

تعیین ماندگاری و سودآوری مشتریان بانک با استفاده از تعمیم مدل درخت تصمیم و جنگل رگرسیون

محمدتقی تقوی فرد^{۱*}، رضا حبیبی^۲، حامد عبداللہی^۳

۱- دانشیار، گروه مدیریت، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران
۲- استادیار، گروه بانکداری، دانشکده بانکداری، موسسه عالی آموزش بانکداری ایران، تهران، ایران
۳- کارشناسی ارشد بانکداری، دانشکده بانکداری، موسسه عالی آموزش بانکداری ایران، تهران، ایران

پذیرش: ۱۳۹۵/۱۲/۷

دریافت: ۱۳۹۵/۸/۱۲

چکیده

در این تحقیق با استفاده از روش‌های جنگل تصادفی و جنگل رگرسیون^۲ که تعمیم‌یافته مدل درخت تصمیم و رگرسیون هستند، عوامل مؤثر بر ماندگاری و سودآوری مشتریان ارزی یک بانک تجاری دولتی، مورد بررسی قرار می‌گیرد. جامعه آماری تحقیق شامل مشتریانی است که دارای یکی از انواع حساب‌های ارزی بانک بوده و به‌طور همزمان از خدمات حواله‌های ارزی، گشایش اعتبارات اسنادی و تسهیلات ارزی استفاده می‌نمایند. همچنین صحت نتایج با رگرسیون لجستیک و بگینگ^۳ مقایسه می‌شود. در روش جنگل رگرسیون، با استفاده از معیار درصد میانگین قدرمطلق خطا (MAPE)^۴، به مقایسه دقت نتایج مدل با روش رگرسیون خطی پرداخته می‌شود. سپس درجه اهمیت هر یک از متغیرهای مستقل بر روی متغیرهای وابسته خرید بعدی، کاهش سطح فعالیت، کاهش سودآوری و تداوم سودآوری تعیین می‌شود.

نتایج نشان می‌دهد در صورت ارائه تسهیلات ارزی بیشتر به مشتریان تولیدی و در نتیجه، گشایش اعتبارات اسنادی یا انجام حواله‌های بیشتر برای این قبیل از مشتریان، احتمال

E-mail: dr.taghavifard@gmail.com

* نویسنده مسئول مقاله:

2. Random Forest & Regression Forest

3. Bagging

4. Mean Absolute Percentage Error

افزایش دوره ماندگاری مشتریان افزایش می‌یابد. همچنین افزایش مبلغ تسهیلات ارزی ارائه‌شده، حواله‌های انجام‌شده و همچنین اعتبارات اسنادی گشایش شده به‌واسطه نرخ سود بازپرداخت تسهیلات، کارمزدها و سود حاصله از فروش ارز و ضمانت‌نامه‌ها در بالا بردن میزان سودآوری مشتریان نقش بسزایی دارند.

کلیدواژه‌ها: ماندگاری مشتریان؛ سودآوری؛ بانک؛ درخت تصمیم؛ جنگل رگرسیون.

۱- مقدمه

برقراری ارتباط با مشتریان به‌منظور شناسایی، حفظ مشتریان کلان و بهینه‌سازی ارائه خدمات در راستای افزودن به دوره ماندگاری مشتریان به سبب افزایش میزان رضایت آن‌ها و همچنین بالا بردن سودآوری مشتریان ناشی از افزایش قدرت رقابتی و درنتیجه، بیشتر نمودن درآمد حاصل از فروش یا ارائه خدمت در موفقیت بانک‌ها نقش بسزایی دارد. ماندگاری مشتریان یک بانک ارتباط مستقیمی با وضعیت بانک از نظر بازگشت سرمایه، سود فروش، رشد سهم بازار، رشد قیمت سهام و کاهش هزینه‌های بانک دارد [۱]. از این رو، بانک‌ها می‌بایست به‌واسطه به‌کارگیری فناوری‌ها و روش‌های اطلاعاتی پیشرفته، ضمن تلاش برای برقراری ارتباط مؤثر با مشتریان، در جهت حفظ طولانی‌مدت این روابط برای تداوم سودآوری ناشی از ارائه خدمات به مشتریان فعالیت نمایند [۲]. لذا ضروری است که سازمان‌ها به مطالعه برای ارزیابی و اجرای استراتژی‌های ماندگاری مشتریان و هدفمند کردن آن‌ها به‌منظور ایجاد و ترویج روابط قوی با مهم‌ترین مشتریان خود بپردازند [۳]. از دست دادن یک مشتری چیزی بیش از دست دادن یک قلم فروش است، زیرا به معنای از دست دادن کل جریان خریدهایی است که مشتری می‌توانست در طول دوره‌ای از زمان انجام دهد. بنابراین، افزایش دوره ماندگاری مشتریان منجر به بیشتر شدن سود آن‌ها خواهد شد [۴]. رویگردانی مشتری از جنبه‌های گوناگونی مورد مطالعه قرار گرفته است؛ مثلاً سائل به اهمیت نقش نشان تجاری در ماندگاری مشتریان اشاره داشته است. در مورد نقش قضاوت مشتری بر روی محصولات و رضایتمندی آن‌ها و ماندگاریشان به مطالعه پرداختند [۵؛ ۶]. پارک و هان این مفهوم را از دیدگاه تنوع محصولات و رضایتمندی مشتریان بررسی نمودند [۷]. در دهه‌های اخیر، شرکت‌های زیادی ماندگاری مشتریان را به‌عنوان یک اصل مهم در تصمیمات مدیریتی و بازاریابی خود

موردتوجه قرار داده‌اند. رقابتی شدن فضای اقتصاد موجب شده است که مشتری به‌عنوان رکن اساسی فعالیت‌های سازمان موردتوجه قرار گیرد. سازمان‌ها به این مهم دست‌یافته‌اند که حفظ مشتریان فعلی ارزان‌تر از جذب مشتریان جدید است و بهترین راه حفظ مشتریان نیز جلب رضایت آن‌هاست. رضایتمندی پشتوانه‌ای برای حفظ روابط ماندگار است و این رابطه به نوع و کیفیت روابط بستگی دارد.

