

**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:  
Sudacevschi Viorica, conf. univ., dr.**

---

„\_\_\_\_\_” \_\_\_\_\_ 2023

# **Mașini cu suport vectorial pentru clasificarea datelor**

**Teză de master**

**Student: Tutunaru Eugenia, MAI-211M**

**Conducător: Moraru Vasile,  
prof. univ., dr.**

**Chișinău, 2023**

## ADNOTARE

### la teza de master cu tema „Mașini cu Suport Vectorial pentru Clasificarea Datelor” a masterandei grupei MAI-211M TUTUNARU Eugenia

Teza de master este constituită din introducere, trei capitole, concluzie, 17 surse bibliografice și 25 figuri.

**Cuvinte-cheie:** mașini cu suport vectorial, inteligența artificială, învățarea automată supravegheată, clasificare, model de clasificare, accident vascular cerebral, python, aplicație web.

Această lucrare are drept scop este aplicarea algoritmului Mașini cu Suport Vectorial din învățarea automată pentru a clasifica în baza caracteristicilor vitale și comportamentale în două clase – predispus, nepredispus.

Întru realizarea scopului a fost stabilit un set de obiective generale. A fost analizată literatura de specialitate privind învățarea automată și în special algoritmul Mașini cu Suport Vectorial. De asemenea s-a efectuat analiza domeniului problematic - accidentul cerebral vascular. La rândul său, analiza domeniului inteligenței artificiale creează o imagine mai clară asupra universului de posibilități care oferă această disciplină și contribuie la identificarea modelelor care ar rezolva numeroase probleme identificate, inclusiv din medicină și, în cazul lucrării date, ale accidentului vascular cerebral. Aceste două obiective au fost atinse în cadrul **Capitolului I** ce poartă denumirea de „**Analiza domeniului**”.

Următoarele obiective vizează nemijlocit crearea modelului de clasificare a datelor ce țin de accidentului vascular cerebral. În această ordine de idei, a fost cercetată literatura de specialitate pentru familiarizarea cu acest algoritm și principiul acestuia de lucru. Luând în considerare importanța datelor la antrenarea modelului, au fost cercetate diverse surse de date. S-a optat pentru setul cel mai relevant identificat pe platforma Kaggle. Având la dispoziție o colecție de date curățite a fost posibilă aplicare a patru algoritmi pentru compararea rezultatelor de clasificare – doi algoritmi creați în cadrul lucrării și doi existenți în librăria ScikitLearn. Compararea modelelor a fost făcută în baza metricilor precum acuratețea, precizia, rata de detecție și scorul F1. **Capitolul II** cu denumirea „**Fundamentare teoretică și implementare**” cuprinde obiectivele care se referă la crearea SVM-ului.

În cele din urmă, în **Capitolul III**, „**Realizarea aplicației web**”, a fost atins obiectivul care constă în integrarea modelului de clasificare în cadrul unei aplicații web pentru a oferi o aplicație simplă, atractivă și user-friendly. Aplicația obținută nu constituie un

produs final destinat utilizării pe larg, ce doar o mostră care permite testarea modelului de clasificare.

## ANNOTATION

**in the master's thesis on the topic "Support Vector Machines for Data Classification" of the master student in the group MAI-211M TUTUNARU Eugenia**

The master's thesis includes an introduction, three chapters, a conclusion, 17 bibliographic sources and 25 figures.

**Keywords:** support vector machines, artificial intelligence, supervised machine learning, classification, classification model, stroke, python, web application.

The purpose of this work is to apply the Support Vector Machines algorithm from machine learning to classify based on vital and behavioural characteristics into two classes – predisposed and non-predisposed.

In order to achieve the goal, a set of general objectives was established. The specialised machine learning literature, especially the Support Vector Machines algorithm, was analysed. Also, the analysis of the problematic area – stroke was carried out. Along with this, the analysis of the field of artificial intelligence creates a clearer picture of the possibilities offered by this discipline. It contributes to identifying models that would solve numerous identified problems, including in medicine and, in the case of the given work, stroke. In **Chapter I, "Domain Analysis"**, these two objectives were achieved.

The following objectives aim to create the classification model for stroke data. Following, the specialised literature was researched for familiarisation with this algorithm and its principle. Various data sources were investigated, considering the importance of data in model training. The most relevant set identified on the Kaggle platform was chosen. It was possible to apply four algorithms, having a collection of cleaned data available, to compare the classification results – two algorithms created within the works and two existing in the ScikitLearn library. The models were compared based on metrics such as accuracy, precision, recall and F1 score. **Chapter II, entitled "Theoretical Foundation and Implementation"**, includes the objectives that refer to the creation of the SVM.

