UNIVERSITATEA "ALEXANDRU IOAN CUZA" DIN IAȘI

FACULTATEA DE ECONOMIE ȘI ADMINISTRAREA AFACERILOR

ȘCOALA DOCTORALĂ DE ECONOMIE ȘI ADMINISTRAREA AFACERILOR

Modele bazate pe machine-learning pentru detectarea fraudelor bancare

Rezumatul tezei de doctorat

Coordonator:

Prof. univ. dr. Gabriela MEȘNIȚĂ

Doctorand:

Adriana-Elena Stan (căs. Mînăstireanu)

Iași, 2020

Cuprins

Cuprinsul tezei de doctorat

Introducere

Cercetarea se dorește a fi o abordare nouă în ceea ce privește detectarea fraudelor bancare în contextul tehnologic actual, prin analiza aprofundată a literaturii de specialitate cu privire la metodele de Machine Learning, a instrumentelor din categoria Data Analysis și Big Data, și crearea unui model stivă de patru clasificatori pentru performanțe predictive ridicate.

Adoptarea rapidă a tehnologiilor digitale, precum cloud computing, machine-learning, data analysis și big data, a condus la modificarea modului în care clientul este implicat în activitatea sectorului bancar. Mai mult, în acest mediu nou al serviciilor bancare, interacțiunile cu clienții nu se mai bazează doar pe derularea tradiționalelor tranzacții bancare, ci pe vânzarea de produse sau servicii de consultanță digitale, prin care clienții dobândesc o poziție și mai importantă față de sistemele tradiționale. Aceștia utilizează tot mai mult sistemele bancare în mediul virtual (internet-banking, mobile-banking, home-banking, robo-banking). Dar, cu cât o tehnologie este mai puternică atunci când este folosită corect, cu atât ea este mai tentantă și periculoasă dacă este folosită ilegal. Iar datele statistice confirmă, din păcate, acest lucru. Un studiu efectuat de Cybersecurity Ventures[1], relevă faptul că daunele pentru infracțiuni informatice vor aduce un prejudiciu lumii în valoare de 6 miliarde de dolari anual până în 2021, cu mai mult decât daunele provocate de dezastrele naturale într-un an și mai profitabile decât comerțul ilegal cu medicamente.

Dinamismul accentuat din cadrul serviciilor bancare actuale, prin digitalizare și inovație, aduce cu sine și o serie de provocări, dintre care cea mai importantă este frauda. Frauda în sistemul bancar, pe lângă scăderea credibilității instituției, poate avea consecințe grave asupra întregii activități. Prin urmare, se fac eforturi pentru a preveni acest fenomen ce se manifestă la nivel economic, politic și social, cu efecte asupra întregii societăți. În acest context, abordările tradiționale de detectare a fraudelor, bazate pe tehnici manuale, s-au dovedit a fi ineficiente datorită dificultăților și complexității operațiunilor în condițiile digitalizării actuale. Aproape în același timp cu implementarea unei tehnologii inovative în sistemul bancar, infractorii reușesc să treacă peste sistemele de securitate, fiind primii care încearcă să exploateze beneficiile noii tehnologii. Acest lucru accentuează necesitatea dezvoltării unui model securizat de detectare a fraudelor.

[1] Cybercrime Magazine – S. Morgan, *2019 Cybersecurity Almanac: 100 Facts, Figures, Predictions And Statistics*, https://cybersecurityventures.com/cybersecurity-almanac-2019/, accesat la data de 1 august 2019

Pe lângă nevoia de a asigura un nivel crescut al securității tranzacțiilor bancare, modelul trebuie să fie eficient și din punct de vedere al costurilor. Investiția în modelul de evaluare a tranzacțiilor și de identificare a fraudelor nu trebuie să fie mai mare decât pierderea datorată fraudei. Tehnicile de machine-learning, pe lângă îmbunătățirea semnificativă a detectării fraudei, ajută și la minimizarea costurilor, datorită capacității de a identifica mici anomalii în seturi foarte mari și dinamice (non-statice) de date.

Machine Learning (ML) reprezintă o metodă de analiză a datelor ce automatizează dezvoltarea unui model analitic. Este o ramură a inteligenței artificiale bazată pe ideea că mașinile (calculatoarele) ar trebui să fie capabile să învețe și să se adapteze prin propria experiență. În ceea ce privește domeniul bancar, Machine Learning este utilizat în două scopuri cheie: identificarea de informații importante în date și prevenirea/detectarea fraudei.

Detectarea fraudelor prin învățare automată poate fi provocator din mai multe puncte de vedere, precum:

- ✓ crearea de tipare pentru detectarea fraudelor din volume foarte mari de date;
- ✓ utilizarea tehnicilor de învățare automată în detectarea și modelarea unor algortimi și strategii pe baza tiparelor de fraudă existente;
- ✓ identificarea de strategii noi asociate comportamentului neobișnuit al infractorilor, ce își perfecționează continuu tehnicile, motiv pentru care există o cerință ca metodele de detectare a fraudei să poată evolua în consecință.

Lucrarea are la bază o analiză extensivă a literaturii de specialitate cu privire la frauda bancară din mediul online, care a constituit fundament în selectarea celor mai bune tehnici și algoritmi de construire a unui model eficient de machine-learning în vederea detectării cu o precizie ridicată a cazurilor frauduloase.

Obiectivele cercetării

Scopul fundamental al tezei este: *Dezvoltarea unui model stivă de patru clasificatori pentru performanțe predictive ridicate în ceea ce privește detectarea fraudelor bancare*. Prin urmare, am urmărit trei obiective principale pentru a atinge scopul tezei:

O1. Identificarea principalelor tipuri de fraudă și a profilului celui care fraudează, pe baza analizei literaturii de specialitate privind fraudele bancare, a aspectelor psihologice care favorizează apariţia fraudelor şi a modalităţilor de manifestare, a statisticilor de fraudă conform cu tiparele identificate.

O2. Crearea cadrului de clasificare a fraudelor prin explorarea potențialului oferit de metodele de învățare automată pentru rezolvarea problemelor de clasificare a fraudelor bancare, a celor legate de distribuția inegală a tranzacțiilor frauduloase și de overfitting.

O3. Crearea unui model eficient și stabil, care să ofere cele mai bune rezultate în conformitate cu complexitățile generate de fraudele bancare din mediul online, precum și să dezvăluie limitările modelelor de prognoză pentru îmbunătățiri și dezvoltări viitoare.

Structura tezei de doctorat

Lucrarea a fost structurată în 5 capitole, abordând atât aspecte conceptuale, cât și pragmatice, experimentale, după cum urmează:

Capitolul 1, **Concepte de bază și cercetări conexe privind frauda bancară**, oferă o privire de ansamblu, pe baza analizei literaturii de specialitate, asupra principalelor tipuri de fraudă, a aspectelor psihologice care favorizează apariţia fraudelor şi a modalităţilor de manifestare. Pentru susținerea acestor aspecte, vor fi prezentate și o serie de statistici privind frauda.

Capitolul 2, **Tehnici și instrumente ale mecanismelor de detectare a tranzacțiilor frauduloase**, delimitează teoretic diferitele tehnici utilizate în analiza tranzacțiilor frauduloase, și anume: tehnici de analiză exploratorie a tranzacțiilor, tehnici hibride de eșantionare în ceea ce privește distribuirea dezechilibrată a tranzacțiilor, tehnici de prevenire a fenomenului de overfitting.

Capitolul 3, **Algoritmi machine-learning folosiți în detectarea fraudei bancare,** prezintă, pe baza literaturii de specialitate, cei patru algoritmi de clasificare utilizați în construirea modelului

final, și anume: XGBoost (eXtreme Gradient Boosting,), Random Forest, LightGBM (Light Gradient Boosting Machine) și MLPClassifier (Multi-Layer Perceptron Classification).

Capitolul 4, **Aplicarea experimentală a algoritmilor machine în detectarea fraudei bancare**, se concentrează pe metodologia de aplicare experimentală a algoritmilor. Metodologia descrie întregul proces de construcție a modelului, și anume: analiza exploratorie a datelor, pregătirea datelor, construirea modelului și evaluarea lui.

Capitolul 5, **Rezultatele cercetării**, sintetizează întreaga activitate de cercetare, inclusiv principalele rezultate și contribuții. De asemenea, aduce în discuție limitările, potențialele îmbunătățiri ale studiului, precum și direcțiile pentru o posibilă extindere a cercetării în viitor.

Concluzii finale, limite ale cercetării și perspective viitoare, prezintă într-o manieră sintetizată rezultatele studiului întreprins în această lucrare.

Metodologia cercetării

Prima parte a tezei (capitolele 1, 2, și 3) are la bază o metodologie de tip **calitativă / documentară** prin care s-au analizat și discutat diferite studii legate de problema fraudei bancare din mediul online.

Demersul științific **calitativ / documentar** are la bază o metodologie a cercetării care a cuprins în mod separat:

- studiul bibliografic al unor materiale generale şi de specialitate (lucrări ştiinţifice, manuale, articole etc. din literatura de specialitate străină şi din ţară, menţionate în bibliografia tezei);
- analiza unor fenomene specifice şi concepte din domeniul lucrării de cercetare (în acest sens, au fost păstrate pe parcursul tezei o serie de termeni în limba engleză întrucât sunt consacrați pentru literatura de specialitate a algoritmilor machine-learning, și anume*: overfitting, oversampling, undersampling, recall, Standard Scaler* etc.);
- prezentarea şi realizarea unei cercetări ştiinţifice de tip documentar-selectiv privind modul de informare, cunoaştere şi diseminare a informaţiilor legate de fraudele bancare şi activitatea de investigare a acestora.

Studiul s-a bazat pe analiza literaturii de specialitate, având ca primă etapă determinarea celor mai relevante surse bibliografice pentru îndeplinirea scopului stabilit. Aceste surse bibliografice au inclus articole științifice din domeniul sistemelor informatice bancare și lucrări publicate în

volumele conferințelor de specialitate (conference proceedings). În acest scop, am căutat în următoarele baze de date:

- ScienceDirect IEEE Xplore
- ACM Digital Library AIS eLibrary
- Google Scholar.

Căutarea s-a efectuat în titlul sau rezumatul lucrării și s-a bazat pe utilizarea următoarelor cuvinte cheie: “internet banking”, “online banking”, “risk”, “bank variables”, “machine-learning algorithms”, “classification algorithms”, “fraud detection clasification issues”, “machine-learning techniques”, “evaluation techniques”. Rezultatele căutării au fost limitate la un număr de 255 de articole din domeniul sistemelor informatice publicate între 2010 – 2020, cu relevanță mai mare pentru detectarea fraudei bancare din mediul online în vederea proiectării modelului de tip stivă de patru clasificatori pentru performanțe predictive ridicate în ceea ce privește detectarea fraudelor bancare. De asemenea, o parte din conținutul tezei a fost validat și prin seria de articole publicate în reviste cotate ISI și BDI (*secțiunea* ***Bibliografie de autor***).

