

میزان کارایی مدل‌های مختلف شناسایی الگو در طراحی و ساخت مدل‌های امتیازبندی اعتباری

محمدفریدون کیانی^۱

مهری محفوظیان^۲

تاریخ پذیرش مقاله: ۱۳۹۱/۵/۲۴

تاریخ دریافت مقاله: ۱۳۹۱/۵/۱

چکیده

تاکنون طیف وسیعی از انواع روش‌های شناسایی الگو^۱ برای تولید و ساخت مدل‌های امتیازبندی اعتباری^۲ مورد توجه محققان این حوزه قرار گرفته است اما آنچه در اغلب این تحقیقات کمتر در نظر گرفته شده توجه کافی به نیازهای مسأله است که آن را از یک مسأله دسته‌بندی^۳ ساده متمایز می‌کند. در امتیازبندی اعتباری لازم است به هزینه

* کارشناس واحد بازرسی بانک کارآفرین، نویسنده مسئول، m.kiyani@karafarinbank.net

** کارشناس بانک کارآفرین، m.mahfoozian@karafarinbank.net

- 1- Pattern Recognition
- 2- Credit Scoring
- 3- Classification

دسته‌بندی نادرست توجه ویژه‌ای معطوف گردد، زیرا در عمل هزینه دسته‌بندی نادرست مشتریان بد بسیار بیشتر از هزینه دسته‌بندی نادرست مشتریان خوب است. از این‌رو در این تحقیق سعی شده با بررسی روش‌های مختلف ارزیابی، میزان کارایی آنها در چندین روش امتیازبندی اعتباری سنجیده شود و در نهایت با ارائه یک روش ارزیابی مناسب مشکلات موجود در آنها مرتفع گردد.

واژه‌های کلیدی: امتیازبندی اعتباری، شناسایی الگو، ارزیابی مدل‌ها، هزینه دسته‌بندی

طبقه‌بندی JEL: G17, E51, E47

۱. مقدمه

بحران‌های مالی اخیر در سطح جهانی که بدترین بحران از دهه ۱۹۲۰ تاکنون محسوب می‌شوند، همگی ریشه در بحث عدم مدیریت صحیح ریسک‌های مالی دارند. در مدیریت ریسک‌های یک بنگاه مالی، ریسک اعتباری اهمیت ویژه‌ای دارد، به نحوی که در آن به موضوع مخاطرات پیش روی یک سازمان در نتیجه عدم بازپرداخت به موقع تعهدات مشتریان پرداخته می‌شود. مدیریت صحیح این فرایند نیازمند ابزارهای مختلفی است که یکی از مهم‌ترین آنها استفاده از سازوکارهای امتیازبندی و رتبه‌بندی مشتریان است (Kim, 2005).

در امتیازبندی اعتباری هدف اصلی شناسایی میزان اهلیت و شایستگی اعتباری مشتری است که با توجه به مجموعه‌ای از ویژگی‌های فرد به انجام می‌رسد. برای انجام درست این فرایند می‌توان با شناسایی مشتریانی که شایستگی لازم برای انجام تعهدات خودشان را ندارند، از به خطر افتادن سرمایه‌های بانک جلوگیری کرد. فرایند پیش‌بینی آینده مالی یک مشتری را می‌توان در زمره روش‌های شناسایی الگو تلقی کرد. در این حالت هدف اصلی تخمین تابعی است که بتواند با توجه به ویژگی‌های هر مشتری به عنوان ورودی‌ها، میزان شایستگی وی را در خروجی تخمین بزند. تاکنون طیف وسیعی از انواع روش‌های شناسایی الگو در این زمینه معرفی و به کار گرفته شده است که برخی بر اساس اصول و سازوکارهای آماری عمل می‌کنند و برخی از روش‌های داده‌کاوی استفاده می‌نمایند. در ادامه به مهم‌ترین روش‌های به کاررفته برای تولید مدل‌های امتیازبندی اعتباری اشاره شده است (Wu, 2008).

از لحاظ قدمت می‌توان گفت اولین بار دوراند^۱ در دهه ۱۹۳۰ با روش تحلیل ممیز و بر اساس ویژگی‌های چندین شرکت تجاری سعی کرد آنها را در دو گروه جداگانه تفکیک نماید (Komorad, 2002). بعد از وی در تحقیقات مختلفی (مانند Yeh, et al, 2007 و Mahlmann, 2004) از این روش برای ایجاد مدل‌های امتیازبندی اعتباری استفاده شد. در کنار روش تحلیل ممیز، روش‌های آماری موفق دیگری نیز معرفی و مورد توجه قرار داده

شدند که مدل‌های لاجیت (Hand & Adams, 2000) و پروبیت (Greene, 1998) از معروف‌ترین آنها به شمار می‌آیند. از آنجا که مدل‌های مذکور بر اساس توابع نرمال گاوسی بنا شده‌اند، حتی امروزه نیز به علت راحتی به‌کارگیری و نحوه ثبت اطلاعات در موارد بسیار زیادی استفاده می‌شوند. فرایند استنتاج به عمل آمده در این سازوکار در نهایت یک رابطه خطی از ضرایب برای هر کدام از ویژگی‌ها را به دست می‌دهد که از طریق آن می‌توان در خصوص میزان اهمیت هر کدام از ویژگی‌ها اطلاع پیدا کرد.

با شکل‌گیری روش‌های جدید در زمینه شناسایی الگو و معرفی ابزارهای داده‌کاوی، تحول مهمی در ایجاد مدل‌های امتیازبندی اعتباری به وجود آمد که در نتیجه تعداد بسیار زیادی از تحقیقات به بررسی نحوه عملکرد این روش‌ها و در بعضی موارد مقایسه آنها با روش‌های قبلی پرداخته‌اند. شبکه‌های عصبی به عنوان پرکاربردترین روش مورد استفاده در این زمینه همیشه مورد توجه کارشناسان قرار گرفته‌اند. (Plantamura, et al, 1993). در این سازوکار می‌توان بر خلاف دیگر روش‌های معمول، الگوهای غیرخطی و پیچیده را با سطوح دقت مختلف شناسایی و مدل‌سازی نمود. این روش مبتنی بر تعدادی متغیر واحد پردازشی به نام نرون است که به صورت موازی عمل کرده و با الهام از فرایند یادگیری در انسان قابلیت آموزش و به‌کارگیری را دارد.

با توجه به ساختارهای مختلف ارائه‌شده در این زمینه ساختار شبکه عصبی پیش‌ران^۱ با نحوه آموزش بازپخش خطا^۲ بیشتر از بقیه مورد توجه قرار داده شده است (Khashman, 2010). با این حال ساختارهای دیگری مانند پرسپترون چندلایه،^۳ شبکه تابع شعاعی،^۴ شبکه عصبی احتمالی،^۵ شبکه عصبی همبسته،^۶ بردار یادگیری کوانتیزه،^۷ بردار ویژگی خودسامانده^۸ در این زمینه معرفی و استفاده شدند. تقریباً در تمامی تحقیقات انجام‌شده دقت شبکه عصبی بهتر از مدل‌های آماری بوده و قابلیت رقابت آن با دیگر روش‌های شناسایی الگو بررسی شده است (Khashman, 2010).