Finally, in **Chapter III, "Web application development"**, the goal of integrating the classification model within a web application to provide a simple, attractive and user-friendly application was achieved. The obtained application is not a final product intended for broad use. It only has one that allows the testing of the classification model.

## CUPRINS

<b>INTRODUCERE</b> .....	<b>8</b>
<b>1 ANALIZA DOMENIULUI</b> .....	<b>9</b>
1.1 Inteligența artificială.....	9
1.2 Învățarea automată .....	12
1.3 Învățarea supervizată.....	14
1.4 Clasificarea și regresia .....	15
1.5 Algoritmi de învățare supervizată .....	16
1.6 Mașini cu suport vectorial.....	18
1.6.1 Clasificator de marjă maximă .....	18
1.6.2 Clasificator cu suport vectorial .....	23
1.6.3 Mașini cu suport vectorial pentru clasificare .....	26
1.7 Problema propusă.....	29
<b>2 FUNDAMENTARE TEORETICĂ ȘI IMPLEMENTARE</b> .....	<b>31</b>
2.1 Baza matematică .....	31
2.2 Implementarea algoritmului SVM .....	39
2.3 Preprocesarea datelor .....	43
2.4 Aplicarea algoritmului .....	46
2.5 Compararea rezultatelor .....	48
2.6 Salvarea modelului.....	50
<b>3 REALIZAREA APLICAȚIEI WEB</b> .....	<b>51</b>
3.1 Instrumente și tehnologii utilizate.....	51
3.2 Dezvoltarea aplicației.....	54
3.3 Vizualizarea rezultatelor .....	60
<b>CONCLUZII</b> .....	<b>63</b>
<b>BIBLIOGRAFIE</b> .....	<b>64</b>

## INTRODUCERE

Inteligența artificială este un domeniu în evoluție rapidă care implică crearea de mașini inteligente care pot îndeplini sarcini care necesită de obicei inteligență umană, cum ar fi înțelegerea limbajului natural, recunoașterea imaginilor și modelelor și luarea deciziilor. Pe când, învățarea automată este un subdomeniu al inteligenței artificiale care implică crearea de algoritmi și modele care permit computerelor să învețe și să facă predicții asupra datelor, fără a fi programate în mod explicit.

Mașinile cu suport vectorial sau SVM este un algoritm puternic de învățare automată utilizat pentru sarcini de clasificare și regresie. Funcționează prin găsirea hiperplanului optim care separă datele în diferite clase. SVM a fost aplicat cu succes în multe aplicații din lumea reală, inclusiv clasificarea imaginilor, clasificarea textului și diagnosticul medical.

Detectarea accidentului vascular cerebral este un domeniu în care clasificarea SVM a arătat rezultate promițătoare. Accidentul vascular cerebral este o cauză principală de dizabilitate și deces la nivel mondial, iar depistarea precoce este crucială pentru un tratament eficient. SVM poate fi utilizat pentru a analiza datele pacientului, cum ar fi imagistica medicală și istoricul pacientului, pentru a diagnostica cu acuratețe accidentul vascular cerebral și a prezice rezultatele pacientului. Acest lucru are potențialul de a îmbunătăți semnificativ îngrijirea și rezultatele accidentului vascular cerebral.

Progresele în inteligența artificială și învățarea automată ajută, de asemenea, la îmbunătățirea altor domenii ale asistenței medicale, inclusiv diagnosticarea bolilor, descoperirea medicamentelor și medicina personalizată. Cu cantitățile mari de date generate de tehnologia medicală modernă, algoritmi de învățare automată pot ajuta la identificarea tiparelor și a relațiilor care ar putea să nu fie evidente pentru cercetătorii umani, conducând la tratamente mai eficiente și la rezultate mai bune pentru pacient.

În această lucrare se cercetează algoritmul SVM și anume ca algoritm de clasificare. Domeniul problematic în care se propune aplicarea algoritmului este medicina și anume accidentul vascular cerebral. Astfel, se propune crearea unui model SVM de clasificare, care va fi antrenat pe date preprocesate ce țin de AVC, iar modelul obținut va fi integrat într-o aplicație web pentru o interfață mai atractivă. Un astfel de model identifică clasa căreia poate aparține omul în dependență de valorile introduse. Scopul de bază a modelului este de a identifica persoanele cu risc sporit de AVC, deoarece cel mai bun tratament este prevenirea bolii.

