

یازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

11th National Congress of
the New Technologies in Sustainable Development of Iran

senaconf.ir

بررسی عوامل موثر بر طبقه بندی مشتریان در شرکت صنایع شیر ایران (پگاه) با استفاده از تکنیک های داده کاوی (مطالعه موردی مشتریان استان گیلان و آذربایجان شرقی)

منصور اسماعیل پور^۱، رضا کاظمی ایمن آبادی^۲

استادیار، گروه مهندسی کامپیوتر، واحد همدان، دانشگاه آزاد اسلامی، ایران، همدان (نویسنده مسئول) esmaeilpour@iauh.ac.ir

دانشجوی دکتری رشته مدیریت فناوری اطلاعات، واحد همدان دانشگاه آزاد اسلامی، ایران، همدان. reza.kazemi.engi@iauh.ac.ir

چکیده

امروزه سازمان ها باید برای حضور قدرتمند در بازار و توانایی رقابت با دیگر شرکت ها توجه بسیاری به مشتریان و بحث طبقه بندی آن ها داشته باشند. لازم است شرکت ها و سازمان ها تمامی مشتریان بالفعل و بالقوه خود را در دسته های مختلف تقسیم نمایند و بر اساس آن برای هر دسته از مشتریان فرآیند های مختلفی را تعریف نمایند، علت اصلی طبقه بندی مشتریان یکی تفاوت نیاز های گروه های مختلف مشتریان است و دوم اینکه علاوه بر ارزشمند بودن تمامی مشتریان برای سازمان، برخی از مشتریان برای سازمان ارزشمند ترند و نیازمند تعاملات ویژه و خاص نیز می باشند. هدف از این تحقیق بررسی عوامل موثر بر طبقه بندی مشتریان در شرکت صنایع شیر ایران (پگاه) با استفاده از تکنیک های داده کاوی است. تحقیق حاضر یک تحقیق کاربردی و از نوع اکتشافی بود، روش گردآوری اطلاعات به صورت میدانی انجام شده است. در این پژوهش از شاخص های مختلفی در طبقه بندی مشتریان استفاده شده است. تعداد جامعه آماری ۳۶۵۹ عدد از مغازه داران خرده فروشی در استانهای گیلان و آذربایجان شرقی بود. برای رسیدن به اهداف این پژوهش از تکنیک های مختلف داده کاوی استفاده گردید. مدل مورد نظر با تعداد ۱۵ عدد ویژگی و تعداد ۳۶۵۹ نمونه اجرا شد. و از الگوریتم های مختلفی استفاده شد، نتایج الگوریتم های مختلف کلاس بندی با در نظر گرفتن آزمون t و بادر نظر گرفتن سطح معناداری ۰,۰۵ مورد مقایسه قرار گرفت و مشخص شد که الگوریتم نیویز، گزینه بهتری نسبت به الگوریتم جی ۴۸ برای حل مسئله می باشد و به جواب بهتری می رسد. سپس از الگوریتم های جانسون و ژنتیک برای حل مسئله استفاده شد و این دو الگوریتم به جواب های یکسانی برای مسئله دست یافتند.

کلیدواژه: طبقه بندی، مشتریان، صنایع شیر ایران، پگاه، تکنیک های داده کاوی.

یازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

11th National Congress of
the New Technologies in Sustainable Development of Iran