- 1- Feed-Forward Neural Networks
- 2- Back-Propagation
- 3- Multi-Layer Perceptron (MLP)
- 4- Radial Basis Function Networks (RBFN)
- 5- Probabilistic Neural Networks (PNN)
- 6- Cascade Correlation Neural Networks (Cascor)
- 7- Learning Vector Quantization (LVQ)
- 8- Self-Organizing Feature Map (SOFM)

ماشین بردار پشتیبان اولین بار توسط وپنیک^۱ در ۱۹۹۵ و به عنوان یک روش جدید یادگیری آماری مبتنی بر شبکه عصبی معرفی شد (Chen, et al, 2005). در سال‌های اخیر توجه بسیار زیادی به استفاده از این روش در علوم مختلف برای حل مسائل دسته‌بندی و نیز بهبود ویژگی تعمیم‌پذیری آن شده است. بر خلاف ماهیت یادگیری در شبکه‌های عصبی که در آن سعی می‌شود میزان ریسک شهودی^۲ کاهش یابد، در ساختار آموزشی ماشین بردار پشتیبان سعی می‌شود ریسک ساختاری^۳ کاهش پیدا کند (Burgess, 1998). این کار با مینیمم کردن یک حد آستانه بالا برای خطای تعمیم‌پذیری به جای خطای آموزشی به انجام می‌رسد. بدین ترتیب امکان حل مسأله بیش تطبیقی^۴ در یادگیری ماشین قابل حل خواهد شد.

در مقایسه با شبکه عصبی، یکی دیگر از ویژگی‌های برتر ماشین بردار پشتیبان به نحوه آموزش آن مربوط می‌شود. مسأله آموزش ماشین بردار پشتیبان را می‌توان به یک مسأله برنامه‌نویسی مرتبه دو محدودشده خطی تبدیل کرد. این آموزش قابلیت ارائه یک راه‌حل یکتا و بهینه سراسری را دارد. یکی از ویژگی‌های مدل‌های مبتنی بر ماشین بردار پشتیبان توانایی بالای آنها در کار با ابعاد بالاست، زیرا بسیاری از مسائل که حل آنها در ابعاد معمولی میسر نیست، در صورتی که به ابعاد بالاتر منتقل شوند، قابلیت حل راحت‌تری را پیدا می‌کنند (Chen., Lin & Scholkopf, 2005). در سال‌های اخیر از مدل ماشین بردار پشتیبان در کاربردهای مختلف مالی و اقتصادی استفاده شده است که بخش مهمی از آن در رابطه با حل مسائل دسته‌بندی و پیش‌بینی سری‌های زمانی^۵ است.

بر اساس اغلب تحقیقات انجام‌شده در حوزه امتیازبندی اعتباری و پیش‌بینی ورشکستگی^۶، ماشین بردار پشتیبان در اکثر مواقع بر دیگر مدل‌های دسته‌بندی آماری و هوشمند برتری دارد. از آن جمله می‌توان به تحقیق یان^۷ و پژوهش تسایی^۸ اشاره کرد که

1- Vapnik

2- Empirical Risk

3- Structural Risk

4- Over/Fitting

5- Time Series Prediction

6- Bankruptcy Prediction

7- Yuan (2007)

8- Tsai (2008)

در آنها دقت مدل‌های ماشین بردار پشتیبان بالاتر از مدل‌های آماری و شبکه عصبی تخمین زده شده است.

شبکه‌های بیزین^۱ در ردیف مدل‌های تصمیم‌گیری گرافیکی و تحلیلگران تصمیم‌گیری قرار می‌گیرد. یک شبکه بیزین قابلیت کار با داده‌های کیفی و کمی را به صورت همزمان دارد، بدین معنا که در این مدل می‌توان هم از داده‌های آماری و هم از تجربه متخصصان تحلیل داده استفاده کرد (Abramowicz et al, 2003).

سابقه استفاده از مدل‌های مبتنی بر شبکه بیزین در شناسایی میزان مخاطرات مالی حتی بیشتر از تحلیل ممیز است. اولین بار بیور^۲ در ۱۹۶۶ ایده به‌کارگیری تحلیل مبتنی بر نرخ‌های مشابهت^۳ بر اساس رویکرد بیزین را معرفی کرد (Abramowicz et al, 2003). وی نشان داد که مسأله پیش‌بینی قصور را می‌توان به عنوان یک احتمال شرایط مخاطره‌آمیز مالی با توجه به یک معیار مالی معین بررسی کرد. در برخی تحقیقات^۴ به میزان دقت مدل‌های مبتنی بر بیزین در مقایسه با روش‌های آماری و مبتنی بر داده کاوی (مانند شبکه عصبی و ماشین بردار پشتیبان) توجه شده است که در پایان، عملکرد مدل‌های بیزین نشان می‌دهد قابلیت رقابت با آنها را داشته است.

در روش‌های مبتنی بر درخت تصمیم با پیاده‌سازی یک استراتژی بالا به پایین تقسیم و غلبه، مجموعه داده مورد بررسی به صورت بازگشتی به بخش‌های کوچک‌تر تقسیم‌بندی می‌شود. این فرایندها پایه مجموعه‌ای از آزمون‌های تعریف‌شده در هر شاخه درخت را شکل می‌دهد. ساختار درخت مانند این روش متشکل از سه جزء ریشه، مجموعه‌ای از گره‌های داخلی و مجموعه‌ای از گره‌های نهایی (برگ) است (Daubie et al, 2002).

کوینلن^۵ اولین بار در ۱۹۸۶ روش قطعه‌بندی متقابل^۶ ۱۹۸۶ و بعد از آن روش C4.5 را در ۱۹۹۳ معرفی کرد (Zhou et al, 2008). این دو روش مهم‌ترین موضوعات مورد بررسی

1- Bayesian Networks

2- Beaver

3- Likelihood

4- McNeil & Wending, (2007); Loffler, et al, (2004)

5- Quinlan

6- Interactive Dichotomizer 3 (ID3)

در یادگیری درخت تصمیم محسوب می‌شوند. در بخش ایجاد درخت تصمیم در روش ID3 سؤال اصلی انتخاب آزمونی است که بتواند بهترین دسته‌بندی را در نمونه‌ها ایجاد کند. در عین حال روش C4.5 از یک تابع ارزیابی میزان انترپوی در داده‌ها به عنوان محدودیت انتخاب استفاده می‌نماید. درخت رگرسیون‌گیری و دسته‌بندی^۱ نوع دیگری از درخت‌های تصمیم است که در اصل یک فرایند آماری است و اولین بار توسط بریمن^۲ در ۱۹۸۴ معرفی شد. از این روش برای دسته‌بندی و یا تحلیل داده‌های پیوسته و اسمی استفاده می‌شود (Lee et al, 2006).

