

Raport Științific și Tehnic 2016

PN-III-P2-2.1-BG-2016-0447

**Soluție suport de decizie pentru managementul riscului și retenția
clienților la BT Direct – DSS-Direct**

Durata proiectului: 01.10.2016-30.09.2018

Faza de execuție: 01.10.2016-31.12.2016

Conținut:

1. Obiective generale ale proiectului
2. Obiectivele perioadei octombrie 2016-decembrie 2016
3. Rezumatul activității desfășurate
4. Descriere științifică și tehnică
5. Concluzii

Bibliografie

Anexe

Avizat,

Coordonator

Reprezentant Legal
Acad. Prof. dr. Ioan Aurel Pop

Director Proiect
Prof. univ. dr. Anca Andreica

Agent Economic

Reprezentant Legal
Gabriel Gogu

Responsabil de proiect
Cezar Chiș-Bodea

1. Obiective generale ale proiectului

Proiectul DSS-Direct reunește echipa Universității Babeș-Bolyai din Cluj-Napoca pentru a desfășura o activitate de transfer de cunoștințe în domeniul Analytics către partenerul BT Direct IFN SA, subsidiară a Băncii Transilvania. BT Direct este orientată către oferirea de împrumuturi către clienți de retail cu profil de risc ridicat pentru finanțarea nevoilor personale.

O soluție de asistarea a deciziei rezultată în urma unui proces de descoperire a cunoștințelor în cadrul datelor colectate de către companie reprezintă o alegere naturală în contextual concurențial din domeniul bancar. Mai mult, cantitatea de date colectată este în continuă creștere, fiind dificil de analizat prin metode tradiționale.

În funcție de gradul de sofisticare a metodelor de analiză, soluțiile Analytics în cadrul companiilor pot fi grupate în 5 mari categorii [1], descrise în figura 1.

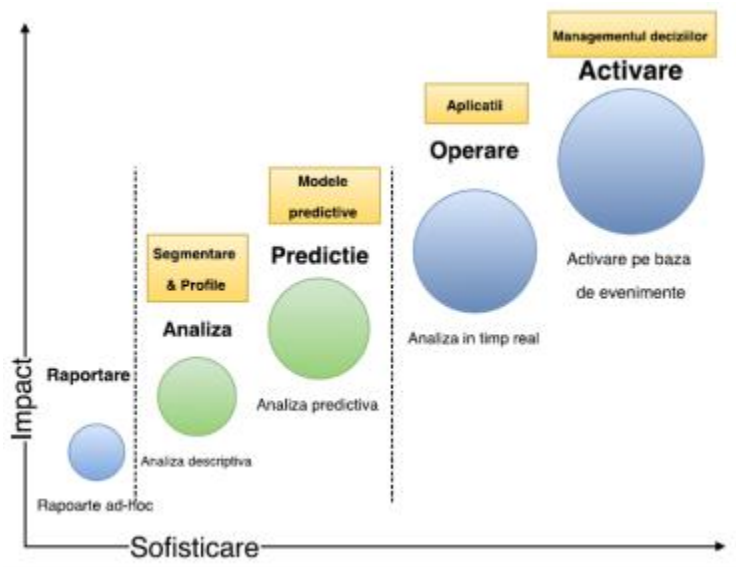


Figura 1 Impactul Analytics în funcție de gradul de sofisticare [1]

În cadrul proiectului DSS-Direct, scopul este acela de a realiza analize descriptive și predictive, structurate în 4 obiective generale:

- 1) Analiza comportamentului clienților și elaborarea de modele privind retenția clienților;
- 2) Conceperea de metode de scoring pentru activitatea de creditare;
- 3) Conceperea de modele pentru estimarea riscului de credit (estimarea probabilității de neplată și a cerințelor de capital aferente);
- 4) Analiza comportamentului clienților care apelează la darea în plată și efectele legii dării în plată.

