



ceos
data

Studie | Retenční model

Predikce odchodu zákazníka ke konkurenci

Udržení stávajících zákazníků primárně v loyality programech je jeden z nejdůležitějších pilířů efektivního podnikání. Nejen v e-commerce platí, že je jednodušší a levnější si udržet stávajícího zákazníka než řadou investic do marketingových aktivit získávat nového. Cílem predikce je zabránit odchodu zákazníků nebo jej co nejvíce omezit a následně jim předložit vhodnou nabídku.

www.ceosdata.com

Manažerské shrnutí

Při návrhu a výrobě retenčního modelu používáme **kombinaci 2 metod** - každá z nich přináší jiný benefit:

1. První metodou je použití **Machine Learning** (dále ML) vycházející z segmentace klientů, tj jejich zařazení do reprezentativních skupin. Model je dále obohacen o různorodé atributy, které ovlivňují jeho přesnost.
2. Druhá metoda je založena na využití tzv. **intervalu spolehlivosti** (dále IS) a narušitel od metody první zkoumá klienta jako suverénní identitu. Tento model je vhodný pro zákazníky s vyšší frekvencí nákupů a umožňuje včasější detekci změny chování.

Díky kombinaci obou metod dokážeme

- » včasně reagovat i na rychlé změny chování (nečekáme pevně dané období 3 měsíců)
- » zároveň využít celou šířku ML modelování pro predikci odchodu v situacích, kde odchod není na první pohled jednoduše detekovatelný běžným pohledem na data

Struktura retenčního modelu CEOS Data

Popis řešení, použité metody a vstupy se zabývá výhodami použitých modelů a parametrů.

Časování projektu představuje časový rámec a termíny dodání funkčního modelu.

Popis implementace modelu a doporučený postup se zabývá vysvětlením jednotlivých fází projektu, které je nutné provést pro doručení modelu a zajištění dodávky jednotlivých fází.

Proces zdokonalování a rozšiřování modelu: možná rozšíření a příklady atributů pro budoucí využití; definice testu pro verifikaci správnosti modelu.

Předpokládaná a garantovaná úspěšnost modelu nabízí

- » vysvětlení úspěšnosti (95%) dosahované naším modelem
- » informace o důležitosti prahových hodnot pro stanovení úrovně „paranoie“ modelu

Požadavky na data, zdroje a součinnost uvádí nutné podmínky pro úspěch projektu.

Popis řešení, použité metody, vstupy

1. Shrnutí

Na základě našich zkušeností doporučujeme realizaci níže uvedených úloh:

- » identifikace odcházejícího zákazníka
- » identifikace zákazníka, který podezřele mění své chování oproti minulosti

Obě úlohy jsou řešitelné podobnými postupy, ale zásadně se liší v definici odezvy. Definice odezvy je přitom klíčová jak z hlediska business využití, tak i způsobu vyhodnocení přesnosti.

V mnoha případech jsme se setkali s požadavkem na pevné x-měsíční časové okno, ve kterém se má testovat, zda klient uskuteční nákup. **Použití takto definovaného pevného okna, podle našeho názoru, není prakticky vhodné.**

Pevné x-měsíčního časové okno je dobře použitelné u klientů, kteří pravidelně nakupují s cca měsíční frekvencí.

Pro klienty, kteří nakupují pravidelně s frekvencí nižší, např. týdně, je však odhadování a následné testování až po uplynutí x-měsíčního lhůty okna příliš statické. Pravděpodobnost, že klient mezitím přešel ke konkurenci, je zde už vysoká. Následný pokus o jejich zpětné získání znamená vyšší náklady, nebo již není možný.

V tomto případě je vhodné včasější varování, a to na bázi týdnů. Ideálním řešením pak je stanovit ad-hoc časové okno (tj. „šité na míru“) každému klientovi zvlášť.

Použití ad-hoc okna bude vykazovat jistou míru ‚paranoie‘ u klienta, který nakupuje standardně každý týden - bude totiž varovat již při několikátýdenní nečinnosti klienta.

