



Credit scoring

Libor Vajbar
Analytik řízení rizik

Home Credit a.s. – přední poskytovatel spotřebitelského financování

Úvěrové produkty

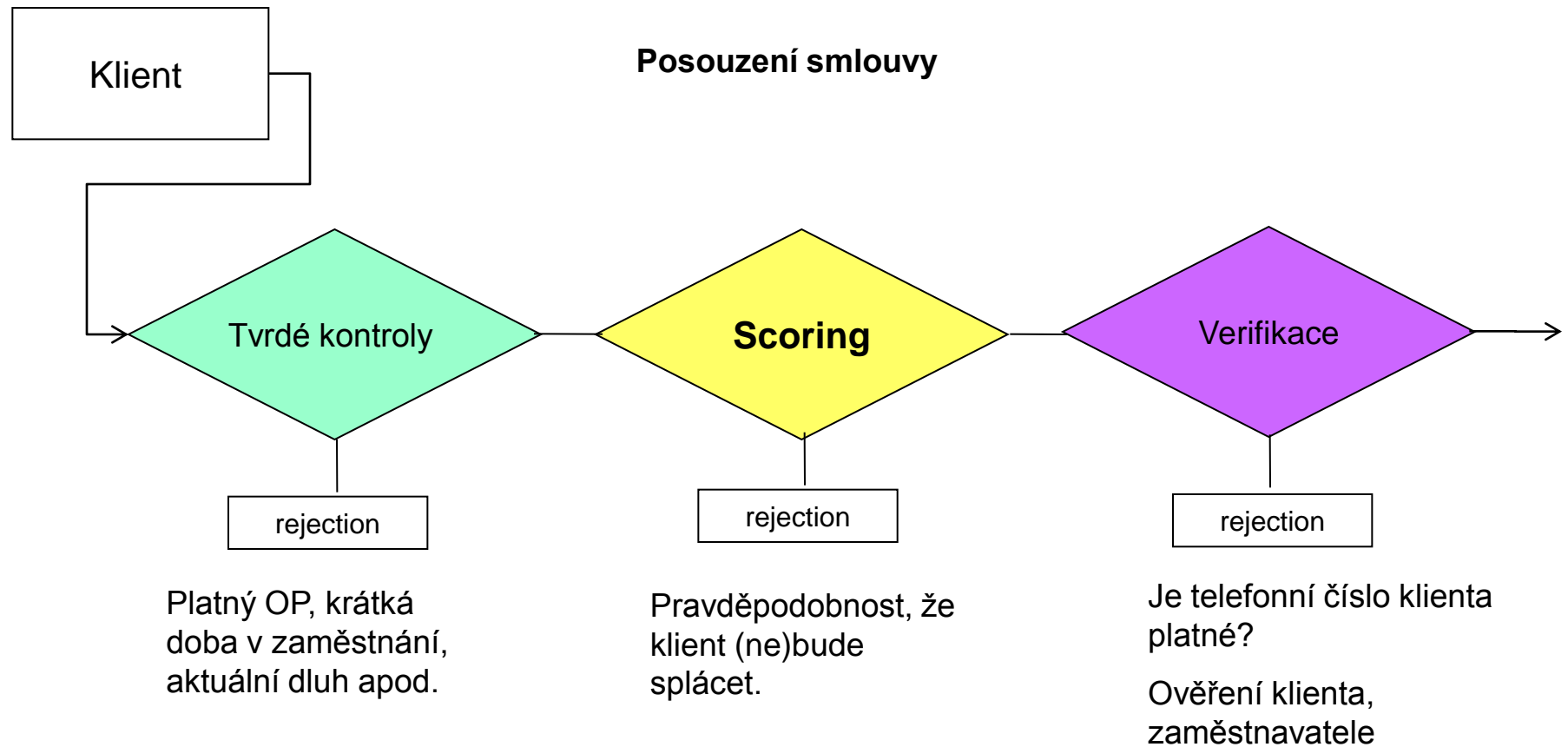
- nákup na splátky u obchodních partnerů
- kreditní karty
- hotovostní půjčky
- úvěry na auta