2. Obiectivele perioadei octombrie 2016-decembrie 2016

În conformitate cu schema de realizare a proiectului, în perioada octombrie-decembrie 2016 au fost incluse componente din 2 pachete de lucru:

PL1. Managementul datelor

PL4. Transfer de cunoștințe și diseminare

PL6. Management de proiect

În cadrul pachetului de lucru Managementul datelor, activitățile desfășurate s-au concentrat în jurul **Caracterizării datelor** disponibile, a **Definirii surselor de date și a variabilelor** necesare analizei, a **Calității datelor** și a **Anonimizării** lor în vederea respectării tratamentului datelor cu caracter personal.

În cadrul pachetului **Transfer de cunoștințe și diseminare**, a fost realizată pagina web a proiectului.

Activitățile din cadrul **Managementului de proiect** se referă la Monitorizarea continuă a activităților, întâlniri lunare pentru analiza progresului și raportare. Au fost agreeate împreună cu reprezentanții BT-Direct condițiile în care acestea urmează să se desfășoare pe perioada de derulare a proiectului, începând cu anul 2017.

3. Rezumatul activității desfășurate

În perioada octombrie-decembrie 2016 echipa de proiect s-a concentrat în jurul activităților înțelegere în profunzime a operațiunilor și practicilor realizate de către partener și de colectare a datelor disponibile de la partener.

În acest scop au fost realizate mai multe întâlniri la sediul BT Direct cu conducerea executivă și cu șeful departamentului de Analiză de credite. Aceste întâlniri au fost realizate pentru a defini modalitatea de colectare a datelor de la partener și a stabili obiective pe termen scurt în privința analizei descriptive. Întâlnirile de lucru au căpătat ulterior un caracter recurent, pe măsură ce au fost identificate neclarități din partea echipei de proiect sau pentru a înțelege în profunzime procese specifice activității BT Direct. Putem astfel identifica 3 tipuri de activități desfășurate împreună cu partenerii: întâlniri de management al proiectului, întâlniri de informare și întâlniri de prezentare și dezbateră a rezultatelor.

Pe de altă parte, echipa de proiect a desfășurat activități de documentare [2,3,4,5] pentru a se asigura de utilizarea celor mai bune practici cu privire la obiectivele avute. Documentarea include studiarea de publicații științifice, economice și/sau tehnice referitoare la Analytics în domeniul financiar-bancar.

În cadrul activităților de analiză a calității datelor primite de la partener s-au avut în vedere utilizarea de metode statistice pentru a identifica, pentru fiecare variabilă anumite caracteristici precum: valori lipsă, intervale de variație, distribuție, relevanță etc.

Până la finalul perioadei s-a realizat un **Raport al calității datelor** și un set de sugestii pentru **Actualizarea datelor**, precum și o prezentare la sediul partenerului care a inclus și o **Analiză descriptivă** a datelor puse la dispoziție.

În cadrul **Managementului de proiect** au fost realizate activități de Monitorizare și Raportare. Au fost realizate **Planul de monitorizare, Rapoarte de progres și Raportul științific și tehnic 2016**. Totodată, a fost realizat și website-ul proiectului, disponibil la adresa <http://www.cs.ubbcluj.ro/bg0447/>.

4. Descriere științifică și tehnică

În activitatea de analiză a datelor s-a utilizat un fișier cu date anonimizate referitoare la clienți oferit de către BT Direct și prelucrat ulterior de echipa de lucru. Au avut loc întâlniri la sediul companiei pentru a înțelege aspecte legate de activitatea de creditare desfășurată de BT Direct și pentru explicații cu privire la semnificația fiecărei variabile din setul de date.

Metodele de analiză utilizate includ statistici descriptive și modele econometrice, în funcție de specificul fiecărui obiectiv.

În funcție de rezultatele obținute, raportul include concluzii și sugestii pentru creșterea posibilităților de analizare a comportamentului clienților și în final a creșterii calității analizei de risc de credit.

4.1 Pregătirea datelor

Setul de date utilizat reprezintă informații legate de comportamentul clienților BT Direct în perioada ianuarie-august 2016, toți acești clienți contractând credit în perioada respectivă. Pentru a putea realiza prelucrările necesare, a fost necesar ca datele să fie aduse la o formă corespunzătoare din punct de vedere al structurii bazei de date.

