نوع حساب، سایر کارت‌های در مالکیت فرد، قیمت تخمینی منزل، تلفن منزل، درآمد درخواست کننده، درآمد همسر، کارت‌های اعتباری اصلی در مالکیت فرد، مخارج استفاده گردید.

اخيراً الگوريتم ژنتيک در کشف داشن^۴ و داده کاوی در عرصه مالي کاربردي مورد توجه قرار گرفته‌اند. يكى از اولىين کاربردهای الگوريتم ژنتيک در انتخاب متغیرهای مناسب جهت استفاده در مدل‌های آماری و شبکه‌های عصبی می‌باشد[16]. هافمن و همکاران سیستم‌های پایگاه قانون فازی ژنتيک^۵ را جهت استنتاج و تحليل اعتبارسنجی مشتریان بکار گرفتند[17]. سیستم‌های پایگاه قانون فازی ژنتيک شامل روش‌های تکاملی جهت تعطیق و بهبود کل پایگاه دانش می‌باشند[18]. چن و همکاران [21] از يك روش ترکيبی برای اعتبارسنجی مشتریان استفاده نمودند به اين ترتيب که ابتدا با استفاده از شبکه عصبی مشتریان را به دو دسته خوب و بد تقسيم نمودند و سپس با استفاده از الگوريتم ژنتيک يك تقریب ترجیحی بر روی مشتریان بد بانک انجام دادند. اما کاربرد اساسی این روش‌ها در توانایی قابل توجه‌شان در يادگیری قوانین طبقه‌بندی فازی جهت مشتریان می‌باشد[22 و 23]. بر اين اساس در اين مقاله سعی شده است پس از تعیین و انتخاب متغیرهای مناسب پیش‌بینی کننده وضعیت مشتریان با استفاده از ابزارهای تصمیم‌گيری چندمعیاره همچون بونیسون فازی به استنتاج قوانین مناسبی پرداخته شود که توانایی بررسی وضعیت مشتریان متقاضی تسهیلات را دارا باشد. برای استنتاج این قوانین از الگوريتم ژنتيک ساده‌ای استفاده شده است.

ساختار اين مقاله به اين شرح است. در بخش دوم با توجه به متغیرهای پیش‌بینی کننده موجود در تحقیقات با استفاده از روش بونیسون فازی و با لحاظ کردن نظرات خبرگان این حیله و معیارهای میزان اهمیت متغیر و همچنین میزان دسترسی به اطلاعات آن متغیرهای مناسب برای این کار تعیین و انتخاب گردد. در بخش سوم چگونگی توسعه بانک اطلاعاتی تشریح و در بخش بعدی با استفاده از الگوريتم

¹ Logistic Regression

² Fuzzy Inference

³ Decision trees

⁴ Knowledge Discovery

⁵ Genetic Fuzzy Rule Based Systems (GFRBS)

ژنتیک یک پایگاه دانش مناسب برای مشتریان کارت‌های اعتباری بانکی توسعه داده می‌شود. در نهایت این مقاله با ارائه نتایج کلی و پیشنهاداتی برای تحقیقات آتی پایان می‌پذیرد.

2- تعیین و انتخاب متغیرهای مناسب با استفاده از روش بونیسون فازی

در این مرحله پس از مطالعه پروژه‌های مختلف اعتبار سنجی در سطح بین المللی، فهرست کاملی از متغیرهای مؤثر بر پروژه‌های اعتبار سنجی انتخاب شد (جدول 1) و سپس با استفاده از فرم‌های نظر سنجی، نظرات بیش از 100 خبره بازار تسهیلات و اعتبارات از بانک‌های مختلف جمع‌آوری گردید و با استفاده از روش بونیسون فازی تعداد هفت متغیر اصلی شامل: سن، وضعیت شغلی، سطح تحصیلات، وضعیت مالکیت مسکن، طول عمر حساب، میزان گردش حساب و میزان درآمد جهت شکل‌دهی پایگاه دانش تعیین شد. در این فرآیند عوامل مؤثر بر اعتبار فرد به دو دسته اهليت مالی و اهليت اجتماعی تقسيم شدند تا کارشناسان اعتباری با دید روشنی نسبت به مسئله، قضاوت خود را انجام دهند. علاوه بر این، ترکیب گسترده‌ای از شعب بانکی در سراسر شهر تهران استفاده شد تا سطح اعتبار قضاوت با توجه به تفاوت‌های اقتصادی و فرهنگی افزایش یابد.

