

تحلیل مشتریان رویگردان از بانک و خوشه بندی سایر مشتریان بر اساس میزان شباهت با مشتریان رویگردان

روح اله کیانی قلعه نو^۱

ام البنین ستوده^۲

تاریخ دریافت: ۱۴۰۱/۰۳/۲۱ تاریخ چاپ: ۱۴۰۱/۰۶/۰۲

چکیده

در نظام بانکی که چرخه ورود و خروج وجوه نقد در قالب سپرده پذیری و پرداخت تسهیلات اصلی ترین فعالیت آن محسوب می شود، حفظ مشتریان اهمیت بالایی دارد. بدین منظور مدیران بانک ها بصورت قابل توجهی به حوزه امور مشتریان تمرکز نموده و در این راستا داده کاوی و روش های تحلیل متنوع آماری می تواند نقش موثری در تصمیم گیری داشته باشد. در این پژوهش روشی پیشنهاد خواهد شد که در آن از یک الگوریتم مبتنی بر ضریب همبستگی، شاخص های موثر بر مشتریانی که اخیراً از بانک خارج شده اند، را شناسایی خواهد نمود و سپس با استفاده از قابلیت رگرسیون خطی، مدل خروج مشتریان استخراج و نهایتاً متناسب با شباهت مشتریان با مدل رگرسیون، مشتریان در ۶ خوشه طبقه بندی خواهند شد. در همین راستا تحلیل هایی درخصوص سایر شاخص ها و انتظارات مدیریتی برای تمرکز و حفظ مشتریان ارائه شده است.

واژگان کلیدی

داده کاوی، رگرسیون خطی، ضریب همبستگی، خوشه بندی

۱. گروه مهندسی صنایع، واحد علی آباد کتول، دانشگاه آزاد اسلامی، علی آباد کتول، ایران. rohollah.ghaleno@yahoo.com

۲. گروه حقوق، واحد زاهدان، دانشگاه آزاد اسلامی، زاهدان، ایران. seode.amir2@gmail.com

۱ مقدمه

با گذر زمان از یک سوی با افزایش تولیدات صنعتی و بنگاه های اقتصادی و از سوی دیگر شاهد بهبود تکنولوژی در بخش های مختلف هستیم، کم شدن فاصله کیفی بین محصولات و پیش رفت همگن محصولات باعث شده است بازارها به سمت رقابتی حرکت نماید، لذا بنگاه ها صرفا با تکیه بر قیمت و کیفیت نمی توانند در جلب رضایت مشتریان موفق عمل نمایند و بدیهی است نیازمند شناسایی صفات ویژه مشتریان و ارائه خدمات متناسب با آن هستند. برای موسسات مالی و اعتباری و بصورت خاص بانک ها، به دلیل وجود حجم قابل توجهی از مشتریان، بررسی صفات نیازمند بهره گیری از ظرفیت های نهفته در علم داده کاوی است، داده کاوی^۱ علم استخراج الگوها، اطلاعات و تحلیل از مجموعه داده های خامی است که در یک سازمان و یا یک جامعه یا هر مجموعه دیگری تولید شده است. داده کاوی که به عنوان ابزاری برای کشف جرایم، ارزیابی میزان ریسک و فروش محصولات به کار می رود، در بر گیرنده ابزارهای تجزیه و تحلیل اطلاعات به منظور کشف الگوهای معتبر و ناشناخته در بین انبوهی از داده هاست.^۲ تکلیف اصلی نظام بانکی ارائه خدمات به مشتریان است و به واسطه آن با آنها ارتباط مستمر برقرار می کند. در همین راستا اطلاعات زیادی از مشتریان به دست می آورند و به تبع آن داده های زیادی برای تحلیل رفتار مشتری در دسترس خواهد بود، تحلیل درست این داده ها، می تواند منجر به حفظ مشتریان شود. داده کاوی می تواند اهداف مورد نظر تصمیم گیرندگان را در قالب حل مسئله تامین نماید. این روش بر روی حجم زیادی از داده ها قابل پیاده سازی بوده و با پیدا کردن ارتباطات بین الگوهای مختلف در نهایت یک سری راه حل های مهم را پیشنهاد می دهد. حجم عظیم داده ها که در نگاه اول کاربرد آن چنانی ندارد با استفاده از قابلیت های داده کاوی می تواند بسیار ارزشمند شود. در شرکت های بزرگ به ویژه بانک ها با استفاده از داده کاوی سعی می کنند این رویکرد را برای تحلیل رقبا و بازار در نظر بگیرند و به واسطه این روش پیش بینی خوبی بر روی فرصت ها و ترندها داشته باشند. این امر باعث خواهد شد فرصت برنامه ریزی برای آینده ایجاد شود و هم جهت با خواسته عموم افراد، از رقبا پیشی گیرد. یکی از مهم ترین مزایای داده کاوی در بانک ها شناخت مشتریان است، این شناخت سبب خواهد شد وفاداری مشتریان به طور دقیق دنبال شود. از سوی دیگر این روش سبب می شود که سبد محصول متناسب با نیاز مشتریان بهینه سازی شود.

۲ ادبیات موضوعی و پیشینه تحقیق

۱.۲ ضریب همبستگی

ضریب همبستگی یکی از ابزاری آماری برای تعیین نوع و رابطه دو متغیر کمی است، و یکی از معیارهای مورد استفاده در تعیین میزان همبستگی دو متغیر است. در ضریب همبستگی شدت رابطه و همچنین نوع رابطه (مستقیم یا معکوس) مشخص خواهد شد. تغییرات عددی این ضریب بین ۱ تا -۱ است و در صورت عدم وجود رابطه بین دو متغیر، مقدار آن

صفر است. همبستگی بین دو متغیر تصادفی X و Y به صورت زیر تعریف می شود:

$$\rho(X, Y) = \frac{cov(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - E(X))(Y - E(Y))]}{\sigma_X \sigma_Y}$$

که در آن E عملگر امید ریاضی، cov به معنای کوواریانس، و σ سیگما نماد انحراف معیار است.

¹ Data Mining

² <https://fa.wikipedia.org>

ضریب همبستگی پیرسون: روشی پارامتری و برای داده‌هایی با توزیع نرمال یا تعداد داده‌های زیاد کاربرد دارد. طبیعتاً تغییرات این ضریب همبستگی بین ۱- و ۱+ است. مقدار مثبت آبیانگر رابطه مستقیم بین دو متغیر است و به این معناست که اگر یکی از متغیرها افزایش (یا کاهش) یابد، دیگری نیز افزایش (یا کاهش) می‌یابد. مقدار منفی ۱ وجود یک رابطه معکوس بین دو متغیر را نشان می‌دهد. و بدان معناست که اگر یک متغیر افزایش یابد متغیر دیگر کاهش می‌یابد و بالعکس. زمانی که ضریب همبستگی برابر صفر است این نشان می‌دهد که بین دو متغیر رابطه خطی وجود ندارد.

۲.۲ رگرسیون

رگرسیون، برای تخمین روابط بین متغیرها استفاده می‌شود و شامل تکنیک‌های زیادی در مدل‌سازی است و اساس کار آن بررسی رابطه بین متغیر وابسته و یک یا چند متغیر مستقل است. در تحلیل رگرسیون پراکندگی متغیر وابسته در اطراف تابع رگرسیون حائز اهمیت است و می‌تواند توسط یک توزیع احتمال تشریح شود. عموماً تحلیل رگرسیون برای پیش‌بینی استفاده می‌شود. شناخت ارتباط میان متغیر مستقل و وابسته و شکل این روابط با استفاده از تحلیل رگرسیون امکان‌پذیر است. تکنیک‌های زیادی برای انجام تحلیل رگرسیون وجود دارد از جمله: روش‌های رگرسیون خطی و حداقل مربعات که پارامتری هستند، و رگرسیون غیر پارامتری که در یک مجموعه مشخص از توابع با احتمال پارامترهای نامحدود مورد استفاده قرار می‌گیرد.

