

MINISTERUL EDUCAȚIEI



---

**UNIVERSITATEA TEHNICĂ**

DIN CLUJ-NAPOCA

FACULTATEA DE AUTOMATICĂ ȘI CALCULATOARE

***Predicting Traffic Volume on US Interstate Highways Using Weather and  
Temporal Data***

PROIECT DE SEMESTRU

Student: **Mihaela-Georgiana Nasta**

Disciplina: **Sisteme bazate pe Cunoaștere**

**2025**



## Cuprins

1.Introducere .....	3
1.Context general.....	3
2.Obiective.....	4
3. Specificații.....	4
2.Cunoașterea și analiza setului de date .....	5
3.Pre-procesarea setului de date .....	6
4.Modelarea sistemului .....	7
5.Concluzii .....	9
1.Rezultate obținute .....	9
2.Direcții de dezvoltare.....	12
6.Bibliografie .....	9



# 1.Introducere

## 1.Context general

Într-o lume în care gestionarea eficientă a traficului devine din ce în ce mai importantă, dezvoltarea unor modele predictive pentru volumul traficului poate juca un rol esențial în optimizarea transportului urban și interurban. Creșterea numărului de vehicule și a complexității infrastructurii rutiere impune necesitatea unor soluții inteligente pentru monitorizarea și prognoza traficului.

Această lucrare se axează pe dezvoltarea și antrenarea unor modele de predicție a volumului de trafic, utilizând tehnici moderne de învățare automată. Prin această implementare, urmărim să demonstrăm impactul condițiilor meteorologice, al sărbătorilor și al orelor din zi asupra fluxului de trafic și să oferim un model capabil să facă previziuni precise.

Poate avea multiple aplicații practice, printre care:

- **Previzionarea volumului de trafic:** Dezvoltarea de modele de regresie pentru a prezice volumul de trafic pe baza condițiilor meteorologice, sărbătorilor și orelor din zi, contribuind la gestionarea și planificarea traficului.
- **Optimizarea fluxului de trafic:** Modelul poate fi utilizat pentru optimizarea sistemelor de control al traficului, reducând congestionarea în orele de vârf sau în condiții meteorologice nefavorabile.
- **Analiza impactului vremii:** Studiarea efectelor diferitelor condiții meteorologice (ploaie, zăpadă, acoperire noroasă) asupra volumului de trafic, sprijinind planificarea urbană și îmbunătățirea infrastructurii.
- **Modele de trafic pentru sărbători:** Identificarea tendințelor de trafic în timpul sărbătorilor și evenimentelor speciale, ajutând autoritățile locale să planifice gestionarea creșterilor de trafic.
- **Integrarea în orașe inteligente:** Utilizarea modelului ca parte a inițiativelor pentru orașe inteligente, integrând predicția traficului cu sistemele inteligente de transport pentru



îmbunătățirea mobilității urbane. Utilizarea modelelor de detecție a obiectelor în timp real poate îmbunătăți semnificativ modul în care analizăm și gestionăm traficul rutier.

## 2. Obiective

Prin această lucrare ne propunem să:

- Dezvoltăm și antrenăm modele de predicție a volumului de trafic.
- Analizăm influența condițiilor meteorologice și a sărbătorilor asupra traficului.
- Comparăm performanțele diferitelor modele de regresie și de analiză a seriilor temporale.
- Oferim o metodologie clară pentru dezvoltarea și evaluarea sistemelor de prognoză a traficului.

## 3. Specificații

Pentru a realiza acest proiect, am definit următoarele cerințe:

- **Funcționalitate:** Modelul trebuie să fie capabil să prevadă volumul de trafic pe baza variabilelor meteorologice și temporale.
- **Precizie:** Modelele trebuie să aibă o acuratețe ridicată în predicțiile lor, utilizând metrici standard precum MAE, MSE și R2 Score.
- **Performanță:** Sistemul trebuie să poată face predicții în timp real sau aproape de timpul real.
- **Fiabilitate:** Datele trebuie să fie preprocesate corect pentru a elimina erorile și valorile lipsă.
- **Scalabilitate:** Modelul trebuie să poată fi extins și adaptat pentru alte regiuni și condiții de trafic.