۲- بیان مسئله

اهمیت ماندگاری بر این فرض بنا نهاده شده است که یک پیوستگی قوی میان ماندگاری مشتری و سودآوری وجود دارد. مشتریان قدیمی خرید بیشتری انجام می‌دهند و در خدمت‌رسانی کم‌هزینه‌تر هستند [۸]. از طرفی، مشتریان قدیمی حساسیت کمتری نسبت به تبلیغات داشته و به‌راحتی هزینه‌ها و قیمت‌ها را می‌پذیرند؛ از طرف دیگر، جایگزینی مشتریان موجود با مشتریان جدید یک خط‌مشی بسیار هزینه‌بر و پرخطر است [۹]. با توجه به ویژگی‌های منحصربه‌فرد بخش خدمات مالی، بانک‌ها و مؤسسات مالی می‌بایست برای حفظ سهم خود در بازار و رشد سودآوری، نسبت به توسعه و تحکیم روابط ماندگار با مشتریان اقدام نمایند [۱۰]. از این‌رو، شناسایی مشتریان برای بانک از منظر ماندگاری و سودآوری، تعیین درجه اهمیت هر یک از خدمات ارائه‌شده در میزان بالا بردن دوره ماندگاری و تأثیرات آن خدمات در افزایش سودآوری مشتریان برای بانک و همچنین تعیین سمت‌وسوی خدمات بر ماندگاری و سودآوری از اهمیت ویژه‌ای برخوردار است. تحقیق حاضر بر این فرض استوار است که به‌کارگیری روش‌های مدیریت ارتباط با مشتریان و حفظ طولانی‌مدت روابط مؤثر و افزایش دوره ماندگاری آن‌ها، موجب افزایش سودآوری مشتریان بانک می‌شود. در ایران نیز طی دهه‌های اخیر با ظهور بانک‌های خصوصی و رقابتی-ترشدن بانکداری در کشور، جلب رضایتمندی مشتریان و افزایش سهم بازار و سودآوری موجب مضاعف شدن اهمیت شناخت و درک مشتریان و همچنین پیش-بینی رفتار و تبیین میزان ماندگاری و سودآوری مشتریان در تصمیم‌گیری مدیران شده است. از این‌رو در این پژوهش، تلاش می‌شود ماندگاری و سودآوری مشتریان در سطح شعب یک بانک دولتی در ایران موردبررسی قرار گیرد. افزون بر این، میزان

اثرگذاری و درجه اهمیت برخی از خدمات ارزی قابل‌ارائه به مشتریان که در افزایش دوره ماندگاری و سودآوری آن‌ها نقش بسزایی دارند نیز مطالعه می‌شود.

۳- مبانی نظری

در این بخش، به تبیین مبانی نظری روش درختان تصمیم و جنگل تصادفی می‌پردازیم. از جمله مزایای این دو روش این است که از ساختار داده‌کاوی برخوردارند و همانند روش‌های استنباط آماری نیاز به برازش فرض‌های توزیعی روی داده‌ها نیست. همچنین این دو روش قواعد تصمیم‌گیری ساده‌ای را برای بخش‌بندی مشتریان رویگردان و ماندگار و تعیین اثرات متغیرها بر ماندگاری مشتریان تبیین می‌نمایند.

۳-۱- درخت تصمیم

درخت تصمیم یک روش داده‌کاوی است و ابزاری ساده و توانمند برای پشتیبانی از تصمیم است که از درختان برای بخش‌بندی مجموعه‌ای از داده‌های نامتجانس به بخش‌های متمایز و متجانس استفاده می‌کند. درخت تصمیم به‌طور خاص در تحلیل تصمیم برای مشخص کردن راهبرد که با بیشترین احتمال به هدف برسد، به کار می‌رود و به کمک آن می‌توان خصوصیات کمی و کیفی متغیرها را به‌طور همزمان پردازش نمود. درختان تصمیم از یک دسته سؤال تشکیل می‌شوند که نمونه آموزشی را به بخش‌های کوچک‌تری تقسیم می‌کند. این امر تا جایی ادامه می‌یابد که مجموعه‌ای از مشاهده‌ها باقی بمانند که آن‌ها را به هیچ طریقی نتوان در گروه‌های جدا از هم قرار داد.

۳-۲- مدل کارت^۱

به‌طورکلی، روش کارت شامل سه مرحله است:

- ۱- ساخت بزرگ‌ترین درخت؛
- ۲- انتخاب اندازه صحیح درخت؛
- ۳- طبقه‌بندی داده‌های جدید با استفاده از درخت ساخته‌شده.

1. CART: Classification and Regression Trees

درختان تصمیم به واسطه مجموعه‌ای از سؤالات (به شکل بله و خیر) که نمونه آموزشی را به بخش‌های کوچک‌تر تقسیم می‌کند، شکل می‌گیرند. کارت در میان کلیه متغیرها بهترین جداکننده را که داده‌ها را به دو بخش با حداکثر میزان همگنی تقسیم می‌کند، جستجو می‌نماید. با تکرار این فرایند به یک درخت پیچیده با سطوح و متغیرهای بسیار دست می‌یابیم که می‌بایست پیش از استفاده از درخت برای طبقه‌بندی داده‌های جدید، اندازه آن بهینه شود. این فرایند در اصطلاح هرس کردن درخت نامیده می‌شود که شامل قطع گره‌ها و حتی زیرشاخه‌های بی‌معنی است. پس از طراحی، درخت تصمیم برای طبقه‌بندی مشاهدات جدید قابل استفاده است. مدل کارت به دو منظور مورد استفاده قرار می‌گیرد. در صورتی که متغیر پاسخ در تحقیق گسسته باشد، از روش درخت طبقه‌بندی و اگر پیوسته باشد، از روش درخت رگرسیون استفاده می‌شود.

۳-۳- مدل جنگل تصادفی

استفاده از درخت‌های تصمیم به سبب سادگی در استفاده، قابلیت تفسیر و همچنین به جهت توانایی آن‌ها در اندازه‌گیری با مقیاس‌های کوواریانس در سطوح مختلف اندازه‌گیری (از جمله متغیرهای اسمی) بسیار رایج شده‌اند [۱۱]. روش جنگل تصادفی، تعمیم‌یافته روش درخت طبقه‌بندی در مدل کارت است؛ با این تفاوت که در این مدل با مجموعه‌ای از درختان تصمیم مواجه هستیم که هر یک، طبقه‌بندی (رأی) ارائه داده و طبقه‌ای که بیشترین رأی را به دست آورد، به‌عنوان نتیجه مدل انتخاب می‌شود. از طرفی، برخلاف مدل کارت، در مدل جنگل تصادفی درخت‌های تصمیم هرس نمی‌شوند و تلاش می‌شود به‌واسطه انتخاب بهترین جداساز در میان زیرمجموعه‌ها، درحالی‌که تنوع درختان زیاد است، دقت پیش‌بینی حفظ شود. این مدل، مقیاسی از اهمیت متغیرها را ارائه می‌دهد که برای تعیین میزان تأثیر هر یک از عوامل به کار می‌رود [۱۲].

مدل جنگل تصادفی به‌واسطه بهره‌مندی از تعداد زیاد درختان از انحراف کمتری نسبت به روش درخت تصمیم دارد. از طرفی به دلیل نمونه‌برداری تصادفی پیش‌بین‌ها در هر گره و ناهمسانی درخت‌ها، همبستگی درخت‌ها در مدل جنگل تصادفی کم است. بهترین روش برای تعیین تعداد درختان موردنیاز، مقایسه

پیش‌بینی انجام‌شده به‌وسیله جنگل با پیش‌بینی‌های حاصل از زیرمجموعه‌های آن است. در صورتی‌که پیش‌بینی زیرمجموعه به‌خوبی پیش‌بینی حاصل از جنگل باشد، تعداد درختان کافی است [۱۳].

الگوریتم این روش به شرح ذیل است:

- ۱- تعداد آزمایشات را N و تعداد متغیرها را M می‌نامیم؛
- ۲- به‌منظور تصمیم‌گیری در هر گره از درخت، m تعداد متغیر را وارد می‌کنیم. مقدار m می‌بایست کمتر از M باشد ($m \leq M$)؛
- ۳- انتخاب داده‌های آزمایش برای این درخت با استفاده از n بار جایگذاری از نمونه‌های N (یعنی انتخاب نمونه‌های خودگردان‌سازی‌شده) 1 و استفاده از بقیه نمونه‌ها برای برآورد خطای درخت با پیش‌بینی طبقات مختلف؛
- ۴- انتخاب تصادفی m متغیر برای هر گره درخت که پایه تصمیم‌گیری در هر گره قرار می‌گیرد. محاسبه بهترین گروه‌ها بر اساس m متغیر موجود در گروه نمونه‌های آزمایش؛
- ۵- هر درخت به‌طور کامل و بدون اینکه هرس شود، بسط داده می‌شود.