رگرسیون خطی، یکی از روش‌های تحلیل رگرسیون برای پیش‌بینی یک متغیر از روی یک یا چند متغیر دیگر است. رگرسیون خطی نوعی تابع پیش‌بینی‌کننده خطی است که به صورت ترکیبی خطی از متغیرهای مستقل پیش‌بینی می‌شود. ساده‌ترین نوع رگرسیون خطی، رگرسیون خطی ساده است که تنها یک متغیر مستقل دارد. در رگرسیون خطی چندمتغیره به جای پیش‌بینی یک متغیر وابسته چندین متغیر وابسته پیش‌بینی می‌شود. رگرسیون خطی در علوم زیستی، رفتاری، اجتماعی، دارایی، اقتصاد و محیط زیست کاربرد دارد و مشتقات آن یکی از ابزارهای شناخته شده و پرکاربرد در یادگیری ماشین هستند.

۳.۲ خوشه بندی^۱

خوشه‌بندی، فرآیندی است که مجموعه‌ای از اشیاء به گروه‌های مجزا افراز می‌شود. هر افراز یک خوشه است. اعضاء هر خوشه متناسب با صفات و ویژگی‌هایی که دارا هستند بسیار شبیه هستند و این شباهت نسبت به خوشه‌ای دیگر کمتر است. هدف از خوشه‌بندی را می‌توان برچسب دهی به اشیاء خواند که هر شیء در خوشه‌ای عضو است. مبنای الگوریتم‌ها یا روش‌های خوشه‌بندی تقریباً یکسان است و تفاوت‌های آنها ناشی از شیوه اندازه‌گیری شباهت، فاصله و یا انتخاب برچسب برای خوشه‌ها است. مهمترین شیوه‌های محاسبه فاصله در الگوریتم‌های خوشه‌بندی فاصله اقلیدسی، منهتن و مینکوفسکی هستند.

خوشه‌بندی تفکیکی^۲: در این روش، براساس تعداد مشاهدات و تعداد گروه‌ها، عملیات خوشه‌بندی انجام می‌شود. تعداد خوشه‌ها یا گروه‌ها از قبل در این الگوریتم مشخص است و با طی مراحل خوشه‌بندی تفکیکی، هر شیء فقط و فقط به یک خوشه تعلق خواهد گرفت و هیچ خوشه‌ای بدون عضو نخواهد بود.

^۱ Clustering

^۲ Partitional Clustering

خوشه‌بندی سلسله مراتبی^۱: برعکس خوشه‌بندی تفکیکی که اشیاء را در گروه‌های مجزا تقسیم می‌شود، در خوشه‌بندی سلسله مراتبی در هر سطح از فاصله، نتیجه خوشه‌بندی مشخص می‌شود. برای نمایش نتایج خوشه‌بندی به صورت سلسله مراتبی از درختواره استفاده می‌شود.

خوشه‌بندی بر مبنای چگالی^۲: روش‌های خوشه‌بندی تفکیکی خوشه‌ها کروی شکل هستند. به این معنی که ب تشخیص خوشه‌ها وقتی مجموعه داده‌ها به شکل‌های محدب باشد خوب عمل می‌کند. اما برای داده‌های مقعر کاربرد خوبی ندارد. در الگوریتم‌ها خوشه‌بندی بر مبنای چگالی، نقاط با تراکم زیاد شناسایی شده و در یک خوشه قرار می‌گیرند.

خوشه‌بندی بر مبنای مدل^۳: در روش خوشه‌بندی بر مبنای مدل یک توزیع آماری برای داده‌ها فرض می‌شود. هدف در اجرای خوشه‌بندی بر مبنای مدل، برآورد پارامترهای توزیع آماری به همراه متغیر پنهانی است که به عنوان برچسب خوشه‌ها در مدل معرفی شده است.

۴.۲ پیشینه تحقیق

امیر شاهی و روان بد (۱۳۸۵) در راستای تقویت اعتماد مشتریان ۱۲ عامل موثر از جمله: مشاوره، کارکنان، خدمات مکمل، شهرت و پاسخگویی بانک را در قبال مشتریان شناسایی و نشان داد این عوامل در حس اعتماد مشتریان در مجموع بیش از ۶۵ درصد دارای اهمیت است. بهرام زاده و مقرب (۱۳۸۶) در راستای میزان سود دهی بانک، سوددهی با میزان اعتماد مشتریان را در یک رابطه مستقیم توصیف نمود و برای آزمون این موضوع، مشتریان بانک‌های پاسارگاد، اقتصاد نوین و استان خوزستان را ارزیابی نمود و نهایتاً وفاداری، رضایت مندی و اعتماد را از مهمترین عوامل موثر در وفاداری مشتریان ذکر نمود. محمدی (۱۳۹۰) در راستای شناسایی عوامل موثر بر وفاداری مشتریان، عوامل کارکنان، خدمات، فرآیندهای کاری، قیمت خدمات و محصولات، ویژگی شعب، فعالیت‌های بازار افزایی را به ترتیب ذکر شده معرفی نمود. نخعی و همکاران (۱۳۹۳) به بررسی عوامل مهم در جذب مشتریان پرداخته‌اند و در همین راستا به ترتیب عوامل نرخ کارمزد پایین، تقویت امکانات و فناوری، مدیریت سرمایه، عوامل فیزیکی و ارتباطات مستمر را معرفی نمودند. هایگورت (۲۰۱۵) هزینه مشتریان جدید به دلیل وجود رقابت‌های شدید را بالا توصیف نمود، لذا حفظ مشتریان و ممانعت از رویگردانی مشتریان را یک استراتژی مقرون به صرفه دانست، برای این منظور مشخصه موثر در رویگردانی، از منظر خبرگان بازار بیمه انتخاب شده است، در نهایت سن و میزان استفاده از امکانات بیمه، مهمترین شاخصه‌های اثر گذار معرفی شده است. هانگ و همکاران (۲۰۱۶) با هدف کنترل ریزش مشتریان یک مدل کلاس بندی فازی ارائه نموده است و به جهت آنکه بر اساس نظر مولفان، روش کلاس بندی مبتنی بر قواعد روشی کارا می‌باشد، با مقایسه نتایج روش کلاس بندی فازی با روش کلاس بندی مبتنی بر قواعد، دقت بیشتری را برای روش فازی نشان داده است. گردینی و وجلیو (۲۰۱۶) از یک سو به تجارت الکترونیک بعنوان یک چارچوب موثر بر آسودگی و آسان بودن معاملات مشتریان اشاره دارد و از سوی دیگر تجارت الکترونیک را در افزایش ریسک رویگردانی مشتریان موثر می‌داند. ایشان در تحقیق خود از ماشین بردار مشتریان برای شناسایی و پیش بینی رویگردانی مشتریان استفاده نمودند. ایشان این روش را به نسبت روش‌های رگرسیون لجستیک و شبکه عصبی کارا تر توصیف نمودند. نجمی و