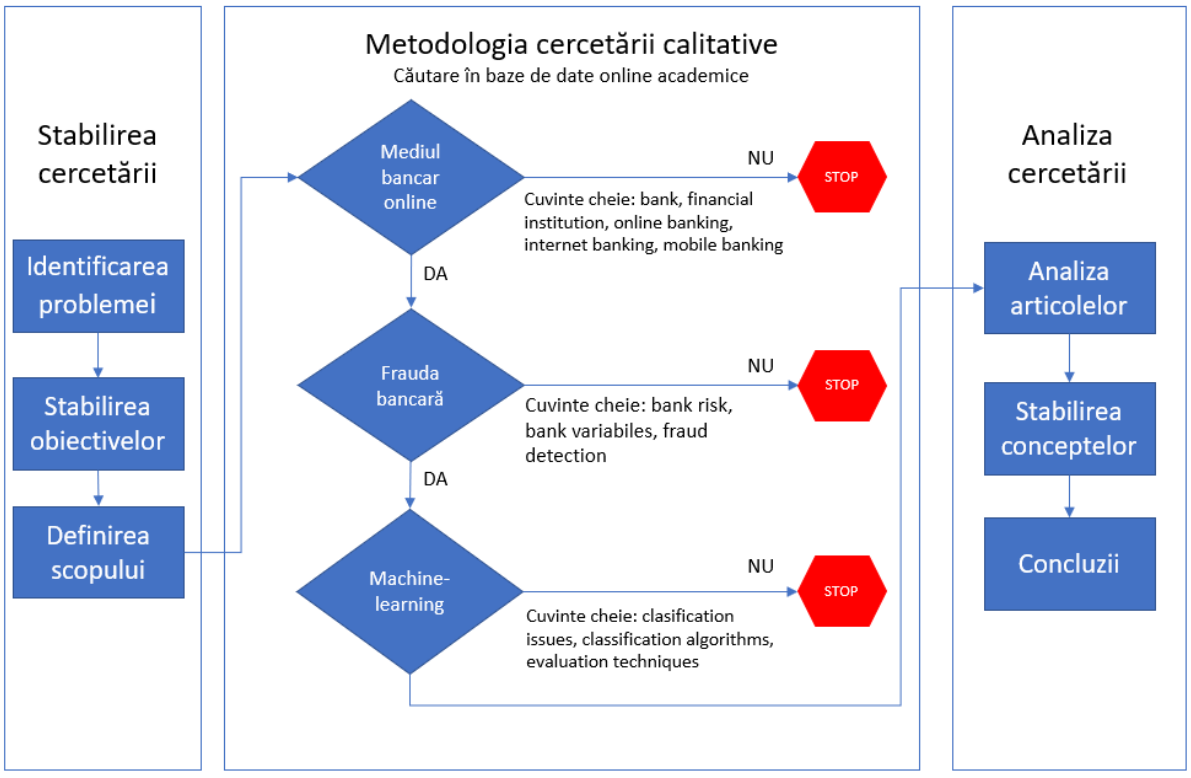

Figura 1 Cadrul metodologiei cercetării calitative / documentare

Figura ***1 Cadrul metodologiei cercetării calitative / documentare*** ilustrează etapele parcurse în cadrul **metodologiei cercetării calitative / documentare**, evidențiind cuvintele cheie folosite în cercetare, precum și criteriile utilizate în clasificarea unui articol relevant pentru cercetare: mediul bancar online, frauda bancară și machine-learning. Cuvintele cheie și criteriile utilizate au determinat o diversitate mai mare a surselor bibliografice selectate.

Plecând de la informațiile colectate prin **metodologia calitativă / documentară**, în partea a doua a tezei, se va aplica o **metodologie de tip cantitativă / experimentală** prin care vor fi prezentate experimentele efectuate, precum și rezultatele obținute.

Demersul științific **cantitativ / experimental** are la bază o metodologie a cercetării care a cuprins:

- planificarea studiului cercetării, dezvoltat în prima parte a tezei prin:
  - identificarea problemei
  - stabilirea obiectivelor și a scopului cercetării
  - stabilirea ipotezelor ce vor sta la baza construirii modelului
- stabilirea cadrului de lucru pentru experiment:
  - analiza problemei din punctul de vedere al literaturii de specialitate
  - identificarea bazelor de date pentru extragerea tranzacțiilor bancare (date referitoare la predicția riscurilor din băncile germane, date referitoare la tranzacții cu carduri bancare etc)
  - analiza metodelor și algoritmilor machine-learning în vederea selectării celor mai potriviți pentru detectarea fraudei bancare
- efectuarea experimentului prin:
  - organizarea și analiza datelor
  - pregătirea datelor ce vor fi utilizate în experiment
  - alegerea modelelor ce răspund cerințelor problemei identificate
  - antrenarea modelelor identificate
  - obținerea rezultatelor în urma antrenării modelelor
  - construirea modelului final
- analiza rezultatelor cercetării prin:
  - evaluarea modelului final

- utilizarea (sau oferind modalități de utilizare a) modelului în practică și / sau în cercetările viitoare.

Plecând de la aceste considerente, în ultima parte a tezei ne vom concentra atenția asupra metodologiei de aplicare a algoritmilor, prin detalierea acesteia în manieră experimentală, precum și asupra analizei rezultatelor cercetării.

Sinteza capitolelor din teză

Capitolul 1 Concepte de bază și cercetări conexe privind frauda bancară

Luând în considerare tendința de expansiune a fenomenului infracțional bancar, analiza literaturii de specialitate din acest capitol va urmări identificarea particularităților și principalelor aspecte ce caracterizează frauda bancară din perspectiva tehnologică, dar și a profilului celui care fraudează. Aceste elemente vor fi folosite la aplicarea experimentală a algoritmilor și dezvoltarea modelului propriu. Pentru susținerea acestor aspecte, au fost analizate și o serie de statistici privind frauda.

Din analiza statisticilor prezentate s-a putut constata faptul că frauda online este mult mai răspândită decât frauda tradițională și reprezintă o provocare mult mai mare. Acest lucru se datorează noilor tehnologii și transformării Internetului într-un instrument cotidian de comunicare și derulare a afacerilor, ce permit comiterea fraudei online la **o scară mai mare**, cu **o viteză accelerată,** având mult mai **multe victime** decât prin frauda tradițională. Astfel, am stabilit următoarele criterii de clasificare a fraudelor:

- după locul realizării acestora, identificate și codificate în teză prin **F1-F4**;
- în funcție de instrumentele folosite, identificate și codificate în teză prin **F5-F10**;
- după mediul de derulare, identificate și codificate în teză prin **F11-F26**;
- în funcție de numărul de participanți, identificate și codificate în teză prin **F27-F29**;
- în funcție de activitatea bancară și de falsificare a documentelor bancare, în vederea delapidarii, identificate și codificate în teză prin **F30-F33**.

În privința comiterii fraudelor **F1-F33** am putut identifica o serie de variabile, codificate prin **V1-V17**. Din aceste variabile, am putut utiliza în modelul experimental, variabilele codificate prin: **V1, V5, V9**, **V14**, **V15**.

În privința aspectelor legate de psihologia socială privind frauda am identificat că accentul cel mai mare se pune pe ***autoritate*** și ***declanșatoare emoționale***. La toate acestea se adaugă **profilul infractorului,** care tinde să fie ***bărbat cu vârsta cuprinsă între 36 și 55 de ani***, lucrează cu organizația "victimă" de mai bine de șase ani și deține o funcție executivă în operațiuni, finanțe sau management general.

Analiza literaturii de specialitate din cadrul acestui capitol a reprezentat o etapă premergătoare, necesară stabilirii obiectivelor cercetării, demersului metodologic și a direcțiilor de cercetare vizate în cadrul tezei.

Capitolul 2 Tehnici și instrumente ale mecanismelor de detectare a tranzacțiilor frauduloase

În cadrul acestui capitol, am urmărit stabilirea particularităților privind instrumentele și tehnicile utilizate în analiza și pregătirea datelor din metodologia de aplicare a algoritmilor. În urma cercetării întreprinse, analiza noastră s-a concentrat pe următoarele trei categorii de instrumente și tehnici:

- **instrumente specifice analizei exploratorie a datelor** – sunt utilizate în prima etapă din metodologia de aplicare a algoritmilor și au rolul de a înțelege datele și de a extrage ipoteze din datele brute, prin diferite reprezentări grafice avansate (***heatmap, boxenplot, distplot, geographic map, dashboard***);
- **tehnici hibride de eșantionare** în ceea ce privește distribuirea dezechilibrată a tranzacțiilor. Sunt aplicate în cea de-a doua etapă a metodologiei de aplicare a algoritmilor. Au rolul de a echilibra distribuția tranzacțiilor bancare în vederea obținerii unor rezultate performante în detectare. În cadrul acestui capitol se va realiza o descriere teoretică generală în ceea ce privește cele mai utilizate tehnici (***sampling, ensembling, cost-based, distance-based, hybrid, cost-sensitive***), însă doar tehnica de tip ensembling va fi folosită în experimentul final.
- **tehnici de prevenire a fenomenului de overfitting** folosite după eșantionarea din a doua etapă a metodologiei de aplicare a algoritmilor. Au rolul de a oferi o generalizare perfectă a modelului atât pe setul de antrenament, cât și pe cel de testare.

Particularitățile și concluziile acestui capitol, obținute pe baza analizei literaturii de specialitate, reprezintă cadrul general necesar aplicării algoritmilor.

Capitolul 3 Algoritmi machine-learning folosiți în detectarea fraudei bancare

În acest capitol, ne-am îndreptat atenția asupra identificării celor mai potriviți algoritmi pentru construirea modelului de tip stivă, având în vedere indicatorii de măsurare a performanței, respectiv: ***spațiul de memorie*** necesar stocării datelor prelucrate de către algoritm, ***timpul de execuție*** necesar prelucrărilor din cadrul algoritmului. Ambii indicatori se bazează pe următoarea ipoteză: ***volumul resurselor de calcul necesare depinde de volumul datelor de intrare (dimensiunea problemei) (I2)***.

Performanța algoritmilor din perspectiva spațiului de memorie presupune dezvoltarea unor structuri de date eficiente pentru ***comprimarea bazei de date*** și parcurgerea ***rapidă printr-o singură scanare a bazei de date***.

Pe lângă acești doi indicatori, un alt indicator cel puțin la fel de important este și ***costul*** necesar prelucrării datelor. Pentru a fi performant, un algoritm trebuie să asigure reducerea acestui cost, iar una din soluții este reprezentată de ***clasificarea tranzacțiilor și evaluarea doar a celor cu potențial fraudulos (I3).***

În contextul acestor indicatori, construirea modelului de tip stivă a avut în vedere următorii algoritmi de clasificare: **XGBoost** (eXtreme Gradient Boosting) din categoria algoritmilor de tip **ensemble**, arbori decizionali pe bază de randomizare (**Random Forest**), **LightGBM** (Light Gradient Boosting Machine) din categoria algoritmilor de tip **ensemble** și **MLPClassifier** (Multi-Layer Perceptron Classification) din categoria de **rețele neuronale**. Analiza acestor algoritmi, pe baza literaturii de specialitate, a urmărit determinarea indicatorilor de performanță, a parametrilor utilizați, utilitatea acestor algoritmi, precum și avantajele oferite în raport cu alți algoritmi din aceeași clasă.

Capitolul 4 Aplicarea experimentală a algoritmilor machine în detectarea fraudei bancare

În acest capitol, abordăm cercetarea experimentală prin care aplicăm algoritmii machine-learning pentru dezvoltarea modelului propriu de tip stivă. În acest scop, am folosit metodologia de aplicare a algoritmilor machine-learning, ce a presupus parcurgerea a patru etape de bază: analiza exploratorie a datelor, pregătirea datelor, construirea modelului și evaluarea acestuia, conform ***Figurii 2 Etapele implicate în demersul de aplicare a algoritmilor***. .

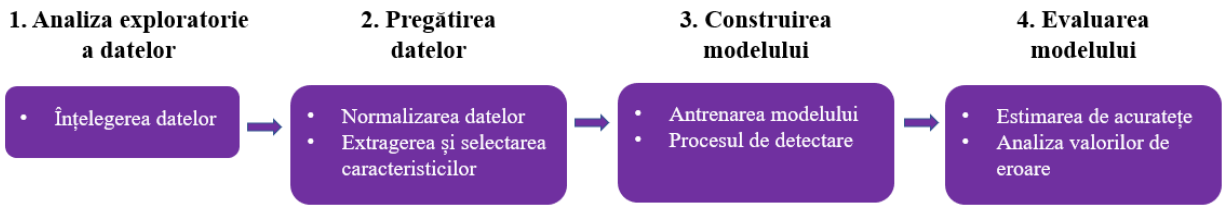

Figura 2 Etapele implicate în demersul metodologic de aplicare a algoritmilor

Descriem, pe scurt, succesiunea etapelor din demersul metodologic de aplicare a algoritmilor, menționând faptul că în cadrul fiecărui subcapitol din teză au fost evidențiate particularitățile și rezultatele experimentelor efectuate:

1. **Analiza exploratorie a datelor** a avut în vedere înțelegerea setului de date utilizat în experiment, precum și ipoteza identificată în capitolul **3 Algoritmi machine-learning folosiți în detectarea fraudei bancare**, conform căreia se ***urmărește reducerea costurilor necesare clasificării tranzacțiilor și evaluarea doar a celor cu potențial fraudulos (I3).***

2. **Pregătirea datelor** a fost realizată, în principal, pentru identificarea și reducerea caracteristicilor irelevante din seturile de date și pregătirea lor pentru detectarea cazurilor de fraudă. Rezultatul etapei de pre-procesare constă în setul de date final care conține doar informații importante pentru modelul nostru de învățare automată. În această etapă se urmărește ipoteza **I2 (**din capitolul **3 Algoritmi machine-learning folosiți în detectarea fraudei bancare,** conform căreia ***volumul resurselor de calcul necesare depinde de volumul datelor de intrare).***

3. **Etapa de construire a modelului de detectare a fraudei,** bazată pe tehnici de învățare automată, necesită o atenție sporită în alegerea celor mai bune tehnici de învățare automată, precum și în setarea valorilor parametrilor ce îmbunătățesc capabilitățile algoritmilor în oferirea de rezultate performante.