سن	12	وضعیت جک برگشتی	2
مقدار قسط	13	مدت فعالیت	3
تعداد شغل	14	نوع وثیقه	4
تعداد کارت‌های اعتباری در ملکیت فرد	15	شغل	5
وضعیت تحصیلی همسر	16	وضعیت تحصیلی	6
وضعیت تأهل	17	نوع مالکیت منزل	7
مقدار وام اخذ شده	18	حقوق ماهیانه	8
شغل همسر	19	گردش عملیات بانکی	9
تعداد فرزندان	20	تعداد وام	10
		حقوق ماهیانه همسر	21

جدول 1: لیست کامل متغیرهای موثر بر اعتبار مشتری

2- روش بونیسون فازی

فرض در این روش بر این است که می‌توان اطلاعات فازی قطعی را در یک تصمیم‌گیری به صورت پارامتریک تقریب زد، به طوری که این تقریب توسط عدد فازی ذوزنقه‌ای به صورت (a, b , α, β) [24]. خواهد بود.

در این روش برای هریک از گزینه‌ها وزن اهمیت و همین‌طور وزن معیارهای مختلف تعیین شده و سپس با استفاده از محاسبات فازی، وزن واحد هر یک از گزینه‌ها محاسبه می‌شود و با استفاده از تکنیک روش بونیسون رتبه‌بندی و مقایسه زوجی صورت می‌گیرد.

در این پژوهش برای نظر سنجی از یک سری متغیر زبانی و عدد فازی ذوزنقه‌ای متناظر استفاده شد. در جدول 2 متغیرهای زبانی و اعداد فازی ذوزنقه‌ای متناظر نمایش داده شده است.

متغیر زبانی معیار اهمیت	متغیر زبانی معیار دسترسی	عدد فازی ذوزنقه‌ای متناظر
خیلی غیر مهم	خیلی کم	(0,0,0,0.2)
غیر مهم	کم	(0,0.1,0,0.2)
تا حدودی غیر مهم	تا حدودی کم	(0.2,0.2,0.2,0.2)
بسیار	مناسب	(0.5,0.5,0.2,0.2)

تا حدودی مهمن	تا حدودی زیاد	(0.8,0.8,0.2,0.2)
مهمن	خوب	(0.9,1,0.2,0)
خیلی مهمن	خیلی خوب	(1,1,0.2,0)

جدول 2: متغیرهای زبانی و اعداد فازی متناظر مورد استفاده

3- بانک اطلاعاتی توسعه داده شده

در این مقاله پس از تعیین متغیرهای مؤثر بر مدل با استفاده از بانک اطلاعاتی Microsoft Access 2007، تعداد 129 رکورد بدون نام و مشخصات فردی تهیه شد، فقدان اطلاعات کافی، یکی از مشکلات اساسی می‌باشد که خبرگان دلایل آن را کمبود پایانه‌های فروش و همین‌طور پایین بودن سطح اعتبار اعطایی اعلام نمودند. بانک اطلاعاتی ایجاد شده نظیر جدول 2 بوده که در آن متغیرهای سن، وضعیت شغلی، سطح تحصیلات، وضعیت مالکیت مسکن، طول عمر حساب، میزان گردش حساب و میزان درآمد، متغیرهای ورودی و متغیر میزان اعتبار فرد، متغیر خروجی در نظر گرفته شده‌اند.

میزان اعتبار فرد	...	وضعیت مالکیت مسکن	سطح تحصیلات	وضعیت شغلی	سن	رکورد
						1
						2
						3

جدول 3: شمایی از بانک اطلاعاتی ایجاد شده

پس از آماده‌سازی بانک اطلاعاتی، یک سری پیش پردازش‌های اولیه لازم است روی داده‌ها صورت گیرد که عبارتنداز:

الف) گسسته‌سازی داده‌ها

مقادیر متغیرهایی نظیر سن، سطح تحصیلات به صورت پیوسته بوده که باید گسسته شوند و به بازه‌های کوچکتری تقسیم شوند. البته باید در گسسته‌سازی داده دقت نمود.

ب) سازگار نمودن داده‌ها

از آنجا که الگوریتم طراحی شده با اعداد کار می‌کند، لذا لازم است به صورت عددی درآید. مثلاً برای متغیر سن به ترتیب (0)، (1) و (2) استفاده می‌شود.

جدول 4 گسسته‌سازی و سازگاری داده‌های متغیرهای منتخب در مدل را نشان می‌دهد.

