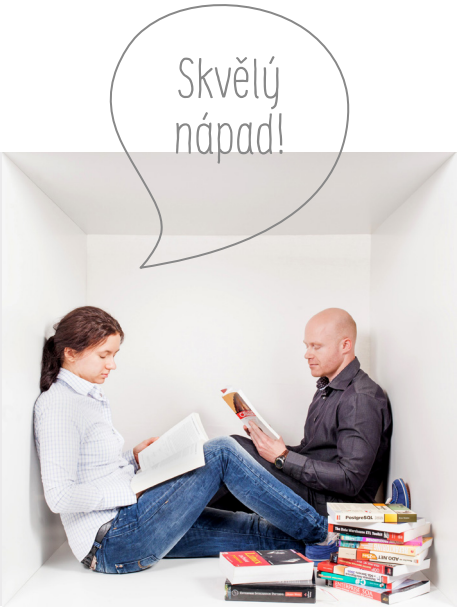
GEM SYSTEM

Odhalte podvody nejen v pojišťovně

Fraud Detection



Často slýcháme o detekci podvodů ve všech oblastech finančních služeb. Výrazy jako prediktivní modely, pokročilé statistické analýzy, sofistikované dataminingové algoritmy nalezneme v každé hodnotné marketingové prezentaci fraud detection systému. Jde o zázračné black boxy, které se samy naučí, jak takový podvod vypadá a pak s elegancí Sherlocka Holmese vypátrají a označí hříšníky v téměř reálnem čase.

PRODUKTY A IT ŘEŠENÍ

Běžná praxe

Pravděpodobně se mnohým z Vás stalo, že jste se pokoušeli koupit zájezd pro celou rodinu a zaplatit ho platební kartou online. Za pár minut volá vyděšený pracovník call centra vaší banky a zeptá se, zda jste to skutečně vy, kdo se pokouší zaplatit tuto horentní částku. Vy ho ujistíte, že ano a pak spokojeně usínáte s dobrým pocitem, že vaše peníze jsou v bezpečí a někdo tam daleko na ně dohlíží. Jde o zázrak nebo je to skutečné tak snadné?

POSKYTNUTÉ SLUŽBY

Z pohledu banky nebo pojišťovny si musíme uvědomit, že Fraud Detection systém není nic jiného, než chytrý vyhledávač anomálií nebo shod se známými vzory, který by měl k této činnosti využít všechny dostupné informace. V praxi neexistují řešení, která vybalíme z krabice, a začnou samy pracovat. V ideálním případě je nutné využít všech stávajících vědomostí o vašem byznysu a zkombinovat je s efektivním IT produktem pro vytvoření komplexního a hlavně užitečného fraud detection systému.

Vyhledávání známých vzorů podvodů je možné nad daty, u kterých jsme na dostatečně velkém vzorku schopni jasně odlišit podvody od korektních případů. Potom je už musíme „jen" prohnat chytrými algoritmy strojového učení a po nezbytném čase ladění získáme klasifikátor (systém, který s určitou pravděpodobností rozeznává podvody, tj. identifikuje hledané entity). Problém je, že tento přístup není možné použít pro všechny typy dat.

V mnoha oblastech nemáme k dispozici předem jasně rozlišená korektní a podvodná data, na kterých by se systém mohl učit, a proto se musíme více spoléhat na starou dobrou statistiku. Když se člověk pokouší identifikovat podvod, nebo podezřelé entity mezi několika příklady, intuitivně se soustřeďuje na ty vybočující od průměru. Situaci komplikuje jen fakt, že informací o entitě je obvykle velké množství a lidská mysl není schopná tyto informace pojmout a komplexně zhodnotit. Tento přirozený způsob je možné pojmout algoritmicky a při chytře nadefinovaném datovém modelu získáme ohromné možnosti, jak naše data využít a sofistikovaně vyhledávat ty anomálie, které jsou objektem našeho zájmu.

Příklad místo řečí

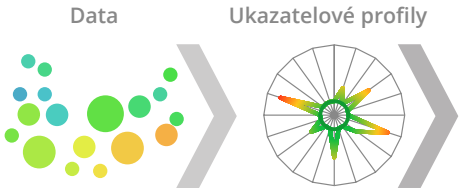
Řekněme, že naším cílem je výběr pěti poboček ze sta, na které se má zaměřit kontrola. Logickým krokem je v tomto případě seřadit všechny pobočky dle stupně „podezřelosti" a zaměřit naší pozornost na pět nejhorších. Zní to jednoduše, ale jak správně tyto pobočky ohodnotit a následně seřadit?

