

به نام خدا

فاز شناخت و فهم مسئله پروژه درس مبانی داده کاوی 1401-02

اعضای گروه:

امیررضا حسینی – پاشا احمدی

عنوان پروژه:

بررسی رفتار کاربران در یک بانک جهت افتتاح حساب

<https://www.kaggle.com/datasets/janiobachmann/bank-marketing-dataset>

الف – مشکل چیست؟ (توضیح: چرا این مسئله یا داده تعریف شده است در واقع اگر این پروژه انجام نمی شد چه مشکلی، حل نشده باقی می ماند.)

امروزه، باتوجه به وجود طیف وسیع مشتریان و تنوع سلاقی و خواسته های گوناگون میان مؤسسات مالی و آنها، بیش از پیش نیاز به برنامه ریزی های کارآمد در مقیاس کلان حس می شود. از این رو، بانک های مختلف، با برنامه ها و پیشنهادهای تجاری زیاد پا به عرصه گذاشته تا گوی رقابت را از یکدیگر برابند. اما یکی از اساسی ترین چالش ها در این بازار تجاری، نبود سیستمی با راهبردی کارآمد در زمینه تبلیغات و کمپین های مختلف تبلیغاتی برای این بازار است. این سیستم باید بتواند بهترین و مفیدترین سناریوی ممکن را در اختیار بانک قرار داده تا بتواند با پیشنهاد دادن آن به هر یک از مشتریان، هم به آنها خدمت کرده و با راضی نگه داشتن آنان، سوددهی بلندمدت خود را به بالاترین حد ممکن برساند که باعث رضایت حداکثری هر دو طرف می شود. در این پروژه اصلی ترین هدف ما تلاش برای ساختن همچنین سیستمی می باشد.

ب- سوال داده کاوی و معیار ارزیابی آن چیست؟ (توضیح: قرار است چه مقدار (ستون یا اطلاعات) را پیش بینی (تخمین یا برآورد) کنید؟ و قصد دارید از چه اقلام اطلاعاتی برای این پیش بینی استفاده کنید؟ معیار ارزیابی چیست و ایده آن چقدر است؟)

یکی از اهداف تجاری که می‌توان در آن از مجموعه داده‌های بازاریابی استفاده کرد، پیش‌بینی نتایج یک کمپین بازاریابی برای هر مشتری و روشن‌شدن عواملی است که بر نتایج کمپین تأثیر می‌گذارند. این به کسب‌وکارها کمک می‌کند تا راه‌هایی برای بهبود کمپین‌های بازاریابی خود پیدا کنند.

هدف ما در این پروژه داده‌کاوی به‌دست‌آوردن مقادیر مناسب برای ویژگی‌هایی مانند: ماه و روز هفته‌ای که تماس انجام می‌شود، تعداد تماس و مدت‌زمان آن و نوع بستر ارتباطی به‌نحوی که منجر به افزایش احتمال و همچنین ترغیب مشتری برای باز کردن سپرده در بانک شود. برای این منظور، می‌توان از اقلام اطلاعاتی موجود در دیتاست مانند سن، شغل، وضعیت تأهل، تحصیلات، میزان موجودی بانکی، داشتن وام و یا تسهیلات بانکی، تاریخ تماس با بانک، تعداد تماس‌های قبلی، مدت‌زمان مکالمه با بانک، و نتیجه تماس قبلی و در نهایت نتیجه آن که آیا باعث باز کردن سپرده توسط مشتری شده است یا نه استفاده کرد. معیار ارزیابی مدل پیش‌بینی معمولاً بر اساس مقایسه بین پیش‌بینی مدل با واقعیت است. برای این منظور، می‌توان از معیارهای مختلفی مانند ماتریس درهم‌ریختگی (confusion matrix)، دقت (accuracy)، بازخوانی (recall)، دقت پیش‌بینی مثبت (precision)، معیار F1 و ... استفاده کرد.

بنابراین، برای حصول این مهم ما قصد داریم تا با انواع تکنیک‌های داده‌کاوی مانند بصری‌سازی، آنالیز، شناسایی الگوها و روابط، یادگیری ماشین و... به تحقق این اهداف دست یابیم تا بهترین و بهینه‌ترین کمک را به سوددهی بانک در استراتژی‌های کمپین‌های تبلیغاتی آتی کرده باشیم.