## 2.Cunoașterea și analiza setului de date

Pentru implementarea acestui proiect, am utilizat următorul mediu de lucru:

- **Limbaj de programare:** Python 3.13.0
- **Biblioteci utilizate:** OpenCV, PyTorch, Matplotlib, Seaborn, Pandas, Scikit-learn
- **Framework-uri utilizate:** YOLOv5 (Ultralytics), Faster R-CNN (Torchvision)

Setul de date utilizat în acest proiect conține date despre volumul orar al traficului de pe autostrada I-94, colectate de Minnesota Department of Transportation între 2012 și 2018. Caracteristicile principale sunt:

- **Holiday:** Indică dacă datele au fost colectate în timpul unei sărbători.
- **Temperatura (temp):** Temperatura medie în Kelvin.
- **Ploaie (rain\_1h):** Cantitatea de precipitații înregistrată în ultima oră (mm).
- **Ninsoare (snow\_1h):** Cantitatea de zăpadă înregistrată în ultima oră (mm).
- **Acoperire nori (clouds\_all):** Procentajul de acoperire cu nori.
- **Condiții meteo:** Descrierea generală și detaliată a vremii.
- **Ora colectării:** Ora la care s-au înregistrat datele.
- **Volumul de trafic:** Numărul de vehicule pe oră.



Fig 2.2: Distribuția procentuală a obiectelor

### 3.Pre-procesarea setului de date

Pre-procesarea este un pas esențial în sarcinile de detecție a obiectelor, asigurând că setul de date este curat, bine structurat și potrivit pentru antrenarea modelelor de învățare profundă. Această secțiune detaliază tehnicile de pre-procesare aplicate setului de date utilizat pentru predicția volumului de trafic, evidențiind metode precum reducerea zgomotului, gestionarea valorilor lipsă, eliminarea valorilor aberante și echilibrarea setului de date.

- **Conversia date\_time și extragerea orei.**

```
df['date_time'] = pd.to_datetime(df['date_time'], errors='coerce')
```

```
df['hour'] = df['date_time'].dt.hour df.drop(columns=['date_time'], inplace=True)
```

- **Eliminarea coloanelor neesențiale și tratarea valorilor lipsă.**

```
df = df.dropna()
```

- **Aplicarea one-hot encoding pentru variabilele categoricale (weather\_main, weather\_description, holiday).**

```
categorical_features = ['weather_main', 'weather_description', 'holiday']
```

```
ohe = OneHotEncoder(handle_unknown='ignore')
```

- **Standardizarea caracteristicilor numerice (temp, rain\_1h, snow\_1h, clouds\_all, hour).**

```
numeric_features = ['temp', 'rain_1h', 'snow_1h', 'clouds_all', 'hour']
```

```
scaler = StandardScaler()
```



## 4. Modelarea sistemului

Pentru a estima volumul de trafic, am antrenat mai multe modele de regresie:

- **Regresie Liniară:** Este cel mai simplu model de regresie, care presupune o relație liniară între variabilele independente (factori meteorologici, ora, sărbători) și variabila dependentă (volumul de trafic).
- **Random Forest Regressor:** Este un model bazat pe arbori de decizie, care utilizează multiple arbori pentru a reduce overfitting-ul și pentru a îmbunătăți acuratețea predicțiilor.
- **Gradient Boosting Regressor:** Model bazat pe boosting, care îmbunătățește succesiv erorile modelului anterior pentru a obține predicții mai precise.
- **Model de serie temporală ARIMA:** Este utilizat pentru analiza seriilor temporale și modelarea relațiilor sezoniere ale traficului.

Un pipeline a fost utilizat pentru pre-procesarea datelor și antrenarea modelelor (metodă utilizată în *scikit-learn* pentru a automatiza procesul de preprocesare a datelor și antrenarea modelului). Acest pipeline ne-a permis să procesăm datele și să antrenăm modelele într-un mod eficient și modular, reducând timpul necesar pentru a pregăti setul de date înainte de fiecare antrenare.

