



پیش بینی ریزش مشتری در تجارت الکترونیک با استفاده از راهبرد هوش تجاری مبتنی بر یادگیری ماشین و یادگیری عمیق

نام و نام خانوادگی نویسنده اول: امیرحسین تعالی

وابستگی سازمانی نویسنده: دانشجو کارشناسی ارشد مهندسی کامپیوتر

نام و نام خانوادگی نویسنده دوم: دکتر مرتضی محمدی زنجیره

وابستگی سازمانی نویسنده: استادیار دانشگاه بین المللی امام خمینی (ره)

چکیده

امروزه ریزش مشتری در تجارت الکترونیک یکی از مهم ترین دغدغه های شرکت ها می باشد. با پیشرفت علم و فناوری، رقابت بین شرکت های بزرگ دنیا برای افزایش درآمد، افزایش تعداد مشتریان و حفظ مشتریان قبلی افزایش یافته است. از این رو شرکت ها همواره می کوشند تا در این بازار رقابتی شدید رضایت مشتریان قبلی خود را با بالابردن کیفیت محصولات و خدمات خود حفظ کنند و جایگاه خود را در صورت ممکن در رتبه جهانی ارتقا دهند یا حفظ کنند تا از ریزش مشتریان خود جلوگیری کنند. از آنجایی که پیدا کردن مشتریان جدید با توجه به هزینه های بسیار سنگین تبلیغات و همچنین زمان بر بودن بسیار سخت تر از حفظ مشتریان فعلی شرکت است پس بیشتر شرکت ها برای کاهش هزینه های خود دست به اقداماتی بزنند. در این پژوهش با استفاده از دادگان کیفیت مجموعه بزرگ دیجیکالا ابتدا با پیش پردازش داده های موجود آن ها را یکدست و نرمال سازی کرده و سپس با استفاده از الگوریتم های یادگیری ماشین مانند بردارهای ماشین پشتیبان، درخت تصمیم، آدا بوست، ایکس جی بوست و الگوریتم های یادگیری عمیق مانند؛ الگوریتم های برت، شبکه عصبی بازگشتی، شبکه عصبی واحد بازگشتی، حافظه کوتاه مدت طولانی، کانولوشن یک بعدی مدل سازی انجام داده و ریزش مشتری را پیش بینی می کنیم. در نهایت بعد از شبیه سازی انجام شده بهترین الگوریتم برای پیش بینی ریزش مشتری الگوریتم برت بهینه شده با دقت ۸۴.۳۲٪ می باشد.

واژگان کلیدی: پیش بینی، ریزش مشتری، هوش تجاری، یادگیری ماشین، تجارت الکترونیک

مقدمه

امروزه با پیشرفت فناوری، فناوری‌های هوشمند به مشتری‌ها وعده می‌دهند که تجربه‌ی آن‌ها را به سطح جدیدی در نسل بعدی ارتقا می‌دهند. شرکت‌ها به‌طور فزاینده‌ای از راهبرد مبتنی بر فناوری استفاده می‌کنند تا تجربه‌ی مشتریان را به‌صورت دیجیتالی افزایش دهند. شرکت‌ها می‌توانند برای ارسال پیام‌های شخصی‌سازی‌شده به مشتریان از دو نوع دستگاه هوشمند استفاده کنند: ۱- گوشی هوشمند خود مشتریان و ۲- صفحه‌نمایش‌های متعلق به فروشگاه‌های شرکت. بر اساس تئوری مجموعه، نقش این دستگاه‌ها در تأثیرگذاری بر تجربه مشتری و در نهایت بر رفتار خرید مصرف‌کننده در نظر گرفته می‌شود. از طریق دو آزمایش و تجزیه و تحلیل تعدیل واسطه، تأثیر متقابل محتوای شخصی و فناوری دستگاه در پاسخ مشتریان به TEP را می‌توان بررسی کرد. نتایج نشان می‌دهد که مصرف‌کنندگان بسته به دستگاهی که از طریق آن پیام منتقل می‌شود، به محتوای پیام واکنش متفاوتی نشان می‌دهند؛ یعنی پیام‌های شخصی‌سازی‌شده (استاندارد شده) بر روی گوشی‌های هوشمند متعلق به مشتری به نسبت صفحه‌نمایش‌های متعلق به فروشگاه مؤثرتر هستند. تجارب مشتری رابطه‌ای، یا میزانی که مشتری احساس ارتباط مثبت با مجموعه‌های فروشگاه‌های می‌کند، بر رفتار خرید تأثیر می‌گذارد؛ بنابراین، برای ایجاد راهبردهای TEP، شرکت‌ها باید از دستگاه‌های هوشمند ادغام‌شده در خود گسترده مشتریان استفاده کنند. (Riger et al., 2022) پیش‌بینی ریزش مشتری^۱ یکی از مشکلات چالش‌برانگیز در صنعت مخابرات است. با پیشرفت در زمینه یادگیری ماشینی و هوش مصنوعی، امکان پیش‌بینی ریزش مشتری به میزان قابل‌توجهی افزایش یافته است. همچنین جهانی‌شدن صنعت مخابرات باعث افزایش تصاعدی رقابت برای فروش محصولات و خدمات بین شرکت‌ها شده است. در عصر رقابت افزایش سود حاصل از فروش برای شرکت‌ها از اهمیت ویژه‌ای برخوردار است که برای آن راهبردهایی معرفی شده‌اند از جمله؛ پیدا کردن مشتریان جدید برای شرکت، افزایش فروش محصولات و خدمات به مشتریان قبلی و حفظ مشتریان قبلی شرکت که راهبرد آخر از نظر هزینه به‌صرفه‌تر از بقیه‌ی آن‌ها می‌باشد. برای اتخاذ این راهبرد شرکت‌ها باید ظرفیت ریزش مشتری برای مراجعه به یک فراهم‌کننده‌ی خدمات دیگر را کاهش دهند. دلیل اصلی ریزش، نارضایتی از خدمات مصرف‌کننده و سیستم پشتیبانی است. برای مؤثر واقع شدن این راهبرد، شرکت باید رفتار مشتری را به‌طور صحیح موردبررسی قرار دهد. مدیریت ریزش مشتری به دو گونه انجام می‌شود: ۱- واکنش پذیر ۲- فعال. در راه اول شرکت صبر می‌کند تا لغو درخواست از سمت مشتری ارسال شود و بعد از آن پیشنهادات ویژه‌ای از سمت شرکت برای حفظ آن مشتری ارسال می‌شود. در راه دوم، ریزش مشتری توسط شرکت پیش‌بینی می‌شود و با توجه به آن راهکارها به مشتریان پیشنهاد می‌شوند. (Lalvati et al., 2021) امروزه با پیشرفت علم یادگیری ماشینی که با استفاده از داده‌ی قبلی و الگوریتم‌های مختلف به پیش‌بینی می‌پردازد، می‌توان به پیش‌بینی ریزش مشتریان پرداخت. در یادگیری ماشین علاوه بر پیش‌پردازش داده‌ها، انتخاب ویژگی نیز نقش بسیار حیاتی دارد به این منظور که با انتخاب ویژگی‌های بهتر، دقت بالاتر و در نتیجه پیش‌بینی دقیق‌تری خواهیم داشت. تاکنون در ایران چند تحقیق بر روی پیش‌بینی ریزش مشتری با کمک شبکه عصبی صورت گرفته اما تمرکز اصلی اکثریت این تحقیقات بالأخص در چند سال اخیر، انجام بررسی و تحلیل بر روی انواع روش‌های موجود پیش‌بینی ریزش مشتری است. نوآوری این پژوهش این است که به پیش‌بینی ریزش مشتریان با کمک یکی از بهترین و جدیدترین الگوریتم‌ها که متناسب با هدف ما است می‌پردازد، در این مقاله سعی می‌شود تا با پیش‌بینی به کمک برت، در ادامه با اعلام گزارش‌هایی از این ریزش در یک محدوده خاص (مثلاً ریزش ۵٪ مشتریان در طول یک ماه) و ارائه راه‌حل‌هایی جدید و متناسب با علایق و نیازهای مشتریان شرکت توسط مدیریت، به حفظ مشتریان قبلی کمک شود؛ که همین امر کمک شایانی به افزایش فروش محصولات و خدمات به مشتریان قبلی و پیدا کردن مشتریان جدید خواهد کرد.