4. Ultimul pas a constat în **evaluarea modelului realizat,** prin compararea rezultatelor cu cele ale algoritmilor individuali (XGBoost, LightGBM, Random Forest, MLPClassifier), având în vedere valorile pentru metrica AUC, precum și cele regăsite în matricea de confuzie. Rezultatele au în vedere ipoteza de cercetare **I1** (din capitolul **2.2.1.2 Tehnici de tip cost-sensitive**), conform căreia se urmărește ***reducerea costurilor datorate de erorile de tip false positive / false negative***.

În încheierea capitolului, au fost prezentate într-o manieră comparativă rezultatele obținute atât în antrenările individuale ale celor patru algoritmi, cât și ale modelului de tip stivă.

Capitolul 5 Rezultatele cercetării

Acest ultim capitol al tezei a avut în vedere analiza rezultatelor obținute în urma aplicării experimentale a algoritmilor din punctul de vedere al ipotezelor identificate pe parcursul cercetării. În acest sens, s-a realizat un studiu extensiv al literaturii de specialitate ce a avut drept scop consolidarea și confirmarea acestor rezultate pentru fiecare algoritm în parte. Analiza efectuată a scos în evidență capacitatea performantă a algoritmilor selectați în ceea ce privește detectarea fraudei bancare online.

Atât rezultatele obținute de noi, cât și cele din literatura de specialitate au indicat faptul că metodologia propusă, ce combină mai mulți algoritmi de învățare automată asistată, **funcționează semnificativ mai bine în comparație cu algoritmii individuali**. Concret, rezultatele obținute de noi au indicat o valoare **a acurateții modelului de tip stivă de 85%** egală cu cea obținută de algoritmul LightGBM. Însă numărul cazurilor de tip *false positive* a fost mult mai mic decât cel obținut de toți algoritmii utilizați în construirea modelului de tip stivă, și anume **841**, în detrimentul creșterii numărului de cazuri de tip *false negative*, unde cel mai mic număr a fost obținut de către algoritmul XGBoost în **XGB-A2**, și anume **656**.

În ceea ce privește algoritmii individuali, s-a putut consta faptul că cele mai bune valori ce satisfac jumătate din ipoteza **I1**, și anume reducerea cazurilor de tip *false negative*, în detrimentul creșterii cazurilor de tip *false positive*, au fost obținute de către algoritmul MLPClassifier (**MLP-A1**), urmat de algoritmul LightGBM (**LGB-A1**). Ipoteza I1 fiind satisfăcută în întregime de algoritmul Random Forest (**RF-A3**). În termeni de acurateție, cele mai bune rezultate au fost generate de algoritmul Random Forest cu un procent de 97%, urmat de LightGBM cu un procent de 85%. Astfel că, forma actuală a modelului prezentat poate fi utilizată în instituțiile bancare pentru ***a reduce costurile generate de alertarea incorectă a clienților asupra acelor cazuri de tip "fraudă" ce în realitate sunt non-fraudă***. Pentru acele cazuri care în realitate sunt fraudă, dar modelul le marchează ca non-fraudă, este nevoie de intervenția angajaților în verificarea / marcarea corectă a acestora. Pentru cercetări viitoare, se urmărește **reglarea hyper-parametrilor meta-clasificatorului** în vederea obținerii ***unei valori mari pentru metrica de acurateție și o scădere semnificativă a cazurilor de tip false positive / false negative***. S-a putut observa faptul că pentru a reduce cazurile de tip *false positive / false negative* trebuie să avem în vedere o adâncime a arborelui cât mai mică (**în cazul nostru a fost egal cu 5**). Cu cât aceasta crește cu atât modelul va

fi mai complex și va tinde către apariția fenomenului de tip overfitting; precum și un număr relativ mic de arbori (**în cazul nostru a fost egal cu 250**).

În ceea ce privește **timpii de excuție,** analiza efectuată a scos în evidență performanța algoritmilor machine-learning (XGBoost, Random Forest, LightGBM), precum și a rețelei neuronale în detectarea fraudelor bancare (***spațiul de memorie alocat și timpul de execuție***). Concret, pentru cei trei algoritmi machine și pentru rețeaua neuronală s-au obținut următorii **timpi de execuție**, și anume: ***XGBoost – 20 minute, Random Forest – 10 minute, LightGBM - 79 minute pentru 10 iterații complete, MLPClassifier – 3 secunde***. Din datele exemplificate, se poate deduce faptul că rețeaua neuronală este cea mai performantă din punctul de vedere al celor doi indicatori, ***spațiu și timp de execuție***, dar și a ***costului*** datorat clasificării eronate a tranzacțiilor, **705**. Fiind și unul dintre motivele pentru care s-a ales această rețea neuronală în construirea modelului final de tip stivă.

Rezultatele obținute pe parcursul cercetării, clasifică **modelul de tip stivă în categoria clasificatorilor ce oferă performanțe predictive ridicate în ceea ce privește detectarea fraudelor bancare,** datorită faptului că valorile obținute pentru cele două tipuri de erori, I și II, sunt destul de mici comparativ cu numărul de tranzacții analizate, și anume: **683 782 tranzacții**.

Concluzii finale, limite ale cercetării și perspective viitoare

Cercetarea întreprinsă a vizat problema detectării fraudelor bancare din mediul online din perspectiva modelelor machine-learning. În acest context, scopul principal al lucrării a fost dezvoltarea unui **model de tip stivă de patru clasificatori pentru performanțe predictive ridicate în ceea ce privește detectarea fraudelor bancare**. Scopul a fost realizat prin examinarea extensivă a literaturii de specialitate cu privire la frauda bancară din mediul online, care a constituit fundament în selectarea celor mai bune tehnici și algoritmi de construire a unui model eficient de machine-learning în vederea detectării cu o precizie ridicată a cazurilor frauduloase.

Metodologia cercetării a avut la bază atât o metodologie de tip **calitativă / documentară** prin care s-au analizat și discutat diferite studii cu referire la frauda bancară și la cei mai eficienți algoritmi machine-learning, cât și o metodologie de tip **cantitativă / experimentală** prin care au fost realizate experimentele și evaluate rezultatele obținute.

Metodologia de tip **calitativă / documentară** a avut în vedere un număr de 255 de articole din domeniul sistemelor informatice publicate între 2010 – 2020, cu relevanță mai mare pentru detectarea fraudei bancare din mediul online în vederea proiectării **modelului de tip stivă de patru clasificatori** pentru performanțe predictive ridicate în ceea ce privește detectarea fraudelor bancare. Astfel că, la baza comiterii fraudelor identificate și codificate prin **F1-F32** am putut distinge o serie de variabilele, codificate prin **V1-V16**. Din aceste variabile, un impact foarte mare asupra fraudei din mediul online a fost prezentat de: *lipsa existenţei unor măsuri adecvate de verificare a tranzacţiilor (V6), accesul la sistemele online și securitatea sistemelor (V5), caracteristicile mediului online de a fi public, partajabil și ușor de „spart" (V11), datele de autentificare (V1), volumul și natura activităților contului de internet banking (V12), versiunea motorului de căutare (V13), versiunea sistemului de operare (V14), suma depusă în cont (V8)*. Variabile de care se vor ține cont în cercetărilor viitoare și în aplicațiile ce vor putea fi dezvoltate din conceptul modelului de tip stivă.

De asemenea, având în vedere provocările complexe ale fraudei bancare, și anume: *distribuția non-staționară a datelor, distribuția extrem de dezechilibrată a claselor și disponibilitatea câtorva tranzacții etichetate cu atributul fraudă de anchetatori*, pe baza analizei literaturii de specialitate am putut identifica și detalia cele mai utilizate instrumente de analiză exploratorie avansate: **heatmap, boxenplot,** și **distplot;** precum și cele mai utilizate tehnici de soluționare a claselor inegal distribuite, și anume: tehnici de tip **resampling**; de tip **cost-sensitive**; **kernel-based** și tehnici de **învățare activă.** Din studiile prezentate s-a putut observa faptul că, deși fiecare instrument și tehnică vine cu o serie de avantaje, nu toate pot fi aplicate într-o manieră eficientă asupra problemei analizate în această lucrare. Plecând de la tehnicile identificate, au fost apoi prezentați în detaliu algoritmii machine-learning ce au putut fi combinați într-un model de tip stivă. Astfel că, au fost prezentați patru algoritmi de învățare automată asistată, și anume: **XGBoost, Random Forest, LightGBM și MLPClassifier.** Pentru fiecare s-a prezentat cadrul conceptual, cei mai importanți parametri care, prin modificarea valorilor, conduc la rezultate performante în termeni de acuratețe și precizie, dar și avantajele oferite.

Având în vedere analiza literaturii de specialitate întreprinsă în metodologia de tip **calitativă / documentară**, metodologia de tip **cantitativă / experimentală** a vizat efectuarea experimentului și analiza rezultatelor cercetării asupra unui set de date de tranzacții de comerț electronic ce aparține celei mai importante companii de servicii de plată din lume, Vesta. Asupra acestui set de

date, au fost realizate o serie de experimente ce au inclus procesele de antrenare (**XGB-A1 – XGB-A4, RF-A1 – RF-A4, LGB-A1 – LGB-A2, MLP-A1 – MLP-A2**) și optimizare individuală a celor patru algoritmi (**XGB-O1 – XGB-O4, RF-O1 – RF-O4, LGB-O1 – LGB-O2, MLP-O1 – MLP-O2**), precum și procesele de antrenare a modelului de tip stivă (**ST-A1 – ST-A9**). Demersul metodologic de aplicare a algoritmilor a avut în vedere patru ipoteze de cercetare:

- ***ML asigură reducerea costurilor datorate de clasificarea incorectă a tranzacțiilor;***
- ***volumul resurselor de calcul necesare depinde de volumul datelor de intrare (dimensiunea problemei);***
- ***ML permite reducerea costurilor prin clasificarea tranzacțiilor și evaluarea doar a celor cu potențial fraudulos;***
- **C*alitatea datelor de intrare influențează calitatea rezultatelor oferite de modelul ML*.**

Rezultatele obținute indică faptul că modelul propus ce combină mai mulți algoritmi de învățare automată, funcționează semnificativ mai bine în comparație cu algoritmii individuali. Concret, valoarea acurateții modelului de tip stivă a fost de **85%** egală cu cea obținută de algoritmul LightGBM, însă numărul cazurilor de tip *false positive* a fost mult mai mic decât cel obținut de toți algoritmii utilizați în construirea modelului de tip stivă, și anume **841**, în detrimentul creșterii numărului de cazuri de tip *false negative*, unde cel mai mic număr a fost obținut de către algoritmul XGBoost în **XGB-A2**, și anume **656**. Deși modelul propus a returnat rezultate bune în ceea ce privește detectarea fraudei bancare din mediul online, există totuși anumite **limitări** și anume:

- dezvoltarea unui algoritm care să acopere pașii descriși în primele două etape din metodologia de aplicare experimentală a algoritmilor, și anume ***analiza exploratorie a datelor*** și ***pregătirea datelor***; algoritm care să transfere modelului de tip stivă o variantă uniformă a datelor
- reglarea hyper-parametrilor meta-clasificatorului în vederea obținerii unei valori mari pentru metrica de acuratețe și o scădere semnificativă a cazurilor de tip *false positive / false negative*
- acces la o bază de date care să conțină tranzacții reale
- resurse de calcul performante în vederea efectuării rapide a proceselor de optimizare a algoritmilor
- combinarea unui număr mai mare de algoritmi în modelul de tip stivă.

