

1. Propunere de proiect

Titlul proiectului: Platformă de analiză a riscului la atac de cord bazat pe Machine Learning

- **Autor:** Szakacs Emma-Evelin
- **Conducator științific:** Conf.dr.ing Anghel Ionuț Manuel

Introducere – Obiectivele și scopul proiectului

Scopul proiectului este dezvoltarea unui sistem care să asiste medicii în detectarea timpurie a atacurilor de cord și gestionarea eficientă a factorilor de risc ce pot duce la apariția acestei afecțiuni.

Obiectivele principale ale proiectului sunt:

- Antrenarea și evaluarea mai multor modele de învățare automată pentru a identifica modelul cu cea mai bună performanță, luând în calcul de mai multe metrice de evaluare.
- Integrarea modelului de învățare automată selectat într-o aplicație web destinată medicilor și asistenților medicali.
- Introducerea unor funcționalități suplimentare în aplicație, cum ar fi vizualizarea progresului pacienților prin grafice personalizate și vizualizarea tendințelor comune în rândul pacienților din sistem.
- Dezvoltarea unei interfețe intuitive și ușor de utilizat, pentru a economisi timpul medicilor și a asistenților medicali.

Descrierea conținutului proiectului

Bolile cardiovasculare rămân principala cauză a mortalității la nivel mondial, fiind responsabile pentru aproximativ 17.9 milioane de decese anual, ceea ce reprezintă 31% din toate decesele globale, conform World Health Organization (WHO). În mod alarmant, aproximativ 85% dintre aceste decese sunt cauzate de atacuri de cord și accidente vasculare cerebrale, evidențiind o nevoie globală pentru îmbunătățirea metodelor de diagnosticare și intervenție timpurie. [1]

Proiectul își propune să redefinească gestionarea riscului cardiovascular, mai specific, al atacurilor de cord, adoptând o abordare bazată pe machine learning. Sistemul integrează un proces de selecție a modelului cu cele mai bune performanțe, analizând diverși algoritmi, precum Random Forest, Decision Trees, Support Vector Machines, Logistic Regression, Artificial Neural Networks, și, nu în ultimul rând, tehnici de ensemble. Această selecție se bazează pe mai multe metrice de evaluare, printre care acuratețea, sensibilitatea, precizia, scorul F1 și log loss.

Setul de date folosit pentru antrenarea și testarea modelelor alese pentru comparație este setul de date Cleveland, disponibil pe UCI Machine Learning Repository [2]. Alegerea acestui set de date este motivată de recunoașterea sa în comunitatea științifică pentru valoarea practică și contribuția în cercetarea medicală.

Pe lângă găsirea modelului cu cele mai bune performanțe în cadrul predicției bolilor cardiovasculare, acest proiect își propune și dezvoltarea unui mediu interactiv cu ajutorul căruia cadrele medicale și pacienții să dispună de soluția propusă. Acest mediu constă într-o aplicație web care să permită existența a trei tipuri de utilizatori: administrator, doctor și asistent medical.

Sistemul va permite asistentilor medicali să introducă în sistem o gamă largă de date clinice pentru pacienți, incluzând analize de sânge, angiografii și informații despre istoricul familial al bolilor de inimă. Aceste date sunt esențiale pentru o evaluare elaborată și personalizată a riscului unui atac de cord. Doctorii vor avea acces la predicția modelului de machine learning, vor putea analiza grafice personalizate care ilustrează evoluția pacienților pe parcursul timpului și vor putea genera rapoarte. Graficele vor reflecta variabile cheie în evaluarea riscului unui atac de cord, precum variațiile nivelului de colesterol și tensiunea arterială. Mai mult, sistemul va oferi posibilitatea de a explora grafice generale, care să identifice corelații între diferiți factori de risc și caracteristici, oferind astfel o perspectivă valoroasă asupra dinamicii bolilor cardiovasculare în populația monitorizată.

Studiu bibliografic

În domeniul diagnosticului și prezicerii bolilor de inimă, numeroase studii au abordat diverse metode de clasificare și predicție, cu scopul de a îmbunătăți precizia și eficiența acestui proces. Deși există o varietate de abordări utilizate în literatura existentă, cercetările recente au evidențiat beneficiile unei abordări care combină rezultatele mai multor modele sau algoritmi de învățare automată.

