

# ACADEMIA DE STUDII ECONOMICE DIN BUCUREȘTI



Școala Doctorală de Informatică Economică

## TITLUL TEZEI DE DOCTORAT:

**Îmbunătățirea performanței companiei utilizând Business Analytics**

### **Coordonator științific:**

Prof. Univ. Dr. BOLOGA Ana Ramona

### **Doctorand:**

ANTAL-VAIDA Claudia

București, Septembrie 2024

## REZUMAT

Teza de doctorat intitulată „Îmbunătățirea performanței companiei utilizând Business Analytics” descrie cadrul economic în continuă transformare care a determinat schimbări majore în modul de operare al industriilor, favorizând adopția și integrarea tehnologiilor emergente pentru a rămâne competitive. Întrucât aceste noi tendințe au condus la generarea unor volume mari de date, dar și la o scădere a posibilității de analiză amănunțită a acestora, transformarea digitală a devenit componentă cheie a strategiei de afaceri, Business Analytics, Inteligența Artificială și Machine Learning devenind concepte din ce în ce mai populare, dar și necesare pentru a susține performanțele unei companii.

Principala contribuție personală a acestei lucrări constă în propunerea unei metodologii de analiză bazată pe tehnici de Machine Learning pentru prelucrarea datelor și obținerea de informații care să faciliteze procesul de luare a deciziilor informate în vederea îmbunătățirii performanțelor unei companii prin optimizarea proceselor operaționale. Această propunere a plecat de la construirea și testarea unor modele de predicție pentru care au fost folosiți zece algoritmi de Machine Learning pentru clasificare, pentru trei seturi de date publice, specifice unor probleme cu care se confruntă instituțiile bancare: identificarea persoanelor cu o probabilitate mare de a nu își plăti creanțele către bancă în luna următoare, identificarea clienților care ar putea renunța la serviciile băncii și identificarea tranzacțiilor de tip fraudă. Pe lângă construirea modelelor pe baza algoritmilor aleși, au fost incluse în metodologia propusă, testate și evaluate diverse metode de optimizare a modelelor, precum selecția variabilelor semnificative, optimizarea hiperparametrilor și tehnici de reeșantionare pentru seturile de date neechilibrate, în vederea propunerii unui model predictiv optim de analiză.

Metodologia propusă poate fi folosită atât în domeniul bancar, pentru problemele evidențiate în această lucrare, dar și pentru alte probleme de clasificare, cât și în alte domenii precum domeniul medical, sportiv, marketing, industrial, etc. Natura flexibilă a acesteia poate evolua în pas cu trendul tehnologic în continua schimbare, asigurând o integrare optimă a tehnicilor evolutive de procesare și analiză a datelor.

**Cuvinte cheie:** *Business Analytics, Machine Learning, probleme de clasificare, performanța companiei, managementul riscului bancar, metodologie de analiză*

## Cuprins

<b>CAPITOLUL 1. INTRODUCERE .....</b>	<b>1</b>
1.1 Structura lucrării .....	3
1.2 Scopul și obiectivele cercetării .....	5
1.3 Metodologia cercetării .....	6
<b>CAPITOLUL 2. ERA DIGITALĂ ȘI EVOLUȚIA TEHNICILOR DE ANALIZĂ A AFACERILOR.....</b>	<b>9</b>
2.1 Era digitală .....	9
2.2 Business Analytics.....	12
2.3 Inteligența Artificială .....	13
2.4 Machine Learning .....	16
2.5 Evoluția Inteligenței Artificiale și Machine Learning .....	18
2.6 Discuții și contribuții personale .....	21
<b>CAPITOLUL 3. STADIUL ACTUAL AL CERCETĂRII – ANALIZA LITERATURII DE SPECIALITATE EXISTENTĂ .....</b>	<b>23</b>
3.1 Tendințe în modele de scorare pentru credite .....	28
3.2 Tendințe în soluționarea problematicei legată de retenția clienților .....	35
3.3 Tendințe în detecția tranzacțiilor frauduloase.....	37
3.4 Tendințe aplicabile altor problematici pentru adresarea riscurilor bancare .....	39
3.5 Discuții și contribuții personale .....	40
<b>CAPITOLUL 4. ALGORITMI DE MACHINE LEARNING FOLOSIȚI PENTRU MANAGEMENTUL RISCULUI ÎN INSTITUȚIILE BANCARE .....</b>	<b>42</b>
4.1 Riscuri la care sunt expuse instituțiile bancare .....	43
4.2 Aplicabilitatea tehnicilor de Business Analytics și Machine Learning .....	45
4.3 Dificultăți în implementarea proiectelor de Machine Learning .....	48
4.4 Nivelul de adopție al tehnicilor de Inteligență Artificială și Machine Learning in domeniul bancar, în timpul pandemiei .....	50
4.4.1 Îmbunătățirea experienței clienților prin modelul băncii conversaționale .....	52
4.4.2 Utilizarea Machine Learning în sistemele anti-spălare a banilor .....	53
4.4.3 Verificarea identității pentru a preveni fraudele și furtul de identitate .....	54
4.4.4 Riscurile cibernetice .....	55
4.4.5 Analiza SWOT a implementării AI/ML în bănci .....	56
4.5 Discuții și contribuții personale .....	58

