

**MINISTERUL EDUCAȚIEI ȘI CERCETĂRII AL REPUBLICII MOLDOVA**

**Universitatea Tehnică a Moldovei**

**Facultatea Calculatoare, Informatică și Microelectronică**

**Departamentul Informatică și Ingineria Sistemelor**

**Admis la susținere**

**Șef departament:**

**conf. univ. dr. V. Sudacevschi**

„\_\_\_\_\_” \_\_\_\_\_ 2023

# **EVALUAREA SIMILARITĂȚII DINTRE FUNCȚII APROXIMATOARE**

**Teză de master în**

**Managementul Aplicațiilor Informaționale**

**Student:** Agachii Victoria ( \_\_\_\_\_ )

**Conducător:** Coandă Ilie ( \_\_\_\_\_ )

**Recenzent:** \_\_\_\_\_ ( \_\_\_\_\_ )

**Chișinău – 2023**

## ADNOTARE

### la teza de master cu tema „Evaluarea similarității dintre funcții aproximatoare”

#### a masterandului gr. MAI-211M, Agachii Victoria

Teza de master este constituită din introducere, trei capitole, **38** surse bibliografice, **21** figuri și concluzie.

**Scopul** prezentei teze de master s-a concentrat pe cercetarea domeniului și evidențierea ulterioară a unor tehnici eficiente de analiză a datelor și dezvoltarea unei aplicații constituite dintr-un set de instrumente informatice, orientată spre asigurarea unui nivel înalt de cercetare a informației.

**Obiectivele** cercetării sunt:

- analiza introductivă a formei, conținutului și principalelor caracteristici ale datelor primare;
- elaborarea unei scheme sau algoritm de curățire-netezire a datelor inițiale;
- efectuarea și prezentarea rezultatelor unor analize-cercetări cu scopul de a prezenta asemănările dintre ele;
- formularea concluziilor din urma rezultatelor.

**Capitolul 1, „Inițializare în procesul de analiză a datelor. Instrumente”** se bazează pe înțelegerea conceptului de analiză a datelor și tehnologiile/metodele de analiză a datelor. În el se evidențiază importanța lucrului calitativ cu datele în scopul îndeplinirii eficiente a obiectivelor pe care și le pune cercetătorul în procesul de lucru cu datele, prelucrarea corectă pentru obținerea rezultatelor adecvate. Pentru asta, au fost descrise tehnicile/metodele cunoscute de lucru cu datele, și au fost enumerate cele utilizate în lucrarea dată.

**Capitolul 2, „Metode, modele, tehnici moderne de studiere a datelor”** dezvoltă descrierea analizei de regresie într-una detaliată pentru a prezenta succint componenta matematică aplicată în lucrare, care se focusează pe procesarea și analiza datelor.

**Capitolul 3, „Evaluarea similarității dintre funcții aproximatoare”** se focusează pe partea practică a lucrării, cu prezentarea funcțiilor matematice și trigonometrice utilizate, care au urmat a fi evaluate ulterior prin tehnici specifice, algoritmi dezvoltați în cadrul tezei.

Teza de master este un produs de care pot beneficia mai mulți utilizatori, întrucât setul de date folosit în lucrare poate fi modificat cu un altul, în funcție de scopurile pe care și-l pun beneficiarii, cum ar fi cercetătorii din domeniul informației și dezvoltatorii de softuri.

**Cuvinte cheie:** Data Mining, analiza datelor, prelucrarea datelor, regresie, bază de date.

## ANNOTATION

### **the master thesis entitled " Evaluation of similarity between approximating functions "** **of the master student gr. MAI-211M, Agachii Victoria**

The master's thesis consists of an introduction, three chapters, 38 bibliographic sources, 21 figures and a conclusion.

The **purpose** of this master's thesis was focused on field research and the subsequent highlighting of effective data analysis techniques and the development of an application consisting of a set of IT tools, aimed at ensuring a high level of information research.

The **objectives** of the research are:

- introductory analysis of the form, content and main characteristics of primary data;
- development of a scheme or algorithm for cleaning-smoothing the initial data;
- performing and presenting the results of some analysis-research with the aim of presenting the similarities between them;
- formulating conclusions from the results.

