Script prezentare licenta

# Deschidere – Slide 1

Bună ziua, numele meu este Lopătaru Mihnea și astăzi voi prezenta lucrarea mea de licență intitulată „CauseIT”, o platforma destinata profesionistilor din domeniul medical, dezvoltată sub coordonarea doamnei Lector Doctor Plajer Ioana Cristina și a domnului Șef de Lucrări Doctor Inginer Danciu Gabriel.

# Cuprins – Slide 2

Voi începe cu o scurtă introducere în contextul și motivația alegerii acestui subiect, urmând să detaliez arhitectura aplicației și modelul de învățare automată utilizat. De asemenea, voi discuta despre rezultatele obținute și metodele de validare ale acestora.

# Introducere – Slide 3, Slide 4

## Slide 3

În ultimii ani, tehnologia a avansat rapid, extinzându-și aplicabilitatea în diverse sectoare, inclusiv în cel medical. Odată cu acest progres, algoritmii de inteligență artificială și învățare automată și-au îmbunătățit considerabil performanțele, fiind din ce în ce mai des integrați în practica medicala, fapt ce se poate observa și din graficul de pe RBC Capital Markets.

## Slide 4

Odată cu digitalizarea, volumul de date medicale a crescut exponențial, așa cum se observă si în graficul de pe Health IT Data. În acest context, abilitatea algoritmilor de învățare automată de a procesa rapid și eficient seturi de date voluminoase, devine esentiala. Platforma CauseIT utilizează astfel de algoritmi, permițând medicilor să încarce date specifice unei afecțiuni, furnizand mai apoi parametrii relevanți și importanța acestora pentru diferitele stadii ale bolii. Astfel, informațiile sunt gestionate eficient, oferind perspective detaliate esențiale pentru deciziile clinice informate.

# Scop si motivatie – Slide 5, Slide 6

## Slide 5

Scopul acestei lucrări este de a explora potențialul algoritmilor de învățare automată în identificarea cauzelor bolilor. Platforma CauseIT, a fost inițial dezvoltată pentru identificarea factorilor declanșatori ai cancerului pancreatic în cadrul proiectului IHelp, în urma succesului obițnut, a fost extinsă pentru a acoperi o gama mai larga de boli. CauseIT servește drept o unealtă pentru medici, oferind informații suplimentare și ajutând la identificarea corelațiilor care ar putea fi omise in cadrul proceselor analitice clasice.

## Slide 6

Motivația acestei lucrări este dată de necesitatea gestionării și analizei volumelor mari de date medicale. Algoritmii de Machine Learning, precum Random Forest, utilizat în cadrul platformei CauseIT, permit procesarea eficientă a acestor seturi de date. Nu in ultimul rand, platforma CauseIT a fost concepută pentru a fi modulară, facilitând analiza unei varietăți de afecțiuni, eficientizând procesele de diagnosticare și tratament.

# Arhitectura aplicatiei – Slide 7

Arhitectura aplicației CauseIT este concepută pentru a fi modulară și ușor de scalat, fiind structurată pe trei straturi fundamentale:

1. **Frontend**: Dezvoltat in Angular, oferă o interfață utilizator accesibila și usor de folosit.
2. **Server Intermediar**: Dezvoltat în ASP.NET, asigură securitatea platformei și valideaza cererile dintre frontend și backend.
3. **Server Backend**: Dezvoltat în Python, utilizează FastAPI, si se ocupa de procesarea (sau curatarea) datelor și execuția algoritmilor ML.

Arhitectura asigură performanța și securitatea necesare domeniului medical, permițând astfel extinderea funcționalităților aplicației, precum si o adaptare rapidă la noi cerințe.

# Modelul de invatare automata – Slide 8, Slide 9, Slide 10, Slide 11

## Slide 8

În cadrul platformei CauseIT, algoritmul Random Forest a fost utilizat pentru clasificare, deoarece stadiul bolii, adică coloana de interes din seturile medicale, este de obicei marcata cu o etichetă indicând prezența, absența sau stadiul bolii respective. Acest algoritm este capabil să analizeze seturi de date voluminoase, furnizând cu un grad ridicat de acuratețe parametrii declanșatori ai bolii și importanta lor.

## Slide 9

Modelul de Random Forest a fost aplicat pe trei tipuri de seturi de date:

1. **Set de Date Artificial:** Creat pentru a simula diverse scenarii ale bolilor, permițând testarea modelului în medii controlate.
2. **Set de Date Reale Publice:** Conține date reale, disponibile publicului, utilizat pentru a valida performanța modelului în condiții reale.
3. **Set de Date Reale Private:** Furnizat de Siemens in cadrul proiectulului Ihelp, realizat în colaborare cu TMU.

## Slide 10

Pentru a evalua performanța modelului Random Forest, s-a utilizat metoda clasică de împărțire train-test. Modelul a demonstrat o acuratețe ridicată în predicția stadiilor bolilor, folosind parametrii relevanți identificați. Aceste rezultate confirmă eficiența modelului în gestionarea datelor medicale reale, validându-l ca un instrument util pentru personalul medical.

## Slide 11

Deoarece modelul oferă importanța parametrilor extrași, s-a utilizat coeficientul Spearman pentru evaluare, in cadrul testarii pe setul de date artificial. Comparativ cu tehnica PCA, Random Forest a obținut rezultate mai bune, evidențiind eficacitatea sa în identificarea caracteristicilor cheie în seturile de date medicale.

# Concluzii – Slide 12

În concluzie, proiectul a demonstrat eficiența algoritmului Random Forest în identificarea corelațiilor între parametrii clinici ai bolilor complexe, utilizând atât seturi de date artificiale, cât și reale. De asemenea, a evidentiat abilitatea de a identifica cu success influenta parametrilor extrasi asupra stadiului bolii.

Pe viitor, se planifică integrarea datelor multi-omice pentru a îmbunătăți acuratețea diagnosticului și dezvoltarea unei aplicații mobile care să permită medicilor acces constant la datele pacienților și la rezultatele predicțiilor instantanee.

# Demo – Slide 13

Acum, aș dori să vă ofer o demonstrație practică a aplicației CauseIT, pentru a vedea direct funcționalitățile și interfața dezvoltata.

# Incheiere – Slide 14

Vă mulțumesc pentru atenția acordată.