

ACADEMIA DE STUDII ECONOMICE DIN BUCUREȘTI

Consiliul pentru Studii Universitare de Doctorat

Școala Doctorală de Finanțe

Machine Learning applied in Finance

Codruț-Florin Ivașcu

PhD Supervisor:

prof. univ. dr. Bogdan Negrea

București, 2024

Summary

This thesis aims to examine the role and applicability of Machine Learning algorithms in the financial sector, offering a balanced assessment between the promises of the technology and its limitations. The work navigates between theoretical enthusiasm and practical realism, attempting to build a bridge between technological capability and the practical needs of the financial sector. The main motivation of the work is to surpass traditional methods by demonstrating how ML can provide superior solutions in numerous financial areas.

In a period defined by the acceleration of technological innovations, Artificial Intelligence and ML have captured the attention of the public and researchers, often presented as omnipotent solutions to economic and social challenges. However, this thesis argues that a deep understanding of the basic principles, potential, and limitations of ML is essential for its effective application in finance. In this direction, the work starts with a detailed discussion of various ML models, from linear models and regularization methods to advanced algorithms based on trees, neural networks, and ensemble methods. The discussion is enhanced by analyzing the advantages and limitations of each model, offering a perspective on their potential to transform financial analysis and decision management.

The applicability of ML is further illustrated through multiple case studies that cover different aspects of finance, demonstrating the technology's ability to surpass traditional methods. The case studies include option valuation, where algorithms such as XGBoost and LightGBM provide more accurate evaluations than established parametric models; portfolio management, where ML techniques optimize asset selection and allocation; credit risk management, which benefits from more precise risk assessment; dividend policy modeling, where ML can identify the nonlinear determinants influencing company policies; modeling the underground economy, which uses ML to understand complex interactions between various government expenditures; and inflation forecasting, showcasing the effectiveness of ML models in inflation prediction compared to traditional econometric methods.

The thesis conclusions reflect on the duality of ML: on one hand, a powerful tool for innovation and efficiency in the financial sector; on the other hand, a field full of technical challenges that requires careful management. The thesis emphasizes the need for a critical and informed approach in applying ML, stressing that its success depends not only on the algorithms used but also on the integrity of the data, the accuracy of the modeling, and the relevance of the application context.

CONTENT

SECTION I – INTRODUCTION	1
1. GENERAL CONSIDERATIONS.....	2
1.1. MOTIVATIONS.....	2
1.2. WHAT IS ML?.....	5
1.3. STRUCTURE OF THE THESIS.....	7
2. HISTORY.....	12
2.1. PREMODERN VISIONS	12
2.2. THE BEGINNINGS OF AI.....	13
2.3. EXPERT SYSTEMS.....	14
2.4. THE MACHINE LEARNING REVOLUTION	16
2.5. BIG DATA AND THE RISE OF DEEP LEARNING	17
2.6. SOCIAL AND ETHICAL IMPACT.....	18
3. FUNDAMENTAL CONCEPTS.....	20
3.1. TYPES OF ML	20
3.2. THE MACHINE LEARNING PROCESS	22
3.3. MODEL EVALUATION AND OPTIMIZATION	23
3.4. MODEL INTERPRETABILITY.....	27
3.4.1. SHAP	28
3.4.2. LIME.....	29
3.4.3. OTHER METHODS	31
SECTION II – MODELS	33
4. LINEAR MODELS	34
4.1. INTRODUCTION	34
4.2. LINEAR REGRESSION.....	35
4.2.1. MATHEMATICAL FORMULATION.....	35
4.2.2. PROPERTIES OF ESTIMATORS.....	36