**Chapter 1, „Getting started in the data analysis process. Tools”** is based on understanding the concept of data analysis and data analysis technologies/methods. It highlights the importance of working qualitatively with the data in order to effectively fulfill the objectives that the researcher sets in the process of working with the data, the correct processing to obtain the appropriate results. For this, the known techniques/methods of working with the data were described, and those used in the given work were listed.

**Chapter 2, „Methods, models, modern techniques of studying data”** develops the description of regression analysis into a detailed one to briefly present the mathematical component applied in the work, which focuses on data processing and analysis.

**Chapter 3, „Evaluation of the similarity between approximating functions”** focuses on the practical part of the work, with the presentation of the mathematical and trigonometric functions used, which were later evaluated through specific techniques, algorithms developed within the thesis.

The master's thesis is a multi-user product, as the data set used in the paper can be modified with another, depending on the goals set by the beneficiaries, such as information researchers and software developers.

**Keywords:** Data Mining, data analysis, data processing, regression, database.

## CUPRINS

<b>INTRODUCERE</b> .....	9
<b>1 INIȚIALIZARE ÎN PROCESUL DE ANALIZĂ A DATELOR. INSTRUMENTE</b> .....	11
<b>1.1 Conceptul de analiză a datelor</b> .....	11
<b>1.2 Tehnologii/metode de analiză a datelor</b> .....	18
<b>2 METODE, MODELE, TEHNICI MODERNE DE STUDIERE A DATELOR</b> .....	32
<b>2.1 Analiza de regresie</b> .....	32
<b>2.1.1 Regresie liniară și regresia liniară multiplă</b> .....	34
<b>2.1.2 Regresie polinomială</b> .....	37
<b>2.1.3 Metoda celor mai mici patrate</b> .....	39
<b>2.1.4 Supraadaptarea și subadaptarea</b> .....	43
<b>3 EVALUAREA SIMILARITĂȚII DINTRE FUNCȚIILE APROXIMATOARE</b> .....	48
<b>3.1 Tehnologii și instrumente de cercetare a datelor</b> .....	48
<b>3.2 Funcții matematice și trigonometrice</b> .....	49
<b>3.3 Rolul derivatelor în studiul funcțiilor</b> .....	51
<b>3.3.1 Derivata funcției</b> .....	51
<b>3.3.2 Intervale de monotonie. Puncte de extrem</b> .....	54
<b>3.3.3 Convexitate și concavitate. Puncte de inflexiune</b> .....	56
<b>3.4 Evaluarea funcțiilor aproximatoare</b> .....	57
<b>CONCLUZII GENERALE ȘI RECOMANDĂRI</b> .....	64
<b>BIBLIOGRAFIE</b> .....	65

## INTRODUCERE

Dezvoltarea Tehnologiilor Informaționale și de Comunicare (TIC) la etapa actuală a format necesitatea de dezvoltare și implementare unor tehnici noi de prelucrare a datelor. S-a observat implementarea progresivă a tehnicilor/tehnologiilor noi argumentate științific în calitate de metode, algoritmi în marea majoritate a întreprinderilor și domeniilor de organizare a activităților de Business.

La etapa ce o precede pe cea actuală, volumul de lucru cu tehnicile de calcul, în mare parte, s-a concentrat pe colectarea și păstrarea datelor. Prelucrarea datelor, însă, a fost plasată pe planul doi, ceea ce a dus în rezultat la volume mari de date care s-au cerut a fi analizate, în dependență de scopurile propuse. În acest mod, au apărut noțiuni precum: *Analiza Datelor*, *Data Mining* (Mineritul de date), *Machine Learning* (Învățarea automată) și *Artificial Intelligence* (Inteligența Artificială). Aceste noțiuni au apărut într-o anumită logică, din punct de vedere a timpului. Noțiunea „Analiza Datelor”, cu temeiul de prelucrarea datelor, de exemplu, a apărut în perioada de timp în care volume mari de date erau deja colectate și stocate.

În perioada actuală, metodele, modelele, algoritmi și alte scheme de calcul din domeniul TIC sunt la un nivel mult mai dezvoltat decât nivelul lor de implementare în domeniu. Astfel, s-a evidențiat intensificarea implicării tehnicilor/tehnologiilor moderne bazate, argumentate științific în activitățile din domeniile care se ocupă cu prelucrarea datelor în esență la niveluri superioare.

