România
Ministerul Apărării Naționale
Academia Tehnică Militară "Ferdinand I"

Facultatea de Sisteme Informatice și Securitate Cibernetică
Ingineria și Securitatea Sistemelor Informatice Militare

Platformă de Analiză Automată a Aplicațiilor Malițioase prin Utilizarea unor Algoritmi de Inteligență Artificială

Conducător Științific
Lect. Univ. Dr. Ing. Alin PUNCIOIU

Absolvent
Sd. Sg. Maj. George-Andrei IOSIF

București
2021

Caz de Utilizare: Context

Caz de Utilizare: Context

- Bancă nou-înființată

Caz de Utilizare: Context

- Bancă nou-înființată
- Nevoia de a-și apăra infrastructura organizațională

Caz de Utilizare: Context

- Bancă nou-înființată
- Nevoia de a-și apăra infrastructura organizațională
- Povara financiară a unei soluții comerciale

Caz de Utilizare: Context

- Bancă nou-înființată
- Nevoia de a-și apăra infrastructura organizațională
- Povara financiară a unei soluții comerciale
- Rigiditatea unei versiuni gratuite a unei soluții comerciale

Caz de Utilizare: Posibilă Soluție

Caz de Utilizare: Posibilă Soluție

- Platformă cu sursă deschisă

Caz de Utilizare: Posibilă Soluție

- Platformă cu sursă deschisă
- Înzestrarea cu maleabilitatea inteligenței artificiale

Caz de Utilizare: Posibilă Soluție

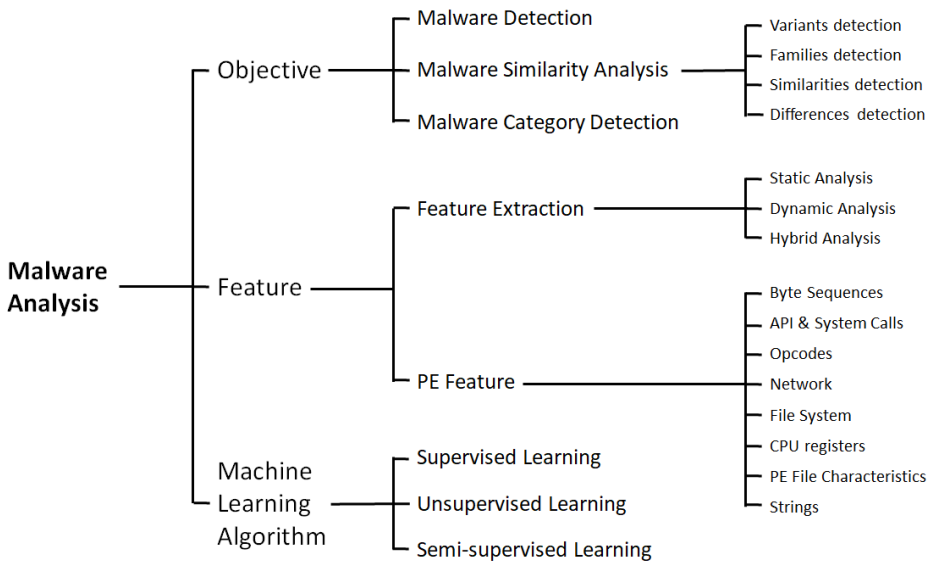
- Platformă cu sursă deschisă
- Înzestrarea cu maleabilitatea inteligenței artificiale
- Condiția unor metrici numerice, obiective

Motivație

- Caz de utilizare prezentat, ca prim motiv de creare a lucrării

Motivație

- Caz de utilizare prezentat, ca prim motiv de creare a lucrării
- Articol "*Survey of machine learning techniques for malware analysis*"



Imagine 1: Taxonomie din Articolul Citat

Objective

Obiective

- Demonstrarea beneficiilor aduse de inteligența artificială analizei de programe malițioase (*academic*)

Obiective

- Demonstrarea beneficiilor aduse de inteligența artificială analizei de programe malițioase (*academic*)
  - Regresia maliției

Obiective

- Demonstrarea beneficiilor aduse de inteligența artificială analizei de programe malițioase (*academic*)
  - Regresia maliției
  - Clasificarea în familii de programe malițioase