۲. مجموعه داده‌های به کار رفته

به منظور ارزیابی درست عملکرد هر کدام از مدل‌های مورد مطالعه در این تحقیق از سه مجموعه داده متفاوت استفاده شده است که برای انتخاب آنها سعی شده ابعاد مختلف مورد توجه قرار داده شوند. بدین منظور دو مجموعه داده محک استاندارد^۳ که در بیشتر مقالات حوزه امتیازبندی اعتباری مورد توجه قرار داده می‌شوند، شامل داده‌های آلمان و استرالیا از پایگاه داده دانشگاه کالیفرنیا^۴ و یک مجموعه داده داخلی از داده‌های اعتباری اشخاص حقوقی متقاضی وام از بانک کارآفرین^۵ که به نام داده‌های ایران شناخته می‌شود، استفاده شده است. در جدول ۱ ویژگی‌های هر کدام از این داده‌ها نشان داده شده است. در هر کدام از این داده‌ها مشتریان بد با یک و مشتریان خوب با صفر نشان داده شده‌اند.

1- Classification and Regression Tree (CART)

2- Breiman

3- Benchmark Datasets

4- UCI Data Repository (<http://archive.ics.uci.edu/ml/datasets.html>)

5- Karafarin Bank

جدول (۱) خصوصیات هر کدام از مجموعه داده‌ها

فیلد خالی	نوع داده‌ها	نسبت نمونه‌های کلاس یک (درصد)	تعداد ویژگی‌ها	تعداد رکوردها	مجموعه داده
ندارد	حقیقی	۳۰	۲۰	۱۰۰۰	آلمان
دارد	حقیقی	۵۰	۱۴	۶۹۰	استرالیا
دارد	حقوقی	۱۵	۶۶	۱۹۹۴	ایران

۳. معیارهای ارزیابی مدل‌ها

در تحقیقات مختلف از ابزارهای گوناگونی برای سنجش میزان کارایی مدل‌ها استفاده شده است. هر کدام از این ابزارها به مسائل ویژه‌ای توجه می‌کنند و از دیدگاه‌های متفاوتی اقدام به بررسی مدل‌ها می‌نمایند، در ادامه به چند مورد از مهم‌ترین معیارهای به کار رفته در مسأله امتیازبندی اعتباری اشاره می‌شود.

۳-۱. خطای دسته‌بندی

یکی از ساده‌ترین و پرکاربردترین معیارهای ارزیابی مدل‌های دسته‌بندی‌کننده استفاده از نمونه‌هایی است که به اشتباه دسته‌بندی شده‌اند. بدین منظور می‌توان برای مشخص کردن میزان خطای مدل دسته‌بندی از تعداد نمونه‌هایی که به اشتباه دسته‌بندی شده‌اند به کل نمونه‌ها استفاده کرد. جهت بررسی بهتر نتایج برای هر دسته‌بندی‌کننده دو کلاسی می‌توان ماتریس درست‌نمایی^۱ در جدول ۲ را ترسیم کرد که در آن میزان درست‌نمایی مدل در هر کدام از کلاس‌ها به صورت جداگانه نشان داده شده است.

1- Confusion Matrix

نمودار (۱) ماتریس درست‌نمایی در یک دسته‌بندی‌کننده دو کلاسی

		کلاس واقعی	
		تعداد نمونه‌های مثبت	تعداد نمونه‌های منفی
کلاس پیش‌بینی‌شده	تعداد کل نمونه‌های مثبت	نمونه‌های مثبت درست	نمونه‌های مثبت نادرست
	تعداد کل نمونه‌های منفی	نمونه‌های منفی نادرست	نمونه‌های منفی درست
	جمع	تعداد کل	تعداد کل

مأخذ: H.-C.Yeh, M.-L.Yang, L.-C.Lee, 2007

با توجه به جدول ۲ می‌توان علاوه بر میزان خطای آزمایشی مدل از عملکرد مدل در ابعاد مختلف نیز اطلاع حاصل کرد. برای مثال در این حالت می‌توان از میزان خطای مدل در دسته‌بندی نمونه‌های کلاس صفر و کلاس یک که به ترتیب خطای نوع اول^۱ و خطای نوع دوم^۲ نامیده می‌شوند، به صورت جداگانه اطلاع پیدا کرد.

$$Error = \frac{TP + TN}{P + N} \quad (۱)$$

$$False\ Positive\ (FP)Rate = \frac{FP}{N} \quad (۲)$$

$$False\ Negative\ (FN)Rate = \frac{FN}{P} \quad (۳)$$

هر یک از معیارهای مورد اشاره را می‌توان به تنهایی برای نشان دادن میزان کارایی مدل‌های مختلف استفاده کرد. معیار خطای آزمایشی در بیشتر تحقیقات تنها معیاری

1- FN Rate

2- FP Rate

است که برای نشان دادن میزان کارایی به آن توجه می‌شود. در حالی که نمی‌توان تنها با تکیه بر آن، میزان توانایی مدل‌ها را تخمین زد، زیرا وقتی تعداد نمونه‌های یکی از کلاس‌ها بسیار بیشتر از دیگری باشد، توانایی این معیار در تعیین میزان درستی به شدت افت خواهد کرد. برای مثال فرض کنید در یک مجموعه داده ۹۵ درصد از داده‌ها در یک کلاس و بقیه در کلاس دیگر قرار داشته باشند. در این حالت دسته‌بندی‌کننده‌ای که بتواند بدون هیچ منطقی کل نمونه‌ها را در کلاس اکثریت قرار دهد، با این معیار دقتی برابر با ۹۵ درصد خواهد داشت.

از سوی دیگر باید در نظر داشت که خطای نوع اول و دوم ماهیتی متضاد دارند و توجه به یکی از آنها موجب تضعیف دیگری خواهد شد. در این راستا معیارهای ارزیابی دیگری مانند سطح زیر منحنی که خود از معیار دیگری به نام منحنی گیرنده ویژگی مشتق شده‌اند نیز معرفی شده که در آنها سعی می‌شود با توجه تقریباً یکسان به هر دو نوع خطای نوع اول و دوم معیار مناسبی ایجاد شود.

در اصل منحنی ROC روشی برای آشکارسازی میزان کارایی یک دسته‌بندی‌کننده در زمینه رعایت توازن بین میزان حساسیت^۱ و تشخیص^۲ است. از این روش اولین بار در تئوری شناسایی سیگنال و برای تعیین میزان دقت یک گیرنده جهت ایجاد تمایز بین یک سیگنال واقعی از نویز استفاده شد. بدین منظور از مختصات دو بعدی که در محور افقی آن FP و در محور عمودی آن TP قرار داده شده است، استفاده می‌شود. هر چه منحنی به نقطه (0,1) تمایل بیشتری نشان دهد، میزان کارایی مدل دسته‌بندی‌کننده نیز افزایش خواهد یافت (Mahlmann, 2004).