senaconf.ir

۱. مقدمه و بیان مسئله

بزرگترین سرمایه های شرکت، مشتریان آنها می باشند، محصولات می آیند و می روند، ولی آنچه امروز برای سازمانها ارزش می آفرینند «ایجاد رابطه مستمر با مشتری» است. بنابراین، سازمانهای موفق، سازمانهایی هستند که قادرند «مشتریان تمام عمر» برای خود ایجاد نمایند (Chan & Ip., 2011). طبقه بندی موجود برای به حداکثر رساندن ارزش مشتری کلیدی برای بقای رقابت شدید در دنیای کسب و کار می باشد که متمایز کردن مشتریان سودآورتر از مشتریان کم سودتر و تمرکز بر طول عمر جای روابط کوتاه مدت با مشتری از استراتژیهای کلیدی کسب و کار برای بقا در بازار رقابتی امروزی به شمار می آید، در نتیجه وفاداری بلندمدت مشتری برای پایداری کسب و کار حیاتی است که ارزش مشتری براساس سود به سبب اهمیت محوری مزایای اش مبنای مهمی برای تقسیم بندی رفتار محسوب می شود. شرکتها با آینده نگری برای حفظ مشتریانی که بیشترین مزایا را برایشان فراهم کرده و هم چنین سودشان را به حداکثر می رسانند (Cuadros & Domínguez., 2014). تدوین هر برنامه ای برای مدیریت مشتریان، ابتدا مستلزم درک و شناخت و نهایتاً گروه بندی آن هاست. وندل آراسمیت تقسیم بندی بازار را به عنوان روندی جهت شناسایی زیرگروه مشتریان (فعلی/ بالقوه) تعریف کرده است که انتظار می رود پاسخ های متفاوتی را به متغیرهای بازاریابی (منافع، قیمت، ترفیع، توزیع، و غیره) بدهند. این تعریف توسط رهبران معاصر تفکر بازاریابی مانند فیلیپ کاتلر نیز پذیرفته شده است (پو یا و فاضل ترشیزی، ۱۳۹۹). تاکنون روش های مختلفی نظیر تحلیل پوششی داده ها، آنالیز ممیزی، رگرسیون خطی و لجستیک، روش های هموارسازی ناپارامتری فرایند تحلیل سلسله مراتب فازی و شبکه های عصبی در زمینه رتبه بندی اعتباری مشتریان مورد استفاده قرار گرفته است (علی حیدری بیوکی و خادمی زارع، ۱۳۹۴). روش های طبقه بندی مشتریان به دو دسته زیر تقسیم می شود: طبقه بندی بر اساس ارزش مشتری و طبقه بندی با استفاده از روش های داده کاوی (ZChe et al., 2010). عموماً استراتژی طبقه بندی براساس ارزش مشتری به سه دسته زیر تقسیم می شود: طبقه بندی براساس ارزش عمر مشتری، طبقه بندی براساس عامل های ارزش عمر مشتری (ارزش فعلی مشتری، ارزش بالقوه، وفاداری مشتری) و طبقه بندی براساس ارزش عمر مشتری و سایر ویژگی های مشتری (کتابی و همکاران، ۱۳۸۴). طبقه بندی های دیگری بر اساس مقدار سود مشتری انجام شده است. برای مثال رده سودآور، مشتریانی را شامل می شود که دارای سوددهی بالا هستند و این مشتریان باید با برنامه های مدیریتی حفظ شوند. طبقه بندی دیگری را می توان بر اساس مشتریان کم سود انجام داد یعنی مشتریانی که هزینه آن ها بیشتر از سودشان است (بیورانی و عظیمی، ۱۳۹۴). صنعت لبنیات یکی از مهمترین بازارهایی است که پیشرفت فناوری ماشین آلات، بهره گیری از تجربه های متخصصان و توسعه و پیشرفت رشته های صنایع غذایی محیطی بسیار رقابتی ایجاد کرده است (Saaty., 2001)، به طوری که هر شرکتی برای ورود به بازار لبنیات و حفظ بقای خود الزم است از یک سلسله معیارهایی چون توان مالی، بخش تحقیق و توسعه قوی، بازاریابی علمی و تخصصی، مدیریت توزیع و فروش نظامند، کارکنان متعهد و باتجربه و روند ارتقای شغلی مشخص برخوردار باشد تا بتواند به راحتی خود را با تغییرپذیری های محیطی سازش و هماهنگی دهد (گلابچی و نورزایی، ۱۳۹۴). صنعت لبنیات در ایران برای استفاده مناسب از ظرفیتهای تولیدی خود و رسیدن به جایگاه شایسته در عرصه رقابتی به ویژه در زمینه بین الملل در آغاز نیاز به موفقیت در بازار داخلی با استفاده از شیوه های بهینه بازاریابی دارد، در نتیجه آگاهی از راهبردهای مناسب بازاریابی با توجه به وضعیت شرکت میتواند به رشد بیشتر این صنعت کمک کند. توجه به رشد روز افزون سازمانها در رقابت برای تحصیل سهم بازار ضرورت وجود برنامه ریزی راهبردی را در سازمانها دو چندان می کند (پیش بهار و همکاران، ۱۳۹۶).

در مجموع با توجه به اهمیت موضوع طبقه بندی مشتریان و نیز با توجه به موضوع و اهمیت صنایع لبنی و آسیب هایی که ممکن است به ساختار های صنایع شیر پگاه وارد آید در این تحقیق به شناسایی و همچنین امکان سنجی عوامل موثر بر طبقه بندی مشتریان شرکت

یازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

11th National Congress of
the New Technologies in Sustainable Development of Iran

senaconf.ir

صنایع شیر ایران با تکنیک های داده کاوی به این سوال پاسخ داده خواهد شد که چه عواملی بر طبقه بندی مشتریان شرکت صنایع شیر ایران با تکنیک های داده کاوی تاثیر دارند.

پیشینه تحقیق

پویا و فاضل ترشیزی (۱۳۹۹) پژوهشی تحت عنوان « بخش بندی مشتریان صنعت بیمه با استفاده از الگوریتم خوشه بندی دو مرحله ای مقیاس پذیر (نمونه موردی یک شرکت بیمه) » انجام دادند. با توجه به تفاوت در سودآوری، نوع خرید، وفاداری، ریسک، بُعد رفتاری و جمعیت شناختی مشتریان در پی ایجاد مرزبندی قابل توجه ای بین آن ها با استفاده از رویکرد بخش بندی مشتریان، هدف این پژوهش، شناخت ویژگی های هریک از این گروه های مختلف، جهت افزایش قدرت رقابتی و موفقیت فعالان این حوزه بود. جامعه آماری پژوهش، مشتریان یک شرکت بیمه در استان خراسان رضوی بود که در نهایت اطلاعات تعداد ۲۲۰ نمونه نهایی وارد فاز تحلیل گردید. باتوجه به تاثیر شاخص های تعیین شده مشتریان در شش خوشه دسته بندی شدند. یافته ها حاکی از آن بود که از نظر متغیر سودآوری تمامی خوشه ها با یکدیگر متفاوت بود. از نظر شیوه جذب خوشه بدحسابها با رهگذران و نور چشمی ها با خوشه های خوش حساب ها متفاوت بود.