¹ Customer churn prediction

مرور ادبیات

Shobana et al. ۲۰۲۳ در پیشگیری از ریزش مشتری در تجارت الکترونیک با استفاده از راهبرد هوش تجاری مبتنی بر یادگیری ماشین از الگوریتم ماشین بردار پشتیبان و طرح راهبرد ترکیبی نگهداری هدفمند برای پیش‌بینی ریزش مشتری استفاده کرده‌اند. همچنین برای شناسایی موثر گروه‌های مجزا از مشتریان از دست رفته و ایجاد راهبرد حفظ مشتری، انواع مختلف مشتریان از دست رفته را با استفاده از اصل (تعداد خریدهای اخیر مشتری و ارزش خرید، بیانگر رضایت مشتریان و ارزش طول عمر آنهاست و تکرار خرید نشانه‌ی حفظ و وفاداری مشتریان است).^۱ Peraba et al. ۲۰۲۳ در تحلیل ریزش مشتری با استفاده از الگوریتم‌های یادگیری ماشین با هدف پیش‌بینی پاسخ مشتریان موجود برای حفظ آنها، داده‌های گنجانده شده که شامل تمام داده‌های مشتریان که به حدود ۹ ماه قبل ریزش آنها می‌رسد با الگوریتم‌های یادگیری ماشین مانند تقویت‌کننده گرادیان تصادفی، جنگل تصادفی، رگرسیون لجستیک و روش‌های k-نزدیک‌ترین همسایگان را آزمایش کرده است. دقت الگوریتم‌های فوق به ترتیب ۸۳.۹٪، ۸۲.۶٪، ۸۲.۹٪ و ۷۸.۱٪ می‌باشد. در نتیجه بهترین الگوریتم از نظر عملکرد تقویت‌کننده گرادیان تصادفی می‌باشد. (Shobana et al. ۲۰۲۳, Riger et al. ۲۰۲۲ در شخصی‌سازی مبتنی بر فناوری: تأثیر انتخاب فناوری هوشمند بر رفتار خرید مصرف‌کننده از طریق دو آزمایش و تجزیه و تحلیل تعدیل واسطه، تأثیر متقابل محتوای شخصی‌سازی شده و فناوری دستگاه را در پاسخ مشتریان به TEP بررسی کردند. نتایج نشان داد که مصرف‌کنندگان بسته به دستگاهی که از طریق آن پیام منتقل می‌شود، واکنش متفاوتی به محتوای پیام نشان می‌دهند؛ یعنی پیام‌های شخصی‌سازی شده (استاندارد شده) روی تلفن‌های هوشمند متعلق به مشتری (صفحه‌نمایش‌های متعلق به خرده‌فروش) مؤثرتر هستند، زیرا آنها در خود گسترده مشتری ادغام می‌شوند (از آن جدا می‌مانند). Kim & Li، ۲۰۲۲ در پیش‌بینی ریزش مشتری در تجارت تأثیرگذار یک کاربرد درختان تصمیم می‌گویند که اینفلوئنسر ها از طریق اس‌ان‌اس^۲ یک محصول یا برند را در شبکه‌های اجتماعی‌شان مانند اینستاگرام، توییتر و... تبلیغ می‌کنند و به فروش می‌رسانند. آنها از داده‌های جمع‌آوری شده توسط آژانس اینفلوئنسر مارکتینگ در کره از آگوست ۲۰۱۸ تا اکتبر ۲۰۲۰ شامل جزئیات خرید مانند اطلاعات مشتری، کالای خرید و مبلغ پرداخت است استفاده کردند. سپس با استفاده از الگوریتم درخت تصمیم به حداکثر دقت پیش‌بینی ۹۰ درصد دست یافتند. بلال و همکاران ۲۰۲۲ در یک رویکرد مبتنی بر مجموعه با استفاده از ترکیبی از الگوریتم‌های خوشه‌بندی و طبقه‌بندی برای افزایش پیش‌بینی ریزش مشتری در صنعت مخابرات یک مدل ترکیبی است که مبتنی بر ترکیبی از الگوریتم‌های خوشه‌بندی و طبقه‌بندی با استفاده از یک مجموعه را پیشنهاد کرده‌اند. ابتدا، الگوریتم‌های مختلف خوشه‌بندی (یعنی کامینز^۳، کامدویدز^۴، ایکس مینز^۵ و خوشه‌بندی تصادفی) به صورت جداگانه بر روی دو مجموعه داده پیش‌بینی ریزش را ارزیابی کردند. لالوانی و همکاران ۲۰۲۱ در سیستم پیش‌بینی ریزش مشتری: رویکرد یادگیری ماشینی در دو فاز اول، پیش‌پردازش داده‌ها و تحلیل ویژگی‌ها انجام شده است. در مرحله سوم، انتخاب ویژگی با استفاده از الگوریتم جستجوی گرانشی مورد توجه قرار می‌گیرد. در مرحله بعد، داده‌ها به دو بخش قطار و مجموعه آزمایشی به ترتیب با نسبت ۸۰ و ۲۰ درصد تقسیم شدند. در فرآیند پیش‌بینی، آنها از مدل‌های رایج پیش‌بینی، یعنی رگرسیون لجستیک، نایو بیز، ماشین بردار پشتیبان، جنگل تصادفی، درخت‌های تصمیم‌گیری و غیره در مجموعه قطار و همچنین روش‌های تقویت و مجموعه برای مشاهده اثر روی دقت مدل‌ها علاوه بر این، اعتبار سنجی متقاطع کافولدز^۶ روی مجموعه قطار برای تنظیم هایپرپارامتر و جلوگیری از برازش بیش از حد مدل‌ها استفاده کردند

¹ RFM-Recency, Monetary value, Frequency

² SNS

³ K-means

⁴ K-medoids

⁵ X-means

⁶ K-folds

روش پیشنهادی

با توجه به اهمیت حفظ مشتری در دنیای رقابتی تجارت الکترونیک، پیش‌بینی ریزش مشتری به یکی از دغدغه‌های اصلی کسب‌وکارهای آنلاین تبدیل شده است. با استفاده از روش‌های هوش تجاری و الگوریتم‌های یادگیری ماشین، می‌توان به مدل‌های پیش‌بینی دقیقی دست یافت که به کسب‌وکارها کمک می‌کند تا مشتریان در معرض خطر ترک را شناسایی کرده و اقدامات لازم برای حفظ آن‌ها را انجام دهند.