برای کمی کردن معیار فوق می‌توان با محاسبه سطح زیر منحنی^۳ میزان کارایی آن را با دقت بیشتری نشان داد. در عین حال از این مفهوم برای ایجاد یک آماره دیگر به نام ضریب جینی^۴ نیز استفاده می‌شود. در واقع ضریب جینی دو برابر سطح زیر نمودار ROC تا خط قطری است.

1- Sensitivity

2- Selectivity

3- Area Under the Curve (AUC)

4- Gini Coefficient

با وجود اینکه معیارهای مشتق‌شده از منحنی ROC به هر دو خطای نوع اول و دوم توجه می‌کنند، ولی در حقیقت اندازه این دو معیار به یک اندازه مورد توجه قرار داده می‌شود، در حالی که ممکن است در بعضی از مسائل میزان اهمیت آنها به یک اندازه نبوده و دارای مقادیر مختلفی باشند. امتیازبندی اعتباری یکی از این موارد است که در آن میزان اهمیت خطای نوع دوم به مراتب بیشتر از خطای نوع اول است. با وجود اهمیت بسیار زیاد این موضوع، عملاً در تعداد کمی از تحقیقات انجام‌شده در حوزه امتیازبندی اعتباری می‌توان مواردی را مشاهده کرد که به این امر توجه کرده باشند.

اولین بار وست (West, 2000) با اندازه‌گیری میزان هزینه دسته‌بندی نادرست هر کدام از کلاس‌ها بیان کرد که به صورت تقریبی میزان هزینه دسته‌بندی نادرست یک نمونه دسته یک (مشتری بد) نسبت به یک نمونه دسته صفر (مشتری خوب) ۵ به ۱ است. پس از آن در تحقیقات دیگری (مانند Abdou et al, 2008) از این نسبت برای تعیین میزان هزینه دسته‌بندی هر کدام از کلاس‌ها استفاده شد.

با فرض درست بودن نسبت‌های به دست آمده در این تحقیق، می‌توان از آن برای ارزیابی میزان دقت هر کدام از مدل‌های دسته‌بندی استفاده کرد. از طرف دیگر با توجه به ماهیت متضاد خطای نوع اول و دوم مسأله به یک مسأله ارضای محدودیت دو بعدی تبدیل می‌شود که در آن با توجه به میزان اهمیت هر کدام از شرایط باید نقطه بهینه مناسب به دست آید. برای رسیدن به این مقصود لازم است تا با ارائه یک تابع هزینه که دربرگیرنده همه جوانب امر باشد، نسبت به مینیمم کردن آن اقدام نمود. یقیناً تابع هزینه شامل هر دو خطای نوع اول و دوم خواهد بود که به نسبت میزان اهمیت هر کدام در نهایت تابع هدف نهایی را شکل خواهند داد اما قبل از عملی کردن این مفهوم لازم است مسأله دیگری نیز مدنظر قرار داده شود.

با وجود اهمیت هر کدام از نمونه‌ها از نظر کلاس هر کدام، باید به فراوانی نمونه‌های هر کدام از دسته‌ها نیز توجه کرد، زیرا توجه بیش از اندازه به میزان درست‌نمایی مدل در دسته‌بندی یک کلاس در صورتی که نمونه‌های موجود در آن در اقلیت باشند، می‌تواند باعث کاهش میزان کارایی مدل دسته‌بندی گردد. از این‌رو باید نسبت نمونه‌های موجود در

هر کدام از کلاس‌ها را نیز در تابع هزینه به دست‌آمده لحاظ نمود. بدین ترتیب می‌توان تابع هزینه نهایی را به صورت رابطه زیر نشان داد:

$$Cost_{Function} = C_{12}f_2FP + C_{11}f_1FN \quad (۴)$$

که در آن $C_{12}=5$ و $C_{11}=1$ قرار داده شده و f_1 و f_2 به ترتیب نسبت نمونه‌های کلاس صفر و یک هستند. در این رابطه FP و FN نیز به ترتیب میزان خطای نوع اول و دوم بعد از دسته‌بندی نمونه‌ها قرار داده شده‌اند.

۲-۳. نتایج عملی

برای تعیین میزان کارایی هر کدام از مدل‌های ارائه‌شده در این تحقیق بعد از عملیات پیش‌پردازش داده‌ها^۱ و حذف داده‌های پرت و ویژگی‌های اضافی اقدام به آموزش و آزمون هر کدام از آنها شده است. به منظور اجتناب از بیش‌تطبیقی مدل‌ها بر داده‌های آموزشی از روش اعتبارسنجی متقاطع k تایی^۲ با مقدار $k=3$ در محیط نرم‌افزار مطلب نسخه R2009a استفاده شده است.

به منظور در اختیار گرفتن یک افق دید مناسب برای ارزیابی نتایج از ۵ معیار ارزیابی استفاده شده است که عبارت‌اند از: خطای آزمایشی، خطای نوع اول، خطای نوع دوم، سطح زیر منحنی و تابع هزینه. با در نظر گرفتن هر کدام از این معیارها می‌توان ارزیابی مناسبی از نحوه عملکرد مدل‌ها در اختیار داشت تا در نهایت نقاط ضعف و قوت هر کدام مشخص شود.

۱- آماده‌سازی داده‌های خام برای استفاده در مدل‌ها (Preprocessing)

2- K-fold Cross Validation

۳-۳. عملکرد مدل‌ها

- مدل‌های آماری

در اینجا دو روش لاجیت و پروبیت از بین روش‌های آماری مهم استفاده شده و نتایج آن در جدول ۲ نمایش داده شده است.

جدول (۲) عملکرد مدل‌های آماری از نظر معیارهای خطای آزمایشی، خطای نوع اول، خطای نوع دوم، سطح زیر منحنی و تابع هزینه برای هر کدام از مجموعه داده‌ها

کشور	Fun	Error	FN	FP	AUC	Cost_Fn
آلمان	لاجیت	۰/۲۳۹۳	۰/۱۱۲۴	۰/۵۳۵۶	۰/۷۸۱۳	۰/۸۸۲
	پروبیت	۰/۲۳۸۷	۰/۱۱۱۴	۰/۵۳۵۶	۰/۷۸۲۴	۰/۸۸۱۳
استرالیا	لاجیت	۰/۱۴۰۸	۰/۱۲۹۸	۰/۱۵۴۳	۰/۹۱۹۸	۰/۴۱۵۵
	پروبیت	۰/۱۳۸۸	۰/۱۳۳۳	۰/۱۴۵۷	۰/۹۲۰۲	۰/۳۹۸
ایران	لاجیت	۰/۱۵۴۴	۰/۰۱۲۷	۰/۹۶	۰/۶۰۴۶	۰/۷۳۰۸
	پروبیت	۰/۱۵۴	۰/۰۱۱۴	۰/۹۶۵	۰/۶۰۵۳	۰/۷۳۳۵

• روش‌های یادگیری ماشین

- نزدیک‌ترین همسایه

یکی از پارامترهای مهم مدل نزدیک‌ترین همسایه تعیین اندازه همسایگی است که با توجه به نتایج به دست آمده از دو مقدار ۱ و ۳ برای آن استفاده شده است. همچنین در کلیه موارد از روش فاصله‌یابی اقلیدسی برای محاسبه نتایج استفاده شده است.

