

# SmartCargo - Optimizarea livrărilor în România

Compania de transport **SmartCargo România** are nevoie de soluții inteligente pentru a îmbunătăți estimările privind durata livrărilor. Trebuie să construiești modele precise care să prezică timpii de livrare între orașele din România.

## Scopul principal

Analizați date istorice privind cursele dintre orașe și înțelegeți cum factori precum distanța, ora din zi, vremea, traficul sau experiența șoferului influențează durata efectivă a livrării. Scopul este să construiți un model de Machine Learning capabil să prezică timpul estimat pentru curse noi.

## Datele de transport

Fiecare rând din fișierele `train_data.csv` și `test_data.csv` reprezintă o livrare între două orașe. Coloanele sunt:

- **ID:** identificator unic al cursei
- **City A:** orașul de plecare
- **City B:** orașul de destinație
- **Distance:** distanța între orașe (km)
- **Time of Day:** momentul zilei exprimat în minute (0-1439)
- **Weather:** vremea în timpul cursei (Clear, Rain, Snow, Fog)
- **Traffic:** nivelul traficului (0.0-1000.0, 1000 = foarte aglomerat)
- **Road Quality:** calitatea drumului (1-1000, 1000 = foarte bun)
- **Driver Experience:** experiența șoferului (1-40 ani)
- **deliver\_time:** timpul real de livrare în minute (prezent doar în `train_data.csv`)

## Misiunea ta

Managerul operațional a stabilit două sarcini esențiale:



## Subtask 1 (20p)

**Situația Bârlad:** un client important a raportat întârzieri frecvente pe cursele în condiții de ceață. Trebuie să determinați câte curse pleacă din **Bârlad** și pe vreme de **Fog** în setul de date de predicție (`test_data.csv`) și să raportați acest număr.

## Subtask 2

**Estimarea timpilor pentru curse necunoscute:** utilizați datele istorice (`train_data.csv`) pentru a antrena modele de predicție și estimați timpii de livrare pentru toate cursele din setul de date de testare (`test_data.csv`).

## Format de ieșire

Fișier `output.csv` cu următoarele coloane:

- **subtaskID:** numărul subtaskului (1 sau 2)
- **datapointID:** corespunde coloanei ID din `test_data.csv`
- **answer:** răspunsul pentru datapoint-ul respectiv și subtask

**Notă:** Pentru subtask-ul 1, afișați o singură linie cu `datapointID = 1`.

## Scor

- Subtask 1: 20 puncte doar dacă răspunsul este corect.
- Subtask 2: evaluare pe baza MAE (Mean Absolute Error) pe setul de date de testare.