نام متغیر	محدوده اطلاعات جمع آوری شده	گسسته سازی اطلاعات
سن	زیر 30 سال بین 30-60 سال بالای 60 سال	جوان میانسال مسن

وضعیت تحصیلی	1 2 3 4 5 6	زیر دیپلم دیپلم فوق دیپلم لیسانس فوق لیسانس دکترا
وضعیت شغلی	1 2	آزاد کارمند
وضعیت مالکیت منزل	1 2	استیجاری ملکیت
میزان گردش بانکی	زیر 30000000 ریال بالای 30000000 ریال	ضعیف مناسب
طول عمر حساب	زیر یک سال بین یک تا سه سال بیش از سه سال	کوتاه متوسط بلند
درآمد	بین 30000000 تا 50000000 ریال بین 5000000 تا 8000000 ریال بیش از 8000000 ریال	ضعیف متوسط خوب

جدول 4- چگونگی سازگاری و گستته سازی داده ها

4- توسعه پایگاه دانش با استفاده از الگوریتم ژنتیک

الگوریتم های ژنتیک، روندهای ریاضی بوده که فرآیند توراث ژنتیک را مورداستفاده قرار می دهند. آنها را می توان برای تعیین الگوهای جالب در زمینه های مختلف کسب و کار به کار گرفت. در این مقاله جهت ایجاد یک پایگاه دانش از یک الگوریتم ژنتیک ساده استفاده می گردد. این الگوریتم روی یک جمعیت با صفات تمایز در قالب مشخص با به کارگیری عملگرهای تقاطع^۱، جهش^۲ و انتخاب^۳ عمل می نماید. فرآیند الگوریتم ژنتیک مورد استفاده عبارتست از:

1. انتخاب تصادفی والدین
 2. تولید مثل به واسطه تقاطع. تولید مثل عملگری است که با انتخاب، موجودات بقاء نسل خواهد کرد. به عبارت دیگر، تعدادی تابع هدف یا صفت مشخص منتخب نیاز است تا بازمانده تعیین شود.
 3. انتخاب بازماندگان برای نسل بعد به واسطه تابع شایستگی^۴.
- این تابع شایستگی با در نظر گرفتن دو پارامتر درصد صحت و درصد پوشش عمل انتخاب را انجام می دهد که به صورت زیر تعریف می شوند:
- درصد پوشش یک قانون را a گویند هرگاه قانون در $a\%$ از کل مجموعه داده صحیح باشد.
 - درصد صحت یک قانون را b گویند در صورتی که اگر قسمت "اگر" قانون درست باشد قسمت آنگاه قانون در $b\%$ موقع صحیح باشد.

¹ Crossover

² Mutation

³ Selection

⁴ Fitness Function

4. جهش، عملیاتی است که به واسطه آن، مشخصه‌های انتخاب شده تصادفی از موجودات انتخاب شده به طور تصادفی در عملیات بعدی تغییر داده می‌شوند.

5. تکرار تازمانی که یک سطح مشخص شایستگی بدست آید یا به تعداد از پیش تعیین شده‌ای تکرار رسیده شود. آنچه از اهمیت زیادی در اجرای فرآیند فوق بخوردار است، طراحی یک ساختار کروموزوم مناسب برای ایجاد قوانین می‌باشد که شکل 1، ساختار این کروموزوم پیشنهادی را نشان می‌دهد.

ژن‌های تشکیل دهنده کروموزوم	{0,1}	{0,1}	{0,1}		{0,1}	{0,1}
نام متغیر			x_1	...		x_k
قانون			قسمت اگر		قسمت آنگاه	

شکل 1: ساختار کروموزوم پیشنهادی

x_k متغیر خروجی است؛ سایر متغیرها، متغیرهای ورودی هستند که پس از انتخاب توسط روش بونیسون فازی و پیش‌پردازش‌های اولیه روی اطلاعات جمع‌آوری شده در بانک اطلاعاتی قسمت اگر قانون را می‌توانند ایجاد نمایند. در این کروموزوم مثلاً x_1 مقادیر (0, 1), (2) می‌گیرد که بیانگر متغیر سن می‌باشد. برای ایجاد پایگاه دانش از یک برنامه تحت C# استفاده شده که به طور خودکار تمام عملیات از جمله پیش‌پردازش‌های اولیه تا الگوریتم ژنتیک را نجات می‌دهد. با اجرای این برنامه تحت مقادیر مختلف پارامترهای الگوریتم ژنتیک نتایجی حاصل شده است که قسمتی از پایگاه دانش بدست آمده را جدول 5 نشان می‌دهد. شبکه کد مربوط به تولید قوانین در پیوست مقاله ارایه شده است.