جدول (۳) عملکرد مدل نزدیک‌ترین همسایه از نظر معیارهای خطای آزمایشی، خطای نوع اول، خطای نوع دوم، سطح زیر منحنی و تابع هزینه برای هر کدام از مجموعه داده‌ها

Cost_Fn	AUC	FP	FN	Error	k	کشور
۰/۸۹۲۷	۰/۶۵۳۲	۰/۵۰۸۹	۰/۱۸۴۸	۰/۲۸۲	۱	آلمان
۰/۹۳	۰/۶۵۵۷	۰/۵۶	۰/۱۲۸۶	۰/۲۵۸	۳	
۱/۱۱	۰/۶۴۴۶	۰/۴۲۸۳	۰/۲۸۲۵	۰/۳۴۷۶	۱	استرالیا
۱/۱۱۴	۰/۶۶۶۶	۰/۴۴۵۷	۰/۲۲۱۱	۰/۳۲۱۴	۳	
۰/۶۷۶	۰/۵۵۸۶	۰/۷۴۲۵	۰/۱۴۰۲	۰/۲۳۰۳	۱	ایران
۰/۶۶۳	۰/۵۶۶۹	۰/۷۳۲۵	۰/۱۳۳۶	۰/۲۲۳۲	۳	

- بیزین ساده

در مدل بیزین ساده می‌توان از انواع مختلف توابع استفاده نمود که در اینجا از دو تابع Uniform و Empirical استفاده شده و نتایج هر کدام از این دو تابع در جدول ۴ نمایش داده شده است.

جدول (۴) عملکرد مدل بیزین ساده تصمیم از نظر معیارهای خطای آزمایشی، خطای نوع اول، خطای نوع دوم، سطح زیر منحنی و تابع هزینه برای هر کدام از مجموعه داده‌ها

Cost_Fn	AUC	FP	FN	Error		کشور
۰/۶۱۹۳	۰/۷۰۳۷	۰/۲۵۵۶	۰/۳۳۷۱	۰/۳۱۲۷	Uniform	آلمان
۰/۷۴۲۳	۰/۶۹۲	۰/۳۸۸۹	۰/۲۲۷۱	۰/۲۷۵۷	Empirical	
۰/۶۹۷۳	۰/۸۱۱۷	۰/۲۹۲۴	۰/۰۸۴۲	۰/۱۷۷۲	Uniform	استرالیا
۰/۷۲۳۹	۰/۸۰۷۴	۰/۳۰۵۴	۰/۰۷۹۸	۰/۱۸۰۶	Empirical	
۰/۸۴۰۸	۰/۵۲۴۴	۰/۸۷۸۸	۰/۲۱۳۸	۰/۳۱۳۳	Uniform	ایران
۰/۸۵۰۲	۰/۵۱۶۵	۰/۹	۰/۲۰۶۲	۰/۳۰۹۹	Empirical	

– شبکه عصبی پیشران

در بین ساختارهای مختلف مدل‌های مبتنی بر شبکه عصبی از ساختار شبکه عصبی پیشران با سه لایه که تعداد نرون‌های لایه اول برابر با تعداد ویژگی‌های هر مجموعه داده می‌باشد، استفاده شده است. در لایه دوم هم با توجه به عملکرد مدل تقریباً از دو برابر تعداد نرون‌های لایه اول استفاده شده است.

جدول (۵) عملکرد مدل شبکه عصبی پیشران از نظر معیارهای خطای آزمایشی، خطای نوع اول، خطای نوع دوم، سطح زیر منحنی و تابع هزینه برای هر کدام از مجموعه داده‌ها

کشور	Error	FN	FP	AUC	Cost_Fn
آلمان	۰/۲۵۸	۰/۱۲۱۹	۰/۵۷۵۶	۰/۶۵۱۳	۰/۹۴۸۷
استرالیا	۰/۱۴۷۶	۰/۱۳۲۵	۰/۱۶۶۳	۰/۸۵۰۶	۰/۴۴۳۵
ایران	۰/۱۵۵۳	۰/۰۰۹	۰/۹۸۷۵	۰/۵۳۶۵	۰/۷۴۸۳

– ماشین بردار پشتیبان

در مدل ماشین بردار پشتیبان به کاررفته در اینجا از تابع کرنل خطی با مقادیر خطای قابل تنظیم که به صورت آزمون و خطا مشخص گردیده، استفاده شده است.

جدول (۶) عملکرد مدل ماشین بردار پشتیبان از نظر معیارهای خطای آزمایشی، خطای نوع اول، خطای نوع دوم، سطح زیر منحنی و تابع هزینه برای هر کدام از مجموعه داده‌ها

کشور	Error	FN	FP	AUC	Cost_Fn
آلمان	۰/۲۴۷	۰/۱۱۴۳	۰/۵۵۶۷	۰/۶۶۴۵	۰/۹۱۵
استرالیا	۰/۱۴۶۶	۰/۱۹۹۱	۰/۰۸۱۵	۰/۸۵۹۷	۰/۲۹۱۹
ایران	۰/۱۵۴۶	۰/۰۰۶۲	۰/۹۹۸۸	۰/۴۹۷۵	۰/۷۵۴۲

- درخت تصمیم

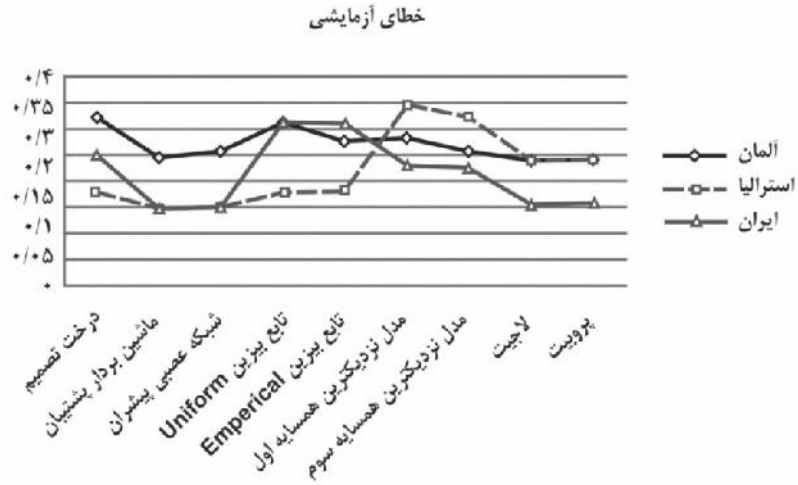
جدول (۷) عملکرد مدل درخت تصمیم از نظر معیارهای خطای آزمایشی، خطای نوع اول، خطای نوع دوم، سطح زیر منحنی و تابع هزینه برای هر کدام از مجموعه داده‌ها