Z důvodu eliminace vysokého počtu planých varování navrhujeme počítat v rámci ad-hoc okna 2 různé varovací ukazatele, a to navzájem nezávislými způsoby:

- » Machine Learning model (dále ML)
- » použitím intervalů spolehlivosti (dále IS)

Kombinací obou ukazatelů dosáhneme spolehlivějších výsledků v různorodější škále časových oken.

V případě, že ze strany zákazníka bude kladen důraz na použití pevného x-měsíčního okna, dokážeme tomuto požadavku vyhovět. ML model je poměrně snadno modifikovatelný pro různé typy odezvy.

2. Podrobnosti o modelech



V řešení aplikujeme dva odlišné přístupy (resp. modely), které se navzájem doplňují.

Přístupy (modely)

- » machine learning [ML]
- » intervaly spolehlivosti [IS]

Výstupy

- » popis současného a budoucího stavu klienta
- » návrhy budoucích rozšíření (např segmentace, retenční faktory)



Hlavní myšlenky kapitoly

- » predikci změny chování klientů řešíme dvěma odlišnými metodami, které se dobře doplňují
- » IS model vychází z nákupní historie každého klienta; určuje pravděpodobnost, že klient k určitému datu nakoupí
- » ML model vychází z agregovaných dat; určuje pravděpodobnost, že klient v jeho specifickém časovém okně nenakoupí
- » v budoucnosti navrhujeme vyzkoušet modely zvlášť na různých segmentech

Machine Learning model

Machine Learning model je běžně používaným analytickým postupem pro identifikaci odcházejících zákazníků. Vstupem jsou agregované ekonomické a socio-demografické údaje vztahované ke klientovi, výstupem pak pravděpodobnost odezvy (odchodu nebo změny chování) jednotlivce.

Odezva v ML modelu může být nastavena několika způsoby, primárně upřednostňujeme postup, že nastavíme období přizpůsobené každému zákazníkovi na míru.

Odezva má pak charakter pravděpodobnosti, že klient nenakoupí v jeho na míru šitém časovém oknu.

Poznámka

Model může být pro testování nějaké hypotézy použit i s následující alternativní odezvou, na základě výše zmíněných důvodů o včasnosti varování to však nepovažujeme za vhodné standardní nastavení.

3. Podrobnosti o modelech

Princip modelu

Nejprve určíme odezvu - zda klient nakoupil v jeho typickém období.

Po definování odezvy přistoupíme k rozdělení dat na tréninkový a testovací set. Na tréninkové sadě matematický model naučíme a jeho úspěšnost pak hodnotíme pomocí druhé sady, kterou model „neviděl“. Pokud je model dostatečně kvalitní, můžeme ho využít pro predikci.

Pro každého klienta tak získáme vyčíslenou pravděpodobnost, že nenakoupí v jeho typickém období. V závislosti na definici odezvy můžeme určit buď pravděpodobnost změny chování nebo odchod zákazníka. Změna chování je ověřitelná spolehlivěji a přesněji, odchod může být z principu ověřen maximálně do (časového) konce datové sady, která je k dispozici.

Možné segmentace

Na počátku sestavíme model pro všechny klienty najednou. Postupem času však předpokládáme rozčlenění klientů do vhodných segmentů a použití speciálního modelu pro každý segment.

Tento postup zpřesní predikční schopnosti modelu. Je totiž pravděpodobné, že každý segment je ovlivněn různými proměnnými predikujícími odezvu. Například klesající trend v obratech kosmetiky u mladých žen může být předzvěstí jejich odchodu.

Návrhy segmentů, na kterých může být model postaven:

- » žádná segmentace (v první fázi jen základní ověření a nastavení modelu)
- » segmentace na bázi nákupní frekvence (každodenní nakupující, měsíční, kvartální)
- » segmentace na bázi již definovaných typů klientů (například již vytvořená segmentace zákazníka)
- » ad-hoc segmentace klientů vhodná pro model (možná do budoucna)

Faktory

Z modelu jsou zároveň patrné faktory a jejich statistická významnost, tj. jak ovlivňují chování zákazníka (v rámci segmentu). Model může být snadno obohacen o novou proměnnou, na které půjde ověřit, zda má vliv na chování zákazníka. K ověření dojde buď prostřednictvím p-hodnoty, nebo srovnáním přesnosti referenčního modelu bez dané proměnné.