Studiile anterioare, precum cele realizate de Wang et al. [3] și Lu și Uddin [4], au demonstrat că utilizarea unei combinații între modelele bazate pe rețele neuronale și cele bazate pe grafuri poate îmbunătăți semnificativ precizia prezicerii bolilor de inimă. De asemenea, cercetările efectuate de Azmi et al. [5] și Sun et al. [6] au evidențiat beneficiile folosirii unor modele care combină atât metode de învățare automată, cât și modele statistice tradiționale.

În ceea ce privește aspectele tehnice ale dezvoltării acestor modele, literatura existentă sugerează că atât extragerea caracteristicilor, cât și ajustarea hiperparametrilor joacă un rol crucial în obținerea rezultatelor optime. Studii precum cele realizate de Patro et al. [7] și Sun et al. [8] au evidențiat importanța unei selecții adecvate a caracteristicilor relevante și a parametrilor optimi pentru performanțe îmbunătățite ale modelului. Astfel, este evident că abordarea care combină mai multe modele și algoritmi de învățare automată, împreună cu o atenție deosebită acordată proceselor de extragere a caracteristicilor și ajustării hiperparametrilor, poate contribui semnificativ la îmbunătățirea predicției și diagnosticului bolilor de inimă.

Resurse necesare

Pentru dezvoltarea aplicației sunt necesare următoarele dependențe:

1. Hardware
 - Procesorul să fie cel puțin generația 5 (Intel, AMD)
 - Memoria RAM minim 8GB
2. Software
 - Pe partea de server:
 - Baza de date PostgreSQL
 - Backend: Java 8.0 utilizând SpringBoot Framework
 - Algoritmi de Machine Learning: Jupyter Notebook cu librării Python
 - Pe partea de client:
 - Vue.js Framework pentru interfața utilizatorilor
 - Librării JavaScript – Axios, Vuex
 - Pe partea de Machine Learning și analiză de date
 - Biblioteca Pandas pentru manipularea datelor
 - Matplotlib și Seaborn pentru vizualizarea datelor
 - Librării ML precum scikit-learn pentru algoritmi de clasificare și regresie
3. Dependențe de Sistem
 - Sistemul de operare: indiferent (cross-platform), deoarece aplicația este accesată prin intermediul unui browser web
 - Baza de date PostgreSQL trebuie să fie pre-configurată și accesibilă de către backend-ul Spring

Rezultate așteptate

Rezultatele așteptate includ:

- Dezvoltarea unui model de învățare automată utilizând ensemble learning, care să atingă o acuratețe de peste 90% în predicția riscului la afecțiuni cardiovasculare. De asemenea, modelul trebuie să aibă un recall foarte mare, deoarece vrem să evităm situația în care un pacient are o boală de inimă, însă modelul prezice contrariul.
- Crearea unei aplicații web eficiente și ușor de utilizat pentru medici, care să optimizeze gestionarea datelor clinice ale pacienților și să contribuie la economisirea timpului acestora.

Planificarea elaborării proiectului pe săptămâni

Week	Period	Research	Implementation	Writing
1 - 6	2 oct – 12 noi	Studiu bibliografic	-	Scurte rezumate a ce am citit
7 - 14	13 noi – 7 ian	Studiu bibliografic	Implementare și testare modele de ML	Analiză comparativă între modelele implementate
14 - 16	8 ian – 21 ian	-	presesiune	-
17 - 21	22 ian – 25 feb	-	sesiune	-
22	26 feb – 10 mar		Ensemble learning	
23	11 mar – 17 mar		Realizare arhitectură conceptuală	
24	18 mar – 24 mar		Realizare mockup-uri	
25 - 26	25 mar – 7 aprilie		Microserviciu de autentificare	Introducere
27 - 28	8 apr – 21 apr		Microserviciu asistenți medicali	Studiu bibliografic
29 - 31	22 apr – 12 mai		Microserviciu medici	Analiză și design
32	13 mai – 19 mai		Performanță și securitate	Implementare
33	20 mai – 26 mai		Îmbunătățiri în cadrul implementarii	Implementare
34	27 mai – 2 iunie			Rezultate și concluzii
35-38	3 iunie – 30 iunie	-	Sesiune și presesiune	-
39	1 iulie – 7 iulie		Testare finală	Îmbunătățiri documentație

Table 1: Timeline-ul proiectului până în momentul predării

2. Cuprins preliminar

Capitolul 1. Introducere

- 1.1 Importanța predicției și diagnosticului timpuriu al bolilor cardiace
- 1.2 Motivația pentru utilizarea machine learning în medicină

