

کشف کلاهبرداری بانکی با استفاده از ماشین بردار پشتیبانی

استاد مربوطه: جناب آقای امیر شکری

دانشجو: فاطمه میرزاده

آذر ۱۴۰۰



چکیده



اشکال مختلف کلاهبرداری را که بانک‌ها در معرض آنها قرار می‌گیرند و ابزارهای داده‌کاوی در معرض آن قرار می‌گیرند، تجزیه و تحلیل می‌کنیم که به داده‌های تشخیص زودهنگام آن که قبلاً در یک بانک انباشته شده است، منجر می‌شود.

از روش‌های یادگیری نظارت‌شده از ماشین‌های بردار پشتیبانی با Spark (SVM-S) برای ساختن مدل‌هایی استفاده می‌کنیم که رفتار عادی و غیرعادی مشتری را نشان می‌دهند و سپس از آن برای ارزیابی اعتبار تراکنش‌های جدید استفاده می‌کنیم

نتایج آزمایش از این مطالعه نشان می‌دهد که :



SVM-S عملکرد پیش‌بینی بهتری نسبت به شبکه‌های انتشار برگشتی (BPN) دارد.

مقدمه

سازمان های خدمات مالی دارای تعدادی اهداف استراتژیک از جمله جذب و حفظ مشتریان جدید و فعلی از طریق بکارگیری روش های مدیریتی مختلف هستند.

بانک ها به عنوان یک مؤسسه مالی، حجم عظیمی از داده ها را از منابع مختلف از مشتریان خود تولید می کنند و این به نیاز به داده های بزرگ کمک کرده است.

ماشین بردار پشتیبانی می تواند به کاهش ریسک و بهبود کیفیت خدمات ارائه شده به مشتریان برای موفقیت در تجارت کمک کند.

مبارزه با کلاهبرداری با تکنیک های پیشگیری سنتی مانند پین، رمز عبور و سیستم های شناسایی همراه است، اما در سیستم های بانکی مدرن ناکافی شده اند.

•

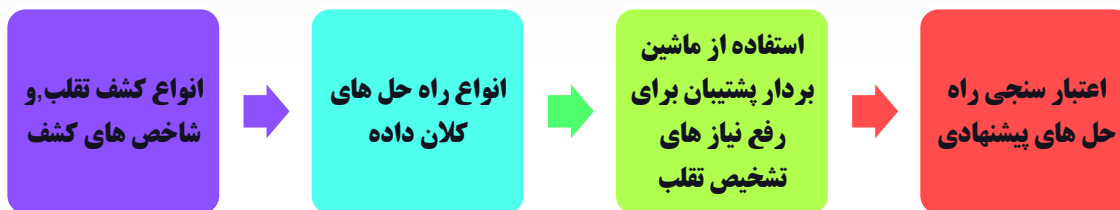
•

•

•



فهرست عناوین :



تکنیک های داده کاوی

برنامه های کاربردی داده های بزرگ با تکنیک های داده کاوی می توانند نقش مهمی در مبارزه با این نوع تقلب ها داشته باشند.

چندین تکنیک برای کشف تقلب توسط بسیاری از محققین پیشنهاد شده و مورد استفاده قرار گرفته است، از جمله تقلب در کارت های اعتباری.

از جمله این تکنیک های داده کاوی، شبکه های بیزی، زنجیره های مارکوف، شبکه های عصبی، رگرسیون خطی، هم ترازی توالی و غیره هستند.