Obchodní model

- získání **nových klientů** přes akviziční produkty prostřednictvím obchodních partnerů – nejčastěji spotřebitelský úvěr
- následné oslovení vhodných klientů (již existuje historie splácení a více informací o klientovi – menší riziko) nabídkou na kreditní kartou/hotovostní úvěr

HC není banka

- nezná příjem klienta
- neví, jak dlouho chodí výplata od stávajícího zaměstnavatele
- neví, jaké částky odchází z účtu (měsíční náklady klienta)



Cíl – schvalovat bonitní klienty, zamezit předlužení, zamezovat podvodům (zároveň udržet vysoké procento schválených klientů)

Nástroj pro **hodnocení bonity klienta** při posouzení úvěrové smlouvy

Score je číselný ukazatel, který se používá k predikci pravděpodobnosti nějaké budoucí události (defaultu)

Scorecard je tabulka proměnných (**prediktorů**), které nejlépe predikují budoucí riziko. Je tvořena proměnnými a jejich možnými hodnotami. Pro každou hodnotu poskytuje scorecard relativní score (počet bodů). Konečné score je součtem bodů udělených pro všechny jednotlivé proměnné.

Příklad: $\text{sum_points} = -0.40 + 0.69 * (\text{prediktor1}) + 0.24 * (\text{prediktor2}) + 0.27 * (\text{prediktor3}) + \dots$

Každé portfolio (HU, SU, KK) má svou vlastní scorecard

Výhody

- automatizovaný proces
- rychlé vyhodnocení
- nízké náklady

Nevýhody

- posouzení není individuální
- nutnost aktualizovat a sledovat výkonnost

Scorecard		Score
Variable Name	Category	
prediktor1	category1	0.69
	category2	0
prediktor2	category1	0.92
	category2	0.81
	category3	0.67
	category4	0.55
	category5	0.41
	category6	0.31
	category7	0.23
	category8	0
prediktor3	category1	0.27
	category2	0
prediktor4	category1	0.41
	category2	0
prediktor5	category1	0.53
	category2	0.26
	category3	0
prediktor6	category1	1
	category2	0.58
	category3	0.68
	category4	0.4
	category5	0.43
	category6	0
	category7	0.12
intercept		-0.4



Krok	Detail
Příprava a analýza dat	<ul style="list-style-type: none">• určení cílové proměnné (vysvětlovaná)• seznam prediktorů (vysvětlující proměnné) – co známe při posouzení smlouvy• uřízneme poslední 2-3 měsíce – out-of-time test• rozdělíme data na vývojový a validační vzorek (v poměru 70:30)
Kategorizace prediktorů	<ul style="list-style-type: none">• kategorizace• analýza diskriminační síly• test stability prediktorů v čase
Vývoj modelu	<ul style="list-style-type: none">• vývojový vzorek – na historických datech se model „naučí“, jak klienta ohodnotit• výkonnost a testování<ul style="list-style-type: none">– out-of-sample test– out-of-time test– kratší proměnná• srovnání s původní scorecard
Nastavení cut off	<ul style="list-style-type: none">• nastavení cut off• simulace na produkci• nasazení
Monitoring scorecard	<ul style="list-style-type: none">• sledování výkonnosti po nasazení na produkci• v případě poklesu výkonnosti se vyvíjí nová scorecard

Načtení dat z databáze

Požadavky na data

- dostatečná kvantita a kvalita
- aktuálnost
- ošetřena sezónnost

Základní analýza a průzkum dat (průměry, minima, maxima, chybějící hodnoty, četnosti, apod.)

Očištění dat – vyřazení nevhodných dat např. duplicity, přeschválené smlouvy, akční produkty, smlouvy, kde se ignoruje score

Příklady:

- u podnikatele vyplněno pole „zaměstnán od“
- 25-letý člověk měl vyplněno, že pracuje již 15 let
- 30-letý člověk bydlí na dané adrese již 35 let

Cílová (vysvětlovaná) **proměnná** vyjadřuje skutečnost, zda je daný úvěr splácen včas nebo ne (smyslem scoringu je ohodnotit bonitu klienta vzhledem ke schopnosti splácet).