بیورانی و عظیمی (۱۳۹۴) پژوهشی تحت عنوان « طبقه بندی مشتریان بانک صادرات براساس ارزش مشتری با استفاده از درخت تصمیم » انجام دادند. هدف از انجام این تحقیق به دست آوردن معیارهای موثر در انتخاب مشتری ارزشمند است که بتوان مشتریان را براساس ویژگی های جمعیت شناختی شان و سایر متغیرهای مربوط به معاملات به طبقات سود خیلی کم، کم سود، سود بالا و سود خیلی بالا طبقه بندی کرد. جامعه آماری در این تحقیق مشتریان دارای حساب جاری فعال نزد بانک صادرات ایران در شهر تبریز بود و مشتریانی در نظر گرفته شد که حداقل یک سال سابقه فعالیت بانکی نزد بانک صادرات دارند. یافته ها نشان داد متغیرهای سن، تحصیلات مشتری و درجه شعبه بانک تاثیر معنی داری بر ارزش مشتری ندارند. تعداد تراکنش مشتری با بانک موثرترین ویژگی مشتری در تشخیص طبقه مشتری می باشد.

پیش بهار و همکاران (۱۳۹۶) پژوهشی تحت عنوان تدوین و اولویت بندی راهبردهای بازاریابی فرآورده های لبنی شرکت پگاه آذربایجان شرقی کاربرد رهیافت تحلیل سلسله مراتبی انجام شده است، با توجه به محیط متحول و در حال تغییر، داشتن راهبرد مناسب میتواند به بنگاه در تداوم حیات و کسب مزیت رقابتی کمک کند؛ چرا که راهبرد انتخابی باید علاوه بر ارضای نیازهای بازار هدف، با هدف های بازاری و سازمانی مؤسسه نیز منطبق باشد. هر شرکت بسته به کسب و کار و وظایفی که بر عهده دارد به تدوین و انتخاب راهبرد بازاریابی مناسب خود میپردازد. در صورت تدوین و انتخاب راهبرد بازاریابی استاندارد، ارزیابی فعالیت های جاری مدیریتی آسان میشود. لذا تلاش شد در این تحقیق با مطالعه منابع کتابخانه ای، مهمترین معیارهای مؤثر بر انتخاب راهبرد مناسب بازاریابی برای فروش فرآورده های لبنی شرکت پگاه آذربایجان شرقی شناسایی شوند. که پرسشنامه هایی توسط ۳۰ کارشناس و خبره در زمینه بازاریابی و صنعت لبنیات تکمیل شدند و مقایسه های زوجی بین معیارها و زیرمعیارها و همچنین ارزیابی گزینه ها به کمک معیارها صورت گرفت. آنگاه در یک جمع بندی کلی تاثیر هر یک از معیارهای به دست آمده بر چهار راهبرد بازاریابی شامل راهبرد یکپارچگی، راهبرد تمرکز، راهبرد تنوع و راهبرد تدافعی بررسی شدند و راهبرد تدافعی به عنوان مناسبترین راهبرد بازاریابی برای شرکت پگاه آذربایجان شرقی تعیین شد. راهبردهای یکپارچگی، تنوع و تمرکز به ترتیب اولویتهای بعدی را تشکیل می دهند. لذا با توجه به اهمیت بالای معیار سازگاری با اصول بهداشتی و محیط زیست در مقایسه با سایر معیارها، پیشنهاد میشود مدیران شرکت پگاه آذربایجان شرقی به جای استفاده از کیسه های نایلونی و بطری ها و ظرف های بسته بندی پلاستیکی، از پاکتهای کاغذی که قابل تجزیه در محیط زیست می باشند استفاده کنند.

یازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

11th National Congress of
the New Technologies in Sustainable Development of Iran

senaconf.ir

روش تحقیق

تحقیق حاضر یک تحقیق کاربردی و از نوع اکتشافی بود. روش گردآوری اطلاعات به شکل میدانی انجام شده است. در این پژوهش از پرسشنامه استاندارد ۱۷ سوالی استفاده شد. تعداد جامعه آماری ۳۶۵۹ عدد بود. برای دستیابی به اهداف این پژوهش از تکنیک‌های مختلف داده‌کاوی استفاده گردید. مدل مورد نظر با تعداد ۱۵ عدد ویژگی و تعداد ۳۶۵۹ نمونه اجرا شد. و از الگوریتم‌های مختلفی استفاده شد. در ادامه نیز با بهره‌گیری از الگوریتم جانسون و الگوریتم ژنتیک استفاده شد و قوانین استخراج شده مشخص گردید.