با در نظر گرفتن جنگل تشکیل‌شده از M درخت، پیش‌بینی درخت m ام از مقدار X ($T_m(X)$) به شرح زیر است:

$$T_m(X) = \sum_{i=1}^n W_{im}(X) Y_i$$

که در آن $W_{im} = 1/k_m$ و X_i در یک برگ مشابه در درخت m ام و k_m تعداد آزمایشی است که در همان برگ به‌عنوان X در درخت m ام قرار دارد. پیش‌بینی کل جنگل $F(X)$ به شکل زیر است:

$$F(X) = \frac{1}{M} \sum_{m=1}^M T_m(X) = \frac{1}{M} \sum_{m=1}^M \sum_{i=1}^n W_{im}(X) Y_i = \sum_{i=1}^n \left(\frac{1}{M} \sum_{m=1}^M W_{im}(X) \right) Y_i$$

1. Bootstrapped

که نشان می‌دهد پیش‌بینی جنگل تصادفی، میانگین موزون Y_i ها بوده و وزن آن یعنی $W_i(X)$ به شرح زیر است:

$$W_i(X) = \frac{1}{M} \sum_{m=1}^M W_{im}(X)$$

همسایه‌های X در این روش نقاط X_i هستند که در یک برگ (X) و حداقل در یکی از درختان جنگل قرار می‌گیرند. در اینجا، همسایگی X بستگی به مجموعه‌های درختان دارد. این بدان معنی است که داده‌های آزمایشی (آزمون) مهم هستند. در مورد درخت و جنگل تصادفی، مراجع مهمی وجود دارد؛ از جمله، کلمانس به رتبه‌بندی و مرتب‌سازی جنگل‌های تصادفی با معیارهای گوناگون پرداخته است [۱۴]. دنیل در مورد نحوه ساختن جنگل‌های تطبیقی و همزمان مطالعه نمودند و سازگاری نتایج را با جنگل‌های متعارف مقایسه نمودند [۱۵]. سورنت و همکاران در مورد سازگاری و رفتارهای مجانبی در جنگل‌های تصادفی بررسی نمودند [۱۶؛ ۱۷]. بایو و همکاران در مورد رفتارهای برآوردهای حاصل از روش جنگل تصادفی و رفتارهای حدی این برآوردها در جنگل تصادفی بررسی نمودند [۱۸]. ابرازی و همکاران به بررسی تأثیر تناسب فرد با سازمان و فرهنگ‌سازمانی و ماندگاری مشتریان مبادرت ورزیدند [۱۹]. سپهری و همکاران به مطالعه کشف دلایل رویگردانی مشتریان از بانکداری الکترونیک پرداختند و با تعیین روش‌های برتر سنجش، ماندگاری آن‌ها در دوره‌های یادگیری الکترونیکی را مدنظر قرار دادند [۲۰]. رضاپور و همکاران سیستمی توصیه‌گر برای بهینه‌سازی سبد تسهیلات بانک بر مبنای درخت تصمیم ارائه دادند [۲۱]. سهرابی و همکاران و نیز کریمی و کاظمی به پیش‌بینی خرابی‌های خودپردازهای بانکی با استفاده از روش درخت تصمیم اقدام ورزیدند [۲۲؛ ۲۳].

۳-۴- مدل جنگل رگرسیون

روش جنگل تصادفی به روش‌های معمول رگرسیونی نیز بسط داده می‌شود. این مدل، توسعه‌یافته روش درخت رگرسیون است که برای پیش‌بینی داده‌های پیوسته از دسته-ای از درختان استفاده می‌نماید. جنگل‌های تصادفی در روش رگرسیون به‌واسطه رشد درختان در یک شاخه (بردار) شکل می‌گیرند؛ به‌گونه‌ای که متغیرهای پیش‌بین هر

درخت، مقادیر عددی در مقابل برچسب‌های طبقات مختلف به خود می‌گیرند. مقادیر متغیر پیش‌بین جنگل‌های تصادفی، همان میانگین تعداد درخت‌های مشخص شده توسط محقق هستند. پیش‌بینی یک درخت رگرسیون به صورت زیر است:

$$\hat{Y} = \frac{1}{M} \sum_{m=1}^M \hat{Y}_i = \frac{1}{M} \sum_{m=1}^M \sum_{i=1}^n W_{im} X_i = \sum_{i=1}^n \frac{1}{M} \sum_{m=1}^M W_{im} X_i$$

در این فرمول $W_{im} = 1/k_m$ و X_i در یک برگ مشابه در درخت m ام و k_m تعداد آزمایشی است که در همان برگ به عنوان X در درخت m ام قرار دارد. همچنین M تعداد متغیرها و m تعداد متغیرها در هر گره می‌باشد ($m \leq M$).

۳-۵- ارزیابی و اعتبارسنجی مدل

در این پژوهش، به منظور ارزیابی مدل جنگل تصادفی، از معیار فضای زیرمنحنی ROC^۱ استفاده می‌شود. بعلاوه، عملکرد جنگل تصادفی در برابر نتایج فضای زیرمنحنی (AUC)^۲ حاصل از مدل‌های رگرسیون لجستیک و بگینگ، محک زده می‌شود. فضای زیر این منحنی برای ارزیابی دقت پیش‌بینی مدل‌های طبقه‌بندی به کار می‌رود [۲۴].

$$TP Rate = \frac{\text{متغیرهای مثبتی که صحیح طبقه بندی شده‌اند}}{\text{کل متغیرهای مثبت}}$$

$$FP Rate = \frac{\text{متغیرهای منفی که اشتباه طبقه بندی شده‌اند}}{\text{کل متغیرهای منفی}}$$

برای مقایسه مدل جنگل تصادفی از روش رگرسیون لجستیک و مدل بگینگ استفاده می‌شود.

-
1. Receiver Operating Characteristic
 2. Area Under the Receiver Operating Characteristic Curve

۳-۶- روش بگینگ^۱

روش بگینگ روشی برای تولید انواع گوناگونی از متغیرهای پیش‌بین است تا با استفاده از آن بتوان به یک متغیر پیش‌بین جامع رسید. تجمیع این اطلاعات با محاسبه میانگین اکثریت آراء در مقادیر پیش‌بینی شده انجام می‌پذیرد. با استفاده از روش بگینگ می‌توان B تعداد نمونه خودگردان‌سازی^۲ شده با نمونه‌گیری تصادفی را با جایگذاری از داده‌های یادگیری به دست آورد. به عبارت دیگر، انواع متعدد و مختلف یک متغیر با تکرار خودگردان‌سازی داده‌های یادگیری به دست می‌آید که از آن‌ها به عنوان نمونه‌های آزمون جدید استفاده می‌شود [۲۵].