V budoucnu je možné rozšířit výstup o vliv proměnných na konkrétního člověka.

IS model

Tento model identifikuje změnu chování či odchod klienta na základě jeho vlastní nákupní historie.

Vstupem jsou data nákupů klienta a jejich obraty. V toku času se mu od posledního nákupu postupně zvyšuje pravděpodobnost dalšího nákupu (plní se mu pomyslný progress bar k dalšímu nákupu), a pokud v mezním čase nenakoupí (překračuje povolenou toleranci), konstatujeme, že u něj došlo ke změně chování.

Odezva v IS modelu má pak charakter míry pravděpodobnosti, že k danému datu klient uskuteční nákup.

4. Vstupy, výstupy, interpretace



Aplikací dříve popsaných modelů získáme výstupy, s nimiž lze dále pracovat.

Struktura vstupních dat

- » agregace na zákazníka [ML model]
- » agregace na zákazníka a nákup [IS model]

Formát hlavního výstupu

- » ID | P_ml | P_is



Hlavní myšlenky kapitoly

- » modely ML a IS budou současně aplikovány na reálných datech, jejich výstupem budou pravděpodobnosti, že:
 - » klient v budoucnosti usne [ML model]
 - » klient v dané chvíli usíná [IS model]

Pravděpodobnosti

Přestože mají oba modely stejný cíl - odhalit změnu chování - pokouší se ji určit principiálně jiným způsobem.

» **Vstupem pro ML model** jsou data agregovaná na klienta s dobře interpretovatelnými prediktory.

Tato data jsou následně znormována délkou aktivity zákazníka, pro použití modelu na ostrých datech. Výstupem bude model, který po použití na ostrých datech odhadne pravděpodobnost odezvy (klient je spáč).

» **Vstupem pro IS model** jsou data agregovaná na klienta a expedici, čímž dostáváme přesnější informace o dynamice jeho nákupů. Výstupem bude odhad pravděpodobnosti, že člověk v dané chvíli usíná.

Přestože modely popisují odlišná fakta (pravděpodobnost změny chování vs. pravděpodobnost nákupu v daném okamžiku), oba by měly dosahovat podobných výsledků.

Zákazník si může zvolit k čemu přihlédne v souladu se strategií oživení klienta, případně vzít zprůměrovanou hodnotu z obou přístupů.

Časování projektu



Očekávaný průběh projektu

- » start projektu
- » validace na straně zákazníka: 3 - 4 měsíce
- » očekávané dokončení a produkční provoz: 5 - 7 měsíců

Fáze implementace - odhad časování

- » konektory, příprava dat pro modely: 3 měsíce
- » tvorba odezvy, příprava proměnných: 1 měsíc
- » vznik a validace samotných modelů: 1 měsíc
- » kroky mohou proběhnout iterativně
- » automatizace předchozích fází: 1 měsíc
- » nasazení do produkce 1 měsíc

Proces zdokonalování a rozšiřování modelu

Výše zmíněné odhady platí pro první základní implementaci a nasazení.

Celý proces vývoje a používání modelu je však kontinuální. Další ladění a vylepšování modelu bude probíhat dlouhodobě v navazujících projektech.

Případné rozšíření o nové proměnné, popř. aplikace segmentace klientů na bázi frekvence nákupu znamená řádově snazší úlohu.

Vznik případné ad-hoc segmentace pro účely retence znamená časově náročnější zásah.

Popis návrhu implementace modelu, doporučení k jeho praktickému využití



Proces implementace modelů provedeme v následujících fázích:

1. příprava dat
2. tvorba odezvy a příprava vhodných proměnných
3. vytvoření modelů a validace
4. automatizace tvorby dat
5. aplikace modelů a export pravděpodobností

Praktické využití výstupů a předpokládaná rozšíření do budoucna.