La etapa inițială de dezvoltare a tehnicilor caracteristice *Analizei Datelor* se considera suficientă dezvoltarea și implementarea unor tehnologii, tehnici și metode relativ simple. Dar cu ridicarea nivelului de eficiență a componentelor domeniului TIC, a apărut posibilitatea dezvoltării și implementării unor tehnici diverse și noi de prelucrare a datelor, caracteristice noțiunii de *Data Mining*, care a fost apoi succedată de *Artificial Intelligence*. Iar la etapa actuală, se consideră că societatea se află în etapa de dezvoltare a TIC-ului, în care se urmărește implementarea intensă a tehnologiilor noi de procesare a datelor, care corespund noțiunii de *Big Data*.

Unele dintre cele mai clasice tehnici de procesare a datelor care au fost utilizate la etapa precedentă, dar care sunt implicate și astăzi în activitățile de analiză a datelor în mod destul de larg și frecvent, sunt:

- Analiza de regresie – utilizată larg în foarte multe domenii de activitate. Metodele analizei de regresie au fost dezvoltate preponderent, inițial, în domeniul economiei. Tehnicile date sunt utilizate eficient și la etapa actuală. Există multe lucrări și publicații cu tematica acestui tip de analiză și metodele sale. Există și aplicații (soft) care pot fi utilizate eficient la etapa inițială de cercetare a informațiilor.
- Tehnici care sunt utilizate cu scopul de a analiza datele, în funcție de unu sau mai mulți factori.

- Tehnici care se bazează pe modele cu complexitate ridicată, precum sunt tehnicile din SPSS sau VECA.

Tehnicile date au oferit posibilitatea de analiză a datelor până la un anumit nivel „de adâncime”. În practica reală, însă, apare des nevoia de a cerceta datele/informațiile la un nivel mult mai superior și eficient, care, la rândul său, rezultă la noi abordări în procesul de analiză a datelor, cum ar fi: de a propune metodologii noi sau de a propune modificări la tehnicile clasice, deja aprobate.

#### ***Actualitatea temei:***

La etapa de dezvoltare actuală, ridicarea nivelului de dezvoltare a echipamentelor TIC au influențat tehnologiile de colectare și analiză a datelor, care au contribuit considerabil la creșterea semnificativă a volumului de date, precum și creșterea nivelului de diversitate și complexitate al lor. Asta și a creat necesitatea de a iniția cercetări în vederea ridicării eficienței modelelor de analiză și evidențierea unor anumite aspecte de îmbunătățire și de integrare al algoritmilor, modelelor utilizate în domeniul de cercetare al informațiilor.

#### ***Valoarea aplicativă:***

Implementarea tehnicilor de analiză a datelor în activitatea întreprinderilor, în perioada dezvoltării accentuate a economiei de piață, a ajutat la soluționarea problemelor referitoare la fundamentarea și luarea deciziilor pe cât se poate de adecvate, care corespund situației reale a întreprinderilor.

#### ***Scopul de bază:***

Lucrarea s-a concentrat pe cercetarea domeniului și evidențierea ulterioară a unor tehnici eficiente de analiză a datelor și dezvoltarea unei aplicații constituite dintr-un set de instrumente informatice, orientată spre asigurarea unui nivel înalt de cercetare a informației.

#### ***Obiective principale:***

- analiza introductivă a formei, conținutului și principalelor caracteristici ale datelor primare;
- elaborarea unei scheme sau algoritm de curățire-netezire a datelor inițiale;
- efectuarea și prezentarea rezultatelor unor analize-cercetări cu scopul de a prezenta asemănările dintre ele;

formularea concluziilor din urma rezultatelor.

