

CZU: 658.512.2:339.138

DOI: <http://doi.org/10.5281/zenodo.5095155>

DESIGNUL UNEI APLICAȚII DE DATA MINING PENTRU ACHIZIȚIONAREA NOILOR CLIENȚI

Alina POPA

Academia de Studii Economice din București (România)

Pentru identificarea de noi clienți, companiile pot să utilizeze aplicații de Data Mining. În această lucrare se prezintă atât designul unei astfel de aplicații, cât și etapele și activitățile care trebuie efectuate. De asemenea, se oferă sugestii cu privire la algoritmi de inteligență artificială potriviți pentru fiecare pas. În cele din urmă, se menționează o suită de metrici care pot fi utilizate în evaluarea sistemului dezvoltat.

Cuvinte-cheie: *inteligență artificială, machine learning, marketing, comunicare.*

DESIGN OF A DATA MINING APPLICATION FOR NEW CLIENT ACQUISITION

In order to identify new customers, companies can use Data Mining applications. This paper presents both the design of such an application and the steps and activities to be performed. Suggestions are also given for the appropriate artificial intelligence algorithms to be used at each step. Finally, there is presented a suite of metrics that can be used in the evaluation of the developed system.

Keywords: *artificial intelligence, machine learning, marketing, communication.*

Introducere

Pentru majoritatea companiilor, mijlocul principal de creștere implică achiziționarea de noi clienți. Acești potențiali clienți sau prospekți s-ar putea să nu cunoască încă produsul sau serviciile companiei sau să nu fi avut nevoie de ele. În categoria prospekților se pot include și clienții pierduți. Indiferent de tipul lor, în mod tradițional, acești clienți potențiali se abordează printr-o combinație de campanii de marketing în masă și campanii de marketing direct. Modul în care aceste campanii sunt create se bazează în mare parte pe cunoștințele pe care le dețin persoanele de comunicare din cadrul respectivei companii. Un rol important joacă și informațiile disponibile despre clienții vizați incluse atât în bazele de date ale companiei, cât și în bazele de date externe [1]. O persoană de marketing cu experiență poate adesea înțelege și alege cele mai relevante caracteristici socio-demografice, însă odată cu creșterea cantității de date și cu schimbarea rapidă a tiparelor comportamentale, procesul devine nefezabil. În același timp, se dorește o înmagazinare a cunoștințelor care în trecut erau deținute cu precădere de un grup de persoane specializate în cadrul proceselor operaționale ale companiei. În acest demers, aplicațiile de Data Mining (DM) ce folosesc algoritmi de Machine Learning (ML) pot fi utilizate pentru a identifica potențiali clienți rentabili în o bază de date, prin găsirea de „clone” ale clienților valoroși existenți în bazele de date ale companiei [2,3].

De cele mai multe ori, o campanie de achiziție de clienți începe cu o listă de prospekți (potențiali clienți). Această listă reprezintă un set de entități (persoane fizice sau juridice) care au fost selectate pe baza unor criterii sociodemografice destul de generale. Dintre aceste entități, unele ar putea fi interesate de produsul/serviciile companiei [1]. Uneori, aceste liste pot să existe deja în companie, exemplu fiind lista clienților pierduți sau bazele de date rezultate după fuziunea a două companii. Alteori, acestea trebuie cumpărate de la companiile specializate în domeniu.

În cadrul acestui articol se va prezenta designul unei aplicații care are ca scop generarea de liste optimizate pentru achiziționarea de noi clienți.

Material și metode

Pentru a asigura că aplicația dezvoltată răspunde cerințelor de afaceri, procesul de design este compus din următorii pași: 1) determinarea obiectivelor sistemului, 2) dezvoltarea design-ului aplicației și 3) identificarea metricilor utilizate pentru evaluarea sistemului.

1) *Determinarea obiectivelor sistemului:*

Astfel, aplicația are următoarele **obiective**:

- Sa se extragă din lista generală de prospekți numai acele entități care au probabilitatea ridicată de a deveni consumatori.

- În același timp, pentru fiecare dintre acești potențiali consumatori este nevoie să se decidă mesajul de comunicare, astfel încât acesta să aibă cel mai mare succes în atragerea clientului.
- Un alt obiectiv este să se abordeze numai acei prospekți care vor deveni consumatori valoroși. Acest lucru poate fi condiționat de faptul că resursele companiei care se alocă pentru menținerea relației cu clientul sunt limitate, iar menținerea relației cu consumatorul este costisitoare, de aceea se dorește maximizarea numărului clienților profitabili.