¹ Hierarchical Clustering

² Density-Bases Clustering

³ Model-Based Clustering

همکاران (۱۳۹۷) فضای صنعتی و اقتصادی فعلی با وجود رقابت های شدید در آن را دلیل آن دانسته که بنگاه های اقتصادی از سمت جذب مشتریان جدید به سوی حفظ مشتریان قبلی تمایل داشته باشند، برای این منظور توسعه روش های شناسایی مشتریان رویگردان و پیش بینی رویگردانی را مهم تلقی نموده اند. در این پژوهش از یک الگوریتم دو مرحله ای استفاده شده است، ابتدا از دو شاخص میانگین موجودی و میانگین تراکنش مشتریان در یک بازه سه ماهه بعنوان ورودی شبکه عصبی در خوشه بندی استفاده نموده است و در مرحله کلاس بندی از داده های مربوط به تراکنش های نقدی و اعتباری به منظور طبقه بندی و پیش بینی استفاده شده است. نتایج بدست آمده بازدهی ۸۰ درصد این روش را در پیش بینی رویگردانی مشتریان نشان می دهد. نوری خان و همکاران (۱۳۹۹) هدف اصلی از انجام پژوهش، طراحی مدل ساختاری رویگردانی مشتریان از فروشگاه های ورزشی است که نتایج نشان داده است، ابعاد تک بعدی با بار عاملی ۰/۸۷، بعد بی تفاوت با بار عاملی ۰/۸۵، بعد اساسی با بار عاملی ۰/۶۸ و بعد جذاب با بار عاملی ۰/۵۹ بیشترین تاثیر را بر رویگردانی مشتریان از فروشگاه های ورزشی داشتند. همچنین، در بعد اساسی، وجود کیفیت نامناسب اجناس فروشگاه با بار عاملی ۰/۸۲، در بعد عملکردی، نبود دسترسی مناسب و راحت به فروشگاه با بار عاملی ۰/۹۴، در بعد انگیزشی، قیمت نامناسب اجناس فروشگاه با بار عاملی ۰/۸۶ و در بعد بی تفاوت، تبلیغات نامناسب فروشگاه در سطح شهر با بار عاملی ۰/۸۷، به عنوان مهم ترین گویه های رویگردانی مشتریان بودند براین اساس، می توان بیان کرد که مدیران همواره باید عوامل تاثیرگذار بر رویگردانی مشتریان خود را به خوبی درک کنند و برای کاهش این عوامل سعی کنند تا بتوانند در رقابت با سایر رقبا به مزیت رقابتی دست یابند و مشتریان وفادارتری داشته باشند. بیضایی و همکاران (۱۴۰۰) در این تحقیق از نظرات ۱۵ نفر از خبرگان برای شناسایی عوامل استفاده شده است. برای شناسایی عوامل از تکنیک دلفی استفاده شده است. براساس نتایج بدست آمده در مجموع ۲۸ عامل شناسایی که در قالب ۶ دسته کلی طبقه بندی گردیدند. براساس نتایج بدست آمده عوامل کیفیت ارتباط با مشتری، خدمات مالی، توسعه خدمات، رویه ها، بازاریابی و فضای خدماتی بر نرخ رویگردانی مشتریان اثرگذار بوده اند.

۳ روش تحقیق

در این پژوهش هدف، پیش بینی خروج یا رویگردانی تعداد ۱۰۰۰۰ مشتری می باشد که تعداد ۲۰۳۷ مشتری تا کنون خارج شده اند. برای تشریح اهداف داده کاوی در راستای هدف کسب و کار پاسخ به سوالاتی از جمله، تحلیل مشتریان باقیمانده به تناسب صفات کیفی آنها چگونه است؟ معادله رگرسیونی خروج مشتریان چیست؟ وضعیت مشتریان باقیمانده چگونه قابل ارزیابی است؟ مشتریان از دست رفته دارای چه ویژگیهایی بوده اند و راهکارهای پیشنهادی برای حفظ مشتریان چیست؟ مد نظر بوده است. برای پاسخ به سوالات فرایند داده کاوی با استفاده از روش کریسپ صورت گرفته است. الگوریتم پژوهش بگونه ای طراحی شده است تا متناسب با شاخص های معرفی شده برای مشتریان در مرحله نخست با بهره گیری از قابلیت ضریب همبستگی، شاخص های موثر بر خروج مشتریان احصا و سپس با استفاده از رگرسیون خطی، میزان اثرگذاری شاخص های موثر تعیین شود و بعنوان مدل شناسایی رویگردانی مشتریان استفاده شود. در انتها ضمن تحلیل شاخص های اثرگذار و سایر شاخص ها، مشتریان باقیمانده خوشه بندی شده اند.

۴ پیاده سازی روش کریسپ

۱.۴ درک و شناخت کسب و کار (فاز ۱ روش کریسپ)

الف. هدف کسب و کار: در این پژوهش هدف، پیش بینی خروج یا رویگردانی تعداد ۱۰۰۰۰ مشتری می باشد که تعداد ۲۰۳۷ مشتری تا کنون خارج شده اند.

ب. تعیین اهداف داده کاوی متناسب با هدف کسب و کار: برای تشریح اهداف داده کاوی در راستای هدف کسب و کار پاسخ به سوالات زیر مورد نظر است.

- تحلیل مشتریان باقیمانده به تناسب صفات کیفی آنها چگونه است؟
- معادله رگرسیونی خروج مشتریان چیست؟
- وضعیت مشتریان باقیمانده چگونه قابل ارزیابی است؟
- مشتریان از دست رفته دارای چه ویژگیهایی بوده اند و راهکارهای پیشنهادی برای حفظ مشتریان چیست؟

۲.۴ درک داده (فاز ۲ روش کریسپ)

الف. جمع آوری داده های اولیه و قالب ذخیره و نگهداری: داده های مورد نیاز پژوهش برای ۱۰۰۰۰ هزار مشتری با ۱۴ فیلد به شرح جدول زیر جمع آوری و جهت بهره برداری در قالب فایل اکسل و ... ذخیره شده است.

RowNumber	CustomerId	Surname	Credit Score	branch	Gender	Age
شماره ردیف	شماره مشتری	نام مشتری	رتبه اعتباری مشتری	شعبه	جنسیت	سن

Tenure	Balance	NumOfProducts	HasCredit Card	IsActiveMember	Estimated Salary	Exited
قدمت مشتری	مانده حساب مشتری	تعداد محصول مورد استفاده مشتری	آیا کارت اعتباری دارد یا خیر	مشتری فعال یا غیر فعال	پیش بینی درآمد مشتری	خارج شده است یا خیر

ب. بررسی مجموعه داده ها و تعیین نوع متغیرها: فیلدهای پژوهش در دو بخش متغیر کیفی شامل: شماره مشتری، نام مشتری، کشور، جنسیت و متغیر کمی شامل: سن، رتبه اعتباری، قدمت، مانده حساب، دارا بودن کارت اعتباری، فعال بودن یا غیر فعال بودن، میزان درآمد پیش بینی شده و از دست رفته یا حفظ شده قابل تقسیم هستند.

از منظر مقیاس سنجش متغیرهای کمی:

- فیلدهای داشتن کارت اعتباری، فعال یا غیر فعال بودن و از دست رفته یا حفظ شده در دسته متغیرهای اسمی هستند.
- فیلد رتبه اعتباری در دسته متغیرهای ترتیبی قرار دارد.
- فیلد های سن و تعداد محصول استفاده شده در دسته متغیرهای فاصله ای قرار دارد.