Hlavní informace poskytované modely:

- » klienti, kteří již spí, nebo změnili chování
- » klienti se blíží k časovému bodu, kdy by měli nejpozději nakoupit
- » klienti, kteří ještě nespí, ale brzy usnou

Návrhy dalšího rozvoje základního modelu (předpokládané rozšiřující prediktory v budoucnu)



Hlavní myšlenky kapitoly

- » implementace modelů do ostrého provozu proběhne v celkem 5 fázích
- » po základní implementaci možno uvažovat o rozšíření/zdokonalení modelu
- » jak použít výstupy z modelů
- » návrhy další práce s výsledky (segmentace zákazníků, ekonomická evaluace klientů, testování hypotéz atd.)

1. Fáze: příprava dat pro model

Cíl	získat historická data o jednotlivých zákaznících a jejich nákupech
Kroky	získání dat pro ML a IS model
Výstupy	základní datová sada pro analýzu

2. Fáze: tvorba odezvy a příprava proměnných

Cíl	mít k dispozici základní data vhodně upravená, obohacená a transformovaná pro potřeby a přesnost modelů
Kroky	<ul style="list-style-type: none">» volba nejvhodnější predikční proměnné» obohacení vstupních dat o odezvu» normování proměnných rostoucích v čase
Výstupy	2 obohacené datové soubory použitelné pro následné modely

3. Fáze: vytvoření modelů a validace

Cíl	vytvořit modely, které pro jednotlivé zákazníky určí, zda v minulosti změnili chování, popř. dokážou změnu predikovat do budoucnosti
Kroky	<ul style="list-style-type: none">» vytvoření, trénování a validování ML modelu, dokud nedosáhne dostatečné použitelnosti» vytvoření IS modelu pro včasnou detekci spáčů
Výstupy	<ul style="list-style-type: none">» „natrénovaný“ ML model včetně vah prediktorů» IS model

4. Fáze: automatizace tvorby dat

Cíl	<ul style="list-style-type: none">» připravit aktuální datový soubor, na kterém probíhá následná analýza a predikce;» data, která byla použita pro tvorbu modelu, budou generována automaticky pravidelně, (např. 1x měsíčně)
Kroky	vytvoření automatizovaných procedur tvořících/upravujících data z fází 1 a 2 na pravidelné bázi
Výstupy	ostrá data, která vstupují do modelů

5. Fáze: aplikace dříve vypočteného modelu

Cíl	<ul style="list-style-type: none">» zajistit automatické přepočítávání výsledných hodnot (pravděpodobnosti) na aktuálních datech» zajistit snadnou dostupnost výsledných dat uživatelům
Kroky	<ul style="list-style-type: none">» vytvoření automatizovaných procedur predikujících na ostrých datech z fáze 4 výsledné hodnoty (pravděpodobnosti), a to ke každému klientovi» aplikace modelů vytvořených ve fázi 3» 2 pravděpodobnosti ke každému klientovi;
Výstupy	<ul style="list-style-type: none">» možné doplnění o ekonomické ukazatele jednotlivého klienta» seznam vlivných prediktorů a ke každému klientovi» odhad míry odlišnosti chování klienta od průměru;

Proces zdokonalování a rozšiřování modelu

Zdokonalování a rozšiřování modelu v následujících fázích:

- » Fáze 1-3 jsou prováděny opakovaně, tj. při vzniku a každém rozšíření modelu, a to až do doby, než je model použitelný a dostatečně úspěšný.
- » Fáze 4 je jednorázová a probíhá po každém rozšíření modelu.
- » Fáze 5 je pravidelně opakovaná v produkčním provozu.

Možná rozšíření modelu vs. fáze jeho implementace

Pokud nastane potřeba obohatit model o další proměnnou, popř. nový segment, obohacený model musí projít všemi 5 implementačními fázemi.