ردیف	درصد بالایی جمعیت اولیه برای ازدواج	Mutation	Crossover	تعداد تکرار	درصد اطمینان پشتیبانی اولیه	قوانین	اطمینان پشتیبانی		
1	500000	1	5	2	50	10	30	1) Account Duration=Long,House Oweness=Own House,Educational level=Diploma,Job Status=Employee,THEN Customer type =good 2) Account Duration=Long,House Oweness=Own House,Educational level=Diploma,Then Customer Type=good Account Duration=Long,House owness= own house, Educational level= Bachelor of science,job Status= Employee, Age= middle Age Income=Good,Bank Account=Good, Then Customer Type: Good	41
2	750000	1	5	3	50	10	30	Customer type =good Duration=Long,House Oweness=Own House,Educational level=Diploma,Then Customer Type=good Account Duration=Long,House owness= own house, Educational level= Bachelor of science,job Status= Employee, Age= middle Age Income=Good,Bank Account=Good, Then Customer Type: Good	49

3	1000000	1	5	3	50	10	30	,House Owness=Own House,Educational level=Diploma,Job Status=Employee, Age= Young, Income= poor, Bank Account= Poor Then Customer Type= Good	39	1	
4	1000000	1	10	3	50	10	30	1)House Owness=Own House ,Income=Poor,Educational level= Bachelor Of Science, ,Age=middle Age, Bank Account= Medium Then Customer Type= Good 2) House Owness=Own House,Age=middle Age Then Customer Type= Good	57	1	
5	1000000	1	10	5	100	10	30	Income=Poor,Account Duration=Long,House Owness=Own House,Educational level=Diploma, Then Custumer Type=Good Income=Poor,House Owness=Own House,Educational level=Diploma,Job Status=Employee,Age=middle Age , Then Custumer Type=Good	36	1	

جدول 5: قسمتی از پایگاه دانش ایجاد شده با مقدار دهی اولیه پارامترهای الگوریتم ژنتیک

5- نتیجه‌گیری و پیشنهادات آتی

امروزه داده‌کاوی در کسب و کار مختلف از اهمیت قابل توجهی برخوردار شده است. داده‌کاوی با تحلیل داده و شناخت الگوهای ارتباطات پنهان موجود در داده‌ها می‌تواند کمک‌های شایانی به کسب و کار از جمله بیمه، بانک، شرکت‌های تولیدی و ارتباطی نماید. در این مقاله با یکپارچه‌سازی چند مأذول از جمله بانک اطلاعاتی، پیش‌پردازش‌های اولیه و الگوریتم ژنتیک پایگاه دانشی ایجاد شد که می‌تواند برای اعتبارسنجی مشتریان کارا باشد. گرچه با بکارگیری الگوریتم ژنتیک در الگوریتم‌های درخت تصمیم و شبکه‌های عصبی ممکن است بتوان داشت بهتری کسب کرد. از سوی دیگر، با بکارگیری مجموعه داده با استفاده از سیستم‌های استنتاج فازی و روش‌های کلاسیک نظیر رگرسیون می‌توان نتایجی را کسب کرد و نتایج حاصله را مورد مقایسه قرار داد و با کاربرد تکنیک‌های آماری اعتبار مدل را بررسی نمود. از جمله نتایج این کار مجموعه داده کوچک و سری بودن الگوریتم ژنتیک بوده که منجر به افزایش زمان اجرا شده است. لذا تحقیقات آتی می‌تواند هم در راستای برطرف کردن ناقصی و هم در جهت استفاده از سایر تکنیک‌ها برای تحلیل این مجموعه داده باشد.