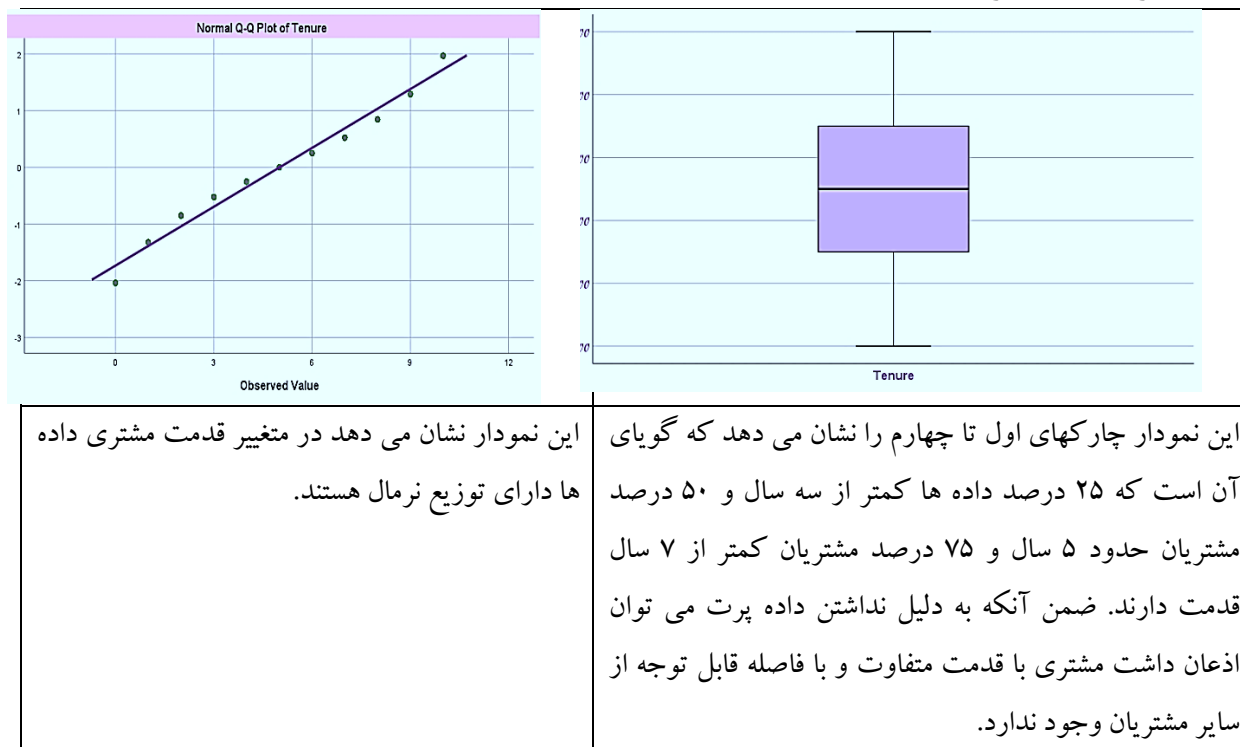
- فیلدهای قدمت، مانده حساب و درآمد پیش بینی شده برای مشتری در دسته متغیرهای نسبی قرار دارند.
- ج. تشریح داده ها: برای تشریح داده ها از خروجی نرم افزار spss استفاده شده است.

Descriptive Statistics

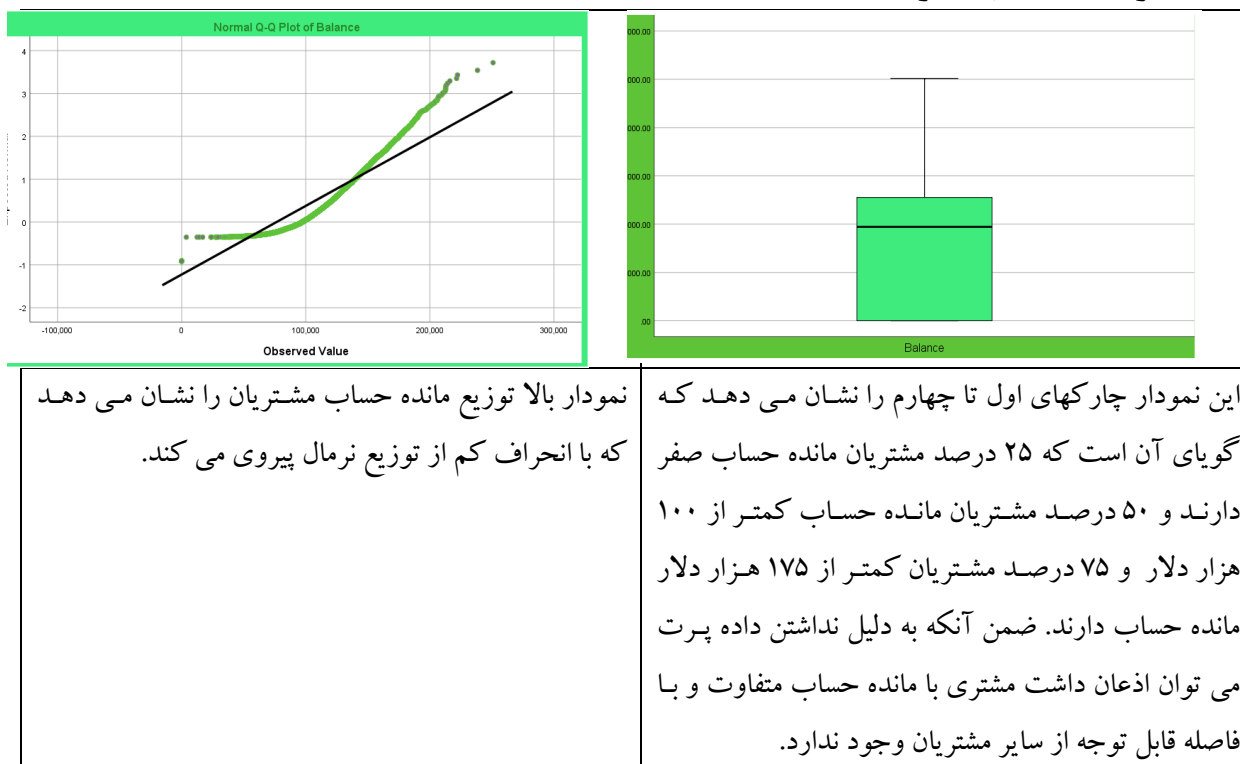
	N	Range	Minimum	Maximum	Mean	Std. Deviation
CreditScore	10000	500.00	350.00	850.00	650.5288	96.65330
Age	10000	74.00	18.00	92.00	38.9218	10.48781
Tenure	10000	10.00	.00	10.00	5.0128	2.89217
Balance	10000	250898.09	.00	250898.09	76485.8893	62397.40520
NumOfProducts	10000	3.00	1.00	4.00	1.5302	.58165
HasCrCard	10000	1.00	.00	1.00	.7055	.45584
IsActiveMember	10000	1.00	.00	1.00	.5151	.49980
EstimatedSalary	10000	199980.90	11.58	199992.48	100090.2399	57510.49282
Exited	10000	1.00	.00	1.00	.2037	.40277
Valid N (listwise)	10000					

