# Spam Detection

Virna Stefan-Alexandru Gavrila Maria-Denisa

January 11, 2024

## 1 Introducere

Detectarea și filtrarea eficientă a e-mail-urilor nedorite, cunoscute sub numele de spam, reprezintă o provocare semnificativă în lumea actuală a comunicațiilor electronice. O abordare eficientă pentru rezolvarea acestei probleme constă în utilizarea algoritmilor de învățare automată, care pot analiza și clasifica automat mesajele în funcție de conținutul lor. În acest raport, analizam implementarea și evaluarea algoritmului ID3 pentru detectarea de spam, bazându-ne pe setul de date Ling-Spam, un set reprezentativ pentru analiza mesajelor textuale.

## 2 Context

E-mailurile spam reprezintă o amenințare persistentă și agasanta pentru utilizatorii de e-mail, având un impact negativ asupra eficienței și experienței utilizatorului. În încercarea de a contracara această problemă, soluțiile de filtrare automată a spamului au devenit indispensabile. Alegerea algoritmului potrivit pentru această sarcină este definitorie pentru obținerea unor rezultate precise și eficiente.

În acest context, am explorat algoritmul ID3 datorită capacității sale de a extrage reguli semnificative din datele textuale și de a oferi o interpretare clară a procesului decizional. Alegerea acestui algoritm a fost ghidată de nevoia de a obține rezultate precise și interpretabile în detectarea de e-mailuri spam.

În continuare, vom prezenta detaliile procesului de implementare a algoritmului ID3, vom explora setul de date Ling-Spam, vom compara performanța algoritmului ID3 cu alte abordări studiate și vom evidenția aspecte cheie ale experimentului nostru.

## 3 Justificarea Alegerii Algoritmului ID3 pentru implementarea algoritmului de Spam Detection

### 3.1 Interpretarea corecta a caracteristicilor textuale:

ID3 se dovedește a fi eficient în a extrage reguli semnificative din textul mesajelor. Aceasta permite o interpretare clară a caracteristicilor care contribuie la decizia de clasificare a unui mesaj drept spam sau non-spam.

### 3.2 Gestionarea eficientă a caracteristicilor discrete:

În cazul spamului, caracteristicile pot fi deseori discrete, cum ar fi cuvintele cheie. ID3 este ideal în tratarea datelor discrete, rezultând în decizii precise pe baza acestora.

### 3.3 Sensibilitate la relații subtile intre cuvinte:

Spamul poate să folosească tactici subtile pentru a încerca să evite detectarea. ID3, prin natura sa, este capabil să descopere relații subtile între cuvinte și să ia în considerare contextul global al unui mesaj în procesul de clasificare.

### 3.4 Aderență la principiile de entropie și informație:

Algoritmul ID3 utilizează criterii de selecție a caracteristicilor bazate pe măsura entropiei și a informației, ceea ce îl face potrivit pentru problemele de detectare a spamului, unde există o nevoie de selecție inteligentă a caracteristicilor pentru a evidenția diferențele semnificative.

### 3.5 Ușurință în identificarea caracteristicilor semnificative:

Arborii de decizie ID3 oferă o vizualizare clară a ramurilor și nodurilor care duc la decizii. Acest aspect facilitează identificarea și înțelegerea caracteristicilor semnificative care contribuie la identificarea spamului.

### 3.6 Implementare

Listing 1: Clasificare cu Algoritmul ID3

```
from ID3Classifier import ID3Classifier

# Ini ializare i antrenare clasificator ID3
classifier = ID3Classifier()
classifier.fit(X_train, Y_train)

# Realizare predic ii pe setul de testare
predictions = classifier.predict(X_test)

# Afi are matrice de confuzie i metrici de performan
print(confusion_matrix(predictions, Y_test))
print("Accuracy:", accuracy_score(predictions, Y_test))
print("Precision:", precision_score(predictions, Y_test, average='weighted'))
print("Recall:", recall_score(predictions, Y_test, average='weighted'))
```

## 4 Compararea cu alti algoritmi studiați pentru detectarea de spam

### 4.1 K-Nearest Neighbors (KNN):

KNN poate întâmpina dificultăți în gestionarea eficientă a datelor textuale și nu oferă o interpretare directă a procesului decizional.