Obiective

- Demonstrarea beneficiilor aduse de inteligența artificială analizei de programe malițioase (*academic*)
  - Regresia maliției
  - Clasificarea în familii de programe malițioase
  - Analiza de similaritate

Obiective

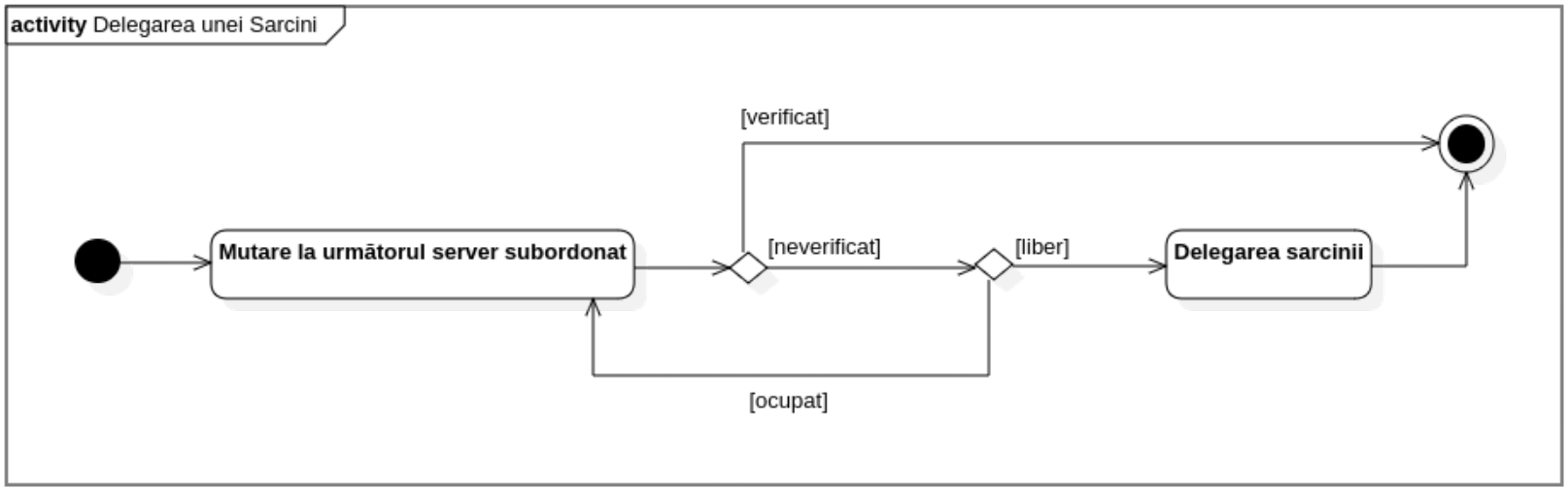
- Demonstrarea beneficiilor aduse de inteligența artificială analizei de programe malițioase (*academic*)
  - Regresia maliției
  - Clasificarea în familii de programe malițioase
  - Analiza de similaritate
- Platformă de automatizare a proceselor de ingineria datelor și analiză de programe (*practic*)