دادگان پژوهش

دادگان مورد استفاده در پژوهش ما دادگان منتشر شده ی فروشگاه اینترنتی دیجیکالا در سال ۱۳۹۸ می‌باشد. این مجموعه شامل ۵ مجموعه داده که مربوط به نظرات مردم درباره ی محصولات پس از خرید، نظرات مردم درباره محصول خریداری شده، سفارشات داده شده و شهر استقرار مشتریان، محصولات فروشگاه و ویژگی های آن‌ها و درنهایت تاریخچه ی خرید مشتریان است، می‌باشد. مجموعه داده ی مورد استفاده ی ما برای تجزیه و تحلیل، مدل‌سازی و پیش‌بینی ریزش مجموعه ی کیفیت می‌باشد. این مجموعه شامل ۱۰۰۰۰۰ سطر یا نمونه و ۱۲ ستون یا ویژگی می‌باشد که ستون اول مربوط به شناسه ی محصول، ستون دوم مربوط به نام محصول، سوم دسته بندی محصول، چهارم شناسه مشتری، پنجم پسندیده شدن نظرات مشتریان، ششم پسندیده نشدن نظرات، هفتمین ستون مربوط به مشتریان احراز هویت شده یا نشده، هشتم خلاصه نظرات مشتریان، نهم توصیه ی مشتریان به بقیه برای خرید محصول مورد نظر، دهم نظرات کامل مشتریان در خصوص محصول خریداری شده، یازدهم و دوازدهم در مورد ویژگی های مثبت و منفی محصول خریداری شده می‌باشد.

1	product_id	product_title_en	user_id	likes	dislikes	verification	recommender	title	comment	advantage	disadvantages
2	3692	IT ماوس بی‌سیم	989472	0	0	verified	\N	من که ازش	واقعاً عالیه. من که ازش		
3	90213	AC شارژر همراه	3862150	4	1	verified	recommender	واقعاً عالیه	من ۲ سال		

شکل ۱: تصویری از نمونه مجموعه داده ی مورد استفاده

محیط تجزیه و تحلیل و مدل‌سازی

تمامی مراحل پژوهش از جمله پیش پردازش دادگان، تجزیه و تحلیل و مدل سازی آن ها در زبان برنامه نویسی پایتون پیاده سازی شده است.

پیش پردازش دادگان

برای آماده سازی دادگان جهت مدل سازی و ارزیابی روی آن ها کار های زیر انجام شده است: در ابتدای کار مجموعه دادگان کیفیت را جهت کار روی آن بارگذاری کردیم، سپس نوبت به پیش پردازش و آماده سازی و انتخاب ویژگی می رسد. پس از مرحله ی انتخاب ویژگی برچسب گذاری داده ها انجام شد از آنجایی که در مجموعه دادگان ما ستون ریزش مشتری به صورت مستقیم قرار نداشت ما ستون های پیشنهاد، پسندیده شدن نظرات و پسندیده نشدن نظرات را به عنوان ستون هدف در نظر گرفتیم به این صورت که براساس پیشنهاد خرید یک محصول توسط مشتری آن به دیگران و مقایسه ی تعداد پسندیده شدن آن نظر نسبت به پسندیده نشدن آن ریزش رو پیش بینی کردیم. پس از اعمال برچسب گذاری ما ستون های عنوان نظر، نظرات، ویژگی های خوب و ویژگی های بد محصول را جهت کار روی آن ها انتخاب کردیم. از آنجایی که دادگان ما به صورت متنی بودند تمامی این ستون ها به صورت رشته درآورده شد تا کار بر روی نمونه ها راحت تر و دقیق تر انجام شود، همچنین از آنجایی که بسیاری از کلمات با توجه به مثبت یا منفی بودنشان و غیره تاثیر بسزایی در مدل سازی و ارزیابی دارند پس ستون های عنوان نظر، نظرات، ویژگی های خوب و ویژگی های بد محصول را با یکدیگر ادغام کردیم تا دقت کار بهتر شود. برای پیش پردازش روی متنی که زبان آن فارسی باشد دو کتابخانه در پایتون معرفی شده اند: هضم و پارسیور. کتابخانه ی هضم بخاطر پردازش داده ی دقیق تر مورد استفاده قرار گرفته است. با استفاده از این کتابخانه ستون ادغامی را نرمال، ریشه یابی، بن یابی کردیم. از آنجایی که تمامی کلمات داخل متن برای مدل سازی ما مناسب نیستند که بیشتر ضمائر، حروف ربط و حروف اضافه هستند و ممکن است دقت آن را پایین بیاورند از داخل جملات حذف شدند. در وهله ی بعد داده های ناهنجار و خالی حذف و با عبارات مطمئن تر جایگزین شدند. در نهایت جهت آماده سازی داده جملات را توکنایز کردیم به این معنی که به هر حرف از کلمه یک عدد متمایز اختصاص داده شد و پس از انجام این کار داده ها را به داده های آموزشی و تست تقسیم شد. در ادامه ی کار پس از آماده سازی دادگان، مدل سازی روی آن ها را با الگوریتم هایی که به شرح زیر می باشد انجام شد:

برت بهینه شده: کلمه BERT مخفف عبارت Bidirectional Encoder Representations from Transformers است. یک الگوریتم هوشمند مبتنی بر شبکه است که برای پردازش و کنکاش زبان های طبیعی پیش آموزشی بکار گرفته می شود. به زبان بسیار ساده می توان اینگونه عنوان کرد که الگوریتم برت تمرکز خود را برای تشخیص ریز کلمات و عبارات در ارائه نتایج بهتر جستجو معطوف داشته است. این مدل برای یادگیری مقدماتی عمیق نمایش متن دو جهته برای استفاده بعدی در مدل های یادگیری ماشین در نظر گرفته شده است. (کورتیو، ۲۰۲۱)

$$P_{BERT} = \frac{1}{|x|} \sum_{\hat{x}_j \in \hat{x}} \max_{x_j \in x} x_i^T \hat{x}_j$$

فرمول ۱:

حافظه کوتاه مدت طولانی: شبکه های عصبی حافظه کوتاه مدت طولانی از لایه هایی از نرون ها تشکیل شده اند. داده ورودی از طریق یک شبکه برای ایجاد پیش بینی انتشار داده می شود. همانند شبکه های عصبی بازگشتی، شبکه های بازگشتی عمیق حافظه کوتاه مدت طولانی هم دارای اتصالات بازگشتی هستند؛ بنابراین وضعیت فعالیت های قبلی نرون از گام زمان قبلی به عنوان بخشی از داده ها برای فرموله کردن ورودی استفاده می شود. (Porzaker، ۱۳۹۸)

شبکه عصبی واحد بازگشتی: این شبکه، نوع خاصی از شبکه های عصبی بازگشتی هستند که توانایی یادگیری وابستگیهای بلندمدت را دارند. شبکه عصبی واحد بازگشتی با تنظیم انواع مختلف واحدهای چرخه، مشکل وابستگی به انواع مقیاسهای زمانی را حل کرده است. شبکه عصبی واحد بازگشتی مانند سازوکار حافظه کوتاه مدت طولانی عمل میکند و فرق آن این است که به جای سه دروازه از دو دروازه استفاده می کند. (Zhang et al., 2020).