## Bibliografie

1. TARAN, Tatiana, ZUBOV Dmytro. *Искусственный интеллект. Теория и приложения* [online]. Lugansk: Видавництво СНУ ім. В.Даля, 2006. [citat 11.02.2023]. Disponibil: [https://www.researchgate.net/profile/Zubov-Dmytro/publication/235686352\\_Artificial\\_Intelligence\\_Theory\\_and\\_Application/links/0fcfd5128ee49188d3000000/Artificial-Intelligence-Theory-and-Application.pdf](https://www.researchgate.net/profile/Zubov-Dmytro/publication/235686352_Artificial_Intelligence_Theory_and_Application/links/0fcfd5128ee49188d3000000/Artificial-Intelligence-Theory-and-Application.pdf)
2. BULLINARIA, John. IAI : The Roots, Goals and Sub-fields of AI [online]. *Universitatea din Birmingham*, 2005 [citat 11 februarie 2023]. Disponibil: <https://www.cs.bham.ac.uk/~jxb/IAI/w2.pdf>
3. *IBM: What is machine learning?* International Business Machines Corporation [citat 11.02.2023]. Disponibil: <https://www.ibm.com/topics/machine-learning>
4. BURKOV, Andriy. *The Hundred-Page Machine Learning Book* [online]. [citat 11.02.2023]. Disponibil: <http://themlbook.com/wiki/doku.php>
5. *IBM: What is supervised learning?* International Business Machines Corporation [citat 17.02.2023]. Disponibil: <https://www.ibm.com/topics/supervised-learning>
6. SHAI, Shalev-Shwartz, SHAI Ben-David. *Understanding Machine Learning: From Theory to Algorithms* [online]. New York: Cambridge University Press, 2014. [citat 17.02.2023] Disponibil: <https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf>
7. JAKKULA, Vikramaditya. School of EECS, Washington State University. In: *Tutorial on Support Vector Machine (SVM)*. 2006, nr. 37 [citat 17.02.2023]. Disponibil: <https://course.ccs.neu.edu/cs5100f11/resources/jakkula.pdf>
8. JAMES, Gareth, WITTEN Daniela, HASTIE, Trevor, TIBSHIRANI, Robert. *An Introduction to Statistical Learning with Applications in R* [online]. New York: Springer Science+Business Media, 2021. [citat 21.02.2023]. Disponibil: [https://hastie.su.domains/ISLR2/ISLRv2\\_website.pdf](https://hastie.su.domains/ISLR2/ISLRv2_website.pdf)
9. *WHO: Stroke, Cerebrovascular accident*. World Health Organization Regional Office for the Eastern Mediterranean [citat 24.02.2023]. Disponibil: <https://www.emro.who.int/health-topics/stroke-cerebrovascular-accident/index.html>

10. SCHOLKOPF, Bernhard, SMOLA, Alexander. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* [online]. The MIT Press, 2018 [citat 10.03.2023]. Disponibil: <https://direct.mit.edu/books/book/1821/Learning-with-KernelsSupport-Vector-Machines>
11. MISHRA, Bhaskar. *SVM Implementation using CVXOPT – Python* [online]. Github Bhaskar's Blog. [citat 15.03.2023] Disponibil: <https://mbhaskar1.github.io/machine%20learning/2019/07/04/svm-using-cvxopt.html>
12. CVXOPT: CVXOPT. [citat 20.03.2023] Disponibil: <https://cvxopt.org/>
13. HAN, Jiawei, KAMBER, Micheline, PEI, Jian. *Data Mining: Concepts and Techniques* [online]. Waltham: Elsevier Inc. [citat 17.03.2023] Disponibil: <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>
14. *Kaggle: Diabetes, Hypertension and Stroke Prediction*. Kaggle Inc. [citat 20.03.2023] Disponibil: <https://www.kaggle.com/datasets/prosperchuks/health-dataset>
15. *Component source: About PyCharm. The Python IDE for professional developers*. ComponentSource [citat 15.04.2023] Disponibil: <https://www.componentsource.com/product/pycharm/about>
16. *Java Tpoint: Python Flask Tutorial*. Java Tpoint. [citat 16.04.2023] Disponibil: <https://www.javatpoint.com/flask-tutorial>
17. *W3C: HTML & CSS*. Worl Wide Web Consortium [citat 16.04.2023] Disponibil: <https://www.w3.org/standards/webdesign/htmlless>