- Arhitectură de tip lider - subordonați

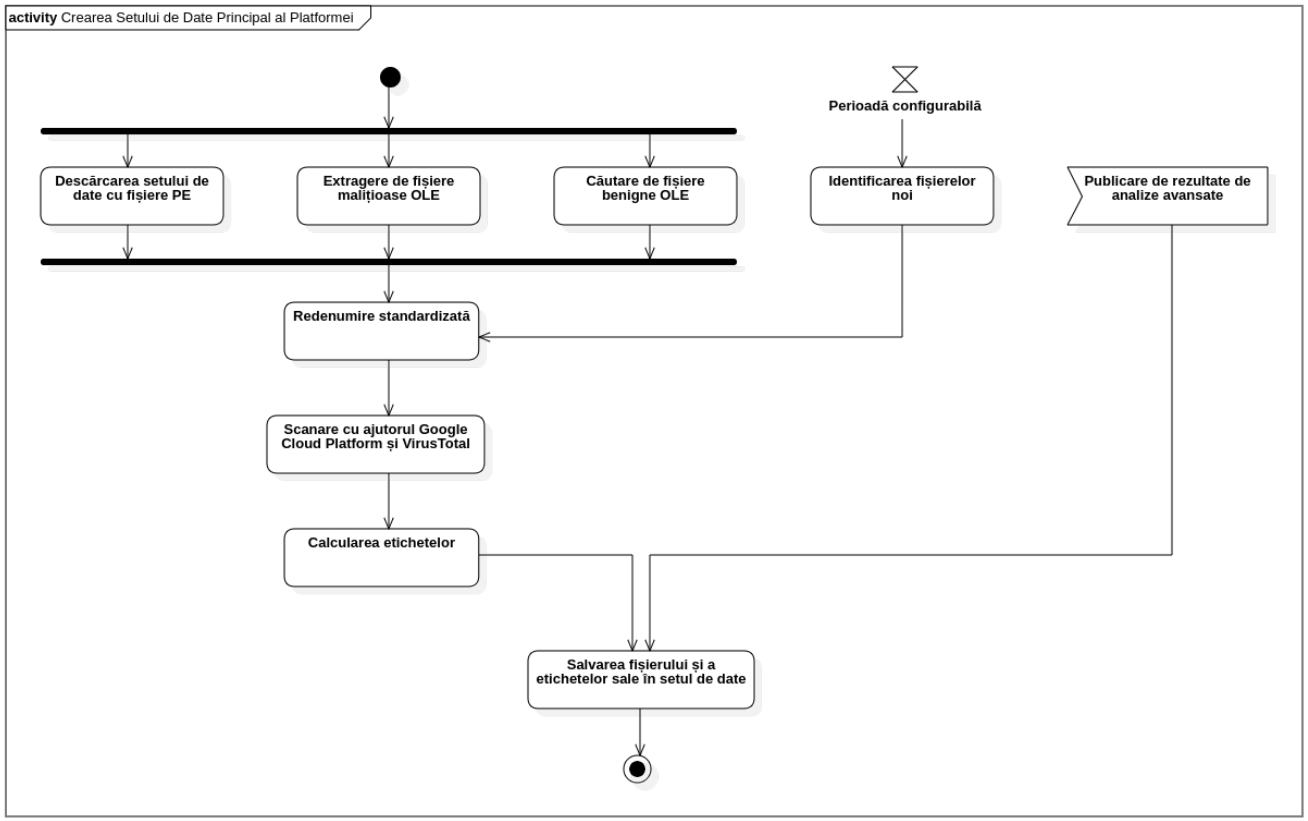
- Arhitectură de tip lider - subordonați
- Servere construite pe fundația modulelor

- Arhitectură de tip lider - subordonați
- Servere construite pe fundația modulelor
- Gruparea modulelor ca o linie de asamblare

- Gestionarea serverelor subordonate
- Etichetarea fișierelor și gestionarea seturilor de date
- Extragerea atributelor din fișiere
- Preprocesarea atributelor
- Gestionarea modelelor de inteligență artificială
- Gestionarea serverelor subordonate