عوامل تاثیرگذار بر مساله عبارتند از :

ردیف	عنوان ویژگی	[ID]	Normalization
1	کد مشتری	[customer management code]	[CMC]
2	مساحت فروشگاه	[store area]	[SA]
3	کد مسیر	[Route code]	[RC]
4	طبقه اقتصادی محدوده مغازه	[Economic floor of the shop area]	[EFSA]
5	موقعیت مکانی فروشگاه	[store location]	[SL]
6	محاسبه امتیاز هر مشتری	[Calculate the points of each customer]	[CPEC]
7	طبقه بندی فروشگاه	[Store classification]	[SC]

جدول ۱: شاخص‌ها اصلی

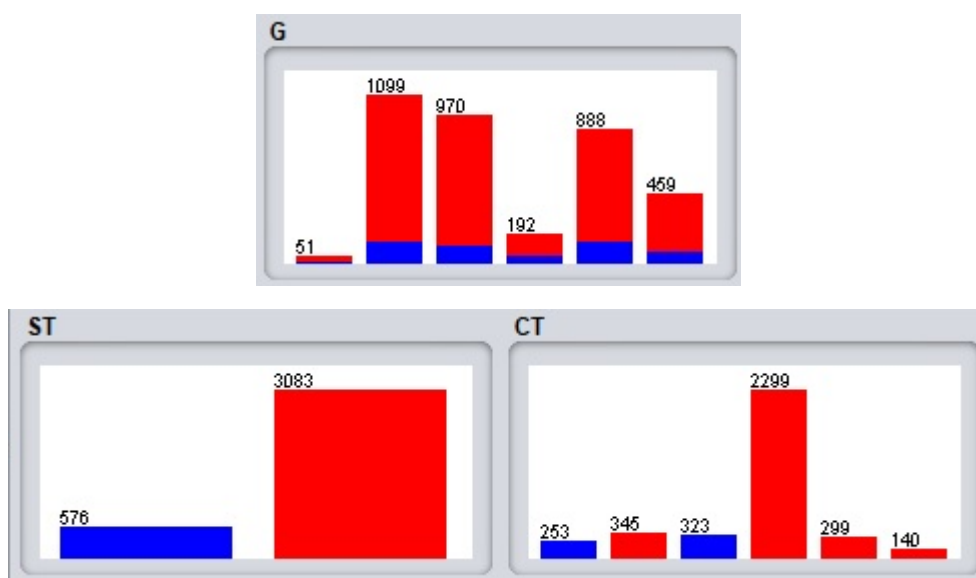
یازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

11th National Congress of
the New Technologies in Sustainable Development of Iran

senaconf.ir

تحلیل یافته ها

مدل مورد نظر با تعداد ۱۵ عدد خصیصه و تعداد ۳۶۵۹ نمونه اجرا می شود سپس داده ها را با توجه به جواب پرسشنامه که کلاس و نتیجه ی پرسشنامه می باشد به دو دسته ی بلی و خیر طبقه بندی می کنیم که نتیجه حاصله در نمودار زیر مشخص شده است:



شکل ۱: نمودارهای شاخص ها

با توجه به نتایج بدست آمده از ۳۶۵۹ عدد مشتری، تعداد هر طبقه بندی در جدول مشخص شده است .

ردیف	طبقه بندی	تعداد مشتریان در هر طبقه	عنوان ویژگی	نام ویژگی
۱	[p+]	۵۱	مشتریان بسیار عالی	Platinum+
۲	[P]	۱۹۲	مشتریان بسیار خوب	Platinum
۳	[G+]	۴۵۹	مشتریان عالی	Gold+
۴	[G]	۹۷۰	مشتریان خوب	Gold
۵	[S+]	۸۸۸	مشتریان متوسط خوب	Silver+
۶	[S]	۱۰۹۹	مشتریان متوسط	Silver

جدول ۲: طبقه بندی مشتریان

یازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

11th National Congress of
the New Technologies in Sustainable Development of Iran

senaconf.ir

منظور از استانداردسازی نرمال نمودن توزیع داده‌ها است. استانداردسازی میانگین داده‌ها را به صفر و واریانس را به ۱ تبدیل می‌کند. هر چه مقادیر داده به سمت بی‌نهایت کشیده شود، واریانس داده‌ها به ۱ نزدیک‌تر می‌شود.

با استاندارد سازی داده‌ها توزیع داده‌ها نرمال می‌شود و میانگین داده‌ها به صفر و واریانس داده‌ها به یک نزدیک می‌شود که نتایج حاصل شده در شکل زیر نشان داده شده است