اشکال کلاهبرداری بانکی و شاخص های آنها

پولشویی

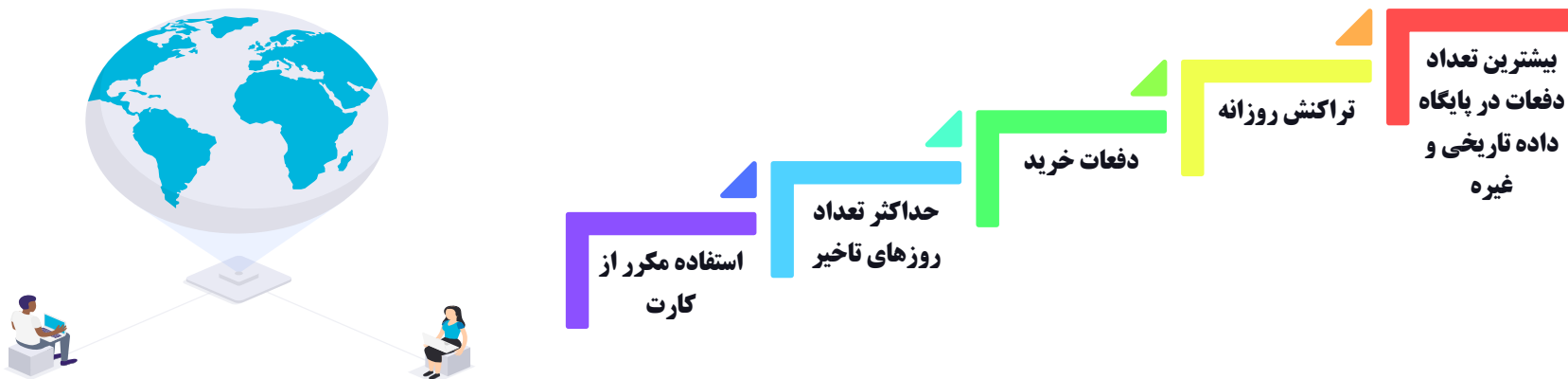
یکی از روش های شناخته شده کلاهبرداری است
مبارزه با پولشویی در صنعت مالی مبتنی بر تجزیه و تحلیل و پردازش اظهارات مربوط
به تراکنش های مشکوک کشف شده توسط موسسات مالی است



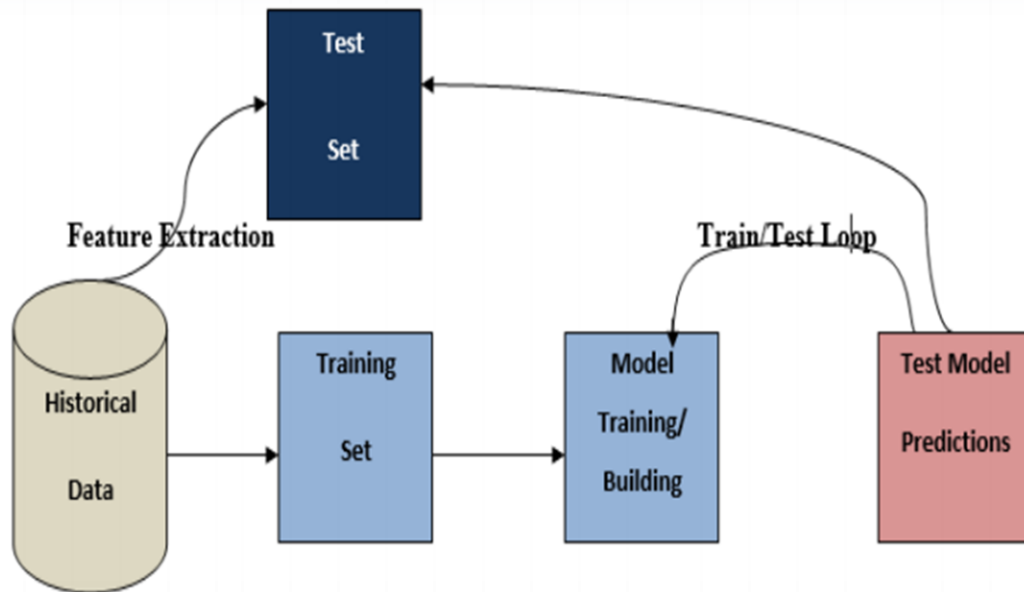
تقلب بر اساس کارت های اعتباری

- کلاهبرداری مبتنی بر کارت اعتباری همچنان در حال افزایش است
- بانک ها و شرکت های مالی سالانه مبالغ هنگفتی را از طریق کلاهبرداری با استفاده از کارت اعتباری از دست می دهند.

شاخص هایی مانند:



مدل تشخیص تقلب



داده های بزرگ برای کشف کلاهبرداری بانکی

به دلیل تغییر سریع و توسعه تکنیک های مورد استفاده توسط کلاهبرداران، ابزارهای داده کاوی دیگر نمی توانند رفتارهای غیرعادی را تجزیه و تحلیل کنند.