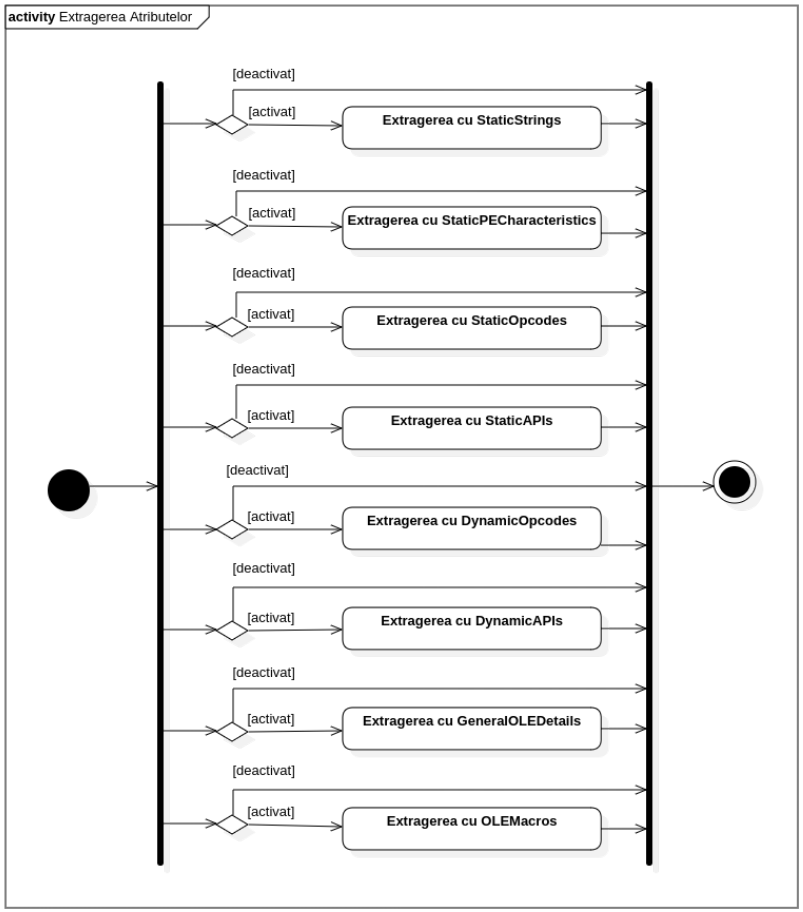


Imagine 2: Arhitectura Modulului pentru Gestionarea Serverelor Subordonate



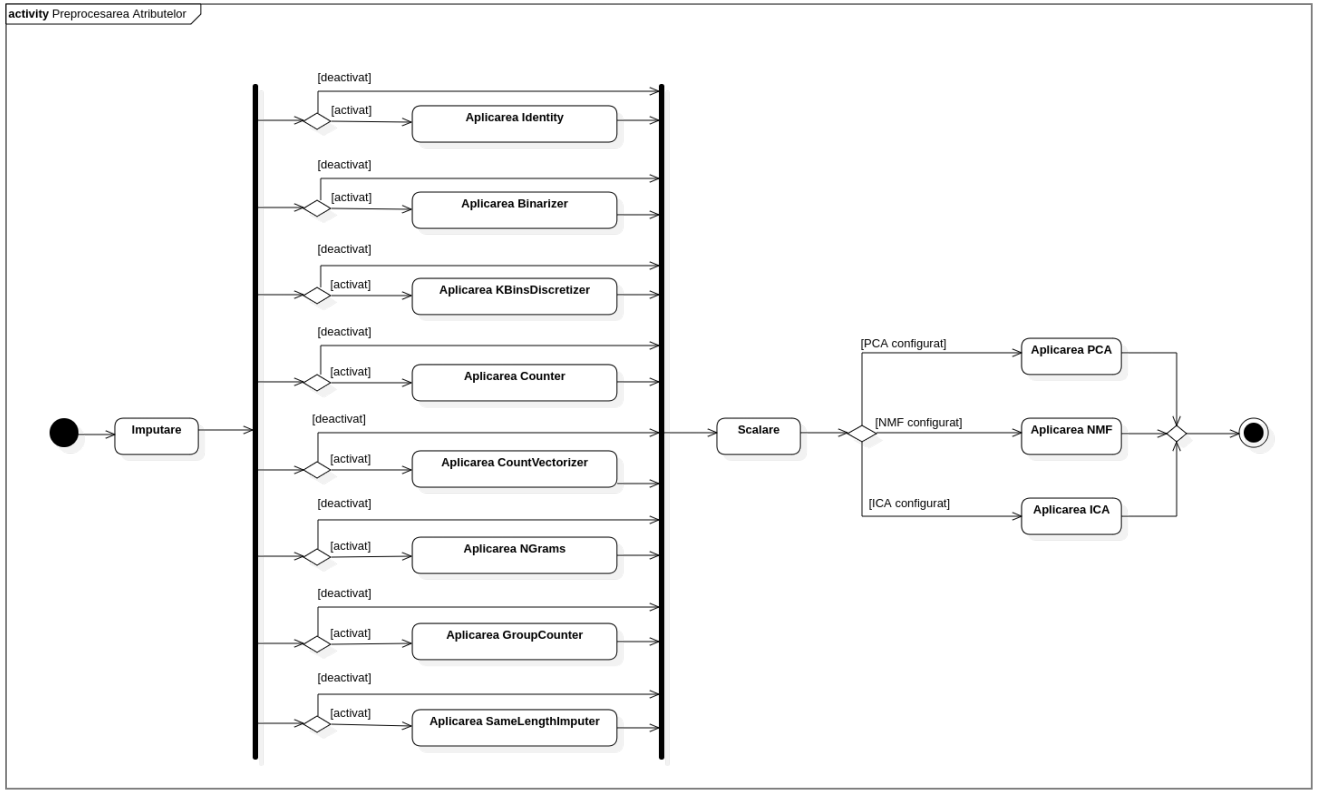
Imagine 3: Arhitectura Modulului pentru Etichetarea Fișierelor și Gestionarea Seturilor de Date