Doporučení a praktické využití

V této chvíli máme ke každému klientovi určeny dvě pravděpodobnosti:

1. vypočtená **ML modelem**: určuje, zda klient změní své chování v budoucnosti.
2. získaná z **IS modelu**: na základě chování klienta v minulosti určuje, zda klient v daném okamžiku uskuteční nákup popř. se přepíná do režimu spánku.

To umožňuje zaměřit pozornost na chování klientů různého typu:

- » klienti, kteří podle modelu **již spí, nebo změnili své chování** (IS model).
- » klienti se blíží naplnění okamžiku, **kdy by měli nejpozději nakoupit** (IS model).
- » klienti, kteří nespí, ale **je jim predikováno, že se tak stane** (ML model)

Doporučení pro další práci s výsledky

Je možné testovat hypotézy pomocí prediktorů, nastavovat karanténu na vzorku 10tis. klientů a cíleně - tj. mnohem přesněji - na nich ověřovat účinnost zvoleného způsobu „oživení ze spánku“. Dalším krokem lze hodnotit ekonomickou sílu klientů (ať už jen obrátově nebo vypočítat CLA, ad-hoc segmentaci klientů vhodnou pro model, sezónnost, cenu a vhodnost retence.

Předpokládaná a garantovaná úspěšnost modelu

Náš odhad úspěšnosti vychází z předběžné analýzy na vzorku dat, na kterém jsme se pokusili validovat oba výše zmíněné přístupy (resp. modely). Validací modelu rozumíme vyhodnocení, zda vytvořený matematický model prokazuje uspokojivou míru shody s reálnými daty.



Souhrn úspěšnosti:

- » dosažený záchyt až 95 % ohrožených klientů
- » úspěšnost záchytu je řízena volitelným parametrem
- » vysvětlení, proč není vhodné cílit pouze na nejvyšší míru záchytu - vzhledem k ceně retence
- » ukázka validace ML i IS modelu

Validace modelu ML

Validace tohoto modelu je přímočará. Známe predikované i skutečné hodnoty a lze tedy jednoznačně změřit přesnost.

Samotná přesnost je dle našeho názoru však jen orientačním kritériem pro posouzení kvality modelu. Vzhledem k tomu, že nepredikujeme jasný odchod, ale změny chování, snažíme se primárně vyhledávat zákazníky, u nichž dochází k snížení frekvence nákupů - tzv. „usnutí“. Model se tedy snaží vyhledávat tyto spáče.

Namísto přesnosti jsou v našem případě prakticky mnohem lépe využitelnější kritéria senzitivity (podíl správně predikovaných spáčů vůči všem skutečným spáčům) a specifity (podíl správně predikovaných nespáčů ze všech skutečných nespáčů).

V rámci experimentální analýzy, kterou jsme provedli na vzorku dat, **vykazoval ML model senzitivitu až 95%, čili zachytil 95% všech spáčů** - za předpokladu vhodně zvoleného dělicího prahu.

Trénovací vzorek obsahoval 10tis. náhodně vybraných klientů nakupujících v období 5 měsíců. Validace byla provedena na datech prodloužených o další rok.

Validace modelu IS

Validace tohoto modelu je komplikovaná. Vzhledem k tomu, že zde nemůžeme použít jasně ověřitelnou odezvu, můžeme dosáhnout různé přesnosti podle toho, jak odezvu definujeme.

Definice požadavků na data



Nevznášíme žádné zvláštní požadavky na přípravu dat ze strany zákazníka.

Preferovaným způsobem je přímý přístup do DB za pomoci některého ze standardních konektorů. Umožnění přímého přístupu znamená nejen usnadnění naší práce, ale i úsporu času a nákladů během projektu.

Požadavky na zdroje a součinnost



Pro zdárný proces vývoje a implementace navrženého řešení požadujeme aktivní součinnost ze strany zákazníka. **Preferujeme dedikovanou osobu, se kterou budeme moci ladit jak očekávání zadavatele, tak průběžné i konečné výsledky.**

Pro představu jsou vhodné konzultace k výběru predikčních proměnných, segmentaci klientů, formě a dostupnosti výsledků projektu (BI, APEX aplikace, WS), atp.