Depășirea acestor limitări ar putea face din acest model de tip stivă, un model potrivit pentru ***dezvoltarea unei aplicații de protejare a cardului bancar***, ce va putea fi utilizată de întreaga populație. Aplicația va putea trimite în timp real notificări pe telefonul utilizatorului pentru absolut toate tranzacțiile ce se vor efectua cu cardul bancar (retrageri de la ATM, tranzacții de plată din cont, tranzacții online etc.). Notificările primite vor trebui aprobate pentru ca tranzacția în cauză să poată fi efectuată. Aprobarea va putea fi efectuată prin diverse metode, și anume: cod pin, amprentă digitală, recunoașterea feței, a vocii, a irisului etc. În această manieră clonarea cardurilor nu va mai putea fi realizată întrucât aceasta va aduce cu sine și o clonare a telefonului utilizatorului.

O altă aplicabilitate a modelului ar putea fi regăsită în ***crearea unui roboțel în Internet Banking*** care să ofere ajutor utilizatorilor în diferite acțiuni întreprinse în contul bancar online precum și un istoric al fiecărei tranzacții efectuate.

În concluzie, se poate menționa faptul că există mai multe oportunități de extindere a activității prezentate în această lucrare prin includerea unor metode suplimentare care să aibă în vedere ***o valoare mare a acurateții și valori mici pentru cazurile de tip false positive / false negative, dar și o securitate sporită în ceea ce privește efectuarea tranzacțiilor bancare.***

Referințe bibliografice

1. *„Managing the Business Risk of Fraud – A Practical Guide”*, the Institute of Internal Auditors, the American Institute of Certified Public Accountants and the Association of Certified Fraud Examiners, 2008
2. Abellán, J. and Castellano, J. G., *A comparative study on base classifiers in ensemble methods for credit scoring*, Expert Systems with Applications, vol 73, pp. 1-10, 2017, doi: https://doi.org/10.1016/j.eswa.2016.12.020
3. AccentureSecurity, *The cost of cybercrime*, https://www.accenture.com/_acnmedia/pdf-96/accenture-2019-cost-of-cybercrime-study-final.pdf, 2019, accesat la data de 15 martie 2019
4. ActionFraud, *Mobile phone fraud*, https://www.actionfraud.police.uk/protect-yourself/mobile-phone-fraud, accesat la data de 16 martie 2018
5. Adali, E. and Mubarek, A. M., *Multilayer perceptron neural network technique for fraud detection*, International Conference on Computer Science and Engineering (UBMK), Antalya, pp. 383-387, 2017, doi: 10.1109/UBMK.2017.8093417
6. AdGlow, *E-Commerce: Men spend more than women*, 2020, https://www.adglow.com/blog/pt-br/e-commerce-men-spend-more-than-women, accesat online la data de 9 mai 2020
7. Alginahi, Y., *Preprocessing techniques in character recognition*, In Character Recognition. InTech, 2010, doi: 10.5772/9776
8. Australian Payments Network, *Fraud statistics*, https://www.auspaynet.com.au/resources/fraud-statistics , 2019, accesat la data de 16 martie 2018
9. Bahnsen, A. C., Stojanovic, A., Aouada, D. and Ottersten, B., *Cost sensitive credit card fraud detection using bayes minimum risk*, Machine Learning and Applications (ICMLA), 2013 12th International Conference on, volume 1, pages 333–338. IEEE, 2013, doi: 10.1109/ICMLA.2013.68
10. Bahnsen, A. C., Villegas, S., Aouada, D. and Ottersten B., *Fraud Detection by Stacking Cost-Sensitive Decision Trees*, Data Science for Cyber-Security, pp 1-15, 2017, doi: https://doi.org/10.1142/9781786345646_012

11. Banerjee , R., Bourlarishi, G., Chen, S., Kashyap, M., Purohit, S. and Battipaglia, J., *Comparative Analysis of Machine Learning Algorithms through Credit Card Fraud Detection*, New Jersey's Governor's School of Engineering and Technology July 27, 2018, https://soe.rutgers.edu/sites/default/files/imce/pdfs/gset-2018/Comparative%20Analysis%20of%20Machine%20Learning%20Algorithms%20through%20Credit%20Card%20Fraud%20Detection.pdf
12. Bart, B., Veronique, V. V. and Wouter, V., *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection*, SAS Institute Inc., Cary, North Carolina, USA, 2015, ISBN: 978-1-119-13312-4
13. Batista, G. E. A. P. A., Bazzan, A. L. C. and Monard, M. C., *Balancing Training Data for Automated Annotation of Keywords: a Case Study*, 2003
14. Batista, G. E. A. P. A., Prati, R. C., and Monard, M. C., *A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data*, ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, pp. 20-29, 2004, doi: https://doi.org/10.1145/1007730.1007735
15. Batista, G., Carvalho, A. and Monard, M., *Applying one-sided selection to unbalanced datasets*, MICAI 2000: Advances in Artificial Intelligence, pp 315–325, 2000
16. Bhattacharyya, S., Jha, S., Tharakunnel, K. and Westland, J.C., *Data mining for credit card fraud: a comparative study*, Decision Support Systems, vol 50, issue 3, pp. 602-613, 2011, doi: https://doi.org/10.1016/j.dss.2010.08.008
17. Biau, G., *Analysis of a Random Forests Model*, Journal of Machine Learning Research, vol 13, pp. 1063-1095, 2012
18. Błaszczynski, J. and Stefanowski, J., *Neighbourhood Sampling in Bagging for Imbalanced Data*, Neurocomputing, vol 150, pp 529–542, 2015, doi: https://doi.org/10.1016/j.neucom.2014.07.064
19. Bondarici, D. and Cătuţi, C., *Criminalitatea bancară în România*, Bucureşti, Ed.Regina din Arcadia, 2005, p.114
20. Bondarici, D. and Ghinea, N., *Utilizarea frauduloasă a instrumentelor de plată*, Bucureşti, Ed.Lucman, 2005, p.60

21. Branco, P., Torgo, L. and Ribeiro, R. P., *A Survey of Predictive Modelling under Imbalanced Distributions*, CoRR abs/1505.01658, 2015, doi: arXiv:1505.01658
22. Breiman, L., *Random forests*, Machine Learning, vol 45, pp. 5–32, 2001
23. Bromium, *Hyper-connected web of profit emerges, as global cybercriminal revenues hit $1.5 trillion annually*, https://www.bromium.com/press-release/hyper-connected-web-of-profit-emerges-as-global-cybercriminal-revenues-hit-1-5-trillion-annually/, 2018, accesat la data de 15 martie 2018
24. Brownlee, J., *How To Prepare Your Data For Machine Learning in Python with Scikit-Learn*, https://machinelearningmastery.com/prepare-data-machine-learning-python-scikit-learn/, 2016, accesat la data de 9 decembrie 2019
25. Brownlee, J., *How to tune the number and size of decision trees with XGBoost in Python* (2016), http://machinelearningmastery.com/tune-number-size-decision-trees-XGBoost-python/
26. Brownlee, J., *Overfitting and Underfitting With Machine Learning Algorithms*, https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/, 2019, accesat la data de 27 martie 2020
27. Bujala, A., *Gender differences in internet usage*, Acta Universitatis Lodziensis, Folia Sociologica 43, 2012
28. Burez, J. and Poel, D. V., *Handling class imbalance in customer churn prediction*, Experts System with Applications, vol. 36, pp. 4626-4636, 2009, doi: https://doi.org/10.1016/j.eswa.2008.05.027
29. Cao, K., Wei, C., Gaidon, A., Arechiga, N. and Ma, T., *Learning imbalanced datasets with label distribution-aware margin loss*, NeurIPS, 2019, doi: arXiv:1906.07413
30. Carmona, P., Climent, F. and Momparler, A., *Predicting failure in the U.S. banking sector: An extreme gradient boosting approach*, International Review of Economics and Finance, vol. 61, pp. 304-323, 2019, doi: https://doi.org/10.1016/j.iref.2018.03.008
31. Carneiro, N., Figueira, G. and Costa, M., *A data mining based system for credit-card fraud detection in e-tail*, Decision Support Systems, vol 95, pp. 91-101, 2017, doi: https://doi.org/10.1016/j.dss.2017.01.002
32. Caruana, R., Lawrence, S. and Giles, L., *Overfitting in neural nets: backpropagation, conjugate gradient, and early stopping*, Advances in Neural Information Processing

Systems 13, Papers from Neural Information Processing Systems (NIPS), Denver, CO, USA, pp. 402-408, 2000

33. Cawley, G. C. and Talbot, N. L. C., *Preventing Over-Fitting during Model Selection via Bayesian Regularisation of the Hyper-Parameters*, Journal of Machine Learning Research 8, pp. 841–861, 2007
34. CCN, *$731 Million Stolen from Crypto Exchanges in 2018: Can Hacks be Prevented?*, https://www.ccn.com/731-million-stolen-from-crypto-exchanges-in-2018-can-hacks-be-prevented/, 2018, accesat la data de 15 martie 2019
35. Chairi, I., Alaoui, S. and Lyhyaoui, A., *Sample selection based active learning for imbalanced data*, International Conference on Signal-Image Technology and InternetBased Systems, pp. 645-651, 2014, doi: 10.1109/SITIS.2014.118.
36. Charleonnan, A., *Credit card fraud detection using RUS and MRN algorithms*, Management and Innovation Technology International Conference, MITicon, IEEE, pp. MIT-73–MIT–76, 2016, doi: 10.1109/MITICON.2016.8025244.
37. Chawla, N. V., Bowyer, K.W., Hall, L. O. and Kegelmeyer, W. P., *SMOTE: Synthetic minority over-sampling technique*, Journal of Artificial Intelligence Research, vol 16, pp. 321-357, 2002, https://arxiv.org/pdf/1106.1813.pdf
38. Chawla, N. V., Japkowicz, N. and Kotcz, A., *Editorial: special issue on learning from imbalanced data sets*, ACM SIGKDD Explorations Newsletter, vol 6, pp. 1–6, 2004, doi: 10.1145/1007730.1007733
39. Chen, J., Shaw, S. L., Yu, H., Lu, F., Chai, Y. and Jia, Q., *Exploratory data analysis of activity diary data: a space-time GIS approach*, Journal of Transport Geography, vol. 19, issue 3, pp. 394-404, 2011, doi: https://doi.org/10.1016/j.jtrangeo.2010.11.002
40. Chen, T. and Guestrin, C., *XGBoost: A Scalable Tree Boosting System*, arXiv e-prints, page arXiv:1603.02754, 2016
41. CloudFlare, *What is DdoS attack?*, https://www.cloudflare.com/learning/ddos/what-is-a-ddos-attack/, accesat la data de 4 august 2029
42. CNet, *Equifax data breach may affect nearly half the US population*, https://www.cnet.com/news/equifax-data-leak-hits-nearly-half-of-the-us-population/ , 2017, accesat la data de 19 decembrie 2019

43. CoinDesk, *$13.5 Million in Crypto Stolen From Token Platform Bancor*, https://www.coindesk.com/token-platform-bancor-goes-offline-following-security-breach/, 2018, accesat la data de 15 martie 2019
44. CoinDesk, *Veritaseum Founder Claims $8 Million in ICO Tokens Stolen*, https://www.coindesk.com/veritaseum-founder-claims-8-million-ico-token-stolen/, 2017, accesat la data de 15 martie 2019
45. Comisia Europeană, *Notă informativă privind indicatorii de fraudă pentru FEDR, FSE şi FC*, pp. 7
46. Conceptul de triunghi al fraudei a fost iniţiat de cercetătorul în domeniul fraudelor Dr. Donald R. Cressey, „The Handbook of Fraud Deterrence", de Harry Cendrowski, James P. Martin şi Louis W. Petro, 2010
47. Cowen, M., *What's next for wearables in 2018?,* https://www.itproportal.com/features/whats-next-for-wearables-in-2018/ , 2018, accesat la data de 15 martie 2019
48. Criminisi, A., Shotton, J. and Konukoglu, E., *Decision Forests for Classification, Regression, Density Estimation*, Manifold Learning and Semi-Supervised Learning, Technical report MSR-TR-2011-114, 2011
49. Cui, Y., Jia, M., Lin, T.-Y., Song, Y. and Belongie, S., *Class-balanced loss based on effective number of samples*, CVPR, 2019, https://openaccess.thecvf.com/content_CVPR_2019/papers/Cui_Class-Balanced_Loss_Based_on_Effective_Number_of_Samples_CVPR_2019_paper.pdf
50. Cybercrime Magazine – S. Morgan, *2019 Cybersecurity Almanac: 100 Facts, Figures, Predictions And Statistics*, https://cybersecurityventures.com/cybersecurity-almanac-2019/, accesat la data de 1 august 2019
51. Das, S., *MyEtherWallet Warns of [Another] Hack, Urges Hola Users to Move Funds*, https://www.ccn.com/myetherwallet-warns-of-another-hack-urges-hola-users-to-move-funds/, 2018, accesat la data de 15 martie 2019
52. De Mast, J. and Kemper, B. P. H., *Principles of Exploratory Data Analysis in Problem Solving: What Can We Learn from a Well-Known Case?*, Quality Engineering, vol. 21, pp. 366-375, 2009, doi: 10.1080/08982110903188276