Implementare: Extragerea Atributelor din Fișiere



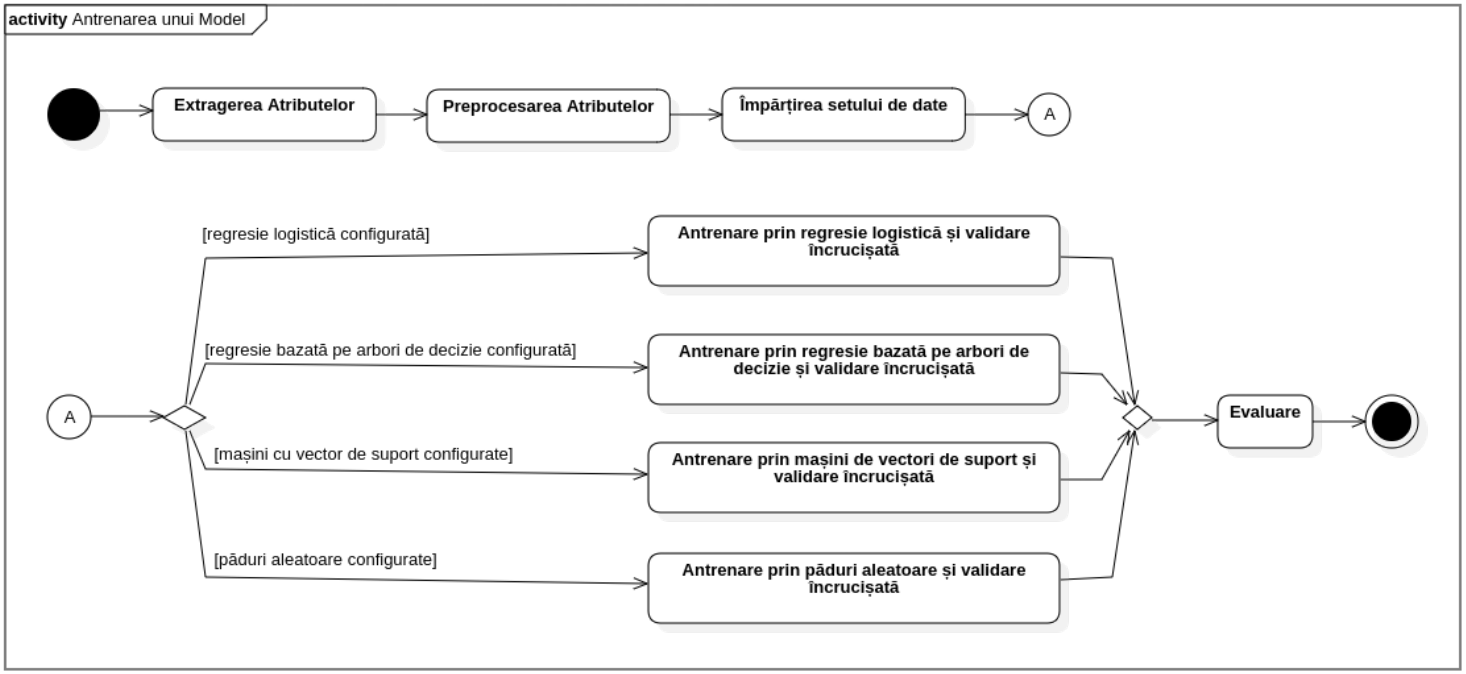
Imagine 4: Arhitectura Modulului pentru Extragerea Atributelor din Fișiere

