

" شناسایی نشانه های تراکنش های جعلی با استفاده از تکنیک های داده کاوی "

مقدمه:

روش های شناسایی تقلب را میتوان به دو دسته ی روش های تشخیص سوء استفاده و تشخیص ناهنجاری تقسیم کرد. تکنیک های مبتنی بر رویکرد تشخیص سوء استفاده زمانی استفاده می شود که تشخیص تقلب به صورت از پیش شناخته شده بوده و بر اساس امضای تقلب می توان رفتار جاری مشتریان را بررسی نمود . طبیعتاً به دلیل شناخت کامل رفتار قبلی مشتریان، دقت شناسایی تقلب در این روش بسیار بالاست. اما نقطه ضعف این روشها، عدم پوشش دهی کامل محدوده ی تقلب می باشد. بدین معنی که فقط تقلب هایی شناسایی میشوند که حداقل یک بار رخ داده و یا امضای آن به سیستم تشخیص تقلب ارائه شده باشد. در مقابل این روش ها، روش های دیگری هم هستند که مبتنی بر تشخیص ناهنجاری اند. یعنی سعی می کنند رفتار آتی مشتری را پیش بینی کرده و به این منظور تاریخچه ی رفتار وی را مورد بررسی قرار میدهند. در اینگونه روشها، قاعده ی ثابت و مشخصی برای تعریف تقلب تعریف نمیشود بلکه رفتار عادی و نرمال مشتری به سیستم تشخیص تقلب آموخته شده و هرگونه انحراف از آن به معنی تقلب فرض میگردد. بدیهی است که دقت در این روش پایین تر ولی گستره ی تقلب های کشف شده بالاتر است. که در این پروژه بر روی پیاده سازی روش اول تمرکز شده است.

معرفی مسئله:

در این پروژه، طبقه بندی (classification) را برای شناسایی علائم حسابهای جعلی و الگوهای معاملات جعلی مورد استفاده قرار میدهم که این قوانین براساس علائم شناسایی شده از قبل برای طراحی یک سیستم نظارتی مورد استفاده قرار میگیرد. علت آن این است که برای تشخیص تقلب نیاز هست که ابتدا به سیستم آموزش داده شود که در چه حالاتی تقلب صورت میگیرد.

حالت کلی این پروژه شامل بخشهای اصلی زیر خواهد بود :

ما از مجموعه داده های کارت های اعتباری آلمان استفاده کردیم این دیتاست 1000 رکورد و 21 ویژگی دارد و فرمت فایل آن ARFF میباشد. بعد از یافتن دیتاست به کاوش و ارزیابی اولیه در میان داده ها می پردازیم که بتوان یک دیدگاه کلی نسبت به داده های موجود بدست آورد که نتایج این کاوش در بخش های بعد آورده خواهد شد. سپس با استفاده از نرم افزار وکا بر روی دیتا ست پیش پردازش انجام دادیم ،چون این دیتا ست داده های واقعی نمیباشد بنابراین داده های miss و نویزی وجود ندارد.

سپس مطابق الگویی نشانه های حسابهای جعلی بدست میآوریم ونشانه های بدست آمده را با نشانه های معرفی شده از قبل مقایسه میکنیم.