53. de Oliveira, F. A., Nobre, C. N. and Zárate, L. E., *Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index–Case study of PETR4*, Petrobras, Brazil. Expert Systems with Applications, vol. 40, issue 18, pp.7596-7606, 2013, doi: https://doi.org/10.1016/j.eswa.2013.06.071
54. Deevy, M., Lucich, S. and Beals, M., *Scams, Schemes and Swindles. A review of consumer financial fraud research*, Stanford Center on Longevity, 2012, http://longevity.stanford.edu/2012/11/19/scams-schemes-and-swindles-a-review-of-consumer-financial-fraud-research/
55. Deloitte, *Profiling the fraudster: understanding the threat of insider fraud*, https://www2.deloitte.com/mt/en/pages/risk/articles/mt-risk-article-profiling-the-fraudster.html, accesat la data de 4 august 2019
56. Devi Meenakshi, B., Janani, B., Gayathri, S. and Indira, N., *Credit card fraud detection using random forest*, International Research Journal of Engineering and Technology (IRJET), vol. 06, Issue 03, Mar 2019, e-ISSN: 2395-0056
57. Dhankhad, S., Mohammed, E. and Far, B., *Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study*, IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, pp. 122-125, 2018, doi: 10.1109/IRI.2018.00025.
58. Ding, S., Mirza, B., Lin, Z., Cao, J., Lai, X., Nguyen, T. V. and Sepulveda, J., *Kernel based online learning for imbalance multiclass classification*, Neurocomputing, vol 277, pp. 139–148, 2018, doi: 10.1016/j.neucom.2017.02.102
59. Domingos, P., *Metacost: a general method for making classifiers cost-sensitive*, Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 155–164. ACM, 1999
60. Donovan, F. and Bernier, K., *Cyber Crime Fighters: Tales from the trenches*, QUE, Indianapolis, Indiana, 2009, p. 18, ISBN:978-0-7897-3922-3
61. Drummond, C. and Holte, R., *Exploiting the cost (in)sensitivity of decision tree splitting criteria*, Proceedings of the 17th International Conference on Machine Learning, pp. 239- 246, 2000
62. ElectronicTransactionsAssociation, *Voice Commerce, Software Solutions Top Trends in 2018*, Finds Market Spotlight, https://www.electran.org/voice-commerce-software-

solutions-top-trends-in-2018-finds-market-spotlight/, 2017, accesat la data de 15 martie 2019

63. Elhassan, T., Aljurf, M., Al-Mohanna, F. and Shoukri, M., *Classification of Imbalance Data using Tomek Link (T-Link) Combined with Random Under-sampling (RUS) as a Data Reduction Method*, Journal of Informatics and Data Mining ISSN 2472-1956, vol.1 no.2:11, pp. 3, 2016, doi: 10.4172/2229-8711.S1111
64. Elite Data Science, *Overfitting in Machine Learning: What It Is and How to Prevent It*, https://elitedatascience.com/overfitting-in-machine-learning, accesat la data de 6 august 2019
65. Elkan, C., *The Foundations of Cost-Sensitive Learning*, Proceedings of the Seventeenth International Joint Conference of Artificial Intelligence, pp. 973-978. Seattle, Washington: Morgan Kaufmann, 2001
66. Embrechts, P., *Quantities Model for Operational Risk Extreme*, Dependence and Aggregation, Journal of Banking Finance, vol 30, issue 10, pp. 2635-2658, 2006
67. Ertekin, S., Huang, J. and Giles, C., *Active learning for class imbalance problem*, ACM Conference on Information Retrieval (SIGIR), 2007
68. Esmaily, J. and Moradinezhad, R., *Intrusion detection system based on multilayer perceptron neural networks and decision tree*, International conference on Information and Knowledge Technology (IKT), Urmia, pp. 1-5, 2015, doi: 10.1109/IKT.2015.7288736.
69. Estabrooks, A., Jo, T. and Japkowicz, N., *A Multiple Resampling Method for Learning from Imbalanced Data Sets*, Computational Intelligence, vol. 20, pp. 18-36, 2004, doi: https://doi.org/10.1111/j.0824-7935.2004.t01-1-00228.x
70. Europol, *Two criminal groups dismantled for laundering eur 2.5 million through smurfing and cryptocurrencies*, https://www.europol.europa.eu/newsroom/news/two-criminal-groups-dismantled-for-laundering-eur-25-million-through-smurfing-and-cryptocurrencies, accesat la data de 15 martie 2019
71. Evaluation Division, *Performance measurement definitions*, https://eca.state.gov/files/bureau/performance_measurement_definitions.pdf, accesat online la data de 12 martie 2020

72. Experian, *Fraud Management Insights 2017*, https://www.experian.com.vn/wp-content/uploads/2017/12/fraud-management-insights-2017.pdf , 2017, accesat la data de 19 decembrie 2019
73. Express, *ANDROID WARNING - Google Play apps infect millions of phones with dangerous malware*, https://www.express.co.uk/life-style/science-technology/854529/Android-warning-Google-Play-malware-ExpensiveWall, 2017, accesat la data de 19 decembrie 2019
74. Fang, Y., Zhang, Y. and Huang, C., *Credit Card Fraud Detection Based on Machine Learning*, Computers, Materials & Continua, vol.61, no.1, pp.185-195, 2019, doi: 10.32604/cmc.2019.06144
75. Fashoto, S. G., Owolabi, O., Adeleye, O. and Wandera, J., *Hybrid Methods for Credit Card Fraud Detection Using K-means Clustering with Hidden Markov Model and Multilayer Perceptron Algorithm*, British Journal of Applied Science & Technology, 13(5): 1-11, 2016, Article no.BJAST.21603 ISSN: 2231-0843, NLM ID: 101664541
76. Federal Reserve Systems, *The Federal Reserve Payments Study: 2017 Annual Supplement*, https://www.federalreserve.gov/newsevents/pressreleases/files/2017-payment-systems-study-annual-supplement-20171221.pdf , 2017, accesat la data de 19 decembrie 2019
77. Fernandez-Navarro, F., Hervas-Martinez, C. and Gutierrez, P. A., *A dynamic over-sampling procedure based on sensitivity for multi-class problems*, Pattern Recognition, vol. 44, issue 8, pp. 1821–1833, 2011, doi: https://doi.org/10.1016/j.patcog.2011.02.019
78. Feroz, E. H., Kwon, T. M., Pastena, V. and Park, K. J., *The efficacy of red flags in predicting the SEC's targets: an artificial neural networks approach*, în revista International Journal of Intelligent Systems in Accounting, Finance, and Management, vol. 29, pp. 145-157, 2000, doi: 10.1002/1099-1174(200009)9:3<145::AID-ISAF185>3.0.CO;2-G
79. Fluent, *Devices & Demographics*, 2016, http://www.fluentco.com/wp-content/uploads/2016/01/Fluent2_DevicesandDemographics_2016.pdf, accesat online la 9 februarie 2020

80. Foley, S., Karlsen, J. R. and Putnins, T. J., *Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed Through Cryptocurrencies?*, Review od Financial Studies, Forthcoming, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3102645, 2018
81. Galar, M., Fernando, A., Barrenechea, E., Business, H. and Herrera, F., *A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches*, IEEE Transactions on Systems, Man, and Cybernetics- Part C: Applications And Review, vol. 42, no. 4, pp. 463-484, 2012, doi: 10.1109/TSMCC.2011.2161285.
82. Galarnyk, M., *Understanding Boxplots*, https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51, 2018, accesat la data de 9 decembrie 2019
83. García, S. and Herrera, F., *Evolutionary Undersampling for Classification with Imbalanced Datasets: Proposals and Taxonomy*, Evolutionary Computation, vol 17, no. 3, pp. 275–306, 2009, doi: 10.1162/evco.2009.17.3.275
84. Geeksforgeeks, *Cross Validation in Machine Learning*, https://www.geeksforgeeks.org/cross-validation-machine-learning/, accesat la data de 27 martie 2020
85. Georgescu, M., *Some issues about risk management for e-banking*, FUTURE OF BANKING AFTER THE YEAR 2000 IN THE WORLD AND IN THE CZECH REPUBLIC, S. Poloucek, D. Stavarek, eds., Karvina: Silesian University, 2005, available at SSRN: https://ssrn.com/abstract=903419
86. Georgescu, M. and Georgescu, I., *The Emergence of Electronic Payment Systems for the Growth of E-Business*, International Symposium Economics and Management of Transformation, 2004, available at SSRN: https://ssrn.com/abstract=903622
87. GlobalWebIndex, *Device. GlobalWebIndex's flagship report on device ownership and usage*, https://libranda.com/wp-content/uploads/2019/07/Report_on_devices_ownership_usage_July2019.pdf , accesat la data de 4 august 2019
88. Gomez, J. A., Arévalo, J., Paredes, R. and Nin, J., *End-to-end neural network architecture for fraud scoring in card payments*, Pattern Recognition Letters, volume 105, pp 175–181, 2018, doi: https://doi.org/10.1016/j.patrec.2017.08.024

89. Goodin, D., *Malicious apps with >1 million downloads slip past Google defenses twice*, https://arstechnica.com/information-technology/2017/09/malicious-apps-with-1-million-downloads-slip-past-google-defenses-twice/, 2017, accesat la data de 19 decembrie 2019
90. Gupta, N., Dhankar, A., Ranawat, K. S. and Nautiyal, S., *Credit Card Fraud Detection using Rough Sets and Artificial Neural Network*, Journal of Computer Applications ISSN: 0974 –1925, vol5, Issue EICA2012-1, February 10, 2012
91. Gupta, P., *Regularization in Machine Learning*, https://towardsdatascience.com/regularization-in-machine-learning-76441ddcf99a , 2017, accesat la data de 27 martie 2020
92. Guyon, I., Bennett, K., Cawley, G., Escalante, H. J., Escalera, S., Ho, T. K., Macià, N., Ray, B., Saeed, M., Statnikov, A. and Viegas, E., *Design of the 2015 ChaLearn AutoML Challenge*, Proceedings of the 2015 International Joint Conference on Neural Networks. ChaLearn, pp. 1–8, 2015, doi: 10.1109/IJCNN.2015.728076
93. Haibo, H. and Garcia, E. A., *Learning from Imbalanced Data*, IEEE Transactions On Knowledge And Data Engineering, vol. 21, no. 9, September 2009, pp. 1263-1284, https://www.cs.utah.edu/~piyush/teaching/ImbalancedLearning.pdf
94. Han, J., Pei, J. and Kamber, M., *Data mining: concepts and techniques*, Elsevier, 2011
95. HashedOut, *33 Alarming Cybercrime Statistics You Should Know in 2019*, https://www.thesslstore.com/blog/33-alarming-cybercrime-statistics-you-should-know/, 2019, accesat la data de 18 decembrie 2019
96. He, H. and Garcia, E., *Learning from Imbalanced Data*, IEEE Transactions on Knowledge and Data Engineering, vol. 21, pp. 1263–1284, September 2009IBM Security, *The inside story on botnets*, https://www.ibm.com/downloads/cas/V3YJVYZX, accesat la data de 4 august 2019
97. Heer, J., Bostock, M. and Ogievetsky, V., *A tour through the visualization Zoo*, ACMqueue. Available at http://queue.acm.org/detail.cfm?id=1805128, 2010
98. Hoens, T. R. and Chawla, N. V., *Imbalanced Datasets: From Sampling to Classifiers*, Imbalanced Learning: John Wiley & Sons, Inc., pp. 43-59, 2013, doi: 10.1002/9781118646106.ch3

99. Hurbean L., *Considerations about implementing an accounting and financial management software*, revista Contabilitate şi informatică de gestiune, ASE Bucureşti, nr. 22/2007

100. https://trustvesta.com/

101. Huang, J., Bottou, L. and Giles, C. L., *Learning on the border: Active learning in imbalanced data classification*, Proceedings of the Sixteenth ACMConference on Conference on Information and Knowledge Management, Lisbon, Portugal, pp. 127–136, November 2007, doi: 10.1145/1321440.1321461

102. InfoSecurity Group – P. Muncaster, *Digital Ad Bot Fraud Set to Reach $6.5 Billion*, https://www.infosecurity-magazine.com/news/digital-ad-bot-fraud-set-to-reach/, accesat la data de 1 august 2019

103. James, G., Witten, D., Hastie, T. and Tibshirani, R., *An introduction to statistical learning*, (6th ed.), Springer, New York (2015)

104. Japkowicz, N. and Stephen, S., *The Class Imbalance Problem: A Systematic Study*, Intelligent Data Analysis, vol 6, pp. 429–449, 2002, doi: 10.3233/IDA-2002-6504

105. Jay Lakshmi, T. and Prasad, C. S. R., *A Study on Classifying Imbalanced Datasets*, First International Conference on Networks and Soft Computing (ICNSC), IEEE, Guntur, pp. 141-145, 2014, doi: 10.1109/CNSC.2014.6906652.