2) Dezvoltarea design-ului aplicației

În mod schematic, o aplicație de generare de liste optimizate este prezentată în Figura 1. Orice aplicație de DM începe prin stabilirea datelor necesare și a surselor de date utilizate. Pentru a optimiza atragerea clienților potențiali, este nevoie de date interne pe care le are compania despre clienții actuali în termeni de 1) informații sociodemografice și 2) comportament de cumpărare. Aceste două tabele vor fi principalele date utilizate în antrenarea modelului. O altă sursă de date pe care se va utiliza modelul dezvoltat este, după cum s-a menționat, lista de prospekți care conține și informații sociodemografice despre aceștia.

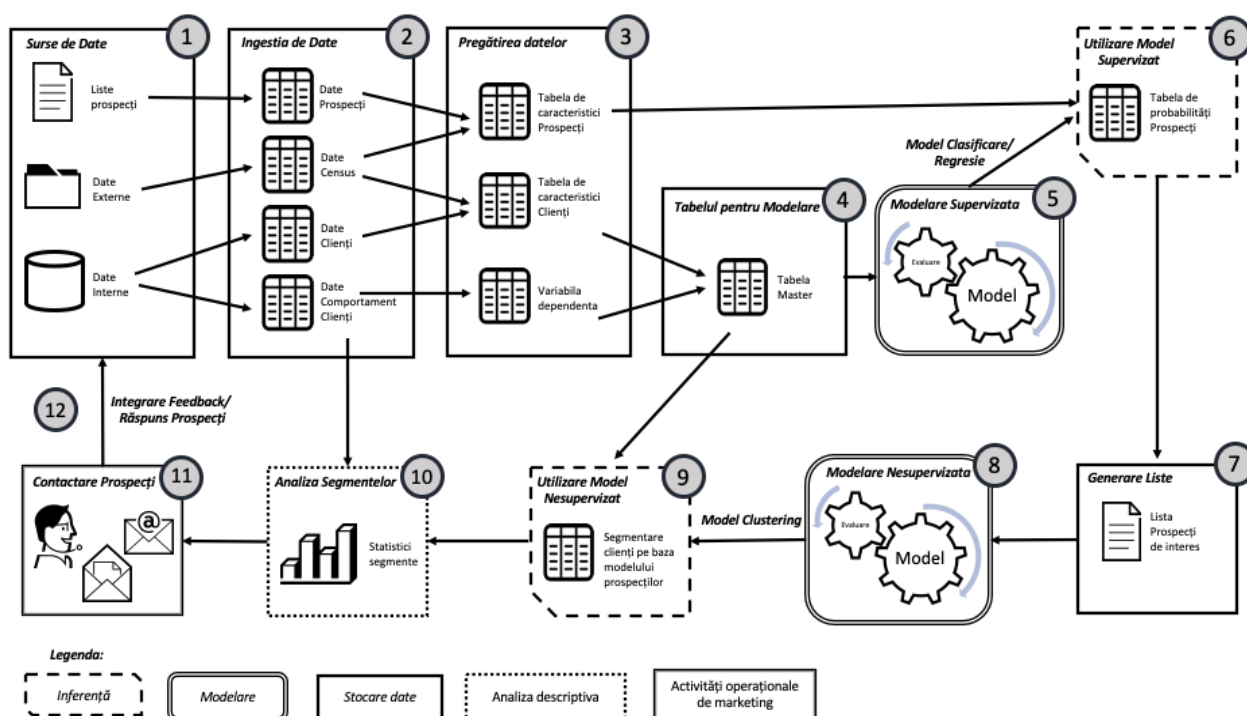


Fig.1. Schema sistemului de Data Mining pentru generarea de liste optimizate.

Pentru a îmbogăți datele existente despre clienți și prospekți, este necesară utilizarea unor surse de date externe, cum ar fi date censuș agregate la nivel de orașe, zone ale țării sau chiar cartiere ale unui oraș. După ce au fost stabilite sursele de date necesare (Fig.1, 1), tabele corespunzătoare sunt extrase și încărcate în baza de date a aplicației (Fig.1, 2).

Următorul pas este pregătirea datelor (Fig.1, 3). Această etapă ia în jur de 80% din timpul necesar dezvoltării aplicației.