کشور	Error	FN	FP	AUC	Cost_Fn
آلمان	۰/۳۱۹۰	۰/۲۴۴۸	۰/۴۹۲۲	۰/۶۳۱۵	۰/۹۰۹۷
استرالیا	۰/۱۷۵۲	۰/۱۴۹۱	۰/۲۰۷۶	۰/۸۲۱۶	۰/۵۴۴۷
ایران	۰/۲۵۴۸	۰/۱۶۵۷	۰/۷۶۱۳	۰/۵۳۶۵	۰/۷۱۱۸

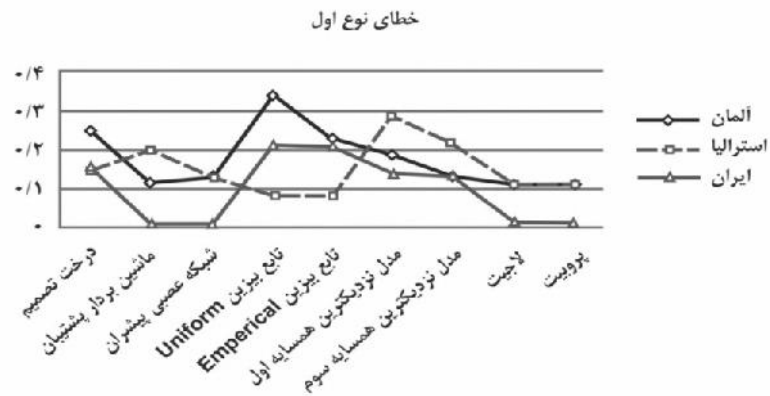
۳-۴. مقایسه عملکرد مدل‌ها

با توجه به نمودارهای ۲ و ۳ می‌توان گفت میزان خطای آزمایشی و خطای نوع اول در مدل‌های شبکه عصبی و ماشین بردار پشتیبان قابل مقایسه با مدل‌های آماری لاجیت و پروبیت است. در این خصوص عملکرد بقیه مدل‌ها چندان مطلوب نیست، در عین حال با توجه به نمودار ۴ خطای نوع دوم دارای الگوی مشخصی نیست و بسته به نوع داده‌های به کاررفته متفاوت است. مثلاً در داده‌های استرالیا عملکرد مدل‌های شبکه عصبی و ماشین بردار پشتیبان بسیار بهتر از مدل‌های دیگر بوده است، در حالی که در داده‌های ایران مدل نزدیک‌ترین همسایه و در داده‌های آلمان مدل بیزین نتایج بهتری را نشان داده است، اما آنچه مسلم می‌باشد، این است که میانگین نتایج مدل‌های آماری ضعیف‌تر از بقیه بوده است.

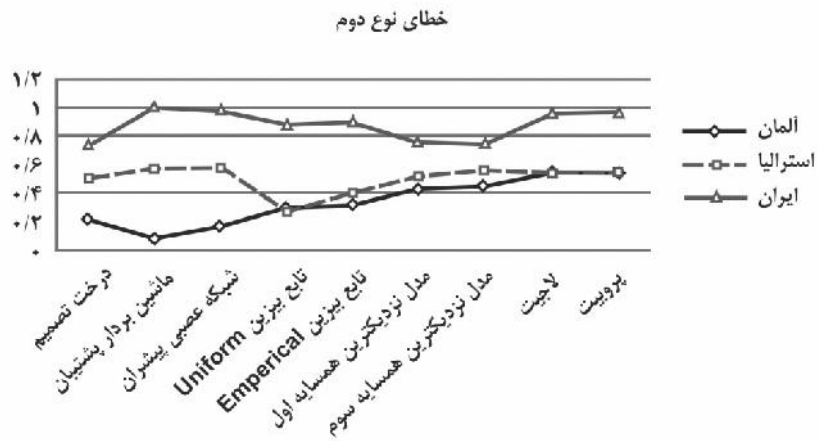
نمودار (۲) اندازه خطای آزمایشی در هر کدام از مدل‌ها



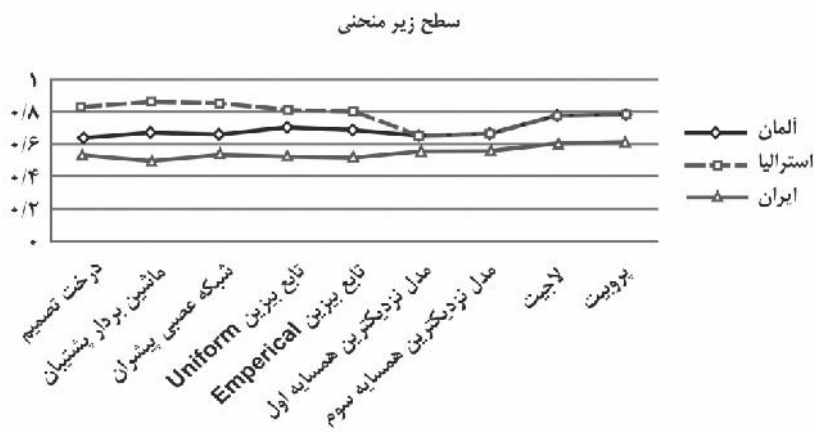
نمودار (۳) اندازه خطای نوع اول در هر کدام از مدل‌ها



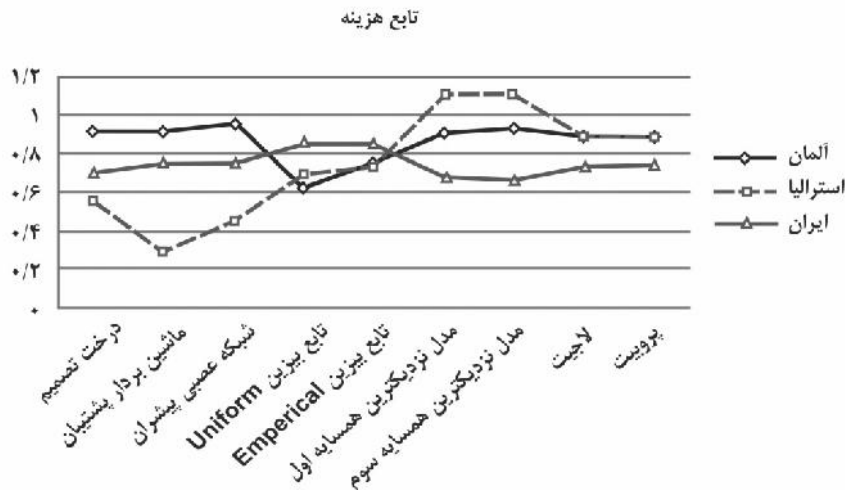
نمودار (۴) اندازه خطای نوع دوم در هر کدام از مدل‌ها