۳-۷- درصد میانگین قدرمطلق خطا

با توجه به متغیر وابسته پیوسته تداوم سودآوری نمی‌توان ارزیابی‌کننده AUC را استفاده نمود، زیرا هر دو مقدار واقعی و پیش‌بینی شده، بیش از دو مقدار دارند (یعنی صفر و یک). تداوم سودآوری بیانگر تغییر در میزان سودآوری مشتری در دوره تحقیق است و در نتیجه می‌تواند دامنه وسیعی از هر دو مقدار مثبت و منفی را داشته باشد. به منظور ارزیابی مقادیر پیش‌بینی شده، درصد میانگین قدرمطلق خطا (MAPE) محاسبه می‌شود:

$$MAPE = \frac{\sum_{t=1}^N \left| \frac{E_t}{Y_t} \right|}{N}$$

۴- پیشینه تحقیق

سودرلاند و ویلگان در تحقیقی با استفاده از مدل ماندگاری نسبت به ارائه یک الگوی ریاضی برای تخمین میزان رضایت‌مندی مشتری اقدام نموده و در ادامه، با تجمیع آن‌ها به محاسبه سهم بازار پرداختند [۲۷].

همچنین ایست در تحقیقی به بررسی تأثیر ماندگاری در پیش‌بینی عناصری مانند جستجو و معرفی مشتری جدید پرداخته‌اند. نتایج تحقیق نشان می‌دهند که ترکیب

1. Bagging
2. Bootstrap

معیارهای ماندگاری به‌عنوان پیش‌بینی‌کننده‌های میزان ماندگاری، در مقایسه با استفاده از یک معیار به‌تنهایی، عملکرد ضعیف‌تری دارد [۲۸].

همچنین در پژوهشی که در این حوزه صورت گرفته است [۱]، پراکسیام به هزینه‌های بالای جذب مشتریان جدید نسبت به هزینه‌های نگهداری مشتریان قدیمی برای ارائه‌کنندگان خدمات ارتباطی^۱ و حیاتی بودن شناسایی مشتریان سودآوری که احتمال قطع دریافت خدمات از جانب آن‌ها در آینده نزدیک وجود دارد، اشاره می‌کند. در مقاله ارائه‌شده، جزئیات این الزامات با تمرکز خاص بر گرایش‌های بازار، مدل پیش‌بینی رویگردانی مشتریان و کاربردهای این مدل روی داده‌های واقعی بیان شده است.

همچنین پژوهش‌های متعددی در حوزه مدیریت ارتباط با مشتریان و فرایند ارزیابی آن صورت گرفته است؛ از جمله لیندگرن و همکاران در تحقیقی ابزاری شامل ده عنصر ارزیابی‌کننده برای ارزیابی مدیریت ارتباط با مشتری پیشنهاد نموده‌اند که به سه زیرمجموعه کلی عناصر راهبردی، عناصر زیرساختی و عناصر فرایندی تقسیم‌بندی می‌شوند. در این مطالعه که به عوامل زیرساختی نیز توجه شده، گامی فراتر از ایجاد یک چارچوب مفهومی با ابزارهای قابل‌سنجش برداشته می‌شود [۲۹].

در تحقیقی دیگر تحت عنوان «سنجش مدیریت ارتباط با مشتری»، بریتون و شیمین ضمن بررسی اهمیت و ضرورت سنجش، مؤلفه‌ها و شاخص‌های اثربخشی آن را معرفی کرده و راهبردهایی را برای بهبود اثربخشی روابط با مشتریان ارائه داده‌اند [۳۰].

در حوزه کاربردهای مدل جنگل تصادفی نیز تحقیقاتی انجام پذیرفته است؛ به‌طور مثال، ونس و کوستا در تحقیقی از مدل جنگل تصادفی برای طبقه‌بندی سرطان بر اساس نوع ژن استفاده نموده‌اند [۳۱].

۵- مدل تحقیق و روش برآورد

در این پژوهش، در ابتدا با استفاده از معیار فضای زیرمنحنی ROC و برای بررسی صحت نتایج حاصل از مدل جنگل تصادفی به مقایسه اندازه فضای زیر منحنی مدل تحقیق با روش‌های رگرسیون لجستیک و بگینگ اقدام شد. همچنین برای سنجش میزان دقت مدل جنگل رگرسیون با استفاده از معیار درصد میانگین قدرمطلق خطا

1. Communication Service Providers

(MAPE)، نتایج با روش رگرسیون خطی مقایسه شد. از طرفی، پس از گردآوری داده‌های تحقیق و دسته‌بندی گروه‌ها به زیرگروه‌ها از حیث نوع فعالیت مشتریان و مبلغ، تعداد و نوع خدمات دریافتی با استفاده از متغیرهای پیش‌بینی‌کننده و برحسب گسستگی و پیوستگی متغیر پاسخ، به ترتیب نسبت به انتخاب مدل‌های جنگل تصادفی و جنگل رگرسیون برای طبقه‌بندی مشتریان بر اساس رأی حداکثری (بیشترین تکرار در طبقه‌بندی) در مدل جنگل تصادفی و میانگین نتایج در مدل جنگل رگرسیون اقدام شد و رفتار مشتریان را از منظر ماندگاری و سودآوری بررسی کردیم و تفاوت اثرات مجموعه مشابهی از متغیرهای مستقل بر روی هر دو متغیر مورد آزمون قرار گرفت. برای متغیرهای خرید بعدی، کاهش سطح فعالیت و کاهش سودآوری که متغیرهایی صفر و یک هستند، از مدل جنگل تصادفی برای پیش‌بینی رفتار و از روش‌های رگرسیون لجستیک و بگینگ برای مقایسه نتایج استفاده می‌شود. همچنین معیار فضای زیر منحنی ROC برای ارزیابی مدل به کار گرفته می‌شود. برای پیش‌بینی تداوم سودآوری نیز از مدل جنگل رگرسیون و برای مقایسه نتایج، روش رگرسیون خطی استفاده خواهد شد. همچنین به واسطه خطی بودن متغیر، از معیار درصد میانگین قدرمطلق خطا برای ارزیابی نتایج استفاده می‌شود.

متغیرهای وابسته: در این تحقیق ماندگاری و سودآوری مشتریان به‌عنوان متغیرهای مستقل به‌طور مجزا مورد ارزیابی قرار گرفته و نتایج هر یک با توجه به تفکیک ابعاد هر یک از متغیرها مورد تحلیل قرار می‌گیرد. در این تحقیق، در واقع، چهار متغیر وابسته به‌طور مستقل مورد بررسی قرار گرفته و در عمل، چهار مرتبه از نتایج روش‌های جنگل تصادفی و جنگل رگرسیون (متناسب با نوع متغیر وابسته) به شرح زیر استفاده می‌شود:

۱- ماندگاری شامل دو متغیر خرید بعدی و کاهش سطح فعالیت است که هر دو، متغیر صفر و یک هستند.

۱-۱- خرید بعدی: بیانگر این مطلب است که آیا یک مشتری محصول دیگری به‌عنوان یک زیرمجموعه مستقل از متغیرهای مستقل خریداری می‌کند. در این تحقیق در صورت استفاده مجدد مشتری از یک خدمت خاص، خرید بعدی در آن خدمت منظور می‌شود.

۲-۱- کاهش سطح فعالیت: بیانگر تصمیم مشتری برای توقف استفاده از یک گروه از خدمات است. عبارت کاهش به این مطلب اشاره دارد که خاتمه یک محصول خاص، لزوماً به معنی خروج کامل یک مشتری نیست و مشتری مجاز به استفاده از سایر محصولاتی است که هنوز ارائه می‌شوند و منقضی نشده‌اند.