شبکه عصبی بازگشتی: شبکه عصبی بازگشتی (RNN) یک شبکه عصبی است که از یک حالت پنهان h و یک خروجی اختیاری y تشکیل شده است که بر روی یک دنباله طول متغیر $X = (X_1, X_2, X_3, \dots, X_T)$ عمل می کند. (Cho et al., 2019).

$$h_{(t)} = f(h_{(t-1)}, x_t) \quad \text{فرمول ۲:}$$

کانولوشن یک بعدی: کانولوشن یک بعدی، به عنوان یک مؤلفه اساسی در شبکه های عصبی کانولوشن که برای مدیریت دادهای متوالی مانند متن طراحی شده است، عمل می کند. این الگوریتم به طور خاص برای کارهایی مانند طبقه بندی متن طراحی شده است و بر روی توالی های ورودی یک بعدی عمل می کند که به طور موثر ماهیت متوالی اطلاعات متنی نشان داده شده توسط توالی کلمات را به تصویر می کشد. (Zhi et al., 2024).

ایکس جی بوست: یک الگوریتم یادگیری تحت نظارت است که فرآیندی به نام تقویت را برای ارائه مدل های دقیق پیاده سازی می کند. یادگیری نظارت شده به وظیفه استنباط یک مدل پیش بینی از مجموعه ای از نمونه های آموزشی برچسب گذاری شده اشاره دارد. ایکس جی بوست در هسته خود یک الگوریتم تقویت درخت تصمیم است. Boosting به روش یادگیری گروهی برای ساختن بسیاری از مدل ها به صورت متوالی با هر مدل جدید تلاش برای اصلاح کاستی های مدل قبلی اشاره دارد. (Mishel & Frank, 2017).

$$\hat{f}(x) = \hat{f}_{(M)}(x) = \sum_{m=0}^M \hat{f}_m(x) \quad \text{فرمول ۳:}$$

آدابوست: روند کلی این الگوریتم به این صورت است که مجموعه ای وزن دار از تعداد زیادی طبقه کننده ی ضعیف را که هر یک معمولاً برحسب برخی ویژگی ها در فضای ویژگی های مسئله تعریف می گردند، به عنوان طبقه بندی کننده ی نهایی انتخاب می کند. (Molahosseini et al., 1388).

$$H(t) = \text{sing} \left(\sum_{t=1}^T \alpha_t h_t(x) \right) \quad \text{فرمول ۴:}$$

جنگل تصادفی^۱: جنگل تصادفی یک طبقه بند مجموعه های متشکل از طبقه بندهای درخت تصمیم است. هر طبقه بند برای هر نمونه ورودی به صورت $h(x, \Theta_k)$ که x نمونه ی ورودی و Θ_k مجموعه ی آموزش برای درخت K ام است. برای هر نمونه x ، هر درخت یک پیش بینی را برای رده ی نمونه ی x ارائه می دهد و در نهایت رده های با بیشترین تعداد رأی درختان روی ورودی x به عنوان رده ی نمونه انتخاب می شود، این فرآیند را جنگل تصادفی می نامند. (Moshrefzade et al., 1400).

$$y(x) = \arg \max_c \left(\frac{1}{k} \sum_{k=1}^k I(h_k(x) = c, x \in OOB_K) \right) \quad \text{فرمول ۵:}$$

¹ Random forest

درخت تصمیم^۱: یادگیری درخت تصمیم روشی برای مدل سازی پیش بینی کننده است که یک مدل را با تقسیم مکرر زیر مجموعه های نمونه های آموزشی (که نمونه ها نیز نامیده می شوند) طبق برخی معیارها یاد می گیرد. القاگرهای درخت تصمیم، یادگیرندگان تحت نظارتی هستند که نمونه های آموزشی برچسب گذاری شده را به عنوان ورودی می پذیرند و مدلی تولید می کنند که ممکن است برای پیش بینی برچسب های نمونه های جدید استفاده شود. (۲۰۱۷، Mishel & Frank)

ماشین بردار پشتیبان: در ماشین بردار پشتیبان با فرض اینکه داده ها به صورت خطی جداپایر باشند، ابر صفحه هائی با حداکثر حاشیه را به دست می آورد که دسته ها را جدا کنند. در مسایلی که داده ها به صورت خطی جداپایر نباشند داده ها به فضای با ابعاد بیشتر نگاشت پیدا می کنند تا بتوان آنها را در این فضای جدید به صورت خطی جدا نمود. ایده ی ماشین بردار پشتیبان برای جداسازی دسته ها این است که دو صفحه مرزی بسازد: دو صفحه مرزی موازی با صفحه دسته بندی رسم کرده و آن دو را آنقدر از هم دور می کنیم که به داده ها برخورد کنند. (۲۰۱۵، Yektaei)

$$w.x + b = 0$$

فرمول ۶:

ارزیابی

در این بخش به بررسی پارامترها و نتایج پرداخته خواهد شد.

Accuracy: مقدار اندازه گیری شده چقدر به مقدار واقعی نزدیک است برای دقت باید صحت بالا باشد ولی برعکسش لزوماً برقرار نیست.

$$ACC = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+TN+FP+FN}$$

فرمول ۷:

Precision: برای اندازه گیری های متوالی از یک مقدار میزان نزدیک بودن مقدارهای اندازه گیری را نشان می دهد.

$$PPV = \frac{TP}{TP+FP} = 1 - FDR$$

فرمول ۸:

Recall: عبارتست از کسری از جوابهای مثبت که درست تشخیص داده شده اند.

$$TPR = \frac{TP}{P} = \frac{TP}{TP+FN} = 1 - FNR$$

فرمول ۹:

F1_Score: برای ارزیابی مسائل طبقه بندی با بیش از دو کلاس هم استفاده می شود.

$$F_1 = 2 \cdot \frac{PPV \cdot TPR}{PPV+TPR} = \frac{2TP}{2TP+FP+FN}$$

فرمول ۱۰:

¹ Decision tree

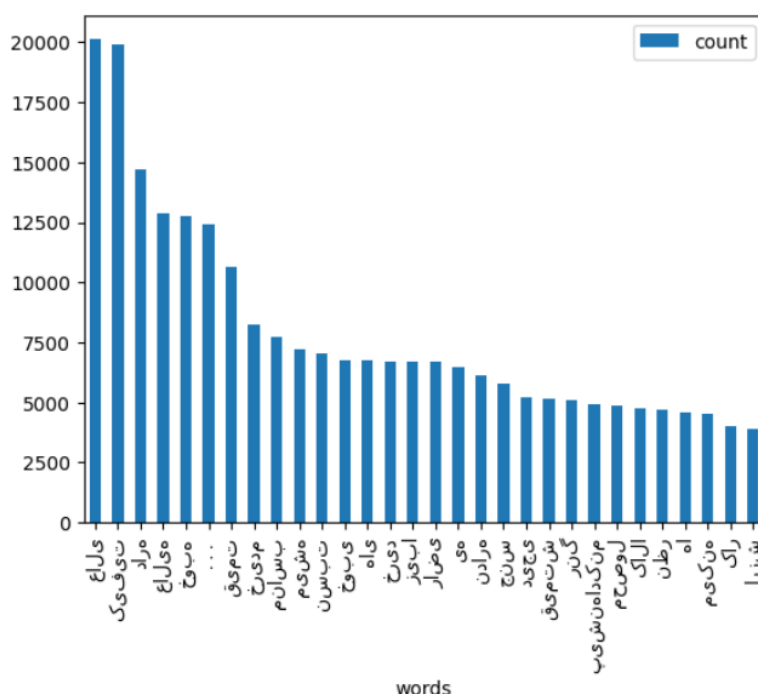
منحنی AUC-ROC: منحنی (ROC) یک معیار ارزیابی برای مسائل طبقه بندی باینری است. این یک منحنی احتمال است که TPR را در برابر FPR در مقادیر آستانه مختلف ترسیم می‌کند؛ و اساساً “سیگنال” را از “نویز” جدا می‌کند. منطقه زیر منحنی (AUC) اندازه گیری توانایی طبقه بندی کننده برای تمایز بین کلاس ها است و به عنوان خلاصه منحنی ROC استفاده می شود. وقتی $AUC = 1$ ، طبقه بندی کننده می تواند کاملاً بین تمام نقاط کلاس مثبت و منفی به درستی تمایز قائل شود. با این حال، اگر $AUC = 0$ بود، طبقه بندی کننده همه منفی ها را به عنوان مثبت و همه مثبت ها را به عنوان منفی پیش بینی می کرد. وقتی $AUC = 0.5$ باشد، طبقه بند قادر به تشخیص نقاط کلاس مثبت و منفی نیست. به این معنی که طبقه بندی کلاس تصادفی، یا کلاس ثابت را برای تمام نقاط داده پیش بینی می کند.

تابع فعالساز: در تابع فعالسازی^۱ که به آن «تابع انتقال» هم گفته می‌شود، در ابتدا مقادیر ورودی گره با یکدیگر ترکیب می‌شوند تا خروجی حاصل به تابع فعالسازی شبکه های عصبی منتقل شود. به عبارتی، مقادیر حاصل از ضرب وزن ها با ویژگی های ورودی با یکدیگر جمع می‌شوند و مقدار حاصل به تابع فعالسازی منتقل می‌شود. تابع فعالسازی در شبکه های عصبی ترکیب خطی ورودی ها را غیرخطی می‌کند و مقادیر ورودی را براساس نوع تابع فعالسازی به فضا با بازه مشخصی نگاشت می‌کند.

نرخ یادگیری: نرخ یادگیری آدر شبکه عصبی، ضریبی است که به وسیله آن می‌توان مقدار تغییر وزن‌ها را تنظیم کرد؛ یعنی اگر نرخ یادگیری بزرگ انتخاب شود، با گام‌های بلندتری به سمت مینیمم حرکت خواهیم کرد؛ اما اگر کوچک باشد، با گام‌های کوچکتری به سمت مینیمم حرکت خواهیم کرد.

نتایج

در این قسمت از پژوهش به تجزیه و تحلیل نتایج بدست آمده می پردازیم:

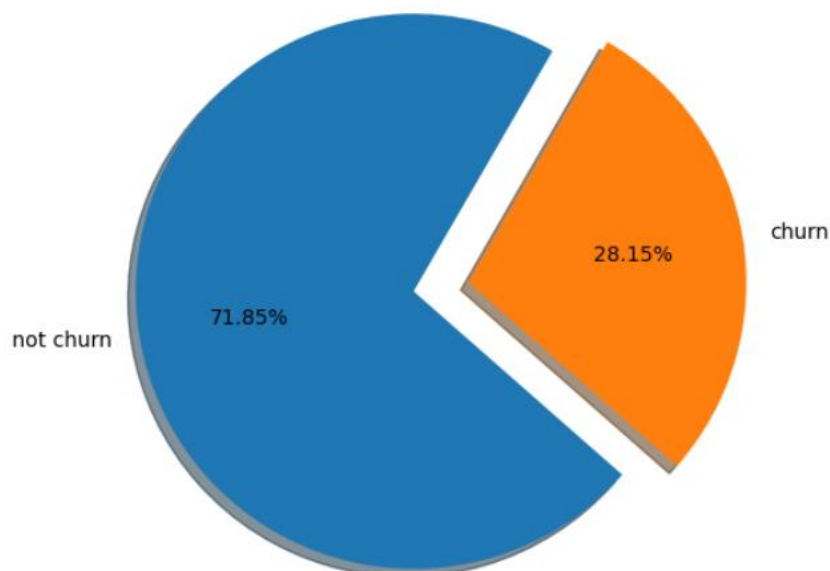


¹ Activation Function

² Learning rate

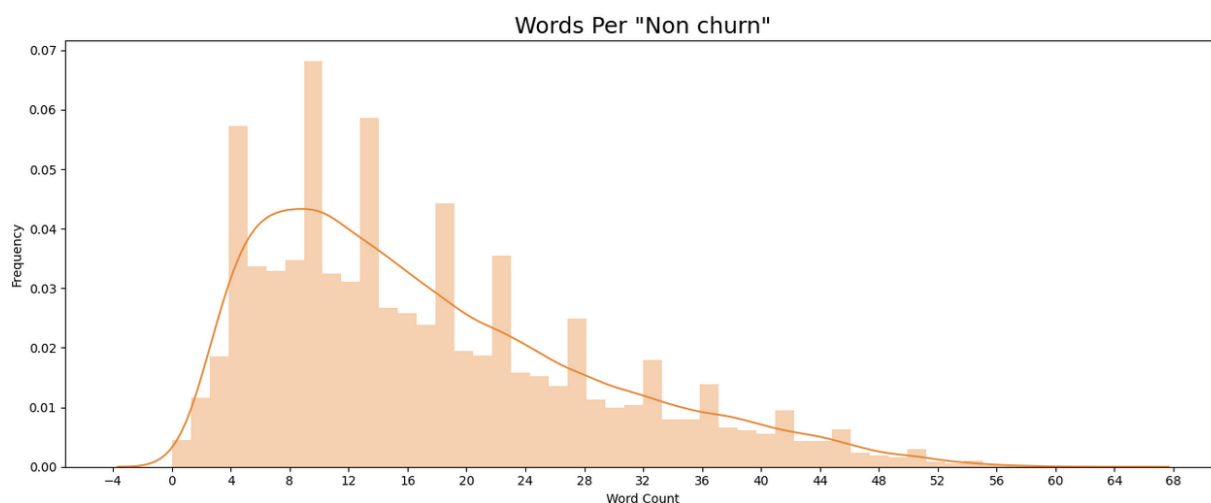
شکل ۲: نمودار فراوانی بیشترین تکرار کلمه در نظرات مردم

از آنجایی که تمامی کلمات به جز کلمات توقف که در فصل پیش گفته شد در مدل سازی نقش بسزایی ایفا می کنند. همانطور که شکل نشان می دهد، محور افقی کلمات آورده شده در نظرات و محور عمودی تکرار آن ها را نشان می دهد. کلمات عالی و کیفیت بیشترین تکرار را در نظرات مردم داشته اند بدین معنی که بیشتر مشتریان از محصول خریداری شده راضی بوده اند.