**CAPITOLUL 5. CONSTRUIREA UNUI MODEL DE CLASIFICARE PENTRU ÎMBUNĂȚĂȚIREA STRATEGIEI UNEI COMPANII ..... 60**

5.1 Algoritmi folosiți în probleme de clasificare .....	64
5.2 Indicatori de evaluare a modelelor .....	71
5.3 Îmbunătățirea algoritmilor de Machine Learning.....	71
5.3.1 Metode de selecție a variabilelor semnificative .....	73
5.3.2 Metode de optimizare a hiperparametrilor .....	77
5.3.3 Metode de reeșantionare .....	79
5.4 Discuții și contribuții personale .....	80

**CAPITOLUL 6. METODOLOGIE DE ANALIZĂ BAZATĂ PE TEHNICI DE MACHINE LEARNING PENTRU DOMENIUL BANCAR..... 81**

6.1 Metodologia propusă .....	81
6.1.1 Definirea problemei .....	84
6.1.2 Descrierea metodologiei propuse .....	84
6.1.3 Funcțiile și librăriile folosite pentru construirea modelelor .....	91
6.1.4 Aplicabilitatea metodologiei propuse.....	93
6.2 Descrierea seturilor de date și procesarea acestora .....	94
6.3 Indicatori de performanță a modelelor .....	103
6.4 Analiza performanței inițiale a algoritmilor .....	104
6.4.1 Analiza rezultatelor obținute pentru setul de date 1 .....	104
6.4.2 Analiza rezultatelor obținute pentru setul de date 2 .....	106
6.4.3 Analiza rezultatelor obținute pentru setul de date 3 .....	107
6.4.4 Interpretare rezultatelor .....	108
6.5 Analiza performanței algoritmilor după aplicarea tehnicilor de selecție a variabilelor semnificative.....	108
6.5.1 Selecția variabilelor pentru setul de date 1 .....	109
6.5.2 Selecția variabilelor pentru setul de date 2.....	113
6.5.3 Interpretarea rezultatelor .....	117
6.6 Analiza performanței algoritmilor după aplicarea tehnicilor de optimizare a hiperparametrilor .....	118
6.6.1 Rezultate obținute pentru setul de date 1.....	118
6.6.2 Rezultate obținute pentru setul de date 2 .....	131
6.6.3 Interpretarea rezultatelor .....	144
6.7 Analiza performanței algoritmilor după aplicarea tehnicilor de reeșantionare .....	144
6.8 Interpretarea rezultatelor .....	154

6.9 Limitări ale studiului .....	158
<b>CAPITOLUL 7. CONCLUZII FINALE, CONTRIBUȚII PERSONALE ȘI DIRECȚII VIITOARE DE CERCETARE .....</b>	<b>159</b>
7.1 Concluzii finale .....	159
7.2 Contribuții personale .....	167
7.3 Direcții viitoare de cercetare .....	168
<b>BIBLIOGRAFIE .....</b>	<b>169</b>
<b>Anexe .....</b>	<b>192</b>
<b>Lista articolelor și a conferințelor .....</b>	<b>210</b>
<b>Lista de abrevieri .....</b>	<b>214</b>
<b>Lista tabelor .....</b>	<b>216</b>
<b>Lista graficelor .....</b>	<b>219</b>