ج- مشخصات دیتاست و ویژگی‌های موجود در آن

این مجموعه داده تشکیل شده از ۳ گروه داده می‌باشد:

داده‌های مربوط به مشتریان شامل:

- سن: (عددی)
- شغل: نوع شغل (دسته‌ای: 'ادمین', 'کارگر', 'کارآفرین', 'خانه‌دار', 'مدیر', 'بازنشسته', 'خود اشتغال', 'خدمات', 'دانشجو', 'تکنسین', 'بیکار', 'نامشخص')
- وضعیت تأهل: وضعیت تأهل (دسته‌ای: 'طلاق گرفته', 'متأهل', 'مجرد', 'نامشخص'),
- تحصیلات: (دسته‌ای: سیکل، دیپلم، عالی، و نامشخص)
- قرض الحسنه: آیا شخص حساب قرض الحسنه دارد؟ (دسته‌ای: 'خیر', 'بله', 'نامشخص')
- وام مسکن: آیا شخص وام مسکن دارد؟ (دسته‌ای: 'خیر', 'بله', 'نامشخص')
- وام شخصی: آیا شخص وام شخصی دارد؟ (دسته‌ای: 'خیر', 'بله', 'نامشخص')
- موجودی: کل موجودی حساب شخص. (یک متغیر عددی که می‌تواند برحسب بدهکاری یا طلبکاری منفی یا مثبت باشد).

داده‌های مربوط به آخرین تماس با هر مشتری در کمپین فعلی:

- نوع بستر ارتباط تماس (دسته‌ای: 'گوشی موبایل', 'تلفن ثابت')
- آخرین ماه تماس در سال (دسته‌ای: 'ژانویه', 'فوریه', 'مارس', ..., 'نوامبر', 'دسامبر')
- آخرین روز هفته تماس (دسته‌ای: 'دوشنبه', 'سه‌شنبه', ..., 'چهارشنبه', 'پنج‌شنبه', 'جمعه')
- مدت‌زمان آخرین تماس گرفته شده با مشتری بر حسب ثانیه (عددی).

- تعداد تماس‌های انجام شده در طول این کمپین و برای این مشتری (عددی، شامل آخرین تماس)
 - تعداد روزهایی که پس از تماس آخرین با مشتری از کمپین قبلی گذشته است. (عددی؛ عدد 999 به معنای عدم تماس قبلی با مشتری است)
 - تعداد تماس‌های انجام شده قبل از این کمپین و برای این مشتری (عددی)
 - نتیجه کمپین بازاریابی قبلی (دسته‌ای: 'شکست'، 'ناموجود'، 'موفقیت')
- برچسب خروجی (Y) شامل:
- آیا مشتری سپرده کوتاه‌مدت باز کرده است؟ (باینری شامل بله و خیر)

د- فعالیت پیش رو و بیان مختصر ایده تیم

در این پروژه قرار است با استفاده از آنالیز داده‌های نامبرده و اعمال الگوریتم‌های داده‌کاوی مختلفی مانند شبکه‌های عصبی، درخت تصمیم و یا SVM روی آنها، الگوهای رفتاری کاربران را پیدا کرده و کشف الگوهای رفتاری بین آنها را بررسی کنیم. سپس با استفاده از رویکردهای دسته‌بندی و خوشه‌بندی و نتیجه‌گیری‌های مرتبط برای برنامه‌ریزی‌های کلان آتی بانکی می‌توانیم به بهینه‌ترین طرح‌ها و پیشنهادات برای طیف وسیع سلايق و خواسته‌های هر مشتری برسیم. در نتیجه با انجام این کار در این پروژه، می‌توانیم سطح رضایت کاربران و تصمیمات مدیریتی و برنامه‌ریزی را باتوجه‌به نیازهای کاربران بهبود بخشیم تا در آینده نیز از استراتژی‌های کارآمدتری برای بازاریابی استفاده کنیم و همچنین سوددهی را برای مؤسسه بانکی بهینه کنیم.