No.	1: a0	2: a1	3: a2	4: a3	5: a4	6: a5	7: a6	8: a7	9: a8	10: a9	11: class
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal
1	true	false	false	false	true	false	true	true	true	true	c0
2	true	true	false	false	false	true	true	false	true	true	c0
3	false	false	true	true	false	true	false	false	false	false	c1
4	false	true	true	true	true	false	true	true	true	false	c0
5	true	true	false	false	false	true	true	false	false	true	c0
6	true	true	true	true	false	false	true	false	true	true	c1
7	true	true	false	true	true	true	false	true	false	false	c0
8	false	false	true	false	true	true	false	true	false	true	c1
9	true	true	false	true	false	true	false	true	true	true	c0
10	false	false	true	true	true	true	true	true	true	false	c1
11	false	true	false	false	false	true	true	false	true	false	c0
12	true	false	true	false	true	false	false	false	false	true	c0
13	true	true	false	true	true	true	true	false	false	false	c1
14	false	false	false	true	true	true	false	false	true	false	c1
15	false	false	false	true	true	false	false	false	true	false	c0
16	true	false	false	true	false	false	true	false	false	false	c0
17	true	true	true	true	false	true	true	true	true	false	c1
18	true	false	false	true	false	true	false	true	false	false	c0
19	true	false	true	false	false	false	false	false	true	true	c0
20	false	false	false	true	true	false	true	true	false	false	c1
21	false	true	false	true	true	true	false	true	true	false	c1
22	false	true	false	true	true	false	false	false	false	true	c0
23	false	true	true	true	false	false	false	false	true	true	c0
24	false	true	false	false	true	false	false	false	false	false	c0
25	true	true	true	false	false	true	false	false	true	true	c0
26	true	true	true	true	false	true	false	true	false	true	c0
27	true	false	false	true	false	true	false	true	true	true	c0
28	true	false	true	true	true	true	true	true	false	false	c1
29	true	true	true	true	true	true	true	true	true	false	c0

شکل ۲: استاندارد سازی داده‌ها

۱- الگوریتم درخت تصمیم گیری 48z:

یکی از دانش‌هایی که در عصر حاضر بسیار مورد استقبال قرار گرفته داده کاوی است. به طور کلی به معنای کاوش در داده‌ها است که به اشکال مختلف برای به دست آوردن الگوها و کسب دانش در مورد الگوی حاکم بر این داده‌ها مورد استفاده قرار می‌گیرد. در فرآیند داده کاوی، ابتدا مجموعه داده‌های بزرگ مرتب می‌شود، سپس الگوها شناسایی می‌شود و روابط و تکنیک‌هایی برای انجام تجزیه و تحلیل داده‌ها و حل مسائل استفاده می‌شود. یکی از پرکاربردترین این روش‌ها، تکنیک طبقه بندی است.

اصولاً از تکنیک‌های طبقه بندی برای طبقه بندی هر داده در مجموعه‌ای از داده‌ها و اختصاص به یکی از مجموعه‌های از پیش تعیین شده کلاس‌ها یا گروه‌ها استفاده می‌شود. روش طبقه بندی از تکنیک‌های ریاضی مانند درخت تصمیم، برنامه ریزی خطی، شبکه عصبی و آمار برای طبقه بندی استفاده می‌کند. به عبارتی طبقه بندی، فرایند یافتن مدلی که توصیف کننده کلاس‌ها و مفاهیم داده است و داده‌ها را به گروه‌های مشخص تفکیک می‌کند. الگوریتم‌های طبقه بندی، قادر به یادگیری از تجربیات گذشته هستند و این یادگیری بر اساس تجربه نشان دهنده یک گام اساسی در تقلید از توانایی‌های استقرایی مغز انسان است که بر اساس این توانایی مغز می‌تواند مسئله‌ی شناسایی یک گروه از دسته‌ها (زیرجمعیت‌ها) را انجام دهد.

یازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

11th National Congress of
the New Technologies in Sustainable Development of Iran

senaconf.ir

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	79	78.2178 %
Incorrectly Classified Instances	22	21.7822 %
Kappa statistic	0.5193	
Mean absolute error	0.2577	
Root mean squared error	0.4363	
Relative absolute error	57.4908 %	
Root relative squared error	92.2338 %	
Total Number of Instances	101	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.821	0.294	0.846	0.821	0.833	0.520	0.737	0.772	c0
	0.706	0.179	0.667	0.706	0.686	0.520	0.737	0.596	c1
Weighted Avg.	0.782	0.255	0.786	0.782	0.784	0.520	0.737	0.713	

=== Confusion Matrix ===

```
a b <-- classified as
55 12 | a = c0
10 24 | b = c1
```

شکل ۳: نتایج حاصل از روش درخت تصمیم گیری J48

درخت تصمیم نقشه ای از نتایج احتمالی یکسری از انتخاب ها یا گزینه های مرتبط بهم است به طوری که به یک فرد یا سازمان اجازه می دهد تا اقدامات محتمل را از لحاظ هزینه ها، احتمالات و مزایا بسنجد. از درخت تصمیم می توان یا برای پیشبرد اهداف و برنامه های شخصی و غیر رسمی یا ترسیم الگوریتمی که بر اساس ریاضیات بهترین گزینه را پیش بینی می کند استفاده کرد.