۲- سودآوری نیز شامل دو متغیر تداوم سودآوری و کاهش سودآوری است که تداوم سودآوری، متغیر وابسته پیوسته و کاهش سودآوری، متغیر صفر و یک گسسته است.

۲-۱- تداوم سودآوری: بیانگر تداوم سودآوری مشتری در طول دوره تحقیق می‌باشد.

۲-۲- کاهش سودآوری: معیاری از سودآوری مشتری در دوره قبل است. متغیر کاهش سودآوری یک متغیر صفر و یک است که نشان‌دهنده سودآوری مشتری در سازمان در پایان دوره تحقیق است.

متغیرهای مستقل:

در این تحقیق، از بین انواع خدمات ارزی قابل ارائه به مشتریان، چهار خدمت عمده که دارای بیشترین تأثیرگذاری از منظر ماندگاری و سودآوری مشتریان برای بانک هستند، انتخاب شده‌اند که به همراه برخی از متغیرها و عوامل مرتبط بدین شرح معرفی می‌شوند: مبلغ، تعداد و سررسید تسهیلات ارزی اعطایی به مشتری؛ مبلغ، تعداد و سررسید اعتبارات اسنادی گشایش شده برای مشتری؛ مبلغ و تعداد حواله‌های ارزی انجام شده برای مشتری؛ تعداد و مانده انواع حساب‌های ارزی مشتری.

همچنین تعداد و ارزش مالی کل خدمات ارائه شده به مشتری و متغیر خرید چندگانه که بیانگر درجه استفاده مشتری از هر یک از انواع خدمات است نیز به کار گرفته شده است.

۶- داده‌ها و نتایج تجربی

به منظور تعیین درجه اهمیت متغیرهای مستقل بر روی متغیرهای وابسته خرید بعدی و کاهش سطح فعالیت که از جمله متغیرهای صفر و یک هستند و هر دو نیز از متغیرهای ماندگاری تحقیق حاضر به‌شمار می‌روند، از مدل جنگل تصادفی استفاده

شده است. از این رو برای به دست آوردن درجه اهمیت هر یک از متغیرهای مستقل، داده‌های متناسب در نرم‌افزار WEKA اجرا شدند. از آنجائی که متغیر کاهش سودآوری یک متغیر صفر و یک است، می‌بایست از مدل جنگل تصادفی نسبت به تعیین درجه اهمیت متغیرهای مستقل اقدام نمود.

فضای زیر منحنی ROC که در این پژوهش به‌عنوان معیار سنجش دقت مدل‌های جنگل تصادفی، رگرسیون لجستیک و روش بگینگ در خصوص متغیرهای صفر و یک انتخاب شده است، برای هر یک از متغیرهای وابسته به شرح جدول ۱ اندازه‌گیری شده است. همان‌گونه که مشاهده می‌شود، در نمونه یادگیری، مدل جنگل تصادفی در مقایسه با مدل‌های رگرسیون لجستیک و بگینگ از سطح زیر منحنی بیشتری برخوردار است که این امر بیانگر بالاتر بودن میزان دقت این روش نسبت به روش رگرسیون لجستیک و بگینگ است. لیکن با کاهش تعداد نمونه‌ها در نمونه آزمون، از میزان دقت مدل جنگل تصادفی نسبت به روش رگرسیون لجستیک کاسته شده و مشاهده می‌شود که نتایج حاصل از روش رگرسیون لجستیک از دقت بالاتری برخوردار هستند؛ البته کماکان در نمونه آزمون، روش جنگل تصادفی از دقت بالاتری نسبت به روش بگینگ برخوردار است.

جدول ۱ فضای زیرمنحنی ROC در مدل‌های جنگل تصادفی، رگرسیون لجستیک و بگینگ

فضای زیر منحنی ROC (AUC)		روش ارزیابی
آزمون	یادگیری	
۰/۸۹۰	۰/۹۸۵	جنگل تصادفی رگرسیون لجستیک بگینگ
۰/۹۰۵	۰/۹۲۰	
۰/۸۶۵	۰/۸۹۷	

مقادیر درصد میانگین قدرمطلق خطا که به‌عنوان معیار اندازه‌گیری صحت ارزیابی مدل جنگل رگرسیون برای متغیر تداوم سودآوری است، در جدول ۲ ارائه شده است. با

توجه به نتایج حاصل از معیار درصد میانگین قدر مطلق خطا، مشاهده می‌شود که مدل جنگل رگرسیون نسبت به روش رگرسیون خطی از میزان دقت بالاتری برخوردار است.

جدول ۲ درصد میانگین قدر مطلق خطا

درصد میانگین قدر مطلق خطا (MAPE)		روش ارزیابی
آزمون	یادگیری	
۱/۰۲	۱/۰۱	جنگل رگرسیون
۲/۴۹	۱/۷	رگرسیون خطی

پس از مقایسه میزان دقت نتایج حاصل از روش‌های جنگل‌های تصادفی و رگرسیون با مدل‌های رگرسیون لجستیک و بگینگ، در نهایت، روش‌های جنگل تصادفی و جنگل رگرسیون به‌عنوان مدل‌های مورد استفاده در این پژوهش انتخاب شد. نتایج به‌دست‌آمده برحسب نوع متغیرهای وابسته از منظر ماندگاری و سودآوری، به ترتیب در جدول‌های شماره ۳ و ۴ ارائه شده است.

جدول ۳ درجه اهمیت متغیرهای مستقل بر روی متغیرهای وابسته

متغیر وابسته کاهش سطح فعالیت			متغیر وابسته خرید بعدی		
درجه اهمیت	نام متغیر مستقل	ردیف	درجه اهمیت	نام متغیر مستقل	ردیف
۰/۸۹	mean.lc.	۱	۳/۱	n.faci.	۱
۰/۷۳	mean.po.	۲	۲/۱۲	n.lc.	۲
۰/۴۱	mean.d.acc.	۳	۱/۵۱	n.acc.	۳
۰/۳۱	mean.c.acc.	۴	۱/۴۸	n.po.	۴
۰/۳۰	n.lc.	۵	۰/۸۴	mean.lc.	۵
۰/۲۸	n.po.	۶	۰/۵۵	mean.faci.	۶
۰/۲۸	mean.day.lc.	۷	۰/۴۵	mean.po.	۷
۰/۲۵	mean.day.d.acc.	۸	۰/۳۸	mean.c.acc.	۸
۰/۰۸	mean.faci.	۹	۰/۲۶	mean.day.lc.	۹
۰/۰۳	n.faci.	۱۰	۰/۲۳	mean.day.d.acc.	۱۰
۰/۰۳	n.acc.	۱۱	۰/۱	mean.d.acc.	۱۱

همان‌گونه که در جداول ۳ و ۴ مشاهده می‌شود، ضریب اهمیت متغیرهای مستقل برای هر یک از متغیرهای وابسته متفاوت است؛ به طوری که در جدول ۳ که اطلاعات مربوط به متغیرهای ماندگاری را نشان می‌دهد، ضریب اهمیت متغیر مستقل تعداد تسهیلات ارزی داده شده (n.faci.) که در متغیر وابسته خرید بعدی دارای بالاترین درجه اهمیت است، در متغیر وابسته کاهش سطح فعالیت، پس از متغیر تعداد حساب‌های ارزی، دارای کمترین درجه اهمیت است. همچنین در جدول ۴ که خروجی‌های متغیرهای مربوط به سودآوری را نمایان می‌سازد، درجه اهمیت متغیر مستقل میانگین معادل دلاری حواله‌های انجام شده (mean.po.) که برای متغیر وابسته کاهش سودآوری دارای بیشترین مقدار است، در جدول ضرایب اهمیت متغیر تداوم سودآوری در جایگاه پنجم قرار دارد.