O operațiune specifică este aceea de curățare a datelor, care presupune validarea fiecărei variabile din punct de vedere a corectitudinii informațiilor. Spre exemplu, au fost identificate înregistrări cu valoarea zero a dobânzii, fiind corectate prin introducerea valorii “missing” în locul valorii “0”. De asemenea, valoarea FICO score a fost corectată prin introducerea valorii “missing” în locul valorii “0” și a valorilor mai mari de “850” (probabil greșeli de introducere de date).

Un alt pas în pregătirea datelor este acela de eliminare a variabilelor redundante (identice sau cu semnificație identică), precum și a variabilelor nominale cu foarte multe valori distincte.

4.2 Analiza comparativă a echipelor de analiză de credite

A fost realizată o analiză a performanțelor celor două echipe (denumite Scoring și Nonscoring) de analiză a riscului de credit din cadrul companiei.

Rezultatele obținute sunt prezentate în graficele de mai jos

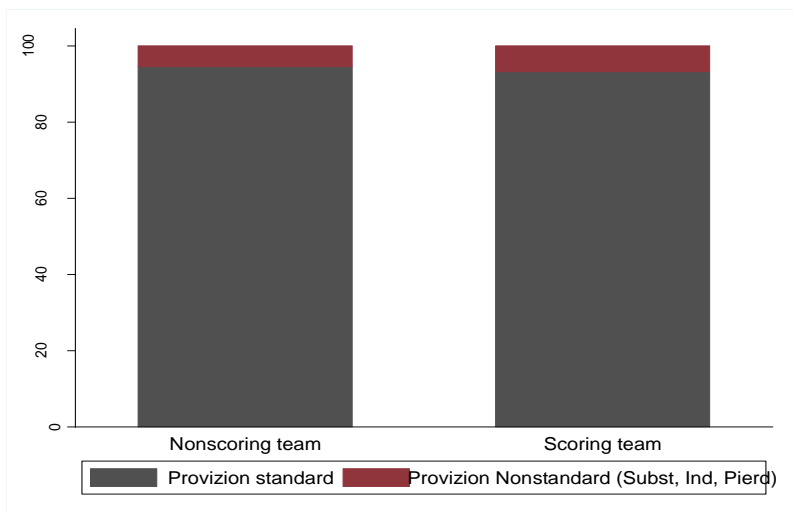


Figura 2 Performanta creditor

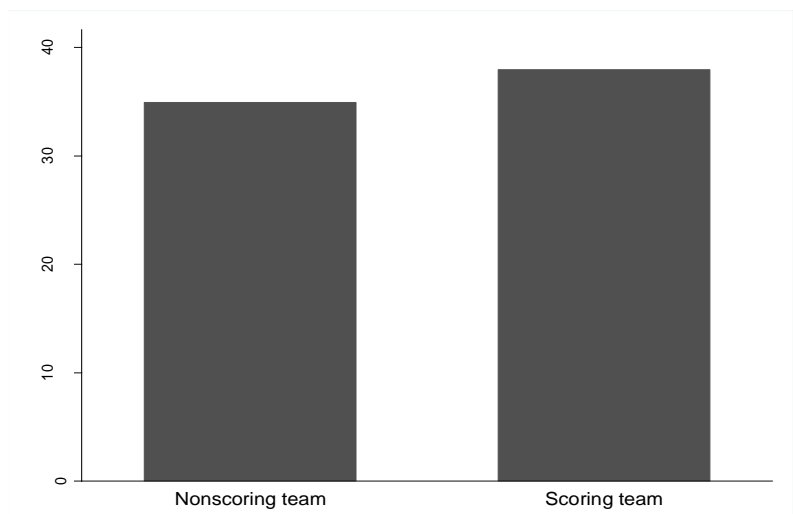


Figura 3 Intarziere la plata

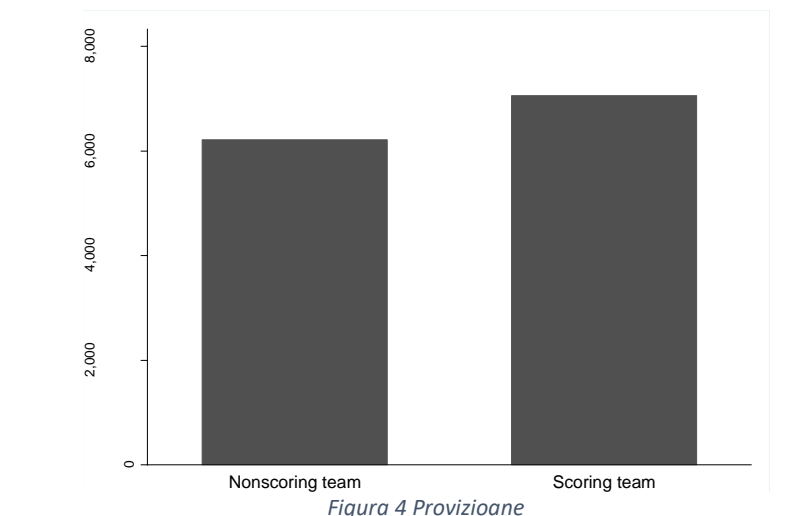


Figura 4 Provizioane

4.3 Identificare elemente pentru construirea unui profil de client

A fost analizată din punct de vedere statistic diferența între clienții încadrați în categoria Standard și categoria Non-Standard. A fost realizată o clasificare a creditelor în funcție de cele două categorii (standard și non-standard) luând în considerare genul, starea civilă și nivelul de educație. Se poate observa că ponderea bărbaților este mai crescută în cadrul grupei non-standard decât în standard. Din punct de vedere al stării civile, persoanele necăsătorite reprezintă un factor mai mare de risc pentru neplata creditelor, având o pondere mult mai mare în rândul rău-platnicilor decât proporția lor naturală în întregul clienților (41% vs. 23.5%). În ceea ce privește educația, nu se observă diferențe majore între distribuția pe cele două categorii și distribuția naturală.