Odată ce tabelul de caracteristici și cel de variabile dependente sunt pregătite, acestea se unesc pe baza identificatorului de client într-un singur tabel, numit și tabel master (Fig.1, 4). Tabelul reprezintă forma finală a datelor pe baza cărora va fi antrenat modelul.

În partea de modelare (Fig.1, 5), tabelul master se împarte în două seturi: de antrenament și de test. Pe setul de antrenament, pentru fiecare variabilă dependentă considerată de interes se aplică diferiți algoritmi de Machine Learning. Se încearcă configurații diverse pentru fiecare tip de algoritm. Pe setul de test se validează performanțele diferitor modele știindu-se valorile variabilei dependente, adică „răspunsurile corecte”.

Dintre algoritmi de ML potriviți pentru modelarea variabilelor numerice se poate alege GLM [4], Arborii de decizie CART [5], Păduri Aleatoare [6] sau Rețelele Neuronale [7]. Pentru variabila binară, setul de metode

se extinde la SVM [8] și KNN [9]. În cazul KNN, este nevoie să se menționeze că modelul nu încearcă să găsească o relație dintre datele de intrare reprezentate prin caracteristicile prospecților și datele de ieșire sub forma probabilității de a deveni client al companiei. Algoritmul identifică direct clienții potențiali într-o bază de date, prin găsirea de entități similare cu actualii clienți ai companiei.

Deoarece înțelegerea profilului unui client vs. non-client al companiei este de interes pentru dezvoltarea mesajului comunicării, se pune accent pe utilizarea de algoritmi care pot fi înțeleși de oameni. De exemplu, un arbore de decizie poate fi descompus într-o serie de reguli „dacă, atunci”.

Pentru acest tip de aplicație, indicatorii de performanță importanți sunt Rădăcina Pătrată din Eroarea Medie Pătratică (eng. Root Mean Square Error, RMSE) și Eroarea Medie Pătratică (eng. Mean Square Error, MSE) [10].

Odată ce modelele ating performanța dorită, acestea se pot utiliza în evaluarea listei de prospecți (Fig.1, 6). Pentru fiecare prospect se va estima o valoare viitoare a tranzacțiilor lunare, număr mediu de tranzacții per lună și o variabilă binară care exprimă probabilitatea să devină client al companiei.

Lista de prospecți cu noile informații se sortează descrescător pe baza produsului dintre probabilitatea de a deveni client și valoarea totală medie tranzacționată pe lună. Se aleg prospecții cu perspectiva ridicată pentru lista finală optimizată (Fig.1, 7) [11].

Următorul obiectiv care trebuie abordat în cadrul aplicației este determinarea produselor de interes sau a mesajului potrivit pentru fiecare prospect din listă. Acest lucru este dificil, avându-se în vedere că nu există un istoric pentru entitățile din lista optimizată. O soluție ar fi segmentarea prospecților pentru extragerea segmentelor principale existente în lista optimizată (Fig.1, 8), după care pentru fiecare segment se va dezvolta o alternativă a mesajului de marketing.

Segmentarea se poate efectua utilizând orice algoritm de grupare. Totuși, deoarece în pasul următor (Fig.1, 9) modelul se aplică pe baza de clienți ai companiei cu scopul de a extrage clienții care se află în aceleași segmente ca și prospecții, se va prioritiza utilizarea unui algoritm de grupare care are noțiunea de zgomot, adică DBSCAN [12].

Rezultatul utilizării modelului de grupare (Fig.1, 9) este reprezentat de o listă de clienți actuali ai companiei care sunt asemănători cu prospecții. Pe baza acestor clienți extrași și alocați fiecărui segment, pentru fiecare segment se extrag statistici cu privire la produsele sau categoriile de produse cumpărate, promoțiile atractive (Fig.1, 10).

În etapele ce urmează se trimit mesajele dezvoltate prin instrumentele de marketing direct (Fig.1, 11) și se integrează feedback-ul obținut de la client. Aceste răspunsuri sunt utilizate în reantrenarea și rafinarea modelului de ML, dar și în evaluarea campaniei de achiziție de clienți.

3) Identificarea metricilor utilizate pentru evaluarea sistemului

Indicatorii de business care exprimă eficiența aplicației sunt Rata de Achiziție (eng. Acquisition Rate, AR) și Randamentul Investițiilor (eng. Return of Investment, ROI) ale aplicației.