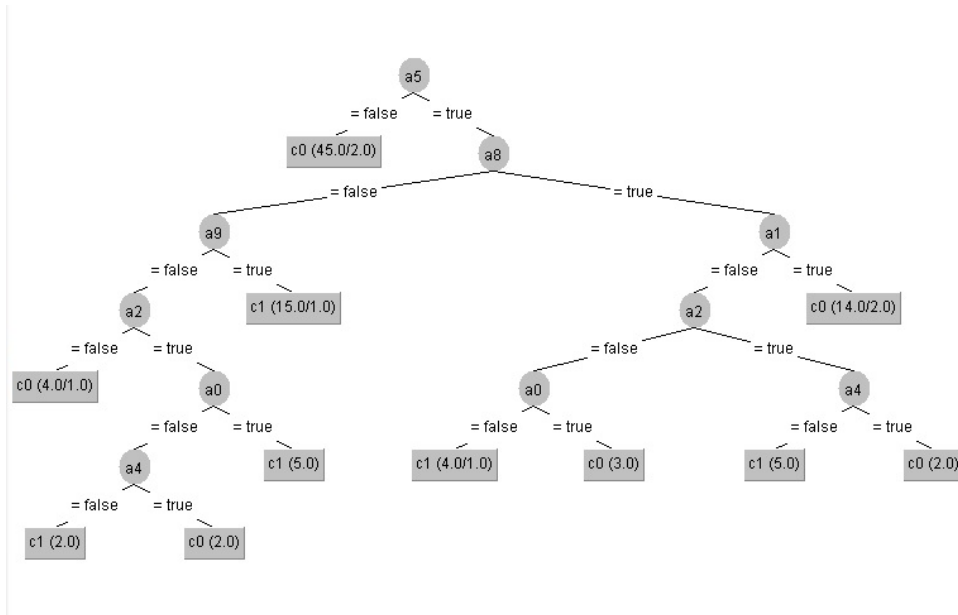
یک درخت تصمیم گیری به طور معمول با یک نود اولیه شروع می شود که پس از آن پیامد های احتمالی به صورت شاخه هایی از آن منشعب شده و هر کدام از آن پیامد هایی به نود های دیگری منجر شده که آن ها همثبه نوبه خود شاخه هایی از احتمالات دیگر را ایجاد می کنند که این ساختار شاخه شاخه سر انجام به نموداری شبیه به یک درخت مبدل می شود.

استخراج قوانین جهت تصمیم گیری از اهمیت بالایی برخوردار است در شکل زیر قوانین استخراج شده از الگوریتم درخت تصمیم گیری مشخص شده است .

یازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

11th National Congress of
the New Technologies in Sustainable Development of Iran

senaconf.ir



شکل ۴: درخت تصمیم گیری

۲- الگوریتم بیز ساده

امروزه برنامه های کاربردی بسیاری وجود دارند که پارامترهای مختلف مربوط به الگوریتم بیز ساده را تخمین می زنند، با استفاده از این ابزارها افراد بدون سروکار داشتن با تئوری بیز می توانند از این امکان در راستای حل مسائل مورد نظر بهره ببرند. با وجود مسائل طراحی و پیش فرض هایی که در خصوص روش بیز وجود دارد، این روش برای طبقه بندی کردن بیشتر مسائل در جهان واقعی، کاملا مناسب است.

برای دسته بندی ساده و تعیین روشی برای تشخیص برچسب اشیا یا نقاط اکثر اوقات از تکنیک دسته بند بیز ساده استفاده می شود. در واقع، برای به کارگیری دسته بند بیز ساده، الگوریتم یکتایی وجود ندارد و در عوض خانواده ای از الگوریتم ها موجود است که با فرض استقلال ویژگی ها یا متغیرها نسبت به یکدیگر عمل می کنند.

یکی از مزایای قابل توجه در دسته بند بیز ساده، امکان برآورد کردن هر یک از پارامترهای مدل با اندازه نمونه کوچک به عنوان مجموعه داده کاوی است.

یازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

11th National Congress of
the New Technologies in Sustainable Development of Iran

senaconf.ir

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	81	80.198 %
Incorrectly Classified Instances	20	19.802 %
Kappa statistic	0.5566	
Mean absolute error	0.3	
Root mean squared error	0.3828	
Relative absolute error	66.9401 %	
Root relative squared error	80.918 %	
Total Number of Instances	101	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.851	0.294	0.851	0.851	0.851	0.557	0.831	0.879	c0
	0.706	0.149	0.706	0.706	0.706	0.557	0.831	0.638	c1
Weighted Avg.	0.802	0.245	0.802	0.802	0.802	0.557	0.831	0.798	

=== Confusion Matrix ===

```
a b <-- classified as
57 10 | a = c0
10 24 | b = c1
```

شکل ۵: نتایج حاصل از پیاده سازی الگوریتم بیز ساده

۳- روش شبکه های عصبی (پرسترون):

قبل از پیدایش شبکه عصبی `mlp`، در سال ۱۹۵۸ فرانک روزنبلات یک شبکه عصبی به نام پرسپترون ابداع کرد. روزنبلات یک لایه ای از نورون ها را تشکیل داد و شبکه حاصل را پرسپترون نامید. اما پرسپترون روزنبلات نیز مشکلات فراوانی داشت. مینسکی و پپرت در سال ۱۹۶۹ کتابی به نام پرسپترون نوشتند. آن ها تمامی توانایی ها و مشکلات پرسپترون را در این کتاب مورد بررسی قرار دادند. مینسکی و پپرت در کتاب خود ثابت کردند که پرسپترون فقط مسائلی را می تواند حل کند که به صورت خطی تفکیک پذیر باشند.

در واقع، پرسپترون یک الگوریتم یادگیری ماشین است که (مانند ماشین بردار پشتیبان) در حوزه یادگیری بانظارت قرار می گیرد. این الگوریتم را از اولین الگوریتم های شبکه عصبی مصنوعی می دانند که در این حوزه مورد استفاده قرار گرفته است. پرسپترون یک نوع الگوریتم دسته بندی دو-دویی تلقی می شود و این به این معناست که این الگوریتم می تواند تصمیم بگیرد که یک عضو متعلق به یک دسته و کلاس خاص می باشد یا خیر.