GEM System a. s. | Na Pankráci 1062/58, 140 00 Praha 4 | sales@gemsystem.cz | www.gemsystem.cz

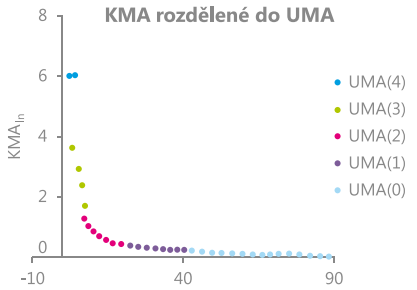
Výsledek

Náš systém jsme vyvíjeli, implementovali a testovali pro hodnocení poskytovatelů zdravotní péče z pohledu zdravotní pojišťovny. Cílem bylo odhalit úmyslné nadměrné vykazování, poskytování nadbytečné péče a různé druhy chyb. Podařilo se nám vydefinovat sadu relevantních ukazatelů vzhledem k způsobu vykazování zdravotní péče v České republice a na kontrolních datech ověřit funkčnost metody.

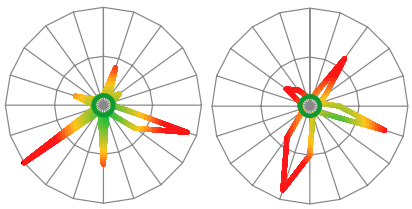
Další výhodou je grafické zobrazování profilů poskytovatelů péče jako velice intuitivní a bohatý zdroj informací, který je možné využít například i při uzavíraní a prodlužování smluv.



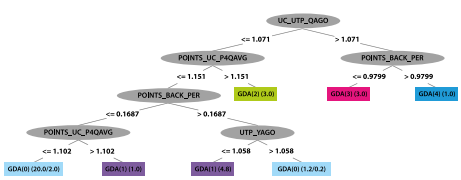
Ohodnocené entity



Profily anomálie



Segmentace



Je horší, pokud pobočka A má o 30 % více zrušených účtů než průměr, pobočka B má o 30 % více nesplacených úvěrů než průměr nebo že pobočka C má o 15 % vyšší oba ukazatele? Evidentně to závisí ještě na dalších aspektech.

A jak se vypořádat s realitou, když těchto ukazatelů je několik desítek až stovek, jsou navzájem propojené a k tomu v čase proměnné? Při hledání odpovědi na tuto otázku jsme se inspirovali výzkumem, provedeném na Jihokorejských univerzitách a testovaném na datech jihokorejské Health Insurance Review Agency.

GEM Fraud Detection

Na základě této studie jsme v GEM System navrhli a otestovali komplexní hodnotící systém, který je schopný dle definovaných ukazatelů (jednotkové náklady, počet zrušených účtů na zaměstnance, atd.) ohodnotit jakékoliv entity (pobočky, klienty, pojištěnce, poskytovatele péče, atd.) dle míry odlišnosti od ostatních a vysvětlit, které ukazatele primárně přispěly do vypočteného hodnocení. Samozřejmostí je porovnávání jednotlivých ukazatelů v čase, protože abnormální růst je často dobrým vodícím znakem pro nekorektní chování.Entity jsou pak seřazeny dle dosažených výsledků a rozděleny do skupin dle míry anomálie. Grafické zobrazení profilů entit napomáhá identifikovat příčiny vysokého skóre, takže kontrolní nebo revizní pracovník hned porozumí hodnocení dané entity.

Popis hodnotící metody

Základem je definice ukazatelů, pro které platí následující pravidlo: čím je ukazatel vyšší, tím spíše se jedná o podezřelé chování. Pro jednotlivé ukazatele se pak v rámci relevantní skupiny spočítají průměry a nadprůměrné hodnoty se statisticky vyhodnotí vzhledem k variabilitě, přičemž hodnoty dále od průměru mají větší vliv a tudíž je jejich míra anomálie vyšší. Výsledným produktem je vážený součet individuálních měr anomálií, které vytváří výslednou kompositní míru anomálie (KMA). Vzhledem k faktu, že ne všechny ukazatele mají stejný vliv na podezřelé chování, je dalším důležitým aspektem relativní váha ukazatelů. Ta se v první fázi stanovuje manuálně na základě expertních odhadů a postupem času je automaticky přizpůsobována dle zpětné vazby od uživatelů. Entity jsou dále pomocí clusteringu rozděleny do pěti skupin dle logaritmu úrovně míry anomálie (UMA). Skupina 4 pak obsahuje nejpodezřelejší entity, na které by se měla zaměřit kontrola. V praxi to není ještě konec. V mnohých případech nás také zajímá jemnější segmentace v rámci skupin dle hodnot ukazatelů. Proto je v poslední fázi vytvořen rozhodovací strom, který entity zařadí do skupin, ne dle měr anomálie, ale dle původních hodnot ukazatelů. To vytvoří segmenty entit s podobným chováním a kontrola pak může být lépe cílená.

GEM System a. s. | Na Pankráci 1062/58, 140 00 Praha 4 | sales@gemsystem.cz | www.gemsystem.cz