106. Jensen, D. D. and Cohen, P. R., *Multiple Comparisons in Induction Algorithms*, Machine Learning, vol. 38, no. 3, pp. 309-338, 2000, doi: https://doi.org/10.1023/A:1007631014630

107. Jo, T. and Japkowicz, N., *Class imbalances versus small disjuncts*, ACM SIGKDD Explorations Newsletter, vol. 6 no. 1, pp. 40–49, 2004, doi: https://doi.org/10.1145/1007730.1007737

108. Jonnalagadda, V., Gupta, P. and Sen, E., *Credit card fraud detection using Random Forest Algorithm*, International Journal of Advance Research, Ideas and Innovations in Technology, vol 5, issue 2, 2019

109. Juniper Research, *IS OEM-PAY THE FUTURE OF CONTACTLESS?*, http://www.kreditwesen.de/system/files/content/inserts/2017/is_oem_pay_the_future_of_contactless_whitepaper_52669.pdf, accesat la data de 4 august 2019

110. Juniper Research, *Online payment fraud whitepaper*, http://www.experian.com/assets/decision-analytics/white-papers/juniper-research-online-payment-fraud-wp-2016.pdf , accesat la data de 19 decembrie 2019

111. Kang, P. and Cho, S., *EUS SVMs: Ensemble of UnderSampled SVMs for Data Imbalance Problems*, Proceeding International Conference on Neural Information Processing, ser. Lecture Notes in Computer Science, pp. 837-846, 2006

112. Karaa, A. and Krichene, A., *Credit-risk assessment using support vectors machine and multilayer neural network models: a comparative study case of a tunisian bank*, Journal of Accounting and Management Information Systems, vol. 11, pp. 587–620, 2012

113. Karla, S., *Young consumers most cautious about online payments*, 2017, https://www.creditcards.com/credit-card-news/young-consumers-most-cautious-online-purchases.php, accesat la data de 9 februarie 2020

114. Kasa, N., Dahbura, A., Ravoori, C. and Adams, S., *Improving Credit Card Fraud Detection by Profiling and Clustering Accounts*, Systems and Information Engineering Design Symposium (SIEDS), doi: 10.1109/SIEDS.2019.8735623, 2019

115. Kassem, R. and Higson, A., *The New Fraud Triangle Model*, Journal of Emerging Trends in Economics and Management Science, pp. 191 – 195, June 2012

116. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. And Liu, T.-Y., *LightGBM: a highly efficient gradient boosting decision tree*, Advances in Neural Information Processing Systems, Vol. 30, pp 3146-3154, 2017

117. Kelly, P. and Hartley, C. A., *Casino gambling and workplace fraud: A cautionary tale for managers*, Management Research Review, vol. 33, issue 3, pp.224-239, 2010, doi: 10.1108/01409171011030381

118. Kerdprasop, N. and Kerdprasop, K., *On the Generation of Accurate Predictive Model from Highly Imbalanced Data with Heuristics and replication Technologies*, International Journal of Bio-Science and Bio-Technology, vol. 4, pp. 49- 64, 2012

119. Khandelwal, S., *Hackers Stole $32 Million in Ethereum; 3rd Heist in 20 Days*, http://thehackernews.com/2017/07/ethereum-cryptocurrency-hacking.html, 2017, accesat la data de 15 martie 2019

120. Khoshgoftaar, T. M., Fazelpour, A., Dittman, D. J. and Napolitano, A., *Ensemble vs. Data Sampling: Which Option Is Best Suited to Improve Classification Performance of Imbalanced Bioinformatics Data?*, Proceedings of the IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), Vietri sul Mare, Italy, pp. 705–712, 9–11 November 2015, doi: 10.1109/ICTAI.2015.106

121. KnowBe4, *KnowBe4 2019 Security Threats and Trends Report – October 2019*, https://blog.knowbe4.com/knowbe4-2019-security-threats-and-trends-report-october-2019, accesat la data de 5 decembrie 2019

122. Koehrsen, W., *Histograms and Density Plots in Python*, https://towardsdatascience.com/histograms-and-density-plots-in-python-f6bda88f5ac0, 2018, accesat la data de 9 decembrie 2019

123. Kolli, C. S. and Devi, T.U., *Isolation Forest and Xg Boosting For Classifying Credit Card Fraudulent Transactions*, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume 8 Issue 8, June 2019

124. Koskivaara, E., *Artificial neural networks in auditing: state of the art*, în revista The ICFAI Journal of Audit Practice 1, pp. 12-33, 2004

125. Koutanaei, F. N., Sajedi, H. and Khanbabaei, M., *A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring*, Journal of Retailing and Consumer Services, vol 27, pp. 11-23, 2015, doi: https://doi.org/10.1016/j.jretconser.2015.07.003

126. KPMG, *Global profiles of the fraudster: Technology enables and weak controls fuel the fraud*, https://assets.kpmg/content/dam/kpmg/pdf/2016/05/profiles-of-the-fraudster.pdf, 2016, accesat la data de 19 decembrie 2019

127. Kuhn, M. and Johnson, K., *Applied Predictive Modeling*, Springer New York, 2013

128. Lacurezeanu, R., Mocean, L. and Lacurezeanu, M., Electronic banking services, STUDIA NEGOTIA - Ediţia nr.1 din 2002, http://studia.ubbcluj.ro/arhiva/abstract.php?editie=NEGOTIA&nr=1&an=2002&id_art=2236

129. Lakshmi, S. V. S. S. and Kavila, S. D., *Machine Learning For Credit Card Fraud Detection System*, International Journal of Applied Engineering Research ISSN 0973-

4562, vol 13, no 24, pp. 16819-16824, 2018, https://www.ripublication.com/ijaer18/ijaerv13n24_18.pdf

130. Laurikkala, J., *Improving Identification of Difficult Small Classes by Balancing Class Distribution*, Procceeding Conference AI in Medicine in Europe: Artificial Intelligence Medicine, pp. 63-66, 2001

131. LightGBM, *Algorithm & comparison with XGBoost*, http://www.ashukumar27.io/LightGBM/, accesat la data de 9 decembrie 2019

132. LightGBM, *Features*, https://lightgbm.readthedocs.io/en/latest/Features.html, accesat la data de 9 decembrie 2019

133. LightGBM, *Parameters Tuning*, https://lightgbm.readthedocs.io/en/latest/Parameters-Tuning.html, accesat la data de 9 decembrie 2019

134. Ling, C. X. and Sheng, V. S., *Cost-sensitive Learning and the Class Imbalanced Prob*lem, Encyclopedia of Machine Learning, Ed. Springer, Berlin, Germany, 2008

135. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y. and Alsaadi, F. E., *A survey of deep neural network architectures and their applications*, Neurocomputing, vol 234, pp. 11-26, 2017, doi: https://doi.org/10.1016/j.neucom.2016.12.038

136. Liu, X. Y., Wu, J. and Zhou, Z. H., *Exploratory Under Sampling for Class Imbalance Learning*, IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS, pp. 965-969, 2006

137. Liu, Y., Yu, X., Huang, J. X. and An, A., *Combining integrated sampling with SVM ensembles for learning from imbalanced datasets*, Information Processing and Management, vol. 47, pp. 617–631, 2011, doi: https://doi.org/10.1016/j.ipm.2010.11.007

138. Lokanan, M. E., *Challenges to the fraud triangle: Questions on its usefulness*, Accounting Forum 39, pp 201-224, 2015

139. Ma, X., Sha, J., Wang, D., Yu, Y., Yang, Q. and Niu, X., *Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning*, Electronic Commerce Research and Applications, vol. 31, pp 24-39, September–October 2018, doi: https://doi.org/10.1016/j.elerap.2018.08.002

140. Maalouf, M. and Trafalis, T. B., *Robust weighted kernel logistic regression in imbalanced and rare events data*, Computational Statistics and Data Analysis, vol. 55, issue 1, pp. 168-183, 2011, doi: https://doi.org/10.1016/j.csda.2010.06.014

141. Maratea, A., Petrosino, A. and Manzo, M., *Adjusted F-measure and kernel scaling for imbalanced data learning*, Information Sciences, vol. 257, pp. 331-341, 2013, doi: https://doi.org/10.1016/j.ins.2013.04.016

142. Marques, J. F. O., *Risk Analysis in Money Laundering A Case Study*, Instituto Superior Técnico Lisbon, Portugal, pp. 3, 2015

143. Memoria, F., *Classic Ether Wallet Falls Victim to a Social Engineering Hacker*, https://www.cryptocoinsnews.com/classic-ether-wallet-falls-victim-to-a-social-engineering-hacker/, 2017, accesat la data de 15 martie 2019

144. Meșniță ,G., Oprea, D., and Dumitriu, F., *Data Processing Cycle in the context of Internet of Things*, IXth International Conference Globalization and Higher Education in Economics and Business Administration, GEBA, 2016

145. Meyer, D., Leisch, F. and Hornik, K., *The support vector machine under test*, Neurocomputing, vol 55, issue 1-2, pp. 169-186, 2003, doi: https://doi.org/10.1016/S0925-2312(03)00431-4

146. Mishra, A., *Metrics to evaluate your Machine Learning Algorithm*, https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234, accesat online la data de 23 martie 2020

147. Mishra, M. K. and Dash, R., *A comparative study of chebyshev functional link artificial neural network, multi-layer perceptron and decision tree for credit card fraud detection*, International Conference on Information Technology, Bhubaneswar, pp. 228-233, 2014, doi: 10.1109/ICIT.2014.25.