داده های بزرگ، در این زمینه، با تکنیک های یادگیری ماشینی برای تشخیص تقلب در پایگاه داده ارائه می شوند که بهترین کار برای مبارزه با کلاهبرداری بانکی است.

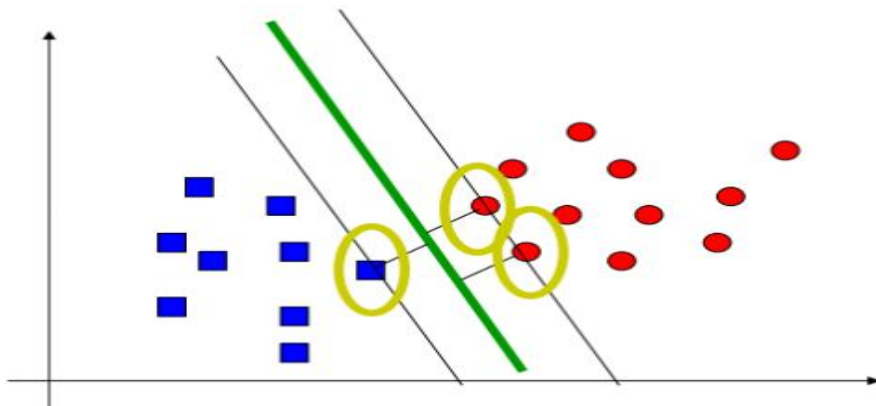
از دو شکل یادگیری خودکار استفاده می شود:

- ▶ نظارت شده
- ▶ بدون نظارت



ماشین بردار پشتیبانی (SVM)

- یک الگوریتم یادگیری ماشینی نظارت شده است که می تواند برای چالش های طبقه بندی و رگرسیون استفاده شود.
- هر آیتم داده را به عنوان یک نقطه در یک فضای بعدی رسم می کنیم



مورد خطی قابل تفکیک

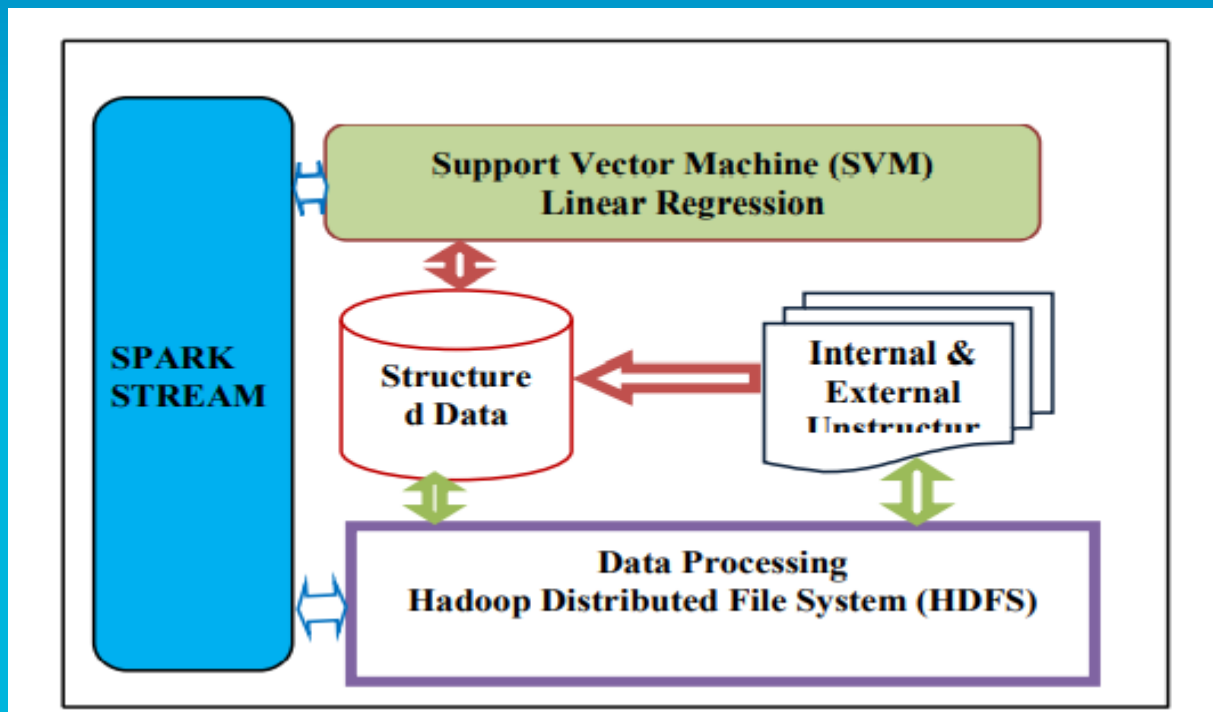
- ▶ در حالت قابل جداسازی خطی، یک یا چند ابرصفحه وجود دارد که ممکن است دو کلاس ارائه شده توسط داده های آموزشی را با ۱۰۰٪ از هم جدا کنند