6. مراجع

- [1] سفری، سید علی اصغر، کلهر، علی اصغر، بهمنی، حسن، هدایتی، محمود، "عملیات بانکی داخلی"، موسسه عالی بانکداری ایران، 1381.
- [2] جمشیدی، سعید "شیوه های اعتبار سنجی مشتریان" پژوهشکده پولی و بانکی، چاپ دوم، تابستان 1384.
- [3] Altman, E. I. (1968). Financial ratios: discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), pp.589–611.
- [4] Sinkey, J. F. Jr., (1992). Commercial bank financial management: In the financialservices industry (4th Ed.). NY: Macmillan Publishing Company.
- [5] Thomas, L. C. (2000). A survey of credit and behavioral scoring: Forecasting financial risk of lending to consumers. *International Journal of Forecasting*, 16, pp.149–172.
- [6] Desai, V. S., Crook, J. N., & Overstreet, G. A. Jr., (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research*, 95, pp.24–37.
- [7] Thomas, L. C. (2000). A survey of credit and behavioral scoring: Forecasting financial risk of lending to consumers. *International Journal of Forecasting*, 16, pp.149–172.
- [8] Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. 2003b). Benchmarking state-of-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54, pp.627–635.
- [9] Goonatilake, S., Treleaven, P. (Eds.). (1995). Intelligent system for finance and business. Chichester: John Wiley & Sons.
- [10] Lee, T. S., Chen, I. F. (2005). A two-stage hybrid credit scoring model using Systems with Applications, 28, 743–752.
- [11] Piramuthu, S., Shaw, M. J., & Gentry, J. A. (1994). A classification approach using multi-layer neural networks. *Decision Support Systems*, 11, 509–525.
- [12] Richeson, L., Zimmermann, R. A., & Barnett, K. G. (1996). Predicting consumer credit performance: Can neural networks outperform traditional statistical methods?

- [13]. Kim, Y. S., & Sohn, S. Y. (2004). Managing loan customers using misclassification patterns of credit scoring model. *Expert Systems with Applications*, 26, pp.567–573.
- [14] Walker, R. F., Haasdijk, E. W., & Gerrets, M. C. (1995). Credit evaluation using a genetic algorithm. In S. Intelligent system for finance and business, pp. 39–59.
- [15] Hsieh, N. C. (2005). Hybrid mining approach in the design of credit scoring models. *Expert Systems with Applications*, 28, pp.655–665.
- [16] West, D. (2000). Neural network credit scoring models *Computers & Operations Research*, 27(11-12), pp. 1131-1152.
- [17] Yobas, M.B, Crook, J.N., Ross, P., (2000) “credit scoring using neural and evolutionary techniques”, *Journal of mathematics applied in business and industry*, 11, pp. 111-125.
- [18] Liu, H-H., Ong, C-Sh. (2008),” Variable selection in clustering for marketing segmentation using genetic algorithms”, *Expert Systems with Applications* 34, pp. 502–510.
- [19] Hoffmann, F., Baesens, B., Mues, C., Van Gestel, T. Vanthienen, J., (2007) “Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms”, *European Journal of Operational Research* 177, pp. 540–555.
- [20] Cordo' n, O., Herrera, F., Hoffmann, F., Magdalena, L. (2001).Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases. *Advances in Fuzzy Systems*, World Scientific, Singapore.
- [21] Chen, M-Ch., Huang, Sh-H. (2003) “Credit scoring and rejected instances reassigning through evolutionary computation techniques”, *Expert Systems with Applications*, 24(4), pp. 433-441.
- [22] Cordo' n, O., Del Jesus, M.J., Herrera, F., (1998) Genetic learning of fuzzy rule-based classification systems cooperating with fuzzy reasoning methods, *International Journal of Intelligent Systems* 13 (10–11) 1025–1053.
- [23] Junco, L., Sanchez, L., (2000) Using the AdaBoost algorithm to induce fuzzy rules in classification problems, in: Proc. Spanish Conference for Fuzzy Logic and Technologies, Sevilla, Spain, pp. 297–301.
- [24] اصغر پور محمد جواد، " تصمیم گیری های چند معیاره "، انتشارات دانشگاه تهران، چاپ چهارم، 1385.

Genome: Array (1 to Variable Count) of integer
Genome[x]=0 means Variable(x) does not affect in this rule
Genome[x]>0 means Group of variable x in this rule predicate

PesudoCode:

Gens= Create Random Genes with random number

//we recommend that for creating better random rules assign 0 with more probalistic ratio (eg .5)

For @ (Iteration Count) times do

```
    IF random number< Mutation_ratio THEN
        newGene=Mutate(Gens(random Index))
        EVALUATE(newGene)
        IF qualify(newGene) THEN add to gens(new gene)
    END IF

    FOR CrossOverRatio*PeopleCount
        FATHER=random gene in top% people
        MOTHER= random gene in top% people
        CROSSOVER(mother,father,out girl, out boy)
        EVALUATE (girl)
        EVALUATE(boy)
        IF qualify(girl) THEN add to gens(girl)
        IF qualify(boy) THEN add to gens(boy)
```