148. Moinescu, B. and Codrilașcu, A., *Administrarea riscului operațional*, pp. 180

149. Moneywise, *How to fight fraudsters: the psychology of scams*, https://www.moneywise.co.uk/work/everyday-life/how-fight-fraudsters-psychology-scams, accesat la data de 4 august 2019

150. Monika, S., Venkataramanamma, K., Pritto Paul, P., and Usha, M., *Credit Card Fraud Detection using Random Forest Algorithm*, International Journal of Research in

Engineering, Science and Management, vol 2, issue 3, ISSN (Online): 2581-5792, 2019 , https://www.ijresm.com/Vol.2_2019/Vol2_Iss3_March19/IJRESM_V2_I3_32.pdf

151. Montague, D., *Essentials of Online Payment Security and Fraud Prevention*, New York, John Wiley & Sons, Inc., 2011, doi:10.1002/9781118386750

152. Muppavarapu, V., Rajendran, A. and Vasudevan, S., *Phishing Detection using RDF and Random Forests*, The International Arab Journal of Information Technology, vol. 15, no. 5, September 2018, https://iajit.org/PDF/September%202018,%20No.%205/10600.pdf

153. Nadim, A. H., Sayem, I.M., Mutsuddy, A. and Chowdhury, M.S., *Analysis of Machine Learning Techniques for Credit Card Fraud Detection*, 2019 International Conference on Machine Learning and Data Engineering (iCMLDE)At: Taiwan Taipei, 2019, doi: 10.1109/iCMLDE49015.2019.00019

154. Nakaya, T., and Yano, K., *Visualising Crime Clusters in a Space-time Cube: An Exploratory Data-analysis Approach Using Space-time Kernel Density Estimation and Scan Statistics*, Transactions in GIS, vol. 14, pp. 223-239, 2010, doi: https://doi.org/10.1111/j.1467-9671.2010.01194.x

155. Nami, S. and Shajari, M., *Cost-sensitive payment card fraud detection based on dynamic random forest and k-nearest neighbors*, Expert System Applied, vol 110, pp. 381–392, 2018, doi: https://doi.org/10.1016/j.eswa.2018.06.011

156. Navlani, A., *Understanding Random Forests Classifiers in Python*, https://www.datacamp.com/community/tutorials/random-forests-classifier-python, 2018, accesat la data de 9 decembrie 2019

157. Neapolitan, A., *Classification Techniques for Noisy and Imbalanced Data*, Dis. Florida Atlantic University, 2009

158. NetGuardians, *Digital Banking Fraud: Best Practice for Technology-Based Prevention*, pp 9

159. Nirusha, K. and Krishnaiah, R. V., *An effective approach towards creditcard deception discovery methods*, International Journal of Computerand Electronics Research, vol. 2, issue 3, June 2013

160. Niu, X., Wang, L. and Yang, X., *A Comparison Study of Credit Card Fraud Detection: Supervised versus Unsupervised*, 2019, doi: arXiv:1904.10604

161. Oprea, D., *Globalizarea şi riscul securităţii informaţiilor*, Fenomene şi procese cu risc major la scară naţională, Editura Academiei Române, Bucureşti, ISBN: 973-27-2250-7, pp. 57-79, 2004

162. Paris, G., Robilliard, D. and Fonlupt, C., *Exploring Overfitting in Genetic Programming*, Artificial Evolution, International Conference, Evolution Artificielle, Ea 2003, Marseilles, France, October 2004. DBLP, pp.267-277, doi: 10.1007/978-3-540-24621-3_22

163. Patil, S., Nemade, V. and Soni, P. K., *Predictive Modelling For Credit Card Fraud Detection Using Data Analytics*, Procedia Computer Science, vol. 132, pp. 385-395, 2018, doi: https://doi.org/10.1016/j.procs.2018.05.199

164. Phua, C., Alahakoon, D. and Lee, V., *Minority Report in Fraud Detection: Classification of Skewed Data*, ACM SIGKDD Explorations Newsletter, vol 6, no 1, 2004

165. Phua, C., Lee, V., Smith, K. and Gayler, R., *A comprehensive survey of data mining based fraud detection research*, Artificial Intelligence Review, pp 1-14, 2005

166. Popa, D., *Document intern. Politica anti-fraudă a unei bănci*, http://hymerion.ro/2011/05/23/document-intern-politica-anti-frauda-a-unei-banci.html, accesat la data de 24 martie 2019

167. Prabhakara, E., Kumar, M. N., Ponnar, K., Suresh, A. and Jayandhiran, R., *Credit card fraud detection using boosted stacking*, South Asian Journal of Engineering and Technology, 2019

168. Puig, A. O. and Mansilla, E. B., *Evolutionary rule-based systems for imbalanced data sets*, Soft Computing, vol. 13, pp. 213-225, 2009, DOI: 10.1007/s00500-008-0319-7

169. PwC, *Fighting fraud: A never-ending battle*, https://www.pwc.com/gx/en/forensics/gecs-2020/pdf/global-economic-crime-and-fraud-survey-2020.pdf, 2020, accesat la data de 4 februarie 2020

170. Radware Blog, *A Quick History of IoT Botnets*, https://blog.radware.com/uncategorized/2018/03/history-of-iot-botnets/, accesat la data de 15 martie 2018

171. Rahimikia, E., Mohammadi, S., Rahmani, T. and Ghazanfari, M., *Detecting corporate tax evasion using a hybrid intelligent system: A case study of Iran*, International Journal of Accounting Information Systems, vol. 25, pp. 1–17, May 2017, doi: https://doi.org/10.1016/j.accinf.2016.12.002

172. Ramentol, E., Verbiest, N., Caballero, Y. and Cornelis, C., *SMOTE-FRST: A New Resampling Method Using Fuzzy Rough Set Theory*, proc: WSPC, 2012, doi: 10.1142/9789814417747_0128

173. Rao, V., *Introduction to Classification & Regression Trees (CART)*, https://www.datasciencecentral.com/profiles/blogs/introduction-to-classification-regression-trees-cart, 2013, accesat la data de 27 martie 2020

174. Raschka, S., *About Feature Scaling and Normalization – and the effect of standardization for machine learning algorithms*, https://sebastianraschka.com/Articles/2014_about_feature_scaling.html, 2014, accesat la data de 9 decembrie 2029

175. Raschka, S., *Mlxtend 0.9.0*, pp. 99-100

176. Rasmussen, C. E. and Williams, C. K. I., *Gaussian Processes for Machine Learning*, MIT Press, pp. 266, ISBN 026218253X, 2006

177. Redman, J., *Hacked South Korean Bitcoin Exchange Yapizon Offers IOU*s, https://news.bitcoin.com/hacked-korean-bitcoin-exchange-yapizon-offers-ious/, 2017, accesat la data de 15 martie 2019

178. Research and Markets, *Analysis on Fraud & Security in Online Payments*, Worldwide (2017-2023), https://www.globenewswire.com/news-release/2019/07/15/1882827/0/en/Analysis-on-Fraud-Security-in-Online-Payments-Worldwide-2017-2023.html , accesat la data de 4 august 2019

179. Richard, O. D., Peter, E. H. and David, G. S., *Pattern classification*, John Wiley & Sons Inc., New York, 2012

180. Roy, B., *All about Categorical Variable Encoding*, https://towardsdatascience.com/all-about-categorical-variable-encoding-305f3361fd02, 2019, accesat la data de 9 decembrie 2019

181. Rusch, J. J., *The "Social Engineering" of Internet Fraud*, https://web.archive.org/web/20080617150031/http://www.isoc.org/isoc/conferences/inet/99/proceedings/3g/3g_2.htm#r19 , accesat la data de 4 august 2019

182. Rushin, G., Stancil, C., Sun, M., Adams, S. and Beling, P., *Horse Race Analysis in Credit Card Fraud—Deep Learning, Logistic Regression, and Gradient Boosted Tree*, Systems and Information Engineering Design Symposium (SIEDS), 2017, doi: 10.1109/SIEDS.2017.7937700

183. Saeed, S. and Ong, H. C., *Performance of SVM with Multiple Kernel Learning for Classification Tasks of Imbalanced Datasets*, Pertanika J. Sci. & Technol, vol. 27, no. 1, pp. 527 – 545, 2019, http://www.pertanika.upm.edu.my/Pertanika%20PAPERS/JST%20Vol.%2027%20(1)%20Jan.%202019/30%20JST-1133-2018.pdf

184. SafeInternetBanking.be, *Fraud techniques*, https://www.safeinternetbanking.be/en/fraud-techniques, accesat la data de 15 martie 2018

185. SafeInternetBanking.be, *Shoulder surfing*, https://www.safeinternetbanking.be/en/fraud-techniques/shoulder-surfing, accesat la data de 16 martie 2018

186. Sanusi, Z. M., Rameli, M. N. F. and Ias, Y. M., *Fraud Schemes in the Banking Institutions: Prevention Measures to Avoid Severe Financial Loss*, Procedia Economics and Finance, vol 28, pp. 107 – 113, 2015, doi: https://doi.org/10.1016/S2212-5671(15)01088-6

187. Sarma, G. and Singh, P. K., *"Internet Banking: Risk Analysis and Applicability of Biometric Technology for Authentication"*, International Journal of Pure and Applied Sciences and Technology, 1(2), pp. 67-78, 2010, ISSN 2229 - 6107

188. Scikit-learn, *sklearn.ensemble.RandomForestClassifier*, https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html , accesat la data de 9 decembrie 2019

189. Scikit-learn, *sklearn.model_selection.GridSearchCV*, https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html, accesat la data de 27 martie 2020

190. Scikit-learn, *sklearn.model_selection.RandomizedSearchCV*, https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html, accesat la data de 27 martie 2020
191. Scikit-learn, *sklearn.model_selection.train_test_split*, https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html, accesat la data de 9 decembrie 2019
192. Scikit-learn, *sklearn.neural_network.MLPClassifier*, https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html, accesat la data de 9 decembrie 2019
193. Seaborn, *seaborn.boxenplot*, https://seaborn.pydata.org/generated/seaborn.boxenplot.html, accesat la data de 9 decembrie 2019
194. SecuredTouch, *Most Common Types of Mobile Fraud*, http://blog.securedtouch.com/most-common-types-of-mobile-fraud, accesat la data de 16 martie 2018
195. SecuredTouch, *Predictions for the Mobile Payment Fraud Landscape*, https://blog.securedtouch.com/predictions-for-the-mobile-payment-fraud-landscape-in-2018, 2017, accesat la data de 15 martie 2019
196. Son, H., Hyun, C., Phan, D. and Hwang, H. J., *Data analytic approach for bankruptcy prediction*, Expert Systems with Application, no 138, pp. 112 – 816, 2019, doi: https://doi.org/10.1016/j.eswa.2019.07.033
197. Spiridon, V., *Fraude cu carduri bancare*, https://cybersecuritytrends.ro/fraude-cu-carduri-bancare/, accesat la data de 15 martie 2018
198. Sputnik, *Un alt furt al miliardului a eșuat din cauza…ortografiei*, https://sputnik.md/world/20160311/5164998.html, accesat la data de 15 martie 2018
199. Statista, *Internet usage rate worldwide in 2019, by gender and region, 2020*, https://www.statista.com/statistics/491387/gender-distribution-of-internet-users-region/, accesat online la data de 9 februarie 2020
200. Statista, *Share of adults in the United States who are online almost constantly as of February 2019, by gender, 2020*, https://www.statista.com/statistics/496929/usa-adults-online-constantly-gender/, accesat la data de 9 februarie 2020

201. Statista, *Share of Americans who used VISA cards in the last 3 months in 2018, by age, 2020*, https://www.statista.com/statistics/350569/users-of-visa-credit-cards-usa/, accesat la data de 9 februarie 2020

202. Statista, *Share of individuals using online banking in the Netherlands from 2012 to 2019, by gender, 2020*, https://www.statista.com/statistics/575705/share-of-individuals-using-internet-banking-in-the-netherlands-by-gender/, accesat la data de 9 februarie 2020

203. Statista, *Share of individuals who used online banking in Belgium from 2006 to 2018, by gender, 2020*, https://www.statista.com/statistics/1088282/online-banking-usage-in-belgium-by-gender/, accesat la data de 9 februarie 2020

204. Stefanowski, J. and Wilk, S., *Improving rulebased classifiers induced by MODLEM by selective pre-processing of imbalanced data*, ECML/PKDD international workshop on rough sets in knowledge discovery (RSKD'2007), pp 54–65, 2007

205. Stoica, O. and Roman, A., *Consequences of the Financial and Economic Crisis on the Banking Activity in CEE Countries*, international conference „Financial and Economic Crisis: Causes, Consequences and the Future", Mendel University, Brno, Cehia, noiembrie 2010, http://vyzc.pef.mendelu.cz/en/konference/konferecnce_vyzc/conf/program;

206. Stoica, O., *Recent Trends in Bank Card Market in the European Union*, Annals of the University of Oradea: Economic Science (Analele Universităţii din Oradea, Seria Ştiinţe Economice), tom XVIII, 2009, volumul III, pp. 658-662, ISSN 1582-5450, indexat DOAJ, REPEC, EBSCO, SCIPIO, Cabell's, http://steconomice.uoradea.ro/anale/volume/2009/v3-finances-banks-and-accountancy/111.pdf;

207. Su, C.-H., Tu, F., Zhang, X., Shia, B.-C. and Lee, T.-S., *An ensemble machine learning based system for merchant credit risk detection in merchant mcc misuse*, Journal of Data Science,17(1), pp. 81 - 106, 2019, doi:10.6339/JDS.201901_17(1).0004

208. Sun, Y., Kamel, M. S., Wong, A. K. C. and Wang, Y., *Cost-sensitive boosting for classification of imbalanced data*, Pattern Recognition, vol. 40, issue 12, pp. 3358 – 3378, 2007, doi: https://doi.org/10.1016/j.patcog.2007.04.009