Každému úvěru je přiřazena 0 nebo 1, kde 0 znamená „dobrý úvěr“, 1 „špatný úvěr“.

Obecně:

Dobrý úvěr – dobrá platební morálka

Špatný úvěr – špatná platební morálka

Definice cílové proměnné záleží především na typu produktu. Cílová proměnná se bude např. lišit u spotřebitelských a hotovostních úvěrů. Rozdíl je v délce splatnosti (SU kolem jednoho roku, HU až sedm let) a výši úvěru.

Příklad

Definice špatného úvěru:

Klient se při splácení prvních 6 splátek dostal do prodlení většího než 90 dnů po splatnosti s dlužnou částkou větší než 100,- Kč

Definice dobrého úvěru:

Klient se při splácení prvních 6 splátek nezpozdil o více než 90 dnů s tolerancí 100,- Kč

Čím ostřeji jsou rozlišeny dobré a špatné úvěry, tím lépe

(volba parametru 90 dnů po splatnosti místo např. 30)

Smyslem je vyvinout model, který dokáže rozlišit „dobré“ a „špatné“ klienty (a to i na jiném než vývojovém vzorku)



Oblast

Příklady

Socio-demografická data

- Věk
- Pohlaví
- Rodinný stav
- Zdroj příjmů (délka zaměstnání)
- Vzdělání

Informace o produktu

- Výše úvěru
- Počet splátek
- Akontace (absolutně, v procentech)

Behaviorální data (pouze pro stávající klienty)

- Počet aktivních úvěrů
- Maximální počet dnů po splatnosti
- Doba od schválení posledního úvěru
- Výše pohledávky

Data z externích zdrojů (NRKI, Solus)

- Zadluženost klienta
- Délka úvěrové historie
- Počet zamítnutých úvěrů
- Počet ukončených úvěrů
- Credit Bureau Score

Prediktory se kategorizují – rozdělí se na vhodné kategorie s ohledem na badrate (podíl špatných úvěrů ku všem úvěrům v dané kategorii).

Smyslem kategorizace je pro každý prediktor vytvořit kategorie, které **co nejlépe rozlišují mezi dobrým a špatným klientem** (co nejvyšší diskriminační síla). Zároveň by prediktor měl být stabilní v čase a dobře rozlišovat i na jiném než vývojovém vzorku. Nemělo by se stát, že kategorie A má nižší badrate než kategorie B v jednom období a v druhém období je tomu naopak.

Pro modelování se používá logistická regrese - odhad pravděpodobnosti jevu (defaultu) na základě známých skutečností, které mohou ovlivnit výskyt tohoto jevu.

$$p(x) = 1 / [1 + \exp(-\beta_0 - \beta_1 x_1 - \beta_2 x_2 - \dots - \beta_n x_n)]$$

Kde $\beta_0, \beta_1, \dots, \beta_n$ jsou odhadované parametry modelu a představují jednotlivé body score.

$$\text{sum_points} = \text{intercept} + \sum_{\text{predictors}} \text{points from predictor}$$

$$\text{sum_points} = 3.45 + 0.12(\text{is women}) - 0.32(\text{is only 24 years old}) + 0.05(\text{married}) \dots$$

$$\text{probability_of_default} = \frac{1}{1 + \exp(-\text{sum_points})}$$



- implementováno v programu SAS
- snadné použití, kontrola a interpretace
- např. pro prediktor věk budeme očekávat: vyšší věk – vyšší hodnota koeficientu (lepší klient), pokud ne, je něco špatně
- málo parametrů

Alternativy:

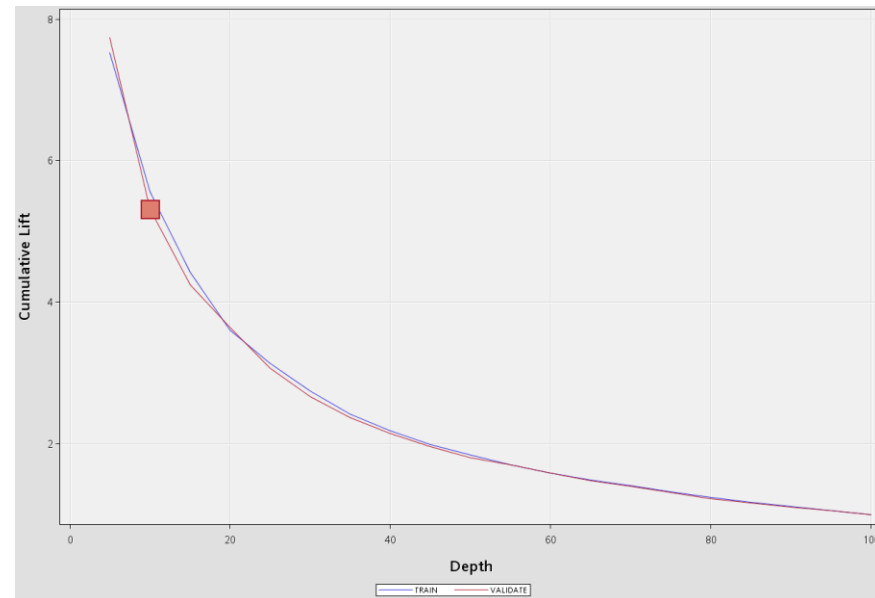
Rozhodovací stromy, neuronové sítě

Index gini - měří stupeň zlepšení separace na základě scorecard oproti náhodnému rozhodování.

Lift n% - vyjadřuje kolikrát je model na dané úrovni zamítání lepší než náhodné rozdělení klientů. Přesněji jde o podíl špatných klientů na množině n% nejhorších (dle score) ku podílu špatných klientů na celém vzorku.

n% - by mělo odpovídat reject rate na score

Statistika **Kolmogorov-Smirnov (K-S)** měří maximální rozdíl v kumulativních procentuálních rozděleních dobrých a špatných jednotlivců. Čím vyšší je hodnota K-S, tím výkonnější je model.



- out-of-sample test (30% dat)
- out-of-time test (poslední 2-3 měsíce)
- „kratší“ cílové proměnné
- srovnání s původní scorecard

Posledním krokem před nasazením scorecard na produkci je nastavení hranice zamítání tzv. cut off.

Jednoduchý přístup:

Klienti se score vyšším než cut off mohou být schváleni, klienti se score na hranici a pod hranicí jsou zamítnuti.

Pokud dochází k nasazení nové scorecard na existujícím portfoliu, zohledňuje se obvykle při nastavení cut off aktuální míra zamítání (počet zamítnutých úvěrů ku všem schvalovaným úvěrům) a objem schválených úvěrů.

Je odhadován dopad do obchodních výsledků.

Nová scorecard je obvykle nasazena s cílem neměnit míru zamítání, ale zlepšit rozlišení klientů na „dobré“ a „špatné“.

Složitější přístup:

Vypočítat očekávaný zisk (ztrátu). Score (očekávaný default) je pouze jedním ze vstupů. Další jsou např. financovaná částka, očekávaná úspěšnost vymáhání, délka úvěru, cena zdrojů. Schválení jsou pouze klienti s kladným očekávaným ziskem.

Po nasazení nové scorecard klesá její prediktivní síla
– způsobeno změnou vstupních dat

Je nutné výkonost scorecard monitorovat

Při výrazném poklesu výkonosti se vyvíjí nové scorecard

Proč se mění vstupní data:

- demografická struktura obyvatel (např. roste počet vzdělaných lidí, roste výše příjmu, apod.)
- změna ekonomické situace
- marketingové akce (jiná skupina klientů)
- databázové chyby