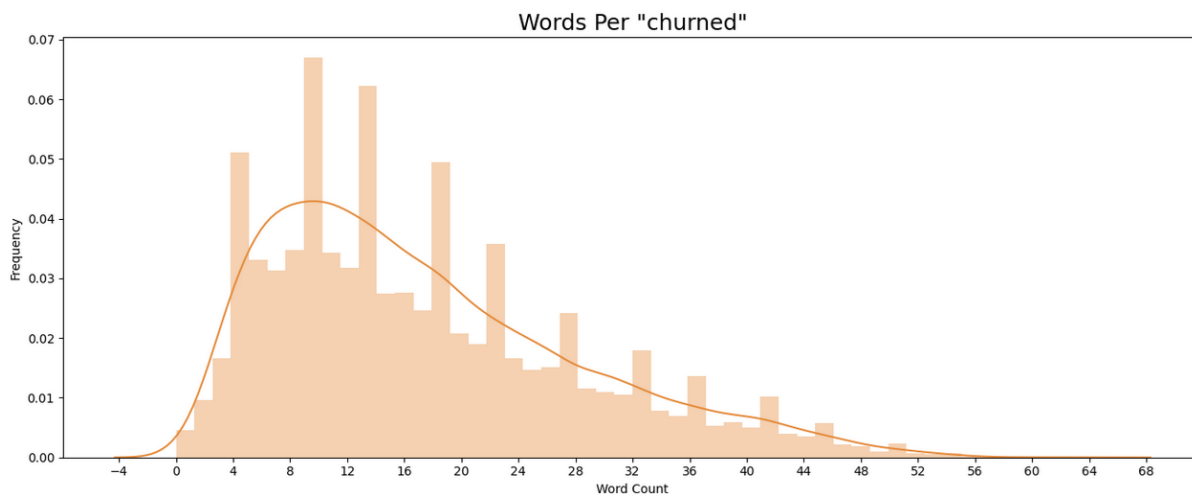
شکل ۳: نمودار ستون های هدف که برچسب گذاری شده

پس از انتخاب ویژگی و انتخاب ستون های هدف براساس پیشنهاد مشتریان به خرید محصول مورد نظر و تعداد پسندیده شدن و نشدن نظر، همانطور که شکل ۳ نشان می دهد ۷۱.۸۵٪ ریزش نمی کنند و ۲۸.۱۵٪ ریزش پیدا خواهند کرد که نشانه ی رضایت حداکثری مشتریان از محصولات خریداری شده است.



شکل ۴: نمودار فراوانی کلمات و تکرار تعداد آن ها در نظراتی که ریزش پیدا نمی کنند

همانطور که شکل ۴ نشان می دهد، محور افقی تعداد کلمات و محور عمودی احتمال وقوع کلمات را در نظرات مشتریانی که ریزش پیدا نمی کنند هستند و وزن یک سری کلمات به دلیل تکرار بیشتر و نرمال نبودن توزیع تاثیر بیشتری در مدل سازی ما دارد.



شکل ۵: نمودار فراوانی کلمات و تکرار تعداد آن ها در نظراتی که ریزش پیدا می کنند

همانطور که شکل ۵ نشان می دهد، در نظرات مشتریانی که ریزش پیدا می کنند وزن یک سری کلمات به دلیل تکرار بیشتر و نرمال نبودن توزیع تاثیر بیشتری در مدل سازی ما دارد که در شکل ۲ دیده شد.

الگوریتم های استفاده شده

الگوریتم برت بهینه شده: در الگوریتم برت بهینه شده، مدل سازی با پارامترهای زیر انجام شد:

جدول ۱: مقادیر پارامترهای استفاده شده در الگوریتم Bert

EarlyStopping	monitor='val_loss'
lr_scheduler	PolynomialDecay
initial_learning_rate	2e-5
end_learning_rate	0
Dropout	0.2
optimizer	adam
metrics	accuracy
validation_split	0.2
Batch-size	24
epoches	5

در نهایت مقادیر زیر محاسبه شدند:

جدول ۲: مقادیر بدست آمده از مدل سازی الگوریتم Bert

accuracy	F1-score	precision	recall	auc
0.8432532462	0.824564523	0.834534654	0.81235354	0.81

با توجه به جدول ۲ مقدار دقت بدست آمده از مدل سازی الگوریتم برت، برابر با ۰.۸۴۳۲۵۳۲۴۶۲ شد که با توجه به فارسی بودن و عامیانه بودن نظرات مقدار خوب و بهینه ای است.

الگوریتم حافظه کوتاه مدت طولانی

جدول ۳: مقادیر پارامترهای استفاده شده در الگوریتم LSTM

Dense	128
dropout	0.5
activation	sigmoid
loss	binary_crossentropy
optimizer	adam
metrics	accuracy
validation_split	0.2
Batch-size	256
epoches	5

جدول ۴: مقادیر بدست آمده از مدل سازی الگوریتم LSTM

accuracy	F1-score	precision	recall	auc
۰.۷۹۵۶۷۴۳	۰.۷۷۵۴۶۵۴۶	۰.۷۸۵۶۷	۰.۷۶۵۷۵۵۷۵	0.76

با توجه به جدول ۴ مقدار دقت بدست آمده از مدل سازی الگوریتم حافظه کوتاه مدت طولانی، برابر با ۰.۷۹۵۶۷۴۳ شد که مقدار کمتری نسبت به الگوریتم برت دارد.

الگوریتم GRU

جدول ۵: مقادیر پارامترهای استفاده شده در الگوریتم GRU

Dense	128
dropout	0.5
activation	sigmoid

loss	binary_crossentropy
optimizer	adam
metrics	accuracy
validation_split	0.2
Batch-size	256
epoches	5

جدول ۶: مقادیر بدست آمده از مدل سازی الگوریتم GRU

accuracy	F1-score	precision	recall	auc
۰.۸۰۴۴۴۸	۰.۷۹۵۶۵۶۸	۰.۷۷۸۸۸۸	۰.۷۸۵۶۵۶۸	0.77

با توجه به جدول ۶ مقدار دقت بدست آمده از مدل سازی الگوریتم GRU، برابر با ۰.۸۰۴۴۴۸ شد که مقدار کمتری نسبت به دو الگوریتم قبلی دارد.

الگوریتم شبکه عصبی بازگشتی

جدول ۷: مقادیر پارامترهای استفاده شده در الگوریتم RNN

Dense	128
Dropout	0.5
activation	sigmoid
loss	binary_crossentropy
Optimizer	adam
Metrics	accuracy
validation_split	0.2
Batch-size	256
epoches	5

جدول ۸: مقادیر بدست آمده از مدل سازی الگوریتم RNN

accuracy	F1-score	precision	recall	auc
----------	----------	-----------	--------	-----



۰.۷۵۶۵۷	۰.۷۴۶۷۶۷	۰.۷۲۶۷۸۷۶	۰.۷۳۶۷۵۶۷	0.75
---------	----------	-----------	-----------	------

همانطور که در جدول ۸ مشاهده می شود، مقدار دقت بدست آمده از مدل سازی شبکه عصبی بازگشتی برابر با ۰.۷۵۶۵۷ است که نسبت به الگوریتم های قبلی استفاده شده مقدار کمتری دارد.