شبکه عصبی پرسپترون چند لایه، دسته ای از شبکه های عصبی مصنوعی پیشخور محسوب می شوند. در یک شبکه عصبی پرسپترون چند لایه، حداقل سه لایه (Layer) از نودها (Nodes) وجود خواهند داشت که این لایه ها عبارتند از:

✓ یک لایه ورودی (Input Layer)

✓ یک لایه نهان (Hidden Layer)

✓ یک لایه خروجی (Output layer)

در این نوع از شبکه عصبی مصنوعی، از خروجی های لایه اول (ورودی)، به عنوان ورودی های لایه بعدی (نهان) استفاده می شود؛ این کار به همین شکل ادامه پیدا می کند تا زمانی که، پس از تعداد خاصی از لایه ها، خروجی های آخرین لایه نهان به عنوان ورودی های لایه

یازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

11th National Congress of
the New Technologies in Sustainable Development of Iran

senaconf.ir

خروجی مورد استفاده قرار می گیرد. همه لایه هایی که بین لایه ورودی و لایه خروجی قرار می گیرند، «لایه های نهان» (Hidden Layers) نامیده می شوند. شبکه های پرسپترون چند لایه، مانند شبکه های عصبی پرسپترون تک لایه، حاوی مجموعه ای از وزن ها نیز هستند که باید برای آموزش و یادگیری شبکه عصبی تنظیم شوند.

Time taken to build model: 0.26 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	84	83.1683 %
Incorrectly Classified Instances	17	16.8317 %
Kappa statistic	0.6204	
Mean absolute error	0.1784	
Root mean squared error	0.3708	
Relative absolute error	39.802 %	
Root relative squared error	78.3853 %	
Total Number of Instances	101	

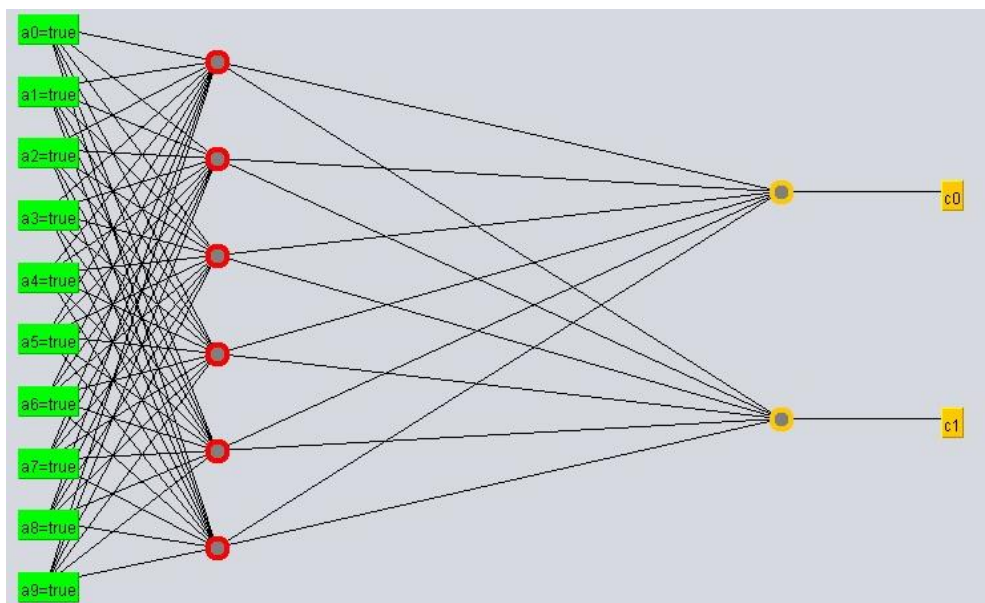
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.881	0.265	0.868	0.881	0.874	0.621	0.886	0.925	c0
	0.735	0.119	0.758	0.735	0.746	0.621	0.886	0.822	c1
Weighted Avg.	0.832	0.216	0.831	0.832	0.831	0.621	0.886	0.890	

=== Confusion Matrix ===

```
a b <-- classified as
59 8 | a = c0
9 25 | b = c1
```

شکل ۶: نتایج حاصل از روش شبکه های عصبی از نوع پرسپترون چند لایه



شکل ۷: شبکه عصبی حاصل از روش پرسپترون چند لایه

یازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

11th National Congress of
the New Technologies in Sustainable Development of Iran

senaconf.ir

مقایسه نتایج :

در جدول شماره ۳ نتایج حاصل از روشهای مختلف بر اساس دقت طبقه بندی و دقت پیش بینی برای طبقه بندی های مختلف مشتریان بررسی شده و برحسب درصد با یکدیگر مقایسه شده است.

با بررسی روشهای مورد استفاده در این پژوهش روش پرسپترون و بیز ساده برای طبقه بندی مشتریان از دقت بالایی نسبت به سایر روشها برخوردار هستند.