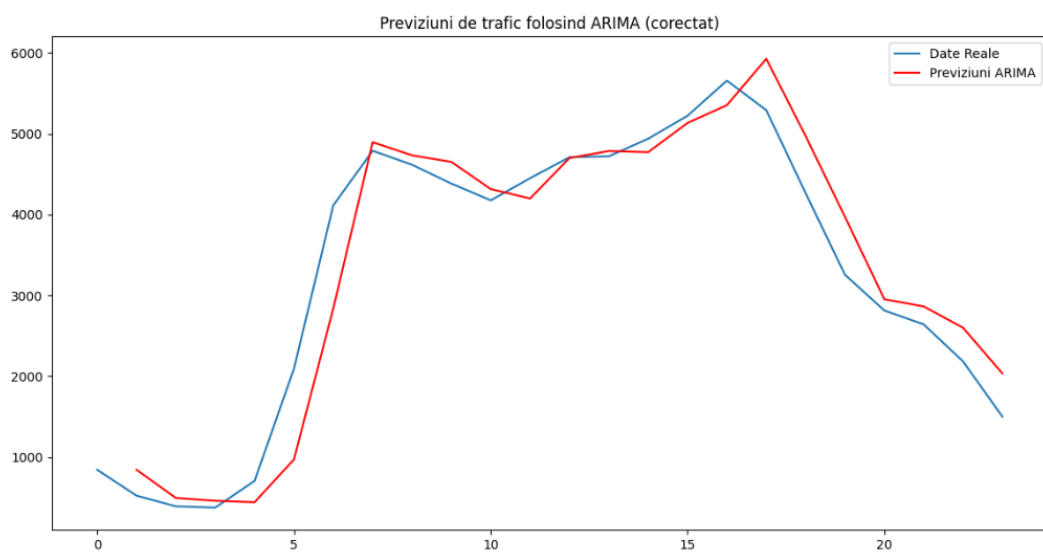




Fig 4.1: ARIMA

Performanțele modelelor sunt măsurate prin următorii indicatori:

- **Eroare Absolută Medie (MAE):** MAE este o metrică de evaluare utilizată pentru a măsura acuratețea unui model de regresie.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

- **Eroare Medie Pătratică (MSE):** MSE este utilizată pentru a calcula media pătratelor diferențelor dintre valorile reale și cele prezise.

$$= \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

- **Rădăcina Erorii Medii Pătratice (RMSE):** RMSE este rădăcina pătrată a MSE și este utilizată pentru a exprima eroarea modelului în aceleași unități ca și datele inițiale.

$$RMSE = \sqrt{\frac{RSS}{n}} = \sqrt{\frac{\sum_{i=1}^n (y_i^{actual} - y_i^{predicted})^2}{n}}$$

- **Coefficientul de determinare (R2 Score):** R2 măsoară cât de bine modelul explică variația variabilei dependente pe baza variabilelor independente. [1]

$$Adjusted R^2 = 1 - \left( \frac{(1 - R^2) \cdot (n - 1)}{n - k - 1} \right)$$





## 5. Concluzii

### 1. Rezultate obținute

- Modelele dezvoltate oferă predicții precise ale volumului de trafic.
- Condițiile meteorologice influențează semnificativ fluxul de vehicule.
- Metodele de regresie și modelele de serii temporale sunt complementare în analiza datelor de trafic.

### 2. Direcții de dezvoltare

- Extinderea dataset-ului pentru a include și alte regiuni.
- Antrenarea unui model de rețele neuronale pentru prognoze avansate.
- Integrarea modelului într-un sistem inteligent de gestionare a traficului.

## 6. Bibliografie

[1]	Predictive modeling I, Curs, Sisteme bazate pe cunoaștere.
[2]	Traffic Volume Prediction, <a href="https://www.kaggle.com/code/xreina8/traffic-volume-prediction/notebook">https://www.kaggle.com/code/xreina8/traffic-volume-prediction/notebook</a>

MINISTERUL EDUCAȚIEI



---

**UNIVERSITATEA TEHNICĂ**

DIN CLUJ-NAPOCA

FACULTATEA DE AUTOMATICĂ ȘI CALCULATOARE

**2025**