## BIBLIOGRAFIE

1. KELLEY, Karin. What is Data Analysis? Methods, Process and Types Explained [online]. Simplilearn, 7 februarie 2023. Disponibil: <https://www.simplilearn.com/data-analysis-methods-process-types-article>
2. HILLIER, Will. A Step-by-Step Guide to the Data Analysis Process [online]. Career Foundry, 30 ianuarie 2023. Disponibil: <https://careerfoundry.com/en/blog/data-analytics/the-data-analysis-process-step-by-step/>
3. HILLIER, Will. What Is Descriptive Analytics? A Complete Guide [online]. Career Foundry, 13 ianuarie 2023. Disponibil: <https://careerfoundry.com/en/blog/data-analytics/descriptive-analytics/>
4. STEVENS, Emily. The 4 Types of Data Analysis [Ultimate Guide] [online]. Career Foundry, 30 noiembrie 2022. Disponibil: <https://careerfoundry.com/en/blog/data-analytics/different-types-of-data-analysis/>
5. DIXON, Michael. Types of predictive analytics models and how they work. In: *Data Analytics, Predictive Analytics* [online]. Selerity Pty Ltd, 12 decembrie 2019. Disponibil: <https://seleritysas.com/blog/2019/12/12/types-of-predictive-analytics-models-and-how-they-work>
6. HARRINGTON, R. Predictive Analytics & Data Mining 101: Clustering [online]. CompassRed, 31 iulie 2017. Disponibil: <https://www.compassred.com/blog/2017/7/31/predictive-analytics-data-mining-101-clustering>
7. COTE, Catherine. WHAT IS PRESCRIPTIVE ANALYTICS? 6 EXAMPLES [online]. Harvard Business School, 2 noiembrie 2021. Disponibil: <https://online.hbs.edu/blog/post/prescriptive-analytics>
8. Variables. In: *Statistics* [online]. Newcastle University, ©2023. Disponibil: [https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/descriptive-statistics/variables.html#Types\\_of\\_Variable](https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/descriptive-statistics/variables.html#Types_of_Variable)
9. Data Analysis. In: *Study Design and Analysis* [online]. Research Connections. ©2023. Disponibil: <https://www.researchconnections.org/research-tools/study-design-and-analysis/data-analysis>

10. *What is factor analysis and how does it simplify research findings?* [online]. Qualtrics, ©2023. <https://www.qualtrics.com/experience-management/research/factor-analysis/>
11. Simple Linear Regression. In: *Statistics* [online]. Newcastle University, ©2023. Disponibil: <https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/simple-linear-regression.html>
12. Types of Correlation. In: *Statistics* [online]. Newcastle University, ©2023. Disponibil: <https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/types-of-correlation.html>
13. Linear Equations. In: *Introduction to Statistics* [online]. Lumen Learning, ©2023. Disponibil: <https://courses.lumenlearning.com/introstats1/chapter/linear-equations/>
14. Regression Analysis. In: *Data Science* [online]. CFI Team, 24 noiembrie 2022. Disponibil: <https://corporatefinanceinstitute.com/resources/data-science/regression-analysis/>
15. Coefficient of Determination, R-squared. In: *Statistics* [online]. Newcastle University, ©2023. Disponibil: <https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/coefficient-of-determination-r-squared.html>
16. Residuals. In: *Statistics* [online]. Newcastle University, ©2023. Disponibil: <https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/residuals.html>
17. FROST, Jim. How To Interpret R-squared in Regression Analysis [online]. Statistics by Jim, ©2023. Disponibil: <https://statisticsbyjim.com/regression/interpret-r-squared-regression/>
18. Nonlinear Regression. In: *Data Science* [online]. CFI Team, 20 decembrie 2022. Disponibil: <https://corporatefinanceinstitute.com/resources/data-science/nonlinear-regression/>
19. *What Is Data Mining: Definition, Benefits, Applications, and More* [online]. Simplilearn, 6 februarie 2023. Disponibil: <https://www.simplilearn.com/what-is-data-mining-article>
20. What is `Data Mining`. In: *Analytics* [online]. The Economic Times, ©2023. Disponibil: <https://economictimes.indiatimes.com/definition/data-mining>
21. *What Is Data Mining: Definition, Examples, Tools, and Techniques (For Beginners)* [online]. Georgia Tech Boot Camps, ©2023. Disponibil: <https://bootcamp.pe.gatech.edu/blog/what-is-data-mining/>
22. *VECA* [online]. vdiscovery, ©2022. Disponibil: <https://vdiscovery.com/technology-solutions/>
23. GOEL, Amita, GUPTA, Aviral, SHARMA Akshay. Review of Regression Analysis Models [online]. International Journal of Engineering Research & Technology. 2017, vol. 6 [citat