جدول ۴ درجه اهمیت متغیرهای مستقل بر روی متغیرهای وابسته تداوم سودآوری و کاهش

متغیر وابسته تداوم سودآوری			متغیر وابسته کاهش سودآوری		
درجه اهمیت	نام متغیر مستقل	ردیف	درجه اهمیت	نام متغیر مستقل	ردیف
۲۹	mean.faci.	۱	۰/۷۶	mean.po.	۱
۲۰/۶۷	mean.lc.	۲	۰/۴۸	mean.d.acc.	۲
۱۸/۵۴	mean.d.acc.	۳	۰/۴۷	mean.faci.	۳
۱۴/۷	mean.day.d.acc.	۴	۰/۴۴	mean.day.lc.	۴
۸/۶۴	mean.po.	۵	۰/۳۶	mean.c.acc.	۵
۲/۲۲	n.po.	۶	۰/۳۰	mean.lc.	۶
۱/۹۴	mean.c.acc.	۷	۰/۲۲	n.lc.	۷
۰/۶۸	n.lc.	۸	۰/۱۵	n.acc.	۸
۰/۶۲	mean.day.lc.	۹	۰/۱۴	mean.day.d.acc.	۹
۰/۲۴	n.faci.	۱۰	۰/۱۳	n.po.	۱۰
۰/۱۷	n.acc.	۱۱	۰/۰۸	n.faci.	۱۱

جداول ۵ و ۶ اطلاعات مربوط به متغیرهای وابسته ماندگاری و سودآوری برای مشتریان حقوقی با نوع فعالیت بازرگانی را نشان می‌دهد. همان‌طور که مشاهده می‌شود ضریب اهمیت متغیرهای مستقل برای هر یک از متغیرهای وابسته متفاوت

است؛ به طوری که در جدول ۵ که اطلاعات مربوط به متغیرهای ماندگاری را نشان می‌دهد، همانند حالتی که برای مشتریان حقوقی با نوع فعالیت تولیدی صادق است، ضریب اهمیت متغیر مستقل تعداد تسهیلات ارزی ارائه شده (n.faci.) که در متغیر وابسته خرید بعدی دارای بالاترین درجه اهمیت است، در متغیر وابسته کاهش سطح فعالیت، دارای کمترین درجه اهمیت است ولی با این تفاوت که در اینجا میانگین حواله‌های انجام شده یا میانگین مانده حساب‌های ارزی برای نگهداری مشتریان بازرگانی (در متغیرهای وابسته خرید بعدی و کاهش سطح فعالیت) نسبت به مشتریان تولیدی از اهمیت بیشتری برخوردار است. این امر می‌تواند ناشی از نوع فعالیت این قبیل از مشتریان باشد که کمتر به دنبال گشایش اعتبارات اسنادی بوده و بیشتر خواستار ارسال حواله‌های ارزی به مقاصد مختلف هستند.

جدول ۵ درجه اهمیت متغیرهای مستقل بر روی متغیرهای وابسته خرید بعدی و کاهش سطح فعالیت (فعالیت بازرگانی)

متغیر وابسته کاهش سطح فعالیت			متغیر وابسته خرید بعدی		
درجه اهمیت	نام متغیر مستقل	ردیف	درجه اهمیت	نام متغیر مستقل	ردیف
۰/۳۶	n.po.	۱	۱/۲	n. faci.	۱
۰/۰۹	mean.day.lc	۲	۰/۲۱	mean.po.	۲
۰/۰۹	mean.d.acc.	۳	۰/۱۱	mean.c.acc.	۳
۰/۰۸	mean.c.acc.	۴	۰/۰۷	mean.lc.	۴
۰/۰۵	n.lc.	۵	۰/۰۴	mean.faci.	۵
۰/۰۵	mean.faci.	۶	۰/۰۳	mean.day.d.acc.	۶
۰/۰۴	mean.day.d.acc.	۷	۰/۰۲	mean.day.lc.	۷
۰/۰۳	mean.lc.	۸	۰/۰۲	mean.d.acc.	۸
۰/۰۳	mean.po.	۹	۰/۰۱	n.lc.	۹
۰/۰۲	n.acc.	۱۰	۰	n.acc.	۱۰
۰	n.faci.	۱۱	۰	n.po.	۱۱

جدول ۶ درجه اهمیت متغیرهای مستقل بر متغیر تداوم سودآوری و کاهش سودآوری در مشتریان با نوع فعالیت بازرگانی

متغیر وابسته تداوم سودآوری			متغیر وابسته کاهش سودآوری		
درجه اهمیت	نام متغیر مستقل	ردیف	درجه اهمیت	نام متغیر مستقل	ردیف
۱۲/۳۶	mean.faci.	۱	۰/۲۱	n.po.	۱
۷/۷۴	mean.lc.	۲	۰/۱۱	mean.d.acc.	۲
۷/۳۳	mean.d.acc.	۳	۰/۱	mean.c.acc.	۳
۶/۳۱	n.lc.	۴	۰/۰۶	mean.faci.	۴
۶/۰۰	mean.day.d.acc.	۵	۰/۰۶	n.acc.	۵
۳/۱۸	n.po.	۶	۰/۰۴	n.lc.	۶
۲/۴۵	mean.po.	۷	۰/۰۳	mean.day.d.acc.	۷
۲/۳۶	mean.c.acc.	۸	۰/۰۲	mean.lc.	۸
۲/۱۰	n.acc.	۹	۰/۰۲	mean.po.	۹
۱/۲۰	mean.day.lc.	۱۰	۰/۰۲	mean.day.lc.	۱۰
۰/۲۳	n.faci.	۱۱	۰	n.faci.	۱۱

در تحقیق حاضر به منظور بررسی نحوه و جهت اثرگذاری متغیرهای مستقل و سنجش معنی دار بودن تأثیر این دسته از متغیرها بر روی متغیرهای وابسته صفر و یک خرید بعدی، کاهش سطح فعالیت و همچنین بر کاهش سودآوری در نظر گرفته شده در تحقیق، مقدار p-value برای متغیرهای مستقل هر یک از متغیرهای وابسته محاسبه شده است. در پژوهش حاضر، اثرگذاری متغیرهای مستقل با معیار p-value سنجیده شد و نتایج آن در جدول شماره ۷ نشان داده شده است.

روش کار به این ترتیب است که به عنوان مثال برای محاسبه مقدار p-value در متغیر خرید بعدی، نمونه‌ها به دودسته دارای خرید بعدی و فاقد آن تقسیم شدند. سپس برای هر یک از متغیرهای مستقل در هر دودسته، مقادیر میانگین و واریانس محاسبه شد. همچنین مقدار آماره (Z) برای متغیرهای مستقل اندازه‌گیری شد.