Tabel 1 Clasificare clienti pe grupe de provizioane

	Provizion - Nonstandard	Provizion - Standard	Total
Gen			
Masculin	59.70%	54.30%	54.64%
Feminin	40.30%	45.70%	45.36%
Total	100.00%	100.00%	100.00%
Stare civila			
Casatorit	44.89%	63.27%	62.15%
Celibatar	41.10%	22.37%	23.52%
Divortat	10.04%	8.14%	8.26%
Vaduv	3.98%	5.99%	5.87%
Nu a declarat	0.00%	0.22%	0.21%
Total	100.00%	100.00%	100.00%
Educatie			
Colegiu	1.70%	3.45%	3.34%
Liceu	43.37%	39.97%	40.18%
Scoala elementara	21.21%	19.44%	19.55%
Scoala tehnica	20.27%	23.88%	23.65%
Universitate	13.45%	12.85%	12.88%
Nu a declarat	0.00%	0.42%	0.39%
Total	100.00%	100.00%	100.00%

4.4 Importanta variabilelor

A fost studiată importanța variabilelor din setul de date în legătură cu probabilitatea de neplată a creditului (Vezi graficul din anexa). Se poate observa că cel mai important factor este FICO score, urmat de tipul produsului, valoarea ratei lunare, a valorii dobânzii etc. Variabila "punctaj scoring" este semnificativ mai puțin importantă decât FICO score în identificarea clienților Good sau Bad.

Au fost utilizați arbori de decizie pentru a ilustra încadrarea clienților într-una din cele două categorii, în funcție de valorile celor mai semnificative variabile.

4.5 Recomandări cu privire la calitatea datelor

În urma analizei, am putut recomanda o serie de îmbunătățiri asupra modului în care sunt colectate datele necesare pentru realizarea analizelor.

R1 - Istoric complet al întârzierilor

Titlu	Istoric complet al întârzierilor
Descriere/ problemă	Un client care a avut întârzieri dar la momentul colectării datelor este la zi cu plata, este un client suspect însă acest aspect nu este reflectat de date. Există mai multe tipuri de comportament pentru întârzieri de plată, cu riscuri diferite pentru LTD. Unele situații pot fi asumate și eficiența per client este încă favorabilă dpdv business.
Decizie/ propunere	Colectare permanentă Periodicitate: zilnic/ săptămânal Colectare automată ? semiautomată?
Prioritate	0