Implementare: Preprocesarea Atributelor



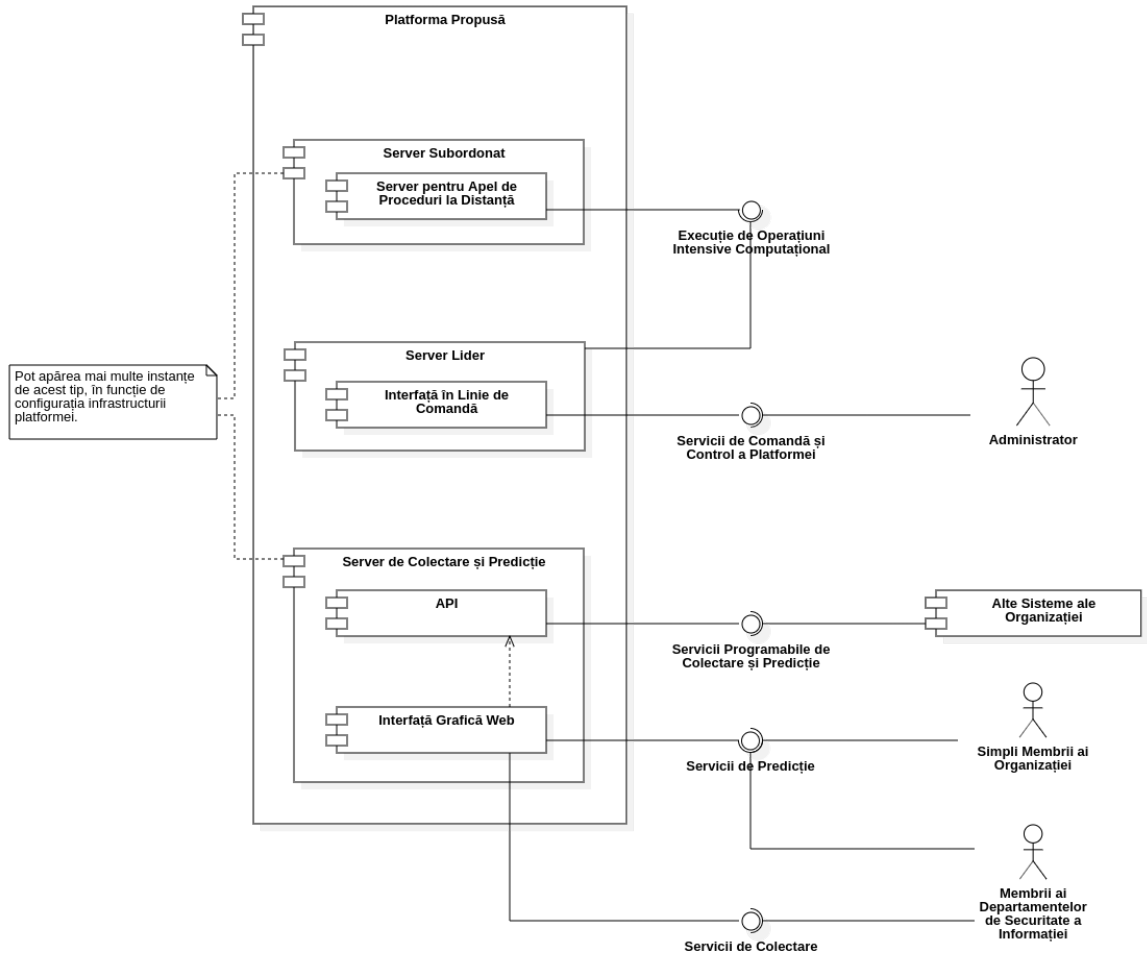
Imagine 5: Arhitectura Modulului pentru Preprocesarea Atributelor

Implementare: Gestionarea Modelelor de Inteligență Artificială



Imagine 6: Arhitectura Modulului pentru Gestionarea Modelelor de Inteligență Artificială

Implementare: Arhitectura de Servere



Imagine 7: Arhitectura de Servere a Platformei

Testare

- Semi-automată, în timpul dezvoltării

- Semi-automată, în timpul dezvoltării
  - De integrare ascendenta

- Semi-automată, în timpul dezvoltării
  - De integrare ascendenta
  - De regresie

- Semi-automată, în timpul dezvoltării
  - De integrare ascendenta
  - De regresie
- În cadrul unui proiect de cercetare al Academiei Tehnice Militare "*Ferdinand I*" București