209. Thach, N. H., Rojanavasu, P. and Pinngern, O., *Cost-sensitive XCS Classifier System Addressing Imbalance Problems*, Fifth International Conference on Fuzzy Systems and Knowledge Discovery, IEEE, Shandong, pp. 132-136, 2008, doi: 10.1109/FSKD.2008.391

210. Thammasiri, D., Delen, D., Meesad, P. and Kasap, N., *A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition*, Expert Systems with Applications, vol. 41, issue 2, 2014, pp. 321-330, doi: https://doi.org/10.1016/j.eswa.2013.07.046

211. The Learning Machine, *Imbalanced Data Over-sampling & Under-sampling*, https://www.thelearningmachine.ai/imbalanced , accesat la data de 6 august 2019

212. The Nilson Report, *Issue 1096 | Oct 2016*, https://nilsonreport.com/publication_newsletter_archive_issue.php?issue=1096 , 2016, accesat la data de 16 martie 2018

213. *The state of crime: fraud, hacking and malware double the number of crimes committed in UK*, Computer Fraud & Security, February 2017, pp 1-3

214. Thiruvadi, S. and Patel, S. C., *Survey of Data-mining Techniques used in Fraud Detection and Prevention*, Information Technology Journal, vol. 10, pp. 710-716, 2011, doi: 10.3923/itj.2011.710.716

215. Tomanek, K. and Hahn, U., *Reducing class imbalance during active learning for named entity annotation*, Proceedings of the fifth International Conference on Knowledge Capture, pp. 105-112, 2009, doi: https://doi.org/10.1145/1597735.1597754

216. Tukey, J. W., *Exploratory data analysis*, Reading, MA: Addison-Wesley, 1977

217. Uyar, A., Bener, A., Ciracy, H. N. and Bahceci, M., *Handling the Imbalance Problem of IVF Implantation Prediction*, IAENG International Journal of Computer Science, 2006

218. Vasu, M. and Ravi, V., *A hybrid under-sampling approach for mining unbalanced datasets: applications to banking and insurance*, International Journal Data Mining Modelling and Management, vol. 3, pp. 75-105, 2011, doi: 10.1504/IJDMMM.2011.038812

219. Vlasselaer, V. V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M. and Baesens, B., *APATE: A novel approach for automated credit card transaction*

fraud detection using network-based extensions, Decision Support System, 75, pp. 38-48, 2015, doi: https://doi.org/10.1016/j.dss.2015.04.013

220. W3Counter, *Browser & Platform Market Share*, April 2020, https://www.w3counter.com/globalstats.php, accesat online la data de 9 mai 2020

221. Wallace, B. C., Small, K., Brodley, C. E. and Trikalinos, T. A., *Class Imbalance*, Redux. 2011 IEEE 11th International Conference on Data Mining, pp 754–763, 2011

222. Wang, B. X. and Japkowicz, N., *Boosting support vector machines for imbalanced data sets*, Knowledge and Information Systems, vol. 25, no. 1, pp. 1–20, 2010, doi: https://doi.org/10.1007/s10115-009-0198-y

223. Wang, D., Zhang, Y. and Zhao, Y., *LightGBM: An effective miRNA classification method in breast cancer patients*, International Conference on Computational Biology and Bioinformatics (ICCBB), pp. 7–11, 2017, doi: https://doi.org/10.1145/3155077.3155079

224. Wang, J., Liao, Y., Tsai, T. and Hung, G., *Technology based financial frauds in Taiwan: issue and approaches,* IEEE Conference on: Systems, Man and Cyberspace, pp 1120-1124, October 2006, doi: 10.1109/ICSMC.2006.384550

225. Wankhede, S. B., *Analytical Study of Neural Network Techniques: SOM, MLP and Classifier-A Survey*, IOSR Journal of Computer Engineering (IOSR-JCE), vol 16, issue 3, pp 86-92, e-ISSN: 2278-0661, 2014

226. Waqas, A., Gilal, A. R., Bhatti, Z. and Mahessar, A. W., *Investigating ANNs and Applications,* Indian Journal of Automation and Artificial Intelligence, vol. 1, issue 2 February 2013

227. Warde-Farley, D., Goodfellow, I. J., Courville, A. and Bengio, Y., *An empirical analysis of dropout in piecewise linear networks*, Computer Science, 2013, doi: arXiv:1312.6197

228. Weiczner, J., *Hackers Just Stole $7 Million in a Brazen Ethereum Cryptocurrency Heist*, http://fortune.com/2017/07/18/ethereum-coindash-ico-hack/, 2017, accesat la data de 15 martie 2019

229. Weiss, G. M. and Provost, F., *The Effect of Class Distribution on Classifier Learning: An Empirical Study*, Technical Report MLTR-43, Department of Computer Science, Rutgers Univ., 2001

230. Weiss, G. M., McCarthy, K. and Zabar, B., *Cost-sensitive learning vs. sampling: Which is best for handling unbalanced classes with unequal error costs?*, DMIN, vol 7, pp. 35–41, 2007

231. Wikipedia, *Kernel density estimation*, https://en.wikipedia.org/wiki/Kernel_density_estimation, accesat la data de 9 decembrie 2019

232. Wikipedia, *Naive Bayes classifiers*, https://en.wikipedia.org/wiki/Naive_Bayes_classifier#Testing , accesat la data de 9 decembrie 2019

233. Wikipedia, *Receiver operating characteristic*, https://en.wikipedia.org/wiki/Receiver_operating_characteristic#cite_note-fawcett-14, accesat la data de 27 martie 2019

234. Wikipedia, *Sparse dictionary learning*, https://en.wikipedia.org/wiki/Sparse_dictionary_learning, accesat la data de 27 martie 2020

235. Wilkinson, L. and Friendly, M., *The History of the Cluster Heat Map*, The American Statistician, vol 63, pp. 179-184, 2009, doi: 10.1198/tas.2009.0033

236. Wong, M. L., Seng, K. and Wong, P. K., *Cost-sensitive ensemble of stacked denoising autoencoders for class imbalance problems in business domain*, Expert Systems With Applications, vol. 141, 2020, doi: https://doi.org/10.1016/j.eswa.2019.112918

237. Wu, H. and Shapiro, J. L., *Does overfitting affect performance in estimation of distribution algorithms*, Conference on Genetic and Evolutionary Computation. ACM, pp.433-434, 2006, doi: https://doi.org/10.1145/1143997.1144078

238. Wu, Y., Xu, Y. and Li, J., *Feature construction for fraudulent credit card cash-out detection*, Decision Support Systems, vol. 127, pp 113 – 155, 2019, doi: https://doi.org/10.1016/j.dss.2019.113155

239. XGBoost, *Notes on Parameter Tuning*, https://xgboost.readthedocs.io/en/latest/tutorials/param_tuning.html, accesat la data de 27 martie 2019

240. XGBoost, *XGBoost Parameters*, https://xgboost.readthedocs.io/en/latest/parameter.html, accesat la data de 27 martie 2019

241. Xia, Y., *A Novel Reject Inference Model Using Outlier Detection and Gradient Boosting Technique in Peer-to-Peer Lending*, IEEE Access, vol. 7, pp. 92893-92907, 2019, doi: 10.1109/ACCESS.2019.2927602.

242. Xiao, J., Xieb, L., Hea, C. and Jiangc, X., *Dynamic classifier ensemble model for customer classification with imbalanced class distribution*, Expert Systems with Applications, vol. 39, issue 3, pp. 3668–3675, 2012, doi: https://doi.org/10.1016/j.eswa.2011.09.059

243. Xiaolei, S., Mingxi, L. and Zeqian, S., *A novel cryptocurrency price trend forecasting model based onLightGBM*, Finance Research Letters, December 2018, DOI: 10.1016/j.frl.2018.12.032

244. Yan, L., Xie, D. and Du, Z., *A new Method of Support vector Machine for Class Imbalance Problem*, International Joint Conference on Computational Science and Optimization, IEEE, Sanya, Hainan, pp. 904- 907, 2009, doi: 10.1109/CSO.2009.169.

245. Yang, C., Yang, J. and Wang, J., *Margin calibration in SVM class-imbalanced learning*, Neurocomputing, vol. 73, pp. 397–411, 2009, doi: https://doi.org/10.1016/j.neucom.2009.08.006

246. Yang, Y. and Ma, G., *Ensemble-based active learning for class imbalance problem*, Journal of Biomedical Science and Engineering, vol. 3, 2010, doi: 10.4236/jbise.2010.310133

247. Yiu, T., *Understanding Random Forest*, https://towardsdatascience.com/understanding-random-forest-58381e0602d2, 2019, accesat la data de 19 decembrie 2019

248. Yong, Y., *The Research of Imbalanced Data Set of Sample Sampling Method Based on K-Means Cluster and Genetic Algorithm*, 2012 International Conference on Future Electrical Power and Energy Systems, Energy Procedia, vol. 17, pp. 164 – 170, 2012, doi: https://doi.org/10.1016/j.egypro.2012.02.078

249. Yu, C. H., *Exploratory Data Analysis in the Context of Data Mining and Resampling*, International Journal of Psychological Research, vol 3, pp. 9-22, 2010, doi: 10.21500/20112084.819

250. Yu, X., *Machine Learning Application in Online Leading Credit Risk Prediction*, 2017, https://arxiv.org/abs/1707.04831

251. Zade, M., *Study of Effective Factors at the E-banking Operational Risk in the Maskan Bank*, The Master of Business Administration, Islamic Azad University, Kermanshah Branch, Iran, 2010

252. Zadrozny, B., Langford, J. and Abe, N., *Cost-sensitive learning by cost proportionate example weighting*, Third IEEE International Conference on Data Mining, Melbourne, FL, USA, pp. 435-442, 2003, doi: 10.1109/ICDM.2003.1250950

253. Zakaryazad, A. and Duman, E., *A profit-driven Artificial Neural Network (ANN) with applications to fraud detection and direct marketing*, Neurocomputing, no 175, pp. 121 – 131, 2016, doi: 10.1016/j.neucom.2015.10.042

254. Zhang, C., Bengio, S., Hardt, M., Recht, B. and Vinyals, O., *Understanding Deep Learning Requires Re thinking Generalization*, Proceedings of the 5th International Conference on Learning Representations, 2017, arXiv:1611.03530

255. Zieba, M. and Tomczak, J., *Boosted svm with active learning strategy for imbalanced data*, Soft Computing, vol. 19, pp. 3357-3368, 2015, doi: https://doi.org/10.1007/s00500-014-1407-5

Bibliografie de autor

ISI Proceedings

1. Elena-Adriana MINASTIREANU and Gabriela MESNITA, *Analysis Of Risk Variables In Online Banking*, European Union Financial Regulation and Administrative Area, 17 mai 2019, ISBN/ISSN 978-606-714-543-4
2. Elena-Adriana MINASTIREANU and Gabriela MESNITA, *Methods of Handling Unbalanced Datasets in Credit Card Fraud Detection*, Broad Research in Artificial Intelligence and Neuroscience, 21 martie 2020, ISSN/ISBN 2067-3957
3. Elena-Adriana MINASTIREANU and Gabriela MESNITA, *Machine Learning algorithms for fraud detection in Internet Banking*, Proceeding of the 17th International Conference on Informatics in Economy Education, Research & Business Technologies, 17 mai 2018, ISBN/ISSN 2247-1480
4. Elena-Adriana MINASTIREANU and Gabriela MESNITA, *Reducing Type II Errors in Credit Card Fraud Detection using XGBoost Classifier*, Proceeding of the 19th International Conference on Informatics in Economy Education, Research & Business Technologies, 21 mai 2020, ISBN/ISSN 2247-1480

BDI Indexed

1. Elena-Adriana MINASTIREANU and Gabriela MESNITA, *An Analysis of the Most Used Machine Learning Algorithms for Online Fraud Detection*, Informatica Economică, vol. 23, no. 1/2019, pp.5-16, 2019, ISBN/ISSN 14531305/23.1.2019.01
2. Elena-Adriana MINASTIREANU and Gabriela MESNITA, *Light GBM Machine Learning Algorithm to Online Click Fraud Detection*, Journal of Information Assurance & Cyber security, vol. 2019 (2019), Article ID 263928, DOI:10.5171/2019.263928, 4 Aprilie 2019, ISBN/ISSN 2165-9923