- رتبه اعتباری مشتریان جامعه آماری در بازه ۳۵۰ تا ۸۵۰ قرار دارد و بصورت میانگین مشتریان دارای رتبه اعتباری ۶۵۰ و با انحراف معیار ۹۶ هستند، دامنه تغییرات در رتبه اعتباری ۵۰۰ می باشد.
- سن مشتریان در بازه ۱۸ تا ۹۲ سال می باشد و سن میانگین مشتریان ۳۸ سال با انحراف ده سال محاسبه شده است. دامنه تغییرات سن مشتریان ۷۴ سال است.
- دامنه تغییرات قدمت مشتریان ۱۰ سال می باشد و از مشتری فاقد سابقه تا مشتری ۱۰ سال سابقه فعالیت در بانک را شاهد هستیم. بطور میانگین مشتریان دارای سابقه ۵ سال با انحراف ۲ سال هستند.
- مشتریان دارای مانده صفر تا ۲۵۰ هزار دلار هستند و بصورت میانگین مانده حساب مشتریان ۷۶ هزار دلار با انحراف معیار ۶۲ دلار می باشد.
- تعداد محصول استفاده شده در بازه ۱ تا ۴ محصول می باشد.
- حدود ۷۰ درصد مشتریان دارای کارت اعتباری و حدود ۳۰ درصد فاقد کارت اعتباری هستند.
- حدود ۵۱ درصد مشتریان فعال و مابقی غیر فعال هستند.
- درآمد پیش بینی شده برای مشتریان در بازه ۱۱ دلار تا ۱۹۹ دلار قرار دارد و بصورت میانگین دارای درآمد ۱۰۰ هزار دلار هستند.
- حدود ۲۰ درصد مشتریان تا کنون در جامعه آماری خارج شده اند.
- به جهت آنکه ۳ متغیر دارای مقیاس سنجش نسبی هستند تشریح از طریق نرمالیزه بودن داده ها بصورت خاص انجام شده است.

متغیر قدمت مشتری:



متغیر مانده حساب مشتریان:



متغیر درآمد مشتری:



۳.۴ آمادہ سازی داده ها (فاز ۳ روش کریسپ)

در این مرحله لازم است داده های پرت، گم شده و تکراری مشخص و متناسب با اهداف حذف یا اصلاح شود، خروجی نرم افزار SPSS نشان می دهد داده ها فاقد داده پرت و گم شده و تکراری هستند.

Statistics

Indicator of each last matching case as Primary

N	Valid	10000
	Missing	0

Indicator of each last matching case as Primary				
		Frequency	Percent	Valid Percent
Valid	Primary Case	10000	100.0	100.0

۴.۴ مدل سازی (فاز ۴ روش کریسپ)

برای مدلسازی مسئله از مدل های یادگیری با نظارت استفاده می شود. مدل رگرسیون از جمله مدل های یادگیری با نظارت است که مدل های احتمالی نیز نامیده می شوند. برای ساخت مدل نظارت، از داده های تاریخی مشتریان به منظور

انجام پیش بینی استفاده می‌شود. مجموعه داده‌های تاریخی شامل برخی متغیرهای در دسترس از مشتریان (متغیرهای مستقل) و متغیر هدف (ستون برجسته) می‌باشد.

مدل پژوهش متناسب با هدفگذاری در چارت زیر ترسیم شده است.



متغیر هدف متغیری است که قصد پیش‌بینی آن را داریم. در واقع می‌خواهیم با استفاده از متغیرهای مستقل و داده‌های در دسترس، به پیش‌بینی آن پردازیم. در این مدل سازی خروج مشتریان بعنوان متغیر هدف در نظر گرفته می‌شود و متغیرهای موثر در خروج مشتریان از داده‌های در دسترس انتخاب می‌شوند. از مجموعه متغیرهای در دسترس آن دسته از متغیرهایی که دارای رابطه منطقی با متغیر هدف هستند احصا خواهند شد و میزان اثرگذاری این دسته از متغیرها با متغیر هدف محاسبه خواهد شد. برای رسیدن به یک الگو مناسب برای پیش‌بینی رفتار جامعه آماری از ظرفیت رگرسیون استفاده خواهد شد و نهایتاً رفتار مشتریان متناسب با ظرفیت‌ها و صفاتی که دارند بر مبنای الگوی تعیین شده خوشه بندی خواهند شد.

۵.۴. بکارگیری مدل برای مطالعه موردی و ارزیابی (فاز ۵ روش کریسپ)

گام ۱- با توجه به اینکه هدف اصلی در این پژوهش پیش‌بینی خروج یا رویگردانی مشتریان می‌باشد، متغیر هدف فیلد خروج یا عدم خروج مشتریان تعیین شده است.

گام ۲- از میان فیلدهای باقی مانده در جدول داده‌ها، آن دسته از فیلدهایی که بعنوان متغیر کیفی می‌باشند و درواقع تشریح صفات کیفی مشتریان است در مدل رگرسیونی دخیل نبوده و سایر متغیرهای کمی بعنوان متغیرهای در دسترس در نظر گرفته می‌شوند.

گام ۳- برای تعیین متغیرهای موثر بر متغیر هدف با استفاده از ضریب همبستگی متغیرهای غیروابسته به متغیر هدف تعیین می‌شود برای این منظور از ظرفیت نرم افزار SPSS استفاده شده است و نتایج در جداول زیر نمایش داده شده

است. به جهت سطح معنی داری کمتر از ۰,۰۵، برای ۵ فیلد رتبه اعتباری، سن، مانده حساب، تعداد محصول استفاده شده و فعال بودن یا غیر فعال بودن مشتری این دسته از متغیرها بعنوان متغیرهای موثر در پیشبینی خروج مشتریان در نظر گرفته می شود.

Correlations										
		CustomerId	CreditScore	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary
Exited	Pearson Correlation	-0.006	-.027**	.285**	-0.014	.119**	-.048**	-0.007	-.156**	0.012
	Sig. (2-tailed)	0.532	0.007	0.000	0.162	0.000	0.000	0.475	0.000	0.226
	N	10000	10000	10000	10000	10000	10000	10000	10000	10000

اعداد مثبت و منفی در محاسبه ضرایب همبستگی به شرح زیر قابل تشریح است.

- با افزایش رتبه اعتباری مشتریان احتمال خروج آنها کمتر می شود.
- با افزایش سن مشتریان، احتمال خروج مشتریان بیشتر می شود.
- مانده حساب و خروج مشتریان با یکدیگر نسبت مستقیم دارند. بدین معنی که مشتریان با مانده حساب بیشتر احتمال خروج آنها بیشتر است.
- تعداد محصول استفاده شده با خروج مشتریان نسبت عکس دارد.
- فعال یا غیر فعال بودن مشتری نیز با خروج آن نسبت عکس دارد. بدین معنی که مشتریان غیر فعال احتمال خروج آنها بیشتر است.

گام ۴- برای تعیین ضریب اثرگذاری هر یک از متغیرهای موثر در متغیر هدف، محاسبات گام قبل برای سه متغیر فعال یا غیر فعال بودن مشتری، مانده حساب مشتری و سن مشتریان انجام می شود که نتایج به شرح زیر است.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.356 ^a	.126	.126	.37654

a. Predictors: (Constant), CreditScore, Age, Balance, IsActiveMember, NumOfProducts

جدول بالا نشان می دهد ضریب همبستگی مدل پیشنهادی مقدار ۰,۳۵۶ است و درواقع نمی توان با احتمال بالا تحلیل رگرسیونی را برای جامعه آماری ارائه نمود.