- 26.03.2023]. ISSN: 2278-0181. Disponibil pe Internet: <https://www.ijert.org/research/review-of-regression-analysis-models-IJERTV6IS080060.pdf>
24. X and Y Intercept Formula. geeksforgeeks. 28 aprilie, 2022 [citat 26.03.2023]. Disponibil pe Internet: <https://www.geeksforgeeks.org/x-and-y-intercept-formula/>
25. CHEN, James. Line of Best Fit: Definition, How It Works, and Calculation [online]. In: Advanced Technical Analysis Concepts [online]. Investopedia, 23 august 2022. [citat 27.03.2023]. Disponibil pe Internet: <https://www.investopedia.com/terms/l/line-of-best-fit.asp>
26. Abhigyan. Understanding Polynomial Regression. In: Analytics Vidhya [online]. Medium, 2 august 2020 [citat 27.03.2023]. Disponibil pe Internet: <https://medium.com/analytics-vidhya/understanding-polynomial-regression-5ac25b970e18>
27. BHARSKAR, Mahesh. Explain the method of least square for fitting a regression line [online]. Sarthaks eConnect, 2020 [citat 28.03.2023]. Disponibil pe Internet: <https://www.sarthaks.com/3214657/explain-the-method-of-least-square-for-fitting-a-regression-line>
28. KENTON, Will. Least Squares Method: What It Means, How to Use It, With Examples. In: Financial Analysis [online]. Investopedia, 28 martie 2022 [citat 28.03.2023]. Disponibil pe Internet: <https://www.investopedia.com/terms/l/least-squares-method.asp>
29. What is overfitting? The International Business Machines Corporation, ©2023 [citat 29.03.2023]. Disponibil pe Internet: <https://www.ibm.com/topics/overfitting>
30. What is underfitting? The International Business Machines Corporation, ©2023 [citat 30.03.2023]. Disponibil pe Internet: <https://www.ibm.com/topics/underfitting>
31. WAINER, Joe. Microsoft's powerful spreadsheet and data management program [online]. Softonic, 2023 [30.03.2023]. Disponibil pe Internet: <https://microsoft-excel.en.softonic.com/>
32. Function, mathematics. Britannica. 01 decembrie 2022 [citat 30.03.2023]. Disponibil pe Internet: <https://www.britannica.com/science/function-mathematics>
33. Trigonometric function. Britannica. 09 decembrie 2022 [citat 30.03.2023]. Disponibil pe Internet: <https://www.britannica.com/science/trigonometric-function>
34. Differentiation. Britannica. 23 decembrie 2022 [citat 30.03.2023]. Disponibil pe Internet: <https://www.britannica.com/science/differentiation-mathematics>
35. Derivatives of Trigonometric Functions. LibreText. [citat 30.03.2023]. Disponibil pe internet: [https://math.libretexts.org/Courses/Laney\\_College/Math\\_3A%3A\\_Calculus\\_1\\_\(Fall\\_2022\)/03%3A\\_Derivatives/3.06%3A\\_Derivatives\\_of\\_Trigonometric\\_Functions](https://math.libretexts.org/Courses/Laney_College/Math_3A%3A_Calculus_1_(Fall_2022)/03%3A_Derivatives/3.06%3A_Derivatives_of_Trigonometric_Functions)

36. Calculus/Extrema and Points of Inflection. WikiBooks. 10 octombrie 2019. [citat 30.03.2023]. Disponibil pe internet: [https://en.wikibooks.org/wiki/Calculus/Extrema\\_and\\_Points\\_of\\_Inflection](https://en.wikibooks.org/wiki/Calculus/Extrema_and_Points_of_Inflection)
37. Convex and Concave Functions And Inflection Points. StudyWell. [citat 30.03.2023]. Disponibil pe Internet: <https://studywell.com/differentiation/convex-concave/>
38. Determinarea intervalelor de convexitate(concavitate) și a punctelor de inflexiune. LectiiVirtuale. [citat 30.03.2023]. Disponibil pe Internet: <https://lectii-virtuale.ro/teorie/determinarea-intervalelor-de-convexitate-concavitate-si-a-punctelor-de-inflexiune>