Evaluare: Exemplu de Rezultate I

Evaluare: Exemplu de Rezultate I

Formatul fișierelor	PE
Numărul de exemplare incluse	1000
Raportul dintre fișiere benigne și malițioase	0.5
Maliție minimă	0.9
Familii de programe malițioase	Toate

Tabel 1: Configurația Setului de Date `LARGE_PE`

Formatul fișierelor	PE
Numărul de exemplare incluse	1000
Raportul dintre fișiere benigne și malițioase	0.5
Maliție minimă	0.9
Familii de programe malițioase	Toate

Tabel 1: Configurația Setului de Date `LARGE_PE`

Descriere	Regresia maliției pentru fișiere PE, cu extractori ce asigură o procesare rapidă a fișierelor
ID-ul setului de date	`LARGE_PE`
Extractori	`STATIC_STRINGS`, `STATIC_PE_CHARACTERISTICS`
Preprocesoare	`IDENTITY`, `COUNTER`, `K_BINS_DISCRETIZER`, `N_GRAMS`

Tabel 2: Configurația Modelului `PE_STATIC_FAST_MALICE`

Eroare maximă	0.7169
Eroare medie absolută	0.0303
Rădăcina erorii medie pătratice	0.0743
Scorul R^2	0.974

Tabel 3: Evaluarea Modelului `PE_STATIC_FAST_MALICE`

Concluzii

Concluzii

- Întrebări la care am răspuns

Concluzii

- Întrebări la care am răspuns
  - De ce?

Concluzii

- Întrebări la care am răspuns
  - De ce?
  - Ce?

Concluzii

- Întrebări la care am răspuns
  - De ce?
  - Ce?
  - Cum?

Concluzii

- Întrebări la care am răspuns
  - De ce?
  - Ce?
  - Cum?
  - Cu ce rezultate?

Concluzii

- Întrebări la care am răspuns
  - De ce?
  - Ce?
  - Cum?
  - Cu ce rezultate?
- Realizări

Concluzii

- Întrebări la care am răspuns
  - De ce?
  - Ce?
  - Cum?
  - Cu ce rezultate?
- Realizări
  - Crearea unui set de date etichetate

Concluzii

- Întrebări la care am răspuns
  - De ce?
  - Ce?
  - Cum?
  - Cu ce rezultate?
- Realizări
  - Crearea unui set de date etichetate
  - Crearea de metode automate de extragere a unor atribute

Concluzii

- Întrebări la care am răspuns
  - De ce?
  - Ce?
  - Cum?
  - Cu ce rezultate?
- Realizări
  - Crearea unui set de date etichetate
  - Crearea de metode automate de extragere a unor atribute
  - Crearea unui proces automatizat de inginerie a datelor

Concluzii

- Întrebări la care am răspuns
  - De ce?
  - Ce?
  - Cum?
  - Cu ce rezultate?
- Realizări
  - Crearea unui set de date etichetate
  - Crearea de metode automate de extragere a unor atribute
  - Crearea unui proces automatizat de inginerie a datelor
  - Integrarea celor menționate în cadrul unei platforme

Concluzii

- Întrebări la care am răspuns
  - De ce?
  - Ce?
  - Cum?
  - Cu ce rezultate?
- Realizări
  - Crearea unui set de date etichetate
  - Crearea de metode automate de extragere a unor atribute
  - Crearea unui proces automatizat de inginerie a datelor
  - Integrarea celor menționate în cadrul unei platforme
  - Integrarea soluției software dezvoltate în cadrul unui proiect de cercetare

Dezvoltare Ulterioară

Dezvoltare Ulterioară

- Îmbunătățiri

Dezvoltare Ulterioară

- Îmbunătățiri
  - Echilibrarea setului de date

Dezvoltare Ulterioară

- Îmbunătățiri
  - Echilibrarea setului de date
- Funcționalități noi

Dezvoltare Ulterioară

- Îmbunătățiri
  - Echilibrarea setului de date
- Funcționalități noi
  - Crearea unor extractori noi cu tehnici de analiză dinamică

Referințe

- Soluția software dezvoltată[1]
- Set de date[2]
- Lucrarea scrisă și prezentare[3]

[1] https://github.com/iosifache/dike

[2] https://github.com/iosifache/DikeDataset

[3] https://github.com/iosifache/BachelorThesis