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.
		B	Std. Error			
1	(Constant)	-.155	.032		-4.878	.000
	IsActiveMember	-.145	.008	-.180	-19.160	.000
	Balance	6.919E-7	.000	.107	10.919	.000
	Age	.011	.000	.297	31.676	.000
	NumOfProducts	-.003	.007	-.004	-.417	.677
	CreditScore	-9.135E-5	.000	-.022	-2.344	.019

a. Dependent Variable: Exited

به جهت آنکه در محاسبه ضرایب رگرسیونی دو متغیر رتبه اعتباری و تعداد محصول استفاده شده، دارای سطح معنی داری بیشتر از ۰,۰۵ هستند حذف آنها از مدل، پیشنهاد می شود.

گام ۵- با توجه به ضرایب تعیین شده فرمول رگرسیونی به صورت زیر می باشد.

$$\text{Exited} = -0.155 - 0.180 * \text{IsActiveMember} + 0.107 * \text{Balance} + 0.297 * \text{Age}$$

گام ۶- با هدف استفاده کاربردی از خوشه بندی مشتریان، خوشه ها در ۶ سطح در نظر گرفته شده است.

- سطح اول: خوشه مشتریان در آستانه خروج از بانک
 - سطح دوم: خوشه مشتریان با احتمال خیلی زیاد برای خروج از بانک
 - سطح سوم: خوشه مشتریان با احتمال بالا برای خروج از بانک
 - سطح چهارم: خوشه مشتریان با احتمال متوسط برای خروج از بانک
 - سطح پنجم: خوشه مشتریان با احتمال کم برای خروج از بانک
 - سطح ششم: خوشه مشتریان با احتمال زیاد برای خروج از بانک
- نتایج خوشه بندی با استفاده از دو نرم افزار SPSS و R انجام و با یکدیگر مقایسه شده است.

بخش اول: استفاده از نرم افزار SPSS

Iteration History^a

Iteration	Change in Cluster Centers					
	1	2	3	4	5	6
1	.147	.062	.146	.006	.165	.065
2	.039	.013	.019	.024	.096	.033
3	.011	.005	.001	.016	.053	.026
4	.008	.006	.000	.011	.038	.021
5	.004	.004	.000	.005	.023	.013
6	.004	.004	.000	.004	.016	.010
7	.002	.002	.000	.004	.011	.009
8	.001	.001	.000	.002	.008	.006
9	.000	.001	.000	.000	.005	.003
10	.000	.000	.000	.000	.005	.003

a. Iterations stopped because the maximum number of iterations was performed. Iterations failed to converge. The maximum absolute coordinate change for any center is .005. The current iteration is 10. The minimum distance between initial centers is .457.

جدول بالا نشان می دهد بعد از ۱۰ تکرار الگوریتم خوشه بندی توانسته است در سطح اطمینان قابل قبول خوشه بندی را انجام دهد.

Distances between Final Cluster Centers

Cluster	1	2	3	4	5	6
1		.189	.533	.619	1.217	.968
2	.189		.344	.809	1.406	1.157
3	.533	.344		1.153	1.750	1.501
4	.619	.809	1.153		.597	.348
5	1.217	1.406	1.750	.597		.249
6	.968	1.157	1.501	.348	.249	

در جدول بالا فاصله مرکز خوشه ها را از یکدیگر نشان می دهد.

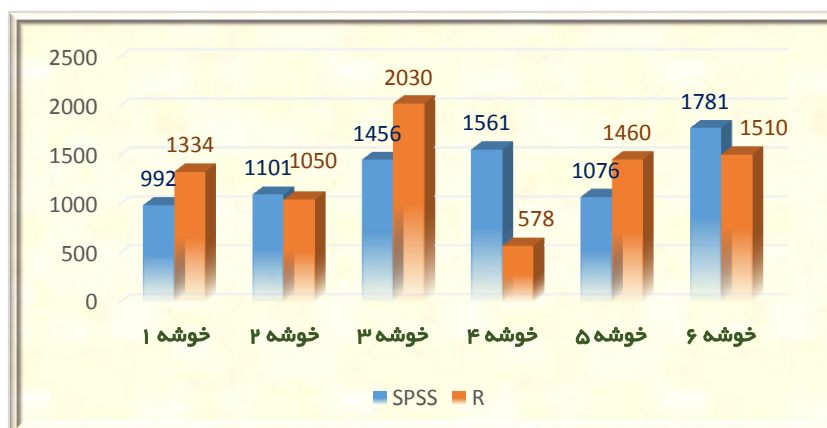
Number of Cases in each Cluster

Cluster	1	992.000	تعداد مشتری در آستانه خروج از بانک
Cluster	2	1101.000	تعداد مشتریان با احتمال خیلی زیاد برای خروج از بانک
Cluster	3	1452.000	تعداد مشتریان با احتمال بالا برای خروج از بانک
Cluster	4	1561.000	تعداد مشتریان با احتمال متوسط برای خروج از بانک
Cluster	5	1076.000	تعداد مشتریان با احتمال کم برای خروج از بانک
Cluster	6	1781.000	تعداد مشتریان با احتمال زیاد برای خروج از بانک
Valid		7963.000	داده ها همه معتبر و فاقد داده گم شده
Missing		0.000	

جدول بالا نتایج خوشه بندی را نشان می دهد.

بخش دوم: مقایسه دو روش

مقایسه دو روش خوشه بندی انجام شده با شرح نمودار زیر توصیف شده است.



۶.۴ تحلیل های کاربردی (فاز ۶ روش کریسپ)

تحلیل مشتریان باقیمانده به تناسب صفات کیفی آنها به شرح زیر قابل توصیف است.

تعداد مشتریان به تفکیک خوشه						
Row Labels	1	2	3	4	5	6
Branch1	444	518	949	1032	448	813

Branch2	345	373	23	23	381	550
Branch3	203	210	480	506	247	418
Grand Total	992	1101	1452	1561	1076	1781

جدول مقابل خوشه بندی مشتریان را به تفکیک شعبه نشان می دهد و جهت نمایش بهتر برای نشان دادن درصد مشتریان هر شعبه در هر طبقه در بخش زیرین درصد از کل محاسبه شده است

درصد از کل مشتریان به تفکیک خوشه						
	1	2	3	4	5	6
Branch1	10.6	12.3	22.6	24.5	10.7	19.3
Branch2	20.4	22	1.36	1.36	22.5	32.4
Branch3	9.84	10.2	23.3	24.5	12	20.3

تعداد مشتریان به تفکیک خوشه						
Row Labels	1	2	3	4	5	6
Female	424	479	629	680	437	755
Male	568	622	823	881	639	1026
Grand Total	992	1101	1452	1561	1076	1781

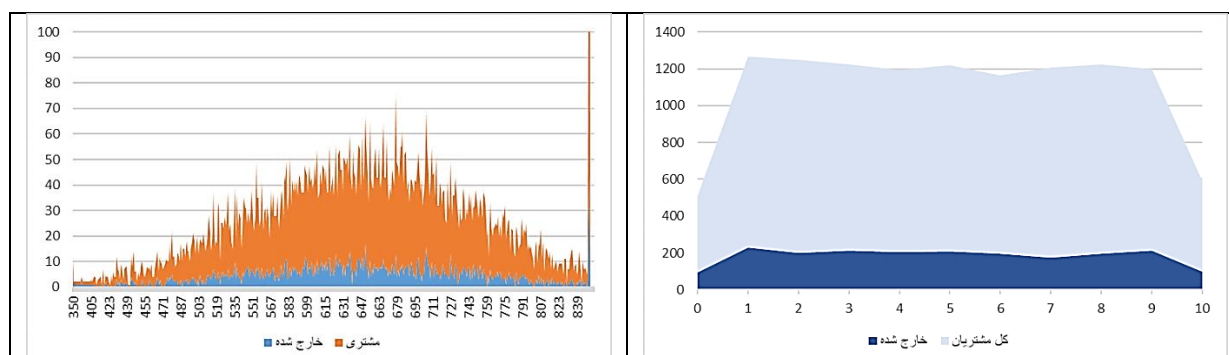
مشابه حالت قبل، با تغییر متغیر کیفی شعبه به متغیر جنسیت محاسبات انجام شده است.