Capitolul 2. Obiectivele proiectului

- 2.1. Specificarea problemei
- 2.2. Cerințe funcționale
- 2.3. Cerințe non-funcționale

Capitolul 3. Studiu bibliografic

- 3.1 Seturi de date utilizate în literatură
- 3.2 Algoritmi și modele de machine learning pentru predicția bolilor cardiace
- 3.2 Tendințe și direcții viitoare

Capitolul 4. Analiză și fundamentare teoretică

- 4.1 Modele
 - 4.1.1 Decision Tree
 - 4.1.2 Random Forest
 - 4.1.3 Logistic Regression
 - 4.1.4 Artificial Neural Networks
 - 4.1.5 Support Vector Machines
 - 4.1.6 Metode de ensemble learning
- 4.2 Metrice de evaluare a performanței modelelor
- 4.3 Fine-tuning
- 4.4 Analiză comparativă

Capitolul 5. Proiectare de detaliu și implementare

- 5.1 Arhitectura conceptuală a aplicației
- 5.2 Mediu de lucru
 - 5.2.1 Resurse necesare
- 5.3 Tehnologii folosite
- 5.4 Design al interfeței utilizator
- 5.5 Implementare

- 5.5.1 Microserviciu de autentificare
- 5.5.2 Microserviciu de gestionare și analiză a pacienților

Capitolul 6. Testare și validare

Capitolul 7. Manual de instalare și utilizare

Capitolul 8. Concluzii

- 8.1 Contribuții și realizări
- 8.2 Analiza rezultatelor
- 8.3 Dezvoltări ulterioare

Bibliografie

3. Bibliografie

- [1] World Health Organization. (2021). "*Cardiovascular diseases (CVDs)*" [Online]. Disponibil la: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- [2] Detrano, R., Janosi, A., Steinbrunn(1989). "*Heart Disease Data Set*". UCI Machine Learning Repository. Disponibil la: <https://archive.ics.uci.edu/ml/datasets/heart+disease>
- [3] Wang, R.; Chang, M.-C.; Radigan, M. Modeling Latent Comorbidity for Health Risk Prediction Using Graph Convolutional Network. In Proceedings of the Thirty-Third International Flairs Conference, North Miami Beach, FL, USA, 17–20 May 2020. [Online]. Disponibil la: <https://cdn.aaai.org/ocs/18459/18459-79399-1-PB.pdf>
- [4] Haohui Lu, Shahadat Uddin. "*Disease Prediction Using Graph Machine Learning Based on Electronic Health Data: A Review of Approaches and Trends*" School of Project Management, Faculty of Engineering, The University of Sydney, Forest Lodge, Sydney, NSW 2037, 2023. [Online]. Disponibil la: <https://www.mdpi.com/2227-9032/11/7/1031>
- [5] Javed Azmi, Muhammad Arif, Md Tabrez Nafis. "*A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data*". Medical Engineering & Physics, Vol. 105, July 2022, 103825. [Online]. Disponibil la: <https://doi.org/10.1016/j.medengphy.2022.103825>
- [6] Sun Z, Dong W, Shi H, Ma H, Cheng L and Huang Z. "*Comparing Machine Learning Models and Statistical Models for Predicting Heart Failure Events: A Systematic Review and Meta-Analysis*". Front. Cardiovasc. Med., 2022, Volume 9. [Online]. Disponibil la: <https://doi.org/10.3389/fcvm.2022.812276>
- [7] Patro, S.P., Padhy, N. & Chiranjeevi, D. "*Ambient assisted living predictive model for cardiovascular disease prediction using supervised learning*". Evol. Intel. 14, 941–969 (2021). [Online]. Disponibil la: <https://doi.org/10.1007/s12065-020-00484-8>
- [8] Zhenchao Sun , Hongzhi Yin , Hongxu Chen. "*Disease Prediction via Graph Neural Networks*".IEEE Journal of Biomedical and Health Informtics, Vol. 25, No. 3, March 2021. [Online]. Disponibil la: <https://ieeexplore.ieee.org/document/9122573>
- [9] Suraj Kumar Gupta, Aditya Shrivastava, S. P. Upadhyay, Pawan Kumar Chaurasia. "*A Machine Learning Approach for Heart Attack Prediction*". International Journal of Engineering and Advanced Technology, Volume-10, Issue-6, August 2021. [Online]. Disponibil la: <https://dx.doi.org/10.35940/ijeat.F3043.0810621>