مورد غیر خطی قابل تفکیک

- ▶ در حالت غیرخطی قابل تفکیک، نمی توان یک ابر صفحه خطی پیدا کرد که تمام مثال های مثبت و منفی را از هم جدا کند.



معماری SVM-S برای تشخیص تقلب در داده های بزرگ



Back Propagation Network

- ▶ استفاده از BPN برای آموزش داده ها نیاز به تنظیم برخی پارامترها دارد. از جمله مهمترین پارامترها می توان به اعداد لایه پنهان N_i ، گره پنهان N_j و دوره های آموزشی N_k ، نرخ یادگیری R_j و نرخ تکانه R_m اشاره کرد.

تست و نتیجه

- داده های مورد استفاده:

پایگاه های داده آلمانی و استرالیایی کارت های اعتباری
مجموعه داده با استفاده از SVM طبقه بندی شد و با رگرسیون خطی و رگرسیون منطقی برای تشخیص
ناهنجاری با استفاده از ویژگی های کارت اعتباری آموزش دید.

Data Set	Classifier Model	Parameters	Definition	Source
1 st Data Set	SVM	C=10 RBF $\epsilon = 0.1$		Djeffal et al. 2014
2 nd Data Set	BPN	C=10 RBF $\epsilon = 0.1$		Soltani et al. 2014

- پارامترهای تنظیم شده برای مجموعه داده

اندازه مجموعه آموزش و آزمون برای نمونه ها

Samples		# of Records	
		Training Set	Test Set
Set -1F-To-1N (SVM-S)	Normal	100	86
	Fraud	25	20
Set -1F-To-1N (BPN)	Normal	150	125
	Fraud	25	12

عملکرد معماری SVM-S

Model/ Data Set	Set-1F-To-1N				Set-1F-To-4N			
	Tra in	Tes t	Bu ild Ti me	Fra ud	Trai n	Tes t	Bu ild Ti me	Fra ud
D1: SVM	98.7 8%	84.3 7%	<2 0m	20	96.3 4%	82.5 4%	<2 0	20
D2: BPN	99.8 6%	85.5 8%	<2 5m	13	97.3 46%	84.6 7%	<2 5	10

نتیجه گیری

دو مورد کلاهبرداری در بانک‌ها را بررسی کردیم:

- کلاهبرداری از کارت اعتباری
- پولشویی.



عملکرد سیستم پیشنهادی بر روی معیارهای دفتر کل، داده های پرداختنی، که شبیه به پایگاه داده بانک ایجاد شده است، آزمایش شد. دقت به دست آمده برای روش SVM تک کلاس حدود ۸۰ درصد بود که نشان دهنده پیشرفت قابل توجهی در مقایسه با مرجع آثار مشابه است. برای این روش، بهبود جزئی در پایگاه های داده امتیازدهی اعتباری به دلیل دشواری دستیابی به پایگاه های داده واقعی بود. نتایج را می توان با مطالعه تأثیر پارامترهای مختلف مورد استفاده در معماری SVM-S بهبود بخشید.

THANKS!

ANY QUESTIONS?

You can find me at:

▶ user@mail.me :

fatemehmirzadeh98@gmail.com