الگوریتم کانولوشن یک بعدی

جدول ۹: مقادیر پارامترهای استفاده شده در الگوریتم Conv1D

dense	128
dropout	0.5
activation	sigmoid
loss	binary_crossentropy
optimizer	adam
metrics	accuracy
validation_split	0.2
Batch-size	256
epoches	5

جدول ۱۰: مقادیر بدست آمده از مدل سازی الگوریتم Conv1D

accuracy	F1-score	precision	recall	auc
۰.۷۶۶۷۷۸	۰.۷۵۶۵۷	۰.۷۴۴۶۶۸	۰.۷۴۳۵۴۳۸	0.77

همانطور که در جدول ۱۰ مشاهده می شود، مقدار دقت بدست آمده از مدل سازی الگوریتم Conv1D برابر با ۰.۷۶۶۷۷۸ است که نسبت به الگوریتم های قبلی استفاده شده مقدار بیشتری دارد.

الگوریتم Conv1D + LSTM

جدول ۱۱: مقادیر پارامترهای استفاده شده در ترکیب الگوریتم های Conv1D+LSTM

dense	128
-------	-----

dropout	0.5
activation	sigmoid
loss	binary_crossentropy
optimizer	adam
metrics	accuracy
validation_split	0.2
Batch-size	256
epoches	5

جدول ۱۲: مقادیر بدست آمده از مدل سازی ترکیب الگوریتم های Conv1D+LSTM

accuracy	F1-score	precision	recall	auc
۰.۷۸۸۷۶۸	۰.۷۵۹۷۶۸۹	۰.۷۶۹۸۹	۰.۷۸۷۶۸۹	0.77

با توجه به جدول بالا مقدار دقت بدست آمده از مدل سازی ترکیب دو الگوریتم Conv1D+LSTM، برابر با ۰.۷۸۸۷۶۸ شد که مقدار بیشتری نسبت به الگوریتم قبلی دارد..

الگوریتم Conv1D+GRU

جدول ۱۳: مقادیر پارامترهای استفاده شده در ترکیب الگوریتم های Conv1D+GRU

dense	128
dropout	0.5
activation	sigmoid
loss	binary_crossentropy
optimizer	adam
metrics	accuracy
validation_split	0.2
Batch-size	256
epoches	5

جدول ۱۴: مقادیر بدست آمده از مدل سازی ترکیب الگوریتم های Conv1D+GRU

accuracy	F1-score	precision	recall	auc
۰.۷۹۵۶۴۵	۰.۷۸۷۷۶۸۹	۰.۷۷۷۴۳	۰.۷۸۷۶۸۹	0.79

همانطور که در جدول ۱۴ مشاهده می شود، مقدار دقت بدست آمده از مدل سازی ترکیب دو الگوریتم Conv1D+GRU برابر با ۰.۷۶۶۷۷۸ است که نسبت به الگوریتم های قبلی استفاده شده مقدار بیشتری دارد.

الگوریتم Conv1D+RNN

جدول ۱۵: مقادیر پارامترهای استفاده شده در ترکیب الگوریتم های Conv1D+RNN

dense	128
dropout	0.5
activation	sigmoid
loss	binary_crossentropy
optimizer	adam
metrics	accuracy
validation_split	0.2
Batch-size	256
epoches	5

جدول ۱۶: مقادیر بدست آمده از مدل سازی ترکیب الگوریتم های Conv1D+RNN

accuracy	F1-score	precision	recall	auc
۰.۷۹۵۶۴۵	۰.۷۸۷۷۶۸۹	۰.۷۷۷۴۳	۰.۷۸۷۶۸۹	0.79

با توجه به جدول ۱۶ مقدار دقت بدست آمده از مدل سازی ترکیب دو الگوریتم Conv1D+RNN، برابر با ۰.۷۹۵۶۴۵ شد که مقدار بیشتری نسبت به الگوریتم قبلی دارد..

الگوریتم ایکس جی بوست

جدول ۱۷: مقادیر بدست آمده از مدل سازی الگوریتم XGBoost



accuracy	F1-score	precision	recall	Auc
۰.۷۶۸۷۶۸	۰.۷۴۹۷۶۸۹	۰.۷۵۵۸۹	۰.۷۵۵۶۸۹	0.75

همانطور که در جدول ۱۷ مشاهده شد، مقدار دقت بدست آمده برابر با ۰.۷۶۸۷۶۸ شد که در مقایسه با بقیه الگوریتم ها به جز RNN مقدار کمتری دارد.

الگوریتم جنگل تصادفی

جدول ۱۸: مقادیر پارامترهای استفاده شده در الگوریتم RF

n_estimators	random_state
100	42

جدول ۱۹: مقادیر بدست آمده از مدل سازی الگوریتم RF

accuracy	F1-score	precision	recall	Auc
۰.۷۴۵۵۵۵	۰.۷۴۹۸۸۷	۰.۷۴۵۶۲	۰.۷۳۵۶۲۳۴	0.75

با توجه به جدول ۱۹ دقت بدست آمده برابر با ۰.۷۴۵۵۵۵ که نسبت الگوریتم های قبلی مقدار کمتری دارد یعنی دقت کمتری در تشخیص ریزش مشتری نسبت به بقیه دارد.

الگوریتم آدا بوست

جدول ۲۰: مقادیر پارامترهای استفاده شده در الگوریتم AdaBoost

n_estimators	random_state
۱۰۰	۴۲

جدول ۲۱: مقادیر بدست آمده از مدل سازی الگوریتم AdaBoost

accuracy	F1-score	precision	recall	auc
۰.۷۲۴۵۳۸	۰.۷۱۴۸۳۸	۰.۷۳۵۶۳۲	۰.۷۲۴۵۶۷	0.75

همانطور که در جدول ۲۱ مشاهده می شود مقدار دقت برابر با ۰.۷۲۴۵۳۸ شد که نسبت عملکرد ضعیف تری نسبت به بقیه الگوریتم ها دارد.

الگوریتم ماشین بردار پشتیبان

جدول ۲۲: مقادیر پارامترهای استفاده شده در الگوریتم SVM

kernel	Random_state
linear	42

جدول ۲۳: مقادیر بدست آمده از مدل سازی الگوریتم SVM

accuracy	F1-score	precision	recall	auc
۰.۶۹۴۳۲۱	۰.۶۷۴۳۴۴	۰.۶۸۵۶۳۵	۰.۶۸۵۴۳۴	۰.۶۹

همانطور که در جدول ۲۳ مشاهده می شود دقت بدست آمده برابر با ۰.۶۹۴۳۲۱ است که عملکرد بسیار ضعیف این الگوریتم در طبقه بندی نظرات مردم را نشان می دهد.

الگوریتم درخت تصمیم

جدول ۲۴: مقادیر بدست آمده از مدل سازی الگوریتم DT

accuracy	F1-score	precision	recall	auc
۰.۶۸۳۴۵۳	۰.۶۲۶۵۸۸	۰.۶۱۶۵۷۴	۰.۶۶۵۶۳۴	۰.۶۶

در جدول ۲۴ مشاهده می شود که مقدار دقت بدست آمده برابر با ۰.۶۸۳۴۵۳ شد که نشان دهنده ی ضعیف ترین الگوریتم برای طبقه بندی نظرات مشتریان است.