نمودار (۵) اندازه سطح زیر منحنی در هر کدام از مدل‌ها



نمودار (۶) اندازه تابع هزینه در هر کدام از مدل‌ها



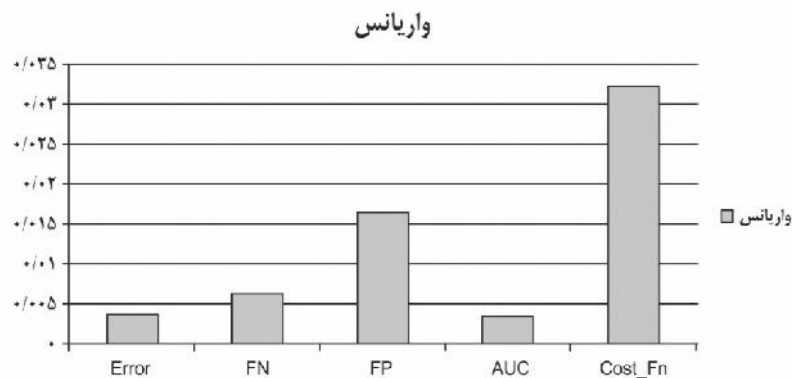
با توجه به نتایج به دست آمده در دیگر شکل‌ها می‌توان مشاهده کرد که در اغلب موارد عملکرد مدل‌ها از نظر خطای آزمایشی و خطای نوع اول با خطای نوع دوم نسبت معکوس دارد. برای تجمیع نتایج هر دو خطای نوع اول می‌توان به نمودار ۵ اشاره کرد که در آن از معیار سطح زیر منحنی استفاده شده است. در این حالت بر خلاف بقیه معیارها هر چه سطح زیر منحنی بیشتر باشد، عملکرد مدل بهتر ارزیابی می‌شود. بدین ترتیب می‌توان گفت اولاً اغلب مدل‌ها تفاوت اندکی را با هم نشان می‌دهند و از طرف دیگر تقریباً همان نتایجی که برای معیارهای خطای آزمایشی و خطای نوع دوم به دست آمده، می‌تواند صحت داشته باشد. دلیل اصلی این امر به این واقعیت برمی‌گردد که در معیار سطح منحنی به دلیل توجه یکسان به هر دو نوع خطای نوع اول و دوم، فراوانی نمونه‌های هر کدام از کلاس‌ها نقش پررنگ‌تری را ایفا می‌کنند و باعث می‌شوند تا نتایج به دست آمده بیشتر به خطای نوع اول شباهت داشته باشد. در عین حال معیار سطح زیر منحنی نتوانسته بین مدل‌های مختلف تفاوت چشمگیری ایجاد نماید.

بررسی نتایج نمودار ۶ نشان می‌دهد که از نظر میزان تابع هزینه باز هم نمی‌توان به صورت کلی در مورد میزان کارایی بهترین مدل‌ها اظهار نظر کرد، در حالی که در داده‌های

استرالیا مدل‌های داده‌کاوی مانند ماشین بردار پشتیبان و شبکه عصبی بهتر از بقیه مدل‌ها بوده‌اند، در داده‌های آلمان بجز مدل‌های مبتنی بر بی‌زین بقیه مدل‌ها تفاوت اندکی را نسبت به هم نشان می‌دهند و در نهایت در داده‌های ایران مدل‌های نزدیک‌ترین همسایه و درخت تصمیم عملکرد بهتری را داشته‌اند.

در عین حال بر خلاف معیار سطح زیر منحنی در معیار تابع هزینه میزان تفاوت نتایج در هر کدام از داده‌ها افزایش پیدا کرده است. برای نشان دادن این موضوع می‌توان از میانگین واریانس هر کدام از معیارها در هر سه مجموعه داده استفاده کرد. با توجه به نتایج نشان داده شده در نمودار ۷ می‌توان گفت تابع هزینه دارای بیشترین میزان واریانس است که نشان می‌دهد معیار مذکور توانایی بالایی در تفکیک مدل‌های مختلف دارد.

نمودار (۷) میانگین مقدار واریانس هر کدام از معیارها در هر کدام از داده‌ها



۴. بحث و نتیجه‌گیری

برای تعیین میزان کارایی هر یک از مدل‌ها باید گفت یقیناً نحوه عملکرد هر کدام وابستگی زیادی به نوع داده‌های به‌کاررفته در آنها دارد. مثلاً در داده‌های استرالیا که تقریباً نسبت یکسانی از داده‌ها در آن مشاهده می‌شود، باید انتظار داشت میزان تفاوت چندان زیادی بین خطای نوع اول و دوم مشاهده نشود، در حالی که در داده‌های آلمان و به خصوص داده‌های ایران این موضوع کاملاً محسوس و تابعی از نسبت داده‌های مورد استفاده خواهد بود.

با وجود این باید در نظر داشت که در این تحقیق هدف تعیین میزان کارایی مدل‌ها نیست، بلکه منظور اصلی بررسی میزان کارایی معیارهای ارزیابی برای نشان دادن بهترین مدل‌ها در مسأله امتیازبندی اعتباری می‌باشد. به نتایج به دست آمده از پیاده‌سازی مدل‌ها مؤید این نکته است که اساساً نمی‌توان مدلی را انتخاب نمود که در همه مجموعه داده‌ها دارای بهترین عملکرد باشد، اما در عین حال می‌توان در مورد انتخاب روش ارزیابی مناسب با دقت بیشتری عمل نمود و آن را به عنوان شاخصی برای انتخاب مدل‌ها مدنظر قرار داد. انجام این امر اهمیت به مراتب بیشتری نسبت به انتخاب خود مدل دارد.

