

Ministerul Educației 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

Învățarea automată în detectarea bolilor

Student: Jalbă Andrea (*Jalbă*)

Conducător: Bobicev Victoria (_____)

Consultanți: _____ (_____)

Recenzent: _____ (_____)

ADNOTARE

la teza de master cu tema “Învățarea automată în detectarea bolilor”, a studentei MAI-211M Andrea Jalbă

Cuvinte cheie: Naïve Bayes, Decision Tree Classifier, Random Forest Tree, HTML, CSS, JavaScript, Bootstrap, Python, Flask.

Teza de master are drept scop crearea unei platforme web care utilizează tehnologia de învățare automată pentru a ajuta pacienții în procesul de diagnosticare a bolilor și oferirea accesibilității și flexibilității în accesarea serviciilor medicale. Pentru a atinge acest scop, sunt definite mai multe obiective principale: prezicerea diagnosticului în baza simptomelor pacientului, prezicerea diagnosticului bolilor de inimă pe baza mai multor indici, vizualizarea spitalelor din apropiere, posibilitatea de a face programări pentru consultații de urgență și posibilitatea de a comunica cu medicul prin intermediul chat-ului. Aceste funcționalități vor ajuta pacienții să primească diagnosticuri mai precise și mai rapide, să găsească spitale și să programeze consultații într-un mod mai eficient, oferindu-le astfel o experiență mai bună în ceea ce privește accesul la serviciile medicale.

În platformă se utilizează algoritmi de învățare automată: Naïve Bayes, Decision Tree Classifier și Random Forest Tree.

Această platformă a fost realizată într-o aplicație complexă Web cu 2 componente, “front-end” și “back-end”. Pe partea de front-end s-au folosit instrumentele: HTML, CSS, JavaScript, Bootstrap, iar pe partea de back-end au fost folosite instrumentele: Python, Flask.

Metodele bazate pe învățarea automată se folosesc pentru a îmbunătăți precizia și acuratețea diagnosticii medicale, reducând astfel erorile și scăzând costurile asociate cu diagnosticul și tratamentul bolilor.

Implicațiile practice ale cercetării sunt semnificative pentru domeniul medical, cu potențiale implicații în dezvoltarea unor tehnologii mai bune de detectare a bolilor, care ar putea fi folosite în viitor pentru îmbunătățirea sănătății și calității vieții oamenilor.

ANNOTATION

The master thesis entitled "Machine Learning in disease detection " of the master student gr. MAI-211M Andrea Jalbă

Keywords: Naïve Bayes, Decision Tree Classifier, Random Forest Tree, HTML, CSS, JavaScript, Bootstrap, Python, Flask.

The master's thesis aims to create a web platform that uses machine learning technology to help patients in the process of diagnosing diseases and providing accessibility and flexibility in accessing medical services. To achieve this goal, several main objectives are defined: predicting the diagnosis based on the patient's symptoms, predicting the diagnosis of heart diseases based on several indices, viewing nearby hospitals, the ability to make appointments for emergency consultations and the ability to communicate with the doctor via chat. These functionalities will help patients receive more accurate and faster diagnoses, find hospitals and schedule appointments more efficiently, thus giving them a better experience in accessing medical services.

The platform uses the machine learning algorithms: Naïve Bayes, Decision Tree Classifier and Random Forest Tree.

This platform was made into a complex Web application with 2 components, "front-end" and "back-end". On the front-end side, the following tools were used: HTML, CSS, JavaScript, Bootstrap, and on the back-end side, the following tools were used: Python, Flask.

Machine learning-based methods are used to improve the precision and accuracy of medical diagnosis, thereby reducing errors and lowering costs associated with disease diagnosis and treatment.

The practical implications of research are significant for the medical field, with potential implications for the development of better disease detection technologies that could be used in the future to improve people's health and quality of life.