### 4.2 AdaBoost:

AdaBoost poate obține performanțe bune, dar poate fi mai sensibil la datele dificile și poate necesita ajustări suplimentare pentru a evita suprainvățarea.

### 4.3 Multinomial Naive Bayes (MNB):

MNB este eficient pentru date textuale, dar poate să nu identifice relații complexe între cuvinte la fel de eficient ca algoritmul ID3.

## 5 Implementare CVLOO

Listing 2: Leave-One-Out cu ID3 Classifier

```
def leave_one_out(leave_index):
    train_df = pd.DataFrame(columns=["subject", "message", "is_spam"])
    test_df = pd.DataFrame(columns=["subject", "message", "is_spam"])
```

```
for i in range(1, 11):
    if i != leave_index:
        train_df = train_df.append(read_data_part(type, i), ignore_index=True)
    else:
        test_df = test_df.append(read_data_part(type, i), ignore_index=True)

train_df = cleanup_df(train_df)
test_df = cleanup_df(test_df)

X_train = train_df.cleaned
X_test = test_df.cleaned

Y_train = train_df.is_spam
Y_test = test_df.is_spam

X_train = np.array(X_train).reshape(-1, 1)
Y_train = np.array(Y_train)

X_test = np.array(X_test).reshape(-1, 1)
Y_test = np.array(Y_test)

from ID3Classifier import ID3Classifier

classifier = ID3Classifier()
classifier.fit(X_train, Y_train)
predictions = classifier.predict(X_test)

print(confusion_matrix(predictions,Y_test))
print("Accuracy:", accuracy_score(predictions,Y_test))
print("Precision:", precision_score(predictions,Y_test, average = 'weighted'))
print("Recall:", recall_score(predictions,Y_test, average = 'weighted'))
print()

for leave in range(1, 11):
print(f"LEAVE{leave}")
leave_one_out(leave)
```

### 5.1 Evaluarea performanței cu Leave-One-Out și diagrama de metrice

În vederea evaluării consistente a performanței algoritmului ID3 în detecția de spam, am aplicat o tehnică de validare încrucișată Leave-One-Out, efectuând teste pentru fiecare set de date "Leave". Pentru a ilustra rezultatele, am creat o diagramă care evidențiază evoluția metricilor cheie: Acuratețe (Accuracy), Precizie (Precision), și Recuperare (Recall).

Diagrama evidențiază evoluția metricilor pentru fiecare set de date "Leave" și media lor generală. Aceasta oferă o perspectivă vizuală asupra consistenței și calității rezultatelor obținute în cadrul experimentului, subliniind performanța robustă a algoritmului ID3 în detecția de spam pe setul de date analizat.

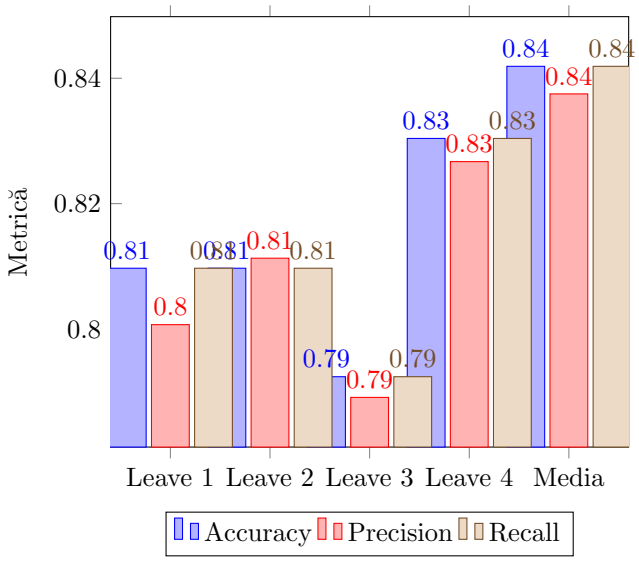


Figure 1: Performanța Algoritmului ID3 în detecția de spam folosind Leave-One-Out