در این تحقیق ضمن معرفی هفت روش مهم برای ایجاد مدل‌های امتیازبندی اعتباری، نتایج به کارگیری آنها در سه مجموعه داده اعتباری مختلف نشان داده شده است. بدین منظور با توجه به معیارهای ارزیابی مهم که برای تعیین میزان کارایی مدل‌ها استفاده می‌شوند، مجموعاً از پنج معیار مهم استفاده شده است که در نهایت نتایج زیر قابل ذکر می‌باشند:

- از نظر کیفی معیار خطای آزمایشی، به خصوص در داده‌هایی که دچار عدم توازن در تعداد نمونه‌های هر دسته می‌باشند، ضعیف‌ترین نتایج را دارد.
- خطای نوع اول و دوم به دلیل توجه یک بعدی به مسأله نمی‌توانند به تنهایی معیار مناسبی برای سنجش کارایی مدل‌ها تلقی شوند.
- سطح زیر منحنی علی‌رغم توجه به هر دو خطای نوع اول و دوم، قابلیت تفکیک‌کنندگی کمی را (به دلیل توجه یکسان به هر دو نوع خطا) دارد، در نتیجه نمی‌تواند معیار مناسبی برای مسائل مبتنی بر تابع هزینه مانند امتیازبندی اعتباری باشد.
- تابع هزینه بر خلاف معیارهای دیگر هم قابلیت اعمال محدودیت‌های چندگانه را دارد و هم می‌توان در آن میزان اهمیت هر کدام از محدودیت‌ها را به دلخواه تعیین کرد. از طرف دیگر با توجه به قابلیت جداسازی بالاتر آن نسبت به معیارهای دیگر مانند سطح زیر منحنی، در هنگام استفاده کارایی بیشتری را به نمایش خواهد گذاشت و از این‌رو مدل‌ها را با دقت بیشتری می‌توان مطالعه کرد.

با توجه به نتایج به دست آمده و با انتخاب درست یک تابع هزینه مناسب می‌توان با اطمینان بیشتری در خصوص میزان کارایی مدل‌های مورد مطالعه اقدام نمود که قدم اول در طراحی هر مدل شناسایی الگو قلمداد می‌شود. در مرحله بعد می‌توان به جوانب دیگر مسأله امتیازبندی اعتباری پرداخت، از جمله اینکه تاکنون بیشتر مدل‌های مورد مطالعه تنها می‌توانند برای دسته‌بندی دو کلاسی مورد استفاده قرار گیرند، در حالی که در امتیازبندی اعتباری باید بتوان علی‌رغم وجود داده‌های اولیه دو کلاسی، نتایج نهایی را به صورت چندکلاسی نیز دسته‌بندی نمود. همچنین بیشتر مدل‌های به دست آمده قابلیت تفسیرپذیری پایینی دارند که برای مسأله امتیازبندی اعتباری مناسب نیست. انجام این مراحل را می‌توان به عنوان کارهای آتی مؤلفین در نظر گرفت.

منابع و مآخذ

- Abdou, H., Pointon, J. & El-Masry, A. (2008). Neural Nets Versus Conventional Techniques in Credit Scoring in Egyptian Banking. *Expert Systems with Applications*, 35, 1275-1292.
- Abramowicz, W., Nowak, M. & Sztykiel, J. (2003). *Bayesian Networks as a Decision Support Tool in Credit Scoring Domain*. The Poznan University of Economics.
- Burges, C. J.C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discovery*, 2, 121-167.
- Chen, P.-H., Lin, C.-J. & Scholkopf, B. (2005). A Tutorial on N-Support Vector Machines, *Applied Stochastic Models in Business and Industry*, 21, 111-136.
- Daubie, M., Levecq, P. & Meskens, N. (2002). A Comparison of The Rough Sets and Recursive Partitioning Induction Approaches: An Application to Commercial Loans. *International Transactions in Operational Research*, 9, 681-694.
- Donga, G., Laib, K. K. & Yen, J. (2010). Credit Scorecard Based on Logistic Regression with Random Coefficients, *Procedia Computer Science*, 1(1), 2463-2468.
- Giudici, P. (2001). Bayesian Data Mining, With Application to Bench Marking and Credit Scoring, *Applied Stochastic Models in Business and Industry*, 17, 69-81.
- Greene, W. (1998). Sample Selection in Credit-Scoring Models. *Japan and the World Economy*, 10, 299-316.
- Hand, D. J. N. & Adams. M. (2000). Defining Attributes for Scorecard Construction in Credit Scoring, *Journal of Applied Statistics*, 27, (5), 527-540.

-
- Huang, S. J. & Yu, J. (2010). Bayesian Analysis of Structural Credit Risk Models with Micro Structure Noises. *Journal of Economic Dynamics & Control*, 34, 2259-2272.
 - Jiang, M. -H. & Yuan, X. -C. (2007). Construction and Application of PSO-SVM Model for Personal Credit Scoring. *Springer-Verlag Berlin Heidelberg, ICCS, Part IV, LNCS 4490*. (pp. 158-161).
 - Khashman, A. (2010). Neural Networks for Credit Risk Evaluation: Investigation of Different Neural Models and Learning Schemes, *Expert Systems with Applications*, 37, 6233-6239.
 - Kiani, M. F. & Mahmoudi, F. (2010). A New Hybrid Method for Credit Scoring Based on Clustering and Support Vector Machine (ClsSVM); *International Conference on Financial Engineering (ICIF, 2010)*, (585-589).
 - Kim, J. (2005). *A Credit Risk Model for Agricultural Loan Portfolios Under the New Basel Capital Accord*, PhD thesis, Texas A&M University.
 - Komorad, K. (2002). On Credit Scoring Estimation. Institute for Statistics and Econometrics Humboldt University, Berlin.
 - Leea, T.- S., Chiub, C.- C., Chouc, Y.- C. & Lud, C. -J. (2006). Mining the Customer Credit Using Classification and Regression Tree and Multivariate Adaptive Regression Splines, *Computational Statistics & Data Analysis*, 50, 1113-1130.
 - Loffler, G., Posch, P. N. & Schone, C. (2004). *Bayesian Methods for Improving Credit Scoring Models*. Department of Finance, University of Ulm, Helmholtzstrasse, Ulm, Germany.
 - Mahlmann, T. (2004). *Classification and Rating of Firms in the Presence of Financial and Non-Financial Information*. Chair of Banking, University of Cologne.
 - McNeil, A. J. & Wendin, J. P. (2007). Bayesian Inference for Generalized Linear Mixed Models of Portfolio Credit Risk. *Journal of Empirical Finance*, 14, (2), 131-149.

-
- Plantamura, V. L., Soucek, B. & Visaggio, G. (1993). The Holographic Fuzzy Learning for Credit Scoring. *Proceedings of 1993 International Joint Conference on Neural Networks*, 729-732.
 - Tsai, C.-F. (2008). Financial Decision Support Using Neural Networks and Support Vector Machines, *Expert Systems*, 25, (4), 380-393.
 - West, D. (2000). Neural Network Credit Scoring Models, *Computers & Operations Research*, 27, 1131-1152.
 - Wu, X. (2008). *Credit Scoring Model Validation*. Master Thesis, Korteweg-de Vries Institute for Mathematics.
 - Yeh, H. -C., Yang, M.- L. & Lee, L. C. (2007). An Empirical Study of Credit Scoring Model for Credit Card, *IEEE*.
 - Zhou, X., Zhang, D. & Jiang, Y. (2008). A New Credit Scoring Method Based on Rough Sets and Decision Tree. *Springer-Verlag Berlin Heidelberg*. 1081-1089