4.2.3. ASSUMPTIONS OF LINEAR REGRESSION	37
4.2.4. ADVANTAGES AND DISADVANTAGES.....	38
4.3. LOGISTIC REGRESSION.....	39
4.3.1. MATHEMATICAL FORMULATION.....	40
4.3.2. MULTINOMIAL LOGISTIC REGRESSION.....	41
4.3.3. MODEL ASSUMPTIONS	41
4.3.4. ADVANTAGES AND DISADVANTAGES.....	42
4.4. REGULARIZATION METHODS	43
4.4.1. LASSO.....	43
4.4.2. RIDGE	46
4.4.3. ELASTIC NET	47
4.5. CONCLUSIONS	48
<u>5. SUPPORT VECTOR MACHINES</u>	<u>49</u>
5.1. INTRODUCTION	49
5.2. HISTORY	49
5.3. MODEL PRESENTATION.....	51
5.3.1. MATHEMATICAL FORMULATION.....	52
5.3.2. KERNEL TECHNIQUES IN SVM.....	54
5.3.3. REGULARIZATION AND OPTIMIZATION	55
5.4. SUPPORT VECTOR REGRESSION (SVR).....	56
5.4.1. MATHEMATICAL FORMULATION.....	57
5.4.2. REGULARIZATION AND OPTIMIZATION	58
5.5. ADVANTAGES AND DISADVANTAGES	59
5.6. CONCLUSIONS	60
<u>6. DECISION TREES</u>	<u>62</u>
6.1. DECISION TREE (MODELUL CART).....	62
6.1.1. INTRODUCTION	62
6.1.2. HISTORY	63
6.1.3. MODEL PRESENTATION	65
6.1.4. TECHNICAL ASPECTS OF THE ALGORITHM.....	70
6.1.5. ADVANTAGES AND DISADVANTAGES.....	71
6.2. RANDOM FOREST	72
6.2.1. INTRODUCTION	72
6.2.2. HISTORY	73
6.2.3. MODEL PRESENTATION	74
6.2.4. TECHNICAL ASPECTS OF THE ALGORITHM.....	77

6.2.5. ADVANTAGES AND DISADVANTAGES.....	78
6.2.6. CONCLUSIONS	79
6.3. XGBOOST.....	80
6.3.1. INTRODUCTION.....	80
6.3.2. HISTORY	81
6.3.3. MODEL PRESENTATION	82
6.3.4. DISTINCTIVE FEATURES OF XGBOOST.....	86
6.3.5. HYPERPARAMETERS.....	90
6.3.6. ADVANTAGES AND DISADVANTAGES.....	91
6.3.7. CONCLUSIONS	92
6.4. LIGHTGBM.....	93
6.4.1. INTRODUCTION.....	93
6.4.2. DISTINCTIVE FEATURES	94
6.4.3. ADVANTAGES AND DISADVANTAGES.....	98
6.4.4. CONCLUSIONS	99
<u>7. RETELE NEURONALE.....</u>	<u>100</u>
7.1. INTRODUCTION	100
7.2. HISTORY	101
7.3. ARTIFICIAL NEURON.....	104
7.4. DEEP NEURAL NETWORK – MULTILAYER PERCEPTRON.....	106
7.5. TRAINING	108
7.5.1. OBJECTIVE FUNCTION	108
7.5.2. BACKPROPAGATION	109
7.5.3. OPTIMIZATION ALGORITHMS	110
7.5.4. WEIGHT INITIALIZATION	113
7.5.5. REGULARIZATION	113
7.6. ADVANTAGES AND DISADVANTAGES	115
7.7. ADVANCED TYPES OF NETWORKS	117
7.7.1. RECURRENT NEURAL NETWORKS (RNN).....	117
7.7.2. LONG SHORT TERM MEMORY (LSTM).....	119
7.8. CONCLUSIONS	122
<u>8. OTHER SUPERVISED MODELS</u>	<u>124</u>
8.1. K-NEAREST NEIGHBOURS (KNN)	124
8.1.1. INTRODUCTION.....	124
8.1.2. MODEL DESCRIPTION	125
8.1.3. ADVANTAGES AND DISADVANTAGES.....	126