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

همان‌گونه که مشاهده می‌شود در اغلب متغیرهای مستقل عنوان شده در جدول ۷، مقدار p-value کمتر از ۵٪ است که بیانگر تأثیر معنی دار بودن این دسته از متغیرهای مستقل بر روی متغیر وابسته مربوطه است.

جدول ۷ نتایج p-value

ردیف	متغیر وابسته خرید بعدی	خرید بعدی دارد	خرید بعدی ندارد	مقدار آماره	p-value
۱	n.acc.	۳/۰۳	۲/۲۵۲۵	۷/۶۶	۰
۲	mean.c.acc.	۱۹۳۹۰/۱۲	۱۶۷۵۸/۰۰	۰	۰/۵
۳	mean.d.acc.	۲۶۵۱۶۰/۴۸	۱۳۳۲۸۰/۱۹	۰	۰/۵
۴	mean.day.d.acc.	۷۷/۲۰	۴۴/۱۸۱۸	-۰/۰۱	۰/۵
۵	n.po.	۵/۶۸۵۷	۴/۱۷۱۷	-۰/۴۳	-۰/۳۳
۶	mean.po.	۹۵۷۷۱/۳۴	۷۶۲۴۲/۵۸	۰	۰/۵
۷	n.lc.	۲/۸۰	۱/۸۶۸۷	۷/۹۷	۰
۸	mean.lc.	۷۹۹۴۰۴/۶۰	۶۰۹۰۰۸/۸۱	۰	۰/۵
۹	mean.day.lc.	۲۰۵/۰۲	۲۱۶/۰۸	-۰/۰۲	۰
۱۰	n.faci.	۲/۰۲۸۶	۱/۱۰۱۰	۸۹/۱۲	۰/۵
۱۱	mean.faci.	۵۵۲۴۹۵/۱۲	۴۶۳۷۸۹/۴۶	۰	۰
	متغیر وابسته کاهش سطح فعالیت	کاهش سطح فعالیت دارد	کاهش سطح فعالیت ندارد	مقدار آماره	p-value
۱	n.acc.	۲/۴۶۵۱	۲/۴۳۷۵	-۰/۲۱	-۰/۴۱
۲	mean.c.acc.	۱۶۹۹۴/۳۷۴۹	۱۸.۲۵۲/۷۶۰۲	۰	۰/۵
۳	mean.d.acc.	۲۰۴۳۰۸/۴۵۵۲	۱۰۲۱۸۳/۹۱۹۸	-۰/۰۱۳	-۰/۴۹
۴	mean.day.d.acc.	۶۱/۹۵۳۵	۳۶/۴۱۶۷	-۰/۰۸۵	-۰/۵۳
۵	n.po.	۴/۴۸۸۴	۴/۷۰۸۳	۰	۰/۵
۶	mean.po.	۲/۰۴۶۵	۸۱۰۶۲/۴۵۳۹	-۱/۰۶	-۰/۸۵
۷	n.lc.	۶۵۰۹۶۵/۴۴۴۷	۲/۲۲۹۲	۰	۰/۵
۸	mean.lc.	۲۲۰/۴۵۶۴	۶۷۲۶۶۶/۷۷۵۳	-۰/۰۳۳	-۰/۴۸
۹	mean.day.lc.	۱/۳۴۸۸	۲۰۰/۱۷۶۰	-۰/۳۶	-۰/۳۵
۱۰	n.faci.	۴۸۲۷۶۲/۸۳۴۶	۱/۳۳۳۳	۰	۰/۵
۱۱	mean.faci.	۴۹۴۴۷۶/۶۹۷۱	۴۹۴۴۷۶/۶۹۷۱	-۱۰۰۶	۰
	متغیر وابسته کاهش سودآوری	کاهش سودآوری دارد	کاهش سودآوری ندارد	مقدار آماره	p-value
۱	n.acc.	۲/۴۷۰۲	۲/۴۳۸۰	-۰/۲۴	-۰/۴
۲	mean.c.acc.	۱۷۰۲۰/۳۴۲۱	۱۸۵۴۶/۳۲۰۱	۰	۰/۵
۳	mean.d.acc.	۲۰۹۳۵۱/۰۲	۱۰۵۶۰۵/۲۱	۰	-۰/۴۱
۴	mean.day.d.acc.	۶۲/۰۳۵۲	۳۶/۷۱۱۲	-۰/۰۱۲	-۰/۵۳
۵	n.po.	۴/۴۸۰۱	۴/۶۷۸۳	-۰/۰۷	-۰/۸۴
۶	mean.po.	۸۲۳۲۹/۰۱۵۳	۸۱۷۲۵/۳۲۶۰	۰	۰/۵
۷	n.lc.	۲/۰۴۶۰	۲/۲۲۰۰	-۱/۰۱	-۰/۴۸
۸	mean.lc.	۶۵۰۸۵۱/۲۶۱۰	۶۷۲۶۶۶/۷۷۵۳	۰	-۰/۳۵
۹	mean.day.lc.	۲۲۱/۲۵۱۴	۱۹۹/۰۲۵۱	-۰/۰۳۶	-۰/۵
۱۰	n.faci.	۱/۳۴۸۷	۱/۳۳۰۳	-۰/۳۶	۰
۱۱	mean.faci.	۴۸۲۰۲۱/۰۲۱۳	۴۸۷۶۵۱/۶۹۰۱	۰	۰

همچنین در ادامه فراوانی هر یک از متغیرهای وابسته خرید بعدی، کاهش سودآوری و کاهش سطح فعالیت را نسبت به یکدیگر موردسنجش قرار داده‌ایم که نتایج در جداول شماره ۸، ۹ و ۱۰ ارائه شده‌اند. مقادیر درصد درج شده در جداول از تقسیم تعداد موارد بر تعداد مجموع متناظر به دست می‌آید. به‌عنوان نمونه، عدد ۳۴/۹۲٪ در جدول ۸ از تقسیم عدد ۴۴ (تعداد مواردی که خرید بعدی و کاهش سطح فعالیت نداشته‌اند) بر عدد ۱۲۶ (تعداد کل مواردی که خرید بعدی نداشته‌اند) حاصل می‌شود. همچنین رقم ۷۳/۳۳٪ در همان جدول از تقسیم عدد ۴۴ (تعداد مواردی که خرید بعدی و کاهش سطح فعالیت نداشته‌اند) بر عدد ۶۰ (تعداد کل مواردی که کاهش سطح فعالیت نداشته‌اند) به دست می‌آید. همان‌گونه که در جدول شماره ۸ مشاهده می‌شود، ۲۶/۶۷٪ از مشتریان، برخوردار از خرید بعدی هستند و خدمتی را قبل از سررسید لغو نکرده‌اند یا ۳۶/۲۳٪ از مشتریان، علی‌رغم کاهش سطح فعالیت، کاهش سودآوری نداشته‌اند (جدول ۱۰).