CUPRINS

INTRODUCERE.....	9
1 ANALIZA DOMENIULUI DE STUDIU	10
1.1 Descrierea domeniului	10
1.2 Actualitatea temei	15
1.3 Sistemele similare și analiza acestora.....	17
1.3.1 Isabel Healthcare	17
1.3.2 Enlitic	19
1.3.3 Zebra Medical Vision	21
1.3.4 TwoXAR	23
1.3.5 Analiza comparativă a sistemelor.....	25
1.4 Scopul și obiectivele.....	26
2 PROIECTAREA SISTEMULUI ȘI TEHNOLOGIILE UTILIZATE.....	28
2.1 Medii și tehnologii de dezvoltare	28
2.1.1 PyCharm	28
2.1.2 HTML.....	29
2.1.3 CSS	29
2.1.4 Java Script	29
2.1.5 Bootstrap.....	30
2.1.6 Python.....	31
2.1.7 Flask	32
2.1.8 Google Maps API.....	33
2.2 Tehnologii de învățare automată	33
2.2.1 Setul de date	33
2.3 Algoritmi de învățare automată	36
2.3.1 Algoritmii Naive Bayes și Decision Tree Classifier.....	36
2.4 Algoritmii Random Forest Classifier.....	37
2.5 Tehnologia de procesare a limbajului natural.....	38
2.6 Proiectarea și crearea bazei de date	38
2.6.1 Tehnologii de dezvoltare a bazei de date.....	38
2.6.2 Proiectarea bazei de date	39
2.6.3 Crearea bazei de date.....	40
2.7 Modelarea și proiectarea sistemului	42
2.7.1 Diagrama de context.....	42
2.7.2 Diagrama Use Case	42
2.7.3 Diagrama de secvență.....	44
2.7.4 Diagrama de comunicare	45

3 DOCUMENTAREA PRODUSULUI REALIZAT	51
CONCLUZII	63
BIBLIOGRAFIE	64
ANEXA	65

INTRODUCERE

Învățarea automată reprezintă un subdomeniu al inteligenței artificiale, care permite sistemelor să învețe din experiență și să îmbunătățească performanța fără a fi programate explicit. Datorită prezenței unei cantități mare de date, putem utiliza învățarea automată în mod eficient pentru a detecta și a diagnostica diverse boli. Acest lucru se realizează prin analizarea cantității imense de date medicale precum istoricul pacientului, informații despre simptome anumite rezultate a analizelor de laborator, analizarea imaginilor medicale pentru a identifica modelele specifice care pot indica prezența unei boli sau a unui risc crescut de boală. NLP poate fi utilizat pentru a analiza anumite rapoarte medicale, conversațiile audio dintre medic și pacienți precum și alte informații bazate pe text pentru a oferi o imagine completă și o concluzie despre starea pacientului. Aceste două instrumente împreună reprezintă o soluție puternică în detectarea bolilor pe baza simptomelor și în personalizarea tratamentelor.

Datorită tehnologiei de învățare automată și procesarea limbajului natural (NLP), pacienții vor economisi timp și bani pentru că nu va trebui să parcurgă distanțe lungi pentru a consulta un medic, dar voi putea comunica cu medicul prin intermediul unui chat.

De asemenea un sistem care folosește învățarea automată și NLP poate oferi consultații virtuale cu medicul. În cazurile de urgență când programarea la medic este inevitabilă și mersul la spital este necesar este nevoie de un asemenea sistem pentru a verifica disponibilitatea medicilor din spitalele din apropiere și programarea la un medic anumit.

Totodată prin analiza datelor și descoperirea de noi modele și relații, se pot identifica medicamente pentru tratarea bolilor.

Este important de a lua în considerare acuratețea și fiabilitatea modelelor pentru garantarea utilizării sigure și eficiente a tehnologiilor de ML și NLP în diagnosticarea și tratarea bolilor. Cu un abord riguros și responsabil, învățarea automată poate oferi un potențial imens pentru transformarea și îmbunătățirea domeniului medical și a sănătății umane.

BIBLIOGRAFIE

1. Stuart J. Russell & Peter Norvig (2010). Artificial Intelligence a Modern Approach Third Edition. Upper Saddle River, New Jersey, SUA: Prentice Hall. p. 2
2. Adriana Birlutiu. Mașini instruibile și recunoașterea formelor. 2018. Universitatea “1 Decembrie 1918” din Alba Iulia. Disponibil pe internet: http://adrianabirlutiu.uab.ro/cursuri/MIRF/note_curs_lab_1.pdf
3. Christopher M. Bishop (2006). Pattern Recognition and Machine Learning. New York, SUA: Springer. Chapter 1. Section 1.2.3. p.5
4. Daniel Jurafsky și James H. Martin (2020). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. New York, SUA: Pearson Education.
5. About PyCharm [online] [citată 2023-03-03]. Disponibil pe internet: <https://www.jetbrains.com/pycharm/>
6. About Bootstrap [online] [citată 2023-04-03]. Disponibil pe internet: <https://www.techtarget.com/whatis/definition/bootstrap>
7. Jake VanderPlas (2016). Python Data Science Handbook. California, SUA: O'Reilly Media. p.1, 117, 225.
8. Cardiovascular disease information [online] [citată 2023-27-03]. Disponibil pe internet: <https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset>
9. Jeffrey Stanton (2013). Introduction to Data Science. Universitatea Syracuse. p. 155
10. Pang-Ning Tan, Michael Steinbach, Vipin Kumar (2018). Introduction to Data Mining. Boston, Massachusetts, SUA: Pearson Education.