درصد از کل مشتریان به تفکیک خوشه						
	1	2	3	4	5	6
Female	12.5	14.1	18.5	20	12.8	22.2
Male	12.5	13.6	18.1	19.3	14	22.5

تبادل در جنسیت در خوشه ها، از موارد جالب در بخش درصد گیری می باشد.

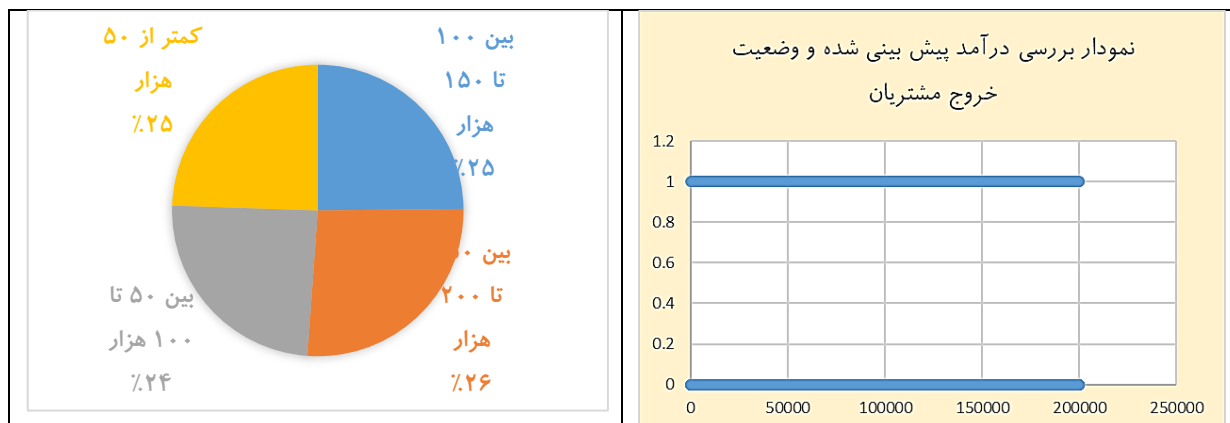
- در داده های جامعه آماری ارتباط

معنی داری بین متغیر قدمت مشتری، داشتن کارت اعتباری، درآمد پیش بینی شده با خروج مشتریان مشاهده نشده است.



نمودار بالا نشان می دهد رتبه اعتباری نمی تواند بعنوان یک استراتژی قابل اتکا برای حفظ مشتریان باشد. چرا که نوسانات خروج مشتریان به تناسب رتبه اعتباری دارای نسبت مشابه است و برای رتبه های اعتباری بالاتر نسبت حفظ مشتریان قابل توجه نمی باشد.

نمودار بالا نشان می دهد، قدمت مشتریان نتوانسته است، بعنوان یک شاخص در ممانعت از خروج مشتریان نقش ایفا کند، پیشنهاد می شود، برای مشتریان با قدمت بالا از اهرم های انگیزشی برای حفظ ایشان استفاده شود.



در دو نمودار بالا نشان داده شده است، وضعیت خروج یا عدم خروج مشتریان با درآمد های پیش بینی شده بدون ارتباط است. در نمودار دایره ای نشان داده شده است با دسته بندی مشتریان به ۴ گروه درآمدی، و سنجش درصد مشتریان خارج شده در هر دسته، نسبت مساوی حدود ۲۵ درصد خروج مشتری دیده می شود. این موضوع می تواند این نگرش را تقویت نماید که استراتژی خاصی برای حفظ مشتریان با درآمد پیش بینی شده بالا وجود نداشته است. و این می تواند بعنوان یک استراتژی هدفمند برای بهره گیری از مشتریان پردرآمد مد نظر قرار گیرد.

از لحاظ صفات کیفی مشتریان، در نمودارهای بعدی نشان داده شده است که کشور آلمان ضریب خروج مشتری بیشتری نسبت به دو کشور دیگر داشته که نیازمند شناسایی چالش های آن کشور می باشد. و خروج مشتریان زن نسبت به مشتریان مرد نیز نیازمند بررسی موضوعی دارد.

Row Labels	خارج شده	کل مشتری	درصد
Branch1	810	5014	16.2
Branch2	814	2509	32.4
Branch3	413	2477	16.7
Grand Total	2037	10000	20.4

Row Labels	خارج شده	کل مشتری	درصد
Female	1139	4543	25.1
Male	898	5457	16.5
Grand Total	2037	10000	20.4

در شعبه ۲ نسبت خروج مشتریان بیشتر از دو شعبه دیگر بوده است بنابراین تمرکز برای حفظ مشتریان شعبه ۲ و بررسی دلیل آن ضروری می باشد.

نسبت خروج مشتریان زن بیش از نسبت خروج مشتریان مرد بوده است. بنابراین تمرکز برای حفظ مشتریان زن بیشتر بایستی مورد توجه قرار گیرد

بحث و نتیجه گیری

با پیاده سازی الگوریتم پژوهش، سه شاخص سن، مانده حساب و فعال بودن مشتریان دارای ارتباط معنی داری با خروج مشتریان داشتند و بعنوان شاخص های موثر مورد توجه قرار گرفتند. با استفاده از رگرسیون خطی مشتریان باقی مانده متناسب با میزان شباهت با مشتریان خارج شده در ۶ خوشه طبقه بندی شدند. برای مشتریان خارج شده، لازم است تصمیم گیرنده با صفات آنها با شرح زیر توجه نماید.

- با افزایش رتبه اعتباری مشتریان احتمال خروج آنها کمتر می شود و می توان اینگونه استنباط کرد که مشتریان با رتبه اعتباری بالاتر تعلق خاطر بیشتری به بانک دارند. البته باید عنوان کرد این موضوع با سطح معنی داری بالایی، قابل اتکا نمی باشد.

- خروج مشتریان با سن بیشتر در جامعه آماری بیشتر مشاهده شده است.

- برای مشتریان با مانده حساب بیشتر خروجی بیشتری در جامعه آماری دیده شده است، بنابراین تمرکز بر مشتریان خرد به رغم آنکه مانده حساب کمتری دارند، می تواند یک استراتژی برای دراز مدت باشد و طبیعتاً حفظ چرخه ورود مشتریان با مانده حساب بالا به دلیل احتمال خروج مشتریان فعلی بعنوان استراتژی کوتاه مدت و میان مدت ضروری می باشد.

- هرچه مشتریان از تعداد محصول بیشتری استفاده نموده اند، تعلق آنها بیشتر شده و خروجی کمتری در جامعه آماری مشاهده شده است. بنابراین ترغیب نمودن مشتریان به استفاده از محصولات بیشتر موجب افزایش حاشیه امنیت برای حفظ مشتریان خواهد بود.