جدول ۸ فراوانی خرید بعدی در کاهش سطح فعالیت

		کاهش سطح فعالیت		مجموع
		ندارد	دارد	
خرید بعدی	ندارد	۴۴ مورد ٪۳۴.۹۲ ٪۷۳.۳۳	۸۲ مورد ٪۶۵.۰۸ ٪۷۵.۲۳	۱۲۶ مورد
	دارد	۱۶ مورد ٪۳۷.۲۱ ٪۲۶.۶۷	۲۷ مورد ٪۶۲.۷۹ ٪۲۴.۷۷	۴۳ مورد
	مجموع	۶۰ مورد	۱۰۹ مورد	۱۶۹ مورد

جدول ۹ فراوانی خرید بعدی در کاهش سودآوری

		کاهش سودآوری		مجموع
		ندارد	دارد	
خرید بعدی	ندارد	۳۹ مورد ٪۳۰.۹۵ ٪۶۲.۹۰	۸۷ مورد ٪۶۹.۰۵ ٪۸۱.۳۱	۱۲۶ مورد
	دارد	۲۳ مورد ٪۵۳.۴۹ ٪۳۷.۱۰	۲۰ مورد ٪۴۶.۵۱ ٪۱۸.۶۹	۴۳ مورد
مجموع		۶۲ مورد	۱۰۷ مورد	۱۶۹ مورد

جدول ۱۰- فراوانی کاهش سطح فعالیت در کاهش سودآوری

		کاهش سودآوری		مجموع
		ندارد	دارد	
کاهش سطح فعالیت	ندارد	۴۴ مورد ٪۷۳.۳۳ ٪۶۳.۷۷	۱۶ مورد ٪۲۶.۶۷ ٪۱۶	۶۰ مورد
	دارد	۲۵ مورد ٪۲۲.۹۴ ٪۳۶.۲۳	۸۴ مورد ٪۷۷.۰۶ ٪۸۴	۱۰۹ مورد
مجموع		۶۹ مورد	۱۰۰ مورد	۱۶۹ مورد

۷- نتیجه‌گیری

در صورت ارائه تسهیلات ارزی بیشتر به مشتریان تولیدی و در نتیجه، گشایش اعتبارات اسنادی یا انجام حواله‌های بیشتر برای این دسته از مشتریان، احتمال افزایش دوره ماندگاری مشتریان افزایش می‌یابد. همچنین صرف افزایش تعداد خدمات ارائه شده به مشتریان، بیانگر افزایش سطح فعالیت آن‌ها نیست، بلکه میانگین مبلغ اعتبارات اسنادی و حواله‌های انجام شده است که در بالا بردن دوره ماندگاری مشتریان از اهمیت بیشتری برخوردار است.

۸- منابع

- [1] Meltzer, M. (2007). A CRM approach to customer profitability. *Downloadable* www.amt.eu.com.
- [2] Gupta. S. (2007). A model to determine customer lifetime value in a retail banking context. *European Management Journal* 25: 221-234.
- [3] Dennis, D. (1998). Customer loyalty strategies. *Journal Consumer Marketing* 5: 434-448.
- [4] Hampe J., and Swatman, F. (2002). Customer relationship management. Case studies of five Swedish. *Technical report*. Luella University of Technologies. Sweden.
- [5] Stahl, F., Heitmann, M., Lehmann, D. R., and Neslin, S.A. (2012). The impact of brand equity on customer acquisition, retention, and profit margin. *Journal of Marketing* 11: 44-63.
- [6] Siu N. Y. M., Zhang, T. J. F., and Yau, C. Y. J. (2013). The roles of justice and customer satisfaction in customer retention: a lesson from service recovery. *Journal of Business Ethics* 114: 675-686.
- [7] Park, S. H., and Han, S. P. (2014). From accuracy to diversity in product recommendations: relationship between diversity and customer retention. *International Journal of Electronic Commerce* 18:51-71.
- [8] Ganesh, J., Arnold, M. J., and Reynolds, K. E. (2000). Understanding the customer base of service providers: An examination of the differences between switchers and stairs. *Journal of Marketing* 64: 65-87.
- [9] Jain R., Jain S., Dhar U. (2003). Measuring customer relationship management. *Journal of Service Research* 2: 97 -109.
- [10] Gilbert, C. D., and Karen, C. C. (2003). Relationship marketing practice in relation to different bank ownership: A study of banks in Hong Kong. *Technical report*. Department of Economics. National University of Hong Kong. Hong Kong.
- [11] Duda, R.O. Hart, P.E. and Stork, D.G. (2001). *Pattern classification*. Wiley.

- [12] Breiman, L. and Adele, C. (2013). Breiman and Cutler's random forests for classification and regression. *MASS* 4: 6-17.
- [13] Payet, N., and Todorovic, S.(2008). Random forest random field. *Technical report*. School of Electrical Engineering and Computer Science Oregon State University. USA.
- [14] Clemencon, M. Depecker, and N. Vayatis. (2013). Ranking forests. *Journal of Machine Learning Research* 14: 39–73.
- [15] Denil, M., Matheson, D. and de Freitas, N. (2013). Consistency of online random forests. In Proceedings of the ICML Conference. *arXiv*:1302.4853.
- [16] Scornet. E. (2014). On the asymptotic of random forests. *arXiv*:1409.2090.
- [17] Scornet, E., Biau, G. and Vert, J. P. (2015). Supplementary materials for : Consistency of random forests. *arXiv*:1510.2180.
- [18] Biau, L., Devroye, M. and Lugosi. G. (2015). Consistency of random forests and other averaging classifiers. *Journal of Machine Learning Research* 9: 2025–2033.
- [19] Abrazi, M., Abbasi, A., HaghShenas, A., and Ansari, M. E. (2009). The effect of the proportion of individuals, organizations and organizational culture on organizational behavior faculty members (Case Study Shaid Bahonar University). *Management Researches in Iran* 14, 1-20. (In Persian).
- [20] Sepehri, M., Noorouzi, A., Teymourpour, B., and Choobdar, S. (2012). Discovering the reasons of customer defection of e-banking services. Data mining methods. *Management Researches in Iran* 17: 139-160. (In Persian).
- [21] Rezapour, M., Sepehri, M. and Rezapour, H. (2014). Determining the best methods of assessing learners in e-learning courses. *Management Researches in Iran* 20:145-161. (In Persian).
- [22] Sohrabi, B., Reisi, I., and Zare, F. (2017). Designing systems to optimize the management of bank lending based on clustering and classification methods. *Modern Researches in Decision Making* 2:53-76. (In Persian).

- [23] Karimi, M. and Kazemi, M. A. (2017). Predicting failure and plan maintenance of ATM. Data mining method. *Modern Researches in Decision Making* 3: 113-129. (In Persian).
- [24] Hanley, J. A., McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29–36.
- [25] Efron, B and Tibshirani. H. (1993). *An introduction to the bootstrap*. Chapman and Hall.
- [27] Soderlund, M., and Vilgon, M. (1999). Customer satisfaction and link to customer profitability. *SSE/EFI Working Paper Series in Business Administration* No.1999:1. USA.
- [28] East, R. (2005). Consumer loyalty: singular, additive or interactive? *Australasian Marketing Journal* 13: 22-35.
- [29] Lindgreen A., Palmer R., Vanhamme J., Wouters J. (2006). A relationship management assessment tool: Questioning, identifying, and prioritizing critical aspects of customer relationships. *Industrial Marketing Management* 35: 57-63.
- [30] Brewton J. and Schiemann W. (2003). Measurement: The missing ingredient in today's CRM strategy. *Cost Management* 17: 5 -14.
- [31] Vens, C. and Costa, F. (2011). Random forest based feature induction. *Technical report*. Department of Computer Science. Katholieke Universiteit. Belgium.