

چکیده

در یک شرکت تجاری ایمیل دهی مستقیم به مشتریان بالقوه می تواند راهی موثر برای فروش محصولات و سرویسها باشد. مهمترین فعالیت بازاریابی از طریق ایمیل شناسایی مشتریان بالقوه به منظور پرهیز از ارسال ایمیل های غیر ضروری و ناخواسته می باشد. هدف این مقاله ارائه راهی برای رویارویی با این مساله در یک شرکت بیمه است. دو فعالیت عمده: 1- پیش بینی مشتریانی که به طور بالقوه علاقه مند به داشتن بیمه *caravan* می باشند. 2- توصیف مشتریهای واقعی و بالقوه بیمه کاروان و بیان تفاوت آنها با دیگر مشتریان. مجموعه داده مجموعه ای نامتوازن شامل رکوردهای مشتریان می باشد که هر رکورد از 86 ویژگی تشکیل شده است. در این بررسی روشهای مختلفی برای کاهش ابعاد داده به منظور کاهش خطا صورت گرفته و برای هر روش مدلهای یادگیری مختلف بررسی شده است. مدلهای مختلف با توجه به معیارهای خاصی مورد مقایسه قرار گرفته و در نهایت مدل *naïve bayse* با 11 ویژگی که به روش *cfsAvaluator* انتخاب شده به عنوان مدل مناسبتر شناخته شده است. برای فعالیت توصیفی نیز از روش *chiSquaredAttributeEval* استفاده نموده ایم. کلیه فعالیتهای مربوطه در محیط *weka* انجام شده است.

کلمات کلیدی

انتخاب ویژگی های مدلهای یادگیری انحراف معیار محیط *weka*