R2 - Localizare geografică client

Titlu	Localizare geografică client
Descriere/ problemă	În prezent nu avem informații legate de localizarea geografică a clientului.
Decizie/ propunere	Abordări colectare date: 1. Încadrarea clientului în una dintre categoriile: 1. Urban reședință de județ, 2. Urban, 3. Rural 2. Cod poștal Mult mai relevant, pentru localități cu populație > 50.000 locuitori se poate realiza localizare până la nivel de stradă sau segment de stradă. Varianta 2 preferabilă. Cost mic de implementare, fezabil dpdv colectare și acuratețe maximă în analize bazate pe localizare geografică.
Prioritate	0

R3 - Localizare geografică intermediari (broker, magazin)

Titlu	Localizare geografică intermediari (broker, magazin)
Descriere/ problemă	În prezent nu există o localizare geografică <u>precisă</u> și consistentă a intermediarilor de credit.
Decizie/ propunere	Cod poștal Fezabilitate bună, datorită numărului relativ redus de entități, colectat la introducerea unui nou broker/ magazin. Consistență cu localizarea clientului: informația trebuie să fie comparabilă ca structură și semantică cu cea care va fi utilizată la localizarea clientului (ex.: cod poștal).

Prioritate	0
-------------------	---

R4 - Profesie, domeniu de activitate, funcție de conducere

Titlu	Profesie, domeniu de activitate, funcție de conducere
Descriere/ problemă	Categoriile actuale de profesie (coloana “profesia”) sunt prea generale, conțin profile profesionale foarte diferite comasate aleator. Coloana “poziție” conține date nestructurate, nu poate fi utilizată deocamdată în analize.
Decizie/ propunere	Definirea unui nomenclator de grupe de profesii. Timpul afectat pentru încadrarea unui client în grupă să fie foarte mic (în cel mai defavorabil caz de implementare să fie egal cu timpul alocat acum pentru colectarea datelor legate de profesie). Datele trebuie să fie structurate și modul de colectare să prevadă valori prestabilite pentru profesie, nu la liber.
Prioritate	0

R5 - Alte produse acordate de către BT clientului

Titlu	Alte produse ale clientului, acordate de BT
Descriere/ problemă	Informație suplimentară în faza de analiză. Dacă clientul are deja întârzieri pe alte produse ale băncii, există șansă relativ semnificativă de a fi rău platnic și pe acest credit. Pot exista informații despre comportamentul clientului în relația cu Banca Transilvania (salariu, depozite, carduri etc.).
Decizie/ propunere	Dacă sunt accesibile în mod automat aceste informații despre client la BT, se poate construi un parametru suplimentar sau grupă de parametri care să reflecte comportamentul clientului.
Prioritate	1

5. Concluzii

În perioada de raportare (octombrie-decembrie 2016), echipa de proiect a urmat graficul de realizare a proiectului, realizând activitățile specifice în colaborare cu partenerii de la BT-Direct, fiind îndeplinite obiectivele de etapă, respectiv: colectarea datelor pentru analiză; realizarea de analize descriptive pentru a putea avea o imagine asupra calității datelor; transmiterea unui raport preliminar de analiză către partener; stabilirea cu partenerul a calendarului cu privire la stagiile de practică; activități de management al proiectului, inclusiv realizarea website-ului proiectului.

Din punct de vedere al colaborării în cadrul proiectului, putem concludiona că toate activitățile s-au desfășurat în condiții de lucru foarte bune, existând o sinergie între membrii echipei și dovedind o colaborare foarte bună și cu reprezentanții BT-Direct, creând astfel premisele pentru o conlucrare de succes și în următoarele etape.

Bibliografie

- [1] Sharda, R., Delen, D., Turban, E., & King, D. (2015). *Business Intelligence and Analytics. Systems for Decision Support*. Pearson Higher Ed.
- [2] Brown, I. (2014). *Developing Credit Risk Models Using SAS Enterprise Miner and SAS/STAT: Theory and Applications*. SAS Institute Inc.
- [3] Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124-136.
- [4] Basel Committee on Banking Supervision (2015). Review of the Credit Valuation Adjustment Risk Framework, *Consultative Document*.
- [5] Basel Committee on Banking Supervision (2016), Reducing variation in credit risk-weighted assets – constraints on the use of internal model approaches, *Consultative Document*.

Anexa