روش	نتایج بدست آمده	دقت الگوریتم
درخت تصمیم گیری	78.217%	0.786
بیز ساده	80.198%	0.802
پرسپترون	83.168%	0.831

جدول ۳: نتایج بدست آمده

نتیجه گیری:

نتایج نشان داد که با استفاده از روشهای داده کاوی، طبقه بندی مشتریان بر اساس دسته بندی های مورد بررسی امکان پذیر است، یعنی توانستیم بر اساس ویژگی های مورد استفاده در پژوهش حاضر با دقت مناسب این کار را انجام دهیم، در این پژوهش با استفاده از الگوریتم های مختلف داده کاوی به بررسی عوامل موثر بر طبقه بندی در مشتریان شرکت صنایع شیر ایران (پگاه) پرداخته شد. با بررسی روشهای مورد استفاده در این پژوهش در روش پرسپترون، بیز ساده، نسبت به الگوریتم درخت تصمیم، برای حل مدل دقت بالاتری داشت که در آن ۸۱ نمونه درست و تنها ۲۰ نمونه اشتباه دسته بندی شده است. لازم به ذکر است که الگوریتم پرسپترون، نسبت به دو الگوریتم دیگر با 83.168% و دقت 0.831، دقت بالاتری داشت. اینکه سازمان چگونه مشتریان خود را طبقه بندی نماید بستگی به معیار های دارد که بر اساس آن مشتریان برای سازمان ایجاد ارزش می کنند بنابراین نحوه طبقه بندی برای هر سازمان ممکن است متفاوت از سایر سازمان ها باشد. این ارزش ها توسط متخصصین کسب و کار و از طریق تحقیقات میدانی شناسایی میگردند و پس از کمی سازی و تعیین شاخص ها مد نظر مشتریان سازمان طبقه بندی می گردند. به طور کلی تحقیق حاضر کار آمدی همه تکنیک های بکار رفته را نشان می دهد، و همچنین مشخص شد، که روشهای داده کاوی برای طبقه بندی مشتریان و نیز دسته بندی آنها کارآمد است. پیشنهاد می شود در تحقیقات آتی از روشهای داده کاوی بیشتری برای طبقه بندی مشتریان در سطوح مختلف استفاده شود.

یازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

11th National Congress of
the New Technologies in Sustainable Development of Iran

senaconf.ir

منابع

- بیورانی، حسین؛ عظیمی، محرم (۱۳۹۴) طبقه بندی مشتریان بانک صادرات براساس ارزش مشتری با استفاده از درخت تصمیم، مدیریت بهره وری، س ۹، ش ۳۳، صص: ۱۳۴-۱۱۹.
- پویا، علیرضا؛ فاضل ترشیزی، داود (۱۳۹۹) بخش بندی مشتریان صنعت بیمه با استفاده از الگوریتم خوشه بندی دو مرحله ای مقیاس پذیر (نمونه موردی یک شرکت بیمه)، پژوهشنامه بیمه، س ۳۵، ش ۲، صص: ۶۵-۳۱.
- پیش بهار، اسماعیل؛ فردوسی، رویا؛ حیاتی، باب اله (۱۳۹۶) تدوین و اولویت بندی راهبردهای بازاریابی فرآورده های لبنی شرکت پگاه آذربایجان شرقی: کاربرد رهیافت تحلیل سلسله مراتبی، اقتصاد کشاورزی، شماره ۴، صص ۱-۱۹.
- علی حیدری بیوکی، طاهره؛ خادمی زارع، حسن (۱۳۹۴) توسعه روش تحلیل پوششی داده ها به منظور خوشه بندی مشتریان اعتباری بانک ها، مجله مدل سازی در مهندسی، س ۱۳، ش ۴۱، صص: ۷۴-۵۹.
- کتابی، سعیده؛ انصاری، محمد اسماعیل؛ ناصری طاهری، مظفر (۱۳۸۴) انتخاب آمیخته بازاریابی مناسب با استفاده از تکنیک AHP با رویکرد برنامه ریزی استراتژیک بازاریابی، مجله دانشکده علوم اداری و اقتصاد، شماره ۱.
- گلابچی، محمدرضا؛ نورزایی، عصمت الله (۱۳۹۴) انتخاب روش مناسب مشارکت خصوصی- دولتی به منظور تامین مالی پروژه های آزاد راه ایران با رویکرد (AHP)، فصلنامه حمل و نقل، شماره ۳، صص ۵۲۳-۵۳۷.
- Chan, S. L., & Ip, W. H. (2011). A dynamic decision support system to predict the value of customer for new product development. *Decision Support Systems*, 52, 178-188
- Cuadros, A. J., & Domínguez, V. E. (2014). Customer segmentation model based on value generation for marketing strategies formulation. *Estudios Gerenciales*, 30(130), 25-30.
- Saaty T.L, 2001. *Decision Making with Dependence and Feedback: The Analytic Network Process*. (2nd ed.). Pittsburg, USA: RWS publications
- ZChe.H., Wang H.S., Chuang C.L., (2010). "A fuzzy AHP and DEA approach for making bank loan decisions for small and medium enterprises in Taiwan", *Expert Systems with Applications*, 37(10), 7189-7199.