

روش‌های آماری کشف تقلب در بیمه

دکتر محمدرضا فریدروحانی

مهدی اولیایی

چکیده

مساله بروز تقلب در ادعاهای بیمه یکی از مشکلات پیش‌روی شرکت‌های بیمه است. از این‌رو بحث کشف این‌گونه تقلبات در انواع بیمه‌ها یکی از موضوعات مورد توجه کارشناسان حوزه‌های مختلف است. در یک تقسیم‌بندی عمومی روش‌های تشخیص تقلب را می‌توان به سه رده نظارتی، نیم‌نظارتی و غیرنظارتی تقسیم کرد. روش‌های نظارتی با استفاده از یک مجموعه داده مشخص که موارد متقلبانه و غیرمتقلبانه در آن مشخص هستند، اقدام به ساخت مدلی برای امتیازدهی درجه تقلب یا برچسب‌گذاری داده‌های جدید می‌کنند. روش‌های خوشه‌بندی آماری از قبیل آنالیز تشخیص خطی و تشخیص لوژستیک، نزدیکترین همسایگی k ام و مارکوف پنهان نظارتی از زمره روش‌های نظارتی محسوب می‌شوند. در روش‌های نیم‌نظارتی با معلوم بودن رفتار یا توزیع مجموعه داده موارد غیرمتقلبانه، اقدام به ساخت مدلی برای امتیازدهی درجه تقلب یا برچسب‌گذاری داده‌های جدید می‌کنند. روش‌های غیرنظارتی زمانی مورد استفاده قرار می‌گیرند که هیچ مجموعه پیشینی از موارد متقلبانه و غیرمتقلبانه نداریم و صرفاً براساس داده‌های مورد آزمون اقدام به کشف موارد متقلبانه می‌کنند. روش‌های مختلفی برای این منظور وجود دارند که می‌توان به روش‌های بنفورد، k - میانگین، تحلیل PRIDIT و تحلیل مولفه‌های اصلی اشاره کرد. بدیهی است هریک از این روش‌ها دارای مزایا و معایبی است، اما روش مورد استفاده براساس نوع داده‌های در اختیار و امکانات رسته‌بندی آنها مشخص می‌شود. در این مقاله با مروری بر روش‌های مزبور، به بررسی نتایج یک مطالعه‌ی شبیه‌سازی مبتنی بر داده‌های واقعی خواهیم داشت.