8.2. GENETIC ALGORITHMS	127
8.2.1. INTRODUCTION	127
8.2.2. MODEL DESCRIPTION	128
8.2.3. ADVANTAGES AND DISADVANTAGES.....	134
<u>9. UNSUPERVISED MODELS</u>	<u>136</u>
9.1. INTRODUCTION	136
9.2. DIMENSIONALITY REDUCTION	137
9.2.1. PRINCIPAL COMPONENT ANALYSIS (PCA).....	138
9.2.2. AUTOENCODERS	140
9.2.3. OTHER MODELS.....	142
9.3. CLUSTERING.....	143
9.3.1. K-MEANS	145
9.3.2. HIERARCHICAL CLUSTERING	146
9.3.3. DBSCAN	147
9.3.4. OTHER MODELS.....	148
<u>SECTION III - CASE STUDIES.....</u>	<u>150</u>
<u>10. CASE STUDY I – OPTION PRICING.....</u>	<u>151</u>
10.1. INTRODUCTION.....	151
10.2. METHODOLOGY	154
10.2.1. DATA	154
10.2.2. BENCHMARKS.....	155
10.2.3. GENERAL CONSIDERATIONS	158
10.3. RESULTS	162
10.4. CONCLUSIONS	167
<u>11. CASE STUDY II – PORTFOLIO MANAGEMENT.....</u>	<u>170</u>
11.1. INTRODUCTION.....	170
11.2. SELECTION	171
11.2.1. RANDOM FOREST.....	172
11.2.2. XGBOOST.....	173
11.2.3. LASSO.....	173
11.2.4. AUTOENCODERS	174
11.3. ALLOCATION	175
11.3.1. OLS.....	176

11.3.2. NEWRAL NETWORKS SENSITIVITY	176
11.4. METHODOLOGY	177
11.5. RESULTS.....	179
11.6. CONCLUSIONS.....	186
11.7. APPENDIX	187
<u>12. CASE STUDY III – RISK MANAGEMENT</u>	<u>193</u>
12.1. INTRODUCTION.....	193
12.2. STATE OF KNOWLEDGE.....	194
12.3. METHODOLOGY	197
12.3.1. DATA	197
12.3.2. REBALANCING TECHNIQUES.....	199
12.3.3. GENERAL CONSIDERATIONS	200
12.3.4. EVALUATION METRICS.....	201
12.4. RESULTS	202
12.5. CONCLUSIONS	209
<u>13. CASE STUDY IV – DIVIDEND POLICY MODELING</u>	<u>210</u>
13.1. INTRODUCTION.....	210
13.2. STATE OF KNOWLEDGE.....	211
13.3. METHODOLOGY	215
13.3.1. DATA ANALYSIS.....	215
13.3.2. GENERAL CONSIDERATIONS	219
13.4. RESULTS	221
13.5. CONCLUSIONS	225
<u>14. CASE STUDY V – SHADOW ECONOMY MODELING</u>	<u>226</u>
14.1. INTRODUCTION.....	226
14.2. STATE OF KNOWLEDGE.....	227
14.3. DATA DESCRIPTION.....	230
14.4. LINEAR ANALYSIS	233
14.5. NONLINEAR ANALYSIS	235
14.6. CONCLUSIONS	241
<u>15. CASE STUDY VI – INFLATION PREDICTION</u>	<u>243</u>
15.1. INTRODUCTION.....	243

15.2. METHODOLOGY	244
15.3. RESULTS	245
15.4. CONCLUSIONS	248
<u>SECTION IV – DISCUSSIONS</u>	<u>249</u>
<u>16. GENERAL CONCLUSIONS.....</u>	<u>250</u>
<u>17. IS MLA PANACEA?.....</u>	<u>257</u>
<u>REFERENCES</u>	<u>262</u>
<u>LIST OF TABLES</u>	<u>280</u>
<u>LIST OF FIGURES</u>	<u>282</u>
<u>GLOSSARY</u>	<u>284</u>