



شناسایی ویژگی‌ها و شاخص‌هایی برای تشخیص پیش‌بینی هوشمند ریزش مشتریان

سید محمد سینا میر عبدالباقی^۱

۱- دانش‌آموخته کارشناسی ارشد، مهندسی صنایع گرایش سیستم‌های اطلاعاتی. دانشگاه شهید بهشتی. تهران. ایران.

چکیده

در این مقاله به طور عمده بر روی روش‌های کاهش ویژگی یا ابعاد مانند تجزیه و تحلیل مولفه اصلی (PCA)، اتوانکدر (AutoEncoder)، تجزیه و تحلیل افتراقی (LDA)، توزیع تی (T-sne)، و همچنین انتخاب ویژگی‌های مهم الگوریتم تقویت‌کننده شدید گرادیان (Xgboost) پرداخته شده است. هدف این مقاله ارائه یک مدل گروهی ترتیبی پیش‌بینی ریزش مشتریان براساس الگوریتم‌های تقویت‌کننده گرادیان (مانند تقویت‌کننده شدید گرادیان (Xgboost) و تقویت‌کننده ضعیف گرادیان (Light GBM)) است. مدل ارائه شده به طور کلی شامل چهار مرحله است: یک انجام پیش‌پردازش داده برای استاندارد کردن و از بین بردن داده‌های فراموش شده و نویزی، دو یک سیستم جامع کاهش ابعاد براساس مدل‌های محبوب کاهش ابعاد و در نتیجه انتخاب بهترین آن‌ها، سه تنظیم کردن ابرپارامترهای الگوریتم‌های تقویت‌کننده با استفاده از بهینه‌سازی بیزین و بهینه‌سازی ژنتیک و چهار تفسیر مدل و تاثیر ویژگی‌ها با استفاده از روش اهمیت ویژگی‌های تقویت‌کننده شدید گرادیان (Xgboost) و شپ (Shap) است. علاوه بر مقایسه و یافتن بهترین کاهش ابعاد دهنده، روش پیشنهادی بر روی چهار مجموعه داده معروف در صنعت پیاده‌سازی و برتری خود را نسبت به روش‌های عادی گروهی و الگوریتم‌های یادگیری ماشین مانند آدابوست (Adaboost)، ماشین پشتیبان بردار (SVM)، درخت تصمیم با استفاده از هفت معیار ارزیابی مهم مانند: دقت، ناحیه زیر منحنی (AUC)، کاپا، بریر (Brier)، ضریب همبستگی متیوز (MCC)، نمره F1 و حداکثر انتظار سودآوری مدل ریزش (EMPC) نشان داده است. همچنین مدل ارائه شده تغییرات بسیار خوبی را در نحوه مواجهه مدل با مجموعه داده نامتوازن را در معیارهای ارزیابی نشان می‌دهد.

واژگان کلیدی: ریزش مشتریان، یادگیری ماشین، الگوریتم تقویت‌کننده، کاهش ابعاد، بهینه‌سازی ابرپارامترها