- در تحلیل های آماری نشان داده شده است که مشتریان غیرفعال خروج بیشتری داشته اند و می توان اینگونه استنباط نمود غیر فعال بودن مشتری زمینه خروج آن را فراهم خواهد ساخت، بنابراین تمرکز بر مشتریان غیر فعال و احیای این مشتریان پیشنهاد می شود. از منظری دیگر می توان اذعان داشت غیرفعال شدن مشتری، زنگ خطر برای خروج آن است.

از بعد تمرکز برای حفظ مشتریان باقیمانده در ۶ سطح می توان مشتریان را طبقه بندی نمود و درصدد کنترل و مدیریت آنها اقدام نمود.

- تعداد ۹۹۲ مشتری در آستانه خروج از بانک

- تعداد ۱۱۰۱ مشتریان با احتمال خیلی زیاد برای خروج از بانک

- تعداد ۱۴۵۲ مشتریان با احتمال بالا برای خروج از بانک

- تعداد ۱۵۶۱ مشتریان با احتمال متوسط برای خروج از بانک

- تعداد ۱۰۷۶ مشتریان با احتمال کم برای خروج از بانک

- تعداد ۱۷۸۱ مشتریان با احتمال زیاد برای خروج از بانک

برای حفظ مشتریان و متناسب با اقتضائات و ظرفیت های مشتریان پیشنهادهای زیر به تصمیم گیرنده که حاصل تحلیل رفتاری ۱۰۰۰۰ مشتری است داده می شود.

- پیشنهاد می شود، برای مشتریان با قدمت بالا از اهرم های انگیزشی برای حفظ ایشان استفاده شود.

- رتبه اعتباری نمی تواند بعنوان یک استراتژی قابل اتکا برای حفظ مشتریان باشد. چرا که نوسانات خروج مشتریان به تناسب رتبه اعتباری دارای نسبت مشابه است و برای رتبه های اعتباری بالاتر نسبت حفظ مشتریان قابل توجه نمی باشد.
- تحلیل ها نشان می دهد در جامعه آماری پژوهش وضعیت خروج یا عدم خروج مشتریان با درآمد های پیش بینی شده بدون ارتباط است. با دسته بندی مشتریان به ۴ گروه درآمدی، و سنجش درصد مشتریان خارج شده در هر دسته، نسبت مساوی حدود ۲۵ درصد خروج مشتری دیده می شود. این موضوع می تواند این نگرش را تقویت نماید که استراتژی خاصی برای حفظ مشتریان با درآمد پیش بینی شده بالا وجود نداشته است. و این می تواند بعنوان یک استراتژی هدفمند برای بهره گیری از مشتریان پردرآمد مد نظر قرار گیرد.
- نسبت خروج مشتریان زن بیش از نسبت خروج مشتریان مرد بوده است. بنابراین تمرکز برای حفظ مشتریان زن بیشتر بایستی مورد توجه قرار گیرد
- در شعبه ۲ نسبت خروج مشتریان بیشتر از دو شعبه دیگر بوده است بنابراین تمرکز برای حفظ مشتریان شعبه ۲ و بررسی دلیل آن ضروری می باشد.

منابع

- نجمی، پ.، و راد، ع.، و شعار، م. (۱۳۹۷). پیش بینی رویگردانی مشتریان بانک با استفاده از روش داده کاوی. فصلنامه مدیریت صنعتی دانشگاه علوم انسانی، دانشگاه آزاد اسلامی واحد سنجند، ۱۳(۴۴).
- بیضایی، سیدحمید و بستام، هادی، (۱۴۰۰)، شناسایی عوامل موثر در کاهش نرخ رویگردانی مشتریان در موسسه اعتباری ملل، اولین کنفرانس بین المللی آزمایشگاه مدیریت و رویکردهای نوآورانه در مدیریت و اقتصاد، تهران، <https://civilica.com/doc/1347828>
- نوری خان یوردی، میثم و بشیری، مهدی و دوستی، مرتضی، (۱۳۹۹)، مدل ساختاری رویگردانی مشتریان از فروشگاه های ورزشی، <https://civilica.com/doc/1211517>
- Amir Shahi, Mirahmad, Siahtiri, Vida, Ravanbod, Fariba. (2010). Identifying the factors of Entrepreneur in Tehran. Management Researches in Iran, 4(64): 61-67 .
- Bahramzadeh, Mohammad Mehdi, Shokati Maghareb, Somayeh. (2010). Identification and ranking of factors affecting customer loyalty in private banks of Khuzestan province .Second International Financial Services Marketing Conference .
- Castanedo, Federico, et al. (2014). Using deep learning to predict customer churn in a mobile telecommunication network.
- Datta, P., Masand, B., Mani, D., & Li, B. (2000). Automated cellular modeling and prediction on a large scale. Artificial Intelligence Review, 14: 485–502.
- Ghanbarterani, Nasim. (2014). Customer clustering based on RFM model and data mining approach to increase customer loyalty. MSc Thesis, Tarbiat Moalem University, Faculty of Engineering.
- Gordini, Niccolò, and Valerio, Veglio. (2017). Customers churn prediction and marketing retention strategies: An application of support vector machines based on the

auc parameter-selection technique in b2b e-commerce industry. *Industrial Marketing Management*, 62: 100-107.

Huang, Bingquan et al. (2016). A Fuzzy Rule-Based Learning Algorithm for Customer Churn Prediction. *Industrial Conference on Data Mining*. Springer International Publishing, 2016.

Huigevoort, Chantine, and Remco, Dijkman. (2015). Customer churn prediction for an insurance company. Diss. M. Sc. Thesis, Eindhoven University of Technology, Eindhoven, Netherland, 2015.

Mutanen, Teemu. (2006). Customer churn analysis—a case study. *Journal of Product and Brand Management*, 14(1): 4-13.

Prasad, U. Devi, and S. Mahdavi. (2012). Prediction of churn behavior of bank customers using data mining tools. *Business Intelligence Journal* 5(1): 96-101.

Sharma, Anuj, Panigrahi, and Prabin Kumar. (2013). A neural network based approach for predicting customer churn in cellular network services. *Ar Xiv preprint Ar Xiv:1(309):39-45*.

Analyzing the customers who turned away from the bank and clustering other customers based on the degree of similarity with the turned away customers

Roohollah Kiani Ghaleno¹
Omalbanin Setoode²

Date of Receipt: 2022/06/11 Date of Issue: 2022/08/24

Abstract

In the banking system, where the cycle of inflow and outflow of cash in the form of depositing and payment of facilities is considered as its main activity, customer retention is very important. For this purpose, bank managers have significantly focused on customer affairs and in this regard, data mining and various statistical analysis methods can play an effective role in decision making. In this research, a method will be proposed in which an algorithm based on the correlation coefficient will identify the indicators affecting customers who have recently left the bank, and then using linear regression, the model of customer withdrawal will be extracted and finally According to the similarity of customers with the regression model, customers will be classified into 6 clusters. In this regard, analyzes of other indicators and managerial expectations have been presented to focus and retain customers.

Keywords

Data mining, linear regression, correlation multiplication, clustering

1. Department of Industrial Engineering, Aliabad Katul Branch, Islamic Azad University, Aliabad Katul, Iran. rohollah.ghaleno@yahoo.com.
2. Department of Law, Zahedan Branch, Islamic Azad University, Zahedan, Iran. setode.amir2@gmail.com