بحث و نتیجه گیری

در دنیای امروز با توجه به پیشرفت روز افزون فناوری و همچنین افزایش تعداد فروشگاه ها به خصوص فروشگاه های آنلاین، باعث پیشرفت بسیار زیادی در حوزه ی تجارت الکترونیک شده است. همین امر موجب به وجود آمدن رقابت بسیار تنگاتنگ در بازار شده به این صورت که تمامی فروشگاه ها و شرکت ها به دنبال افزایش فروش محصولات خود بازار جهانی، افزایش مشتریان، کاهش هزینه های خود و مهمترین عامل یعنی ریزش مشتریان می باشند. در این میان ریزش مشتری مهمترین عامل رقابت بازار کنونی در تجارت الکترونیک می باشد که می تواند به عوامل زیادی از

جمله؛ ناراضی بودن مشتریان از خدمات و قیمت ها، جذب مشتریان به صورت اشتباه، تبلیغات نامناسب تجارت، عدم اعتماد مشتری به فروشگاه و دلایل بسیار دیگری بستگی داشته باشد. پیش‌بینی ریزش مشتریان به کسب و کار کمک می‌کند که؛ مشتریان در حال ترک سازمان یا شرکت شناسایی و از ریزش آن‌ها جلوگیری شود، باعث بهبود تجربه کاربر می‌شود، به شناسایی نقاط ضعف کسب و کار کمک بسزایی می‌کند، باعث به وجود آمدن فرصت افزایش سود و درآمد می‌گردد و همچنین باعث انتخاب اهداف بهتر برای کسب و کار می‌گردد. این پژوهش برای پیش‌بینی ریزش مشتری استفاده از الگوریتم‌های برت، شبکه عصبی بازگشتی، شبکه عصبی واحد بازگشتی، حافظه کوتاه‌مدت طولانی، کانولوشن یک بعدی، ایکس جی بوست، آدابوست، بردار پیش‌تیبان، درخت تصمیم و درنهایت جنگل تصادفی را پیشنهاد می‌دهد که درنهایت در بین این‌ها برت بهترین عملکرد را ارائه می‌دهد. در این پژوهش تجزیه و تحلیل روی مجموعه دادگان کیفیت دیجیکالا که شامل نظرات ۱۰۰ هزار مشتری در مورد کیفیت ۱۰۰ هزار محصول خریداری شده انجام شد و پس از حذف داده‌های ناهنجار، خالی و نرمالسازی متن نظرات و تقسیم داده‌ها به داده‌های تست و آموزشی مدل‌سازی روی آن‌ها با الگوریتم‌های ذکر شده در بالا انجام شد. به طور کلی از نظر معیار دقت الگوریتم‌ها به ترتیب زیر عملکرد خود را نشان دادند؛

Bert>GRU>LSTM>Conv1D + RNN = Conv1D + GRU>Conv1D + LSTM> XGBoost> Conv1D> RNN> AdaBoost> RF> SVM> DT

برای جلوگیری از ریزش مشتریان شرکت‌ها و فروشگاه‌های اینترنتی می‌توانند کارهای زیر را انجام دهند:

- حفظ مشتریان فعلی بهترین گزینه برای حفظ تجارت الکترونیک است، زیرا تلاش برای جذب مشتری جدید تقریباً پنج برابر گران‌تر از حفظ آن است. علاوه بر این، یک مشتری جدید باید پس از یک دوره طولانی تر و تعداد بیشتری از تراکنش‌ها ارزشمند تلقی شود.
- حجم عظیمی از اطلاعات قابل دسترسی برای سازمان‌های تجارت الکترونیکی را می‌توان برای یافتن استانداردهای رفتاری تحت پوشش استخراج کرد و هرگونه انحراف از الگوها یا نمونه‌های مورد انتظار را می‌توان به عنوان یک ریزش مشتری بالقوه در نظر گرفت.
- شناسایی مشتریان وفاداری که بیشتر از بقیه خرید انجام می‌دهند و ارائه تخفیفات ویژه به آن‌ها
- ارتباط مداوم و متناسب با مشتریان و پشتیبانی مناسب از محصولات
- شناسایی مشتریان در حال ریزش و ارائه پکیج‌های ویژه به آن‌ها

منابع

- Bilal, S. F., Almazroi, A. A., Bashir, S., Khan, F. H., & Almazroi, A. A. (2022). An ensemble-based approach using a combination of clustering and classification algorithms to enhance customer churn prediction in telecom industry. *PeerJ Computer Science*, 8, e854.
- Brândușoiu, I., Todorean, G., & Beleiu, H. (2016, June). Methods for churn prediction in the pre-paid mobile telecommunications industry. In 2016 International conference on communications (COMM) (pp. 97-100). IEEE.
- Brzezinski, D., & Stefanowski, J. (2017). Prequential AUC: properties of the area under the ROC curve for data streams with concept drift. *Knowledge and Information Systems*, 52, 531-562.
- Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36(3), 4626-4636.
- Chu, Y., Fei, J., & Hou, S. (2019). Adaptive global sliding-mode control for dynamic systems using double hidden layer recurrent neural network structure. *IEEE transactions on neural networks and learning systems*, 31(4), 1297-1309.
- Dahiya, K., & Bhatia, S. (2015, September). Customer churn analysis in telecom industry. In 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions) (pp. 1-6). IEEE.
- Gürsoy, U. Ş. (2010). Customer churn analysis in telecommunication sector. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 39(1), 35-49.



Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, 34(10), 2902-2917. Hall, 2000

Huang, Y., Zhu, F., Yuan, M., Deng, K., Li, Y., Ni, B., ... & Zeng, J. (2015, May). Telco churn prediction with big data. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data* (pp. 607-618).

Kim, S., & Lee, H. (2022). Customer churn prediction in influencer commerce: An application of decision trees. *Procedia Computer Science*, 199, 1332-1339.



Predicting customer churn in e-commerce using a business intelligence strategy based on machine learning and deep learning

First Author :amirhosein taali

Affiliation:masters degree student in computer engineering

Second Author :dr morteza mohammadi zanjireh

Affiliation: Assistant professor in imam khomeini international university

Abstract

Today, customer churn in e-commerce is one of the most important concerns of companies. With the advancement of science and technology, competition between large companies in the world to increase revenue, increase the number of customers, and retain existing customers has increased. Therefore, companies always try to maintain the satisfaction of their existing customers in this fiercely competitive market and improve or maintain their position in the global ranking to prevent customer churn. Since finding new customers is much more difficult than retaining the company's current customers due to the very high costs of advertising and also being time-consuming, most companies take steps to reduce their costs. In this research, using the quality data of the large Digikala collection, first the existing data is homogenized and normalized by preprocessing them, and then using machine learning algorithms such as support vector machines, decision trees, Ada Boost, XG Boost, and deep learning algorithms such as; We modeled and predicted customer churn using BERT algorithms, recurrent neural network, recurrent unit neural network, long short-term memory, and one-dimensional convolution. Finally, after the simulation, the best algorithm for predicting customer churn is the optimized BERT algorithm with an accuracy of 84.32%.

Keywords: Forecasting, Customer Churn, Business Intelligence, Machine Learning, E-Commerce