Rata de achiziție reprezintă raportul dintre numărul de prospecți achiziționați și numărul de prospecți care au fost abordați. Randamentul investițiilor se calculează ca raport dintre a) diferența dintre venituri din prospecții convertiți în clienți și cheltuielile pentru dezvoltarea aplicației și b) cheltuielile pentru dezvoltarea aplicației. Veniturile clienților convertiți reprezintă creșterea generată de prospecți deveniți clienți într-o perioadă t. Cheltuielile cu aplicația reflectă atât costurile fixe ce țin de cumpărarea listelor de prospecți, tehnologiei, serviciilor de consultanță IT, cât și cheltuielile operaționale pentru menținerea sistemului aplicației în perioada t.

O altă modalitate de a estima eficiența campaniei de achiziție este să se realizeze în paralel o campanie de test sau control sub forma unui experiment controlat.

Rezultate și discuții

În literatura de specialitate nu este prezentat în detalii modul în care o companie ar aborda dezvoltarea unui sistem de achiziționare de clienți cu accent pe activitățile individuale, procesele și algoritmi de analiză a datelor. Modelul conceptual este bazat pe experiența autorului de realizare a unor proiecte similare. Acesta nu este rezultatul unor analize cantitative pe un eșantion de date empirice rezultate în urma implementării abordării, ci al unor analize calitative de ansamblu și „lecții învățate” în urma dezvoltării de sisteme similare. De aceea, adițional, este nevoie și de determinarea unor rezultate empirice sub forma indicatorilor de succes ai sistemului din cadrul diferitor proiecte comerciale. Aceste rezultate ar facilita decizia unei companii de a implementa un astfel de sistem.

Concluzii

În această lucrare a fost prezentat un cadru conceptual de dezvoltare a unui sistem de achiziționare de clienți care poate fi adaptat la o gamă largă de utilizări. De asemenea, s-a explicat în mod simplu și cuprinzător care sunt componentele, etapele și modul în care se poate dezvolta un astfel de sistem. Nu în ultimul rând, au fost oferite niște metrici care pot fi utilizate la evaluarea sistemului dezvoltat.

Cercetări ulterioare sunt necesare pentru a implementa această abordare și pentru a o utiliza în situații din lumea reală cu scopul de a identifica limitările și avantajele practice.

Referințe:

1. MAIMON, O., ROKACH, L. *Data Mining and Knowledge Discovery Handbook*. Springer, 2010.
2. CHORIANOPOULOS, A. *Effective CRM using predictive analytics*. John Wiley & Sons, 2016.
3. KADIYALA, S.S., & SRIVASTAVA, A. Data Mining For Customer Relationship Management. In: *International Business & Economics Research Journal (IBER)*, 2002, no.1(6).
4. NELDER, J.A., & WEDDERBURN, R.W. Generalized linear models. In: *Journal of the Royal Statistical Society: Series A (General)*, 1972, no.135(3), p.370-384.
5. BREIMAN, L., FRIEDMAN, J.H., OLSEN, R.A., & STONE, C.J. *CART. Classification and Regression Trees*. Wadsworth and Brooks/Cole, Monterey, CA, 1984.
6. BREIMAN, L. Random forests. In: *Machine learning*, 2001, no.45(1), p.5-32.
7. HINTON, G.E. How neural networks learn from experience. In: *Scientific American*, 1992, no.267(3), p.144-151.
8. VAPNIK, V., GUYON, I., & HASTIE, T. Support vector machines. In: *Mach. Learn.*, 1995, no.20(3), p.273-297.
9. FIX, E. *Discriminatory analysis: nonparametric discrimination, consistency properties*. Vol.1. USAF school of Aviation Medicine, 1985.
10. KANTARDZIC, M. *Data mining: concepts, models, methods, and algorithms*. 3rd edition, John Wiley & Sons, 2019.
11. TSIPTISIS, K.K., & CHORIANOPOULOS, A. *Data mining techniques in CRM: inside customer segmentation*. John Wiley & Sons, 2011.
12. ESTER, M., KRIEGEL, H. P., SANDER, J., & XU, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Kdd*, 1996, vol.96, no.34, p.226-231.

Date despre autor:

Alina POPA, doctorandă, Școala doctorală *Marketing*, Academia de Studii Economice din București (România).

E-mail: popa.alina.alexci@gmail.com

ORCID: 0000-0001-8727-1687

Prezentat la 16.04.2021