**Motivație**

Potrivit Comisiei Europene, un oraș inteligent este un loc în care rețelele și serviciile tradiționale sunt eficientizate prin utilizarea tehnologiei informației și a comunicațiilor în beneficiul locuitorilor săi și ale afacerilor lor [1]. Scopul sau este crearea unei zone de dezvoltare economică durabilă și o calitate ridicată a vieții prin excelența în mai multe domenii-cheie cum ar fi: economie, mobilitate, mediu, oameni, viață și guvernare. Dezvoltarea orașelor inteligente in tot spațiul Uniunii Europene este un dintre prioritățile stringente ale Comisiei Europene.

Conceptul de oraș inteligent a evoluat mult în România în ultimii ani, iar astăzi vedem nenumărate astfel de inițiative în diverse stadii de implementare. Relativ la Bucuresti, traficul extrem de dificil este una dintre prioritățile acute ale locuitorilor săi. Conform cu clasamentul Global Congestion Impact publicat de INRIX [2], București este orașul cu traficul cel mai aglomerat din lume. Clasamentul pe primele 5 poziții in topul pe 2020 prezentat in raport este următorul: București (134 de ore pierdute în trafic), Bogota (133 de ore), New York (100 de ore), Moscova (100 de ore) și Philadelphia (94 de ore).

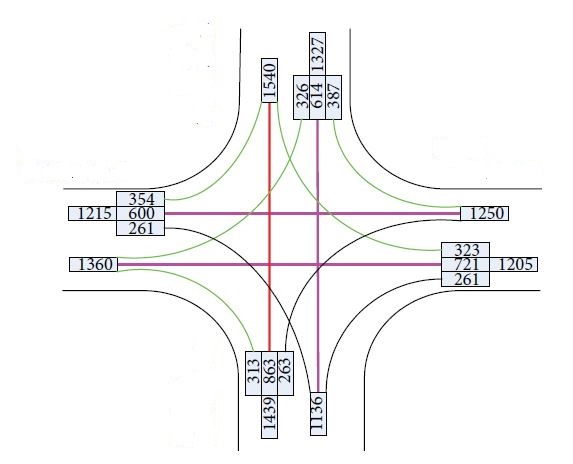
**Descriere generala a setului de date**

Setul de date care va fi utilizat in cercetare este furnizat de Caltrans Performance Measurement System (PeMs) [3], un sistem de monitorizare a traficului operat de Departamentul Transporturilor din California. Acest departament are sediul central din Sacramento și principala sa responsabilitate este managementului străzilor din California.

Caltrans PeMs colectează date de la senzoriii montați în rețeaua de străzi răspândita în tot statul californian. De asemenea, colaborează cu diverse alte agenții care furnizează date suplimentare. Platforma disponibila online stochează datele brute colectate de la senzori și poate fi folosit de cercetătorii interesați pentru o mai bună înțelegere a problemelor de trafic, astfel încât să fie generate idei care ar putea ajuta la reducerea congestiei traficului sau care ar putea îmbunătăți eficienţa traficului în general.

Principalul motiv pentru alegerea setului de date Caltrans PeMs (în comparație cu alte seturi de date) se datorează faptul că este gratuit și datele pot fi în orice moment accesate de către cercetători. *Aici ar putea fi discutat ceva despre experiența cu PMB.*

**Descriere tehnica a setului de date**



**Descrierea generala a studiilor efectuate**

Previziunile precise ale fluxului de trafic sunt necesare pentru programarea și funcționarea unui sisteme inteligente de transport. De exemplu, se pot utiliza previziunile de trafic pentru diferite momente ale zilei (*predicții pe termen lung*) pentru determinarea timpilor cu impact minim pentru închiderea traficului in scopul efectuării lucrărilor de mentenanţă. Un alt exemplu este utilizarea sistemelor care stabilesc limite de viteză variabile in funcție de fluxul de trafic. Astfel de sisteme funcționează în timp real și se bazează pe datele privind fluxul de trafic măsurate de senzorii de trafic. Un sistem proactiv care poate anticipa condițiile fluxului de trafic în viitorul apropiat (*predicții pe termen scurt*) și regla limitele de viteză sau semafoarele înainte ca debitul să se deterioreze este mai valoros decât un sistem reactiv care se bazează numai pe valorile fluxului trecut.

Următoarele modele de prognoză a fluxului de trafic vor fi studiate.

1. Predicția infrastructurii folosind doar datele culese din trafic.
2. Prognoza pe termen lung a traficului pe un senzor folosind datele istorice de trafic pe intervale zilnice, având ca scop predicțiile săptămânale de trafic.
3. Prognoza pe termen scurt a traficului pe un senzor folosind date de trafic temporale culese în timp real cu 60, 45, 30, 15 și 5 minute în avans.
4. Prognoza pe termen scurt a traficului pe doi sau mai mulți senzori folosind date de trafic temporale și spațiale culese în timp real cu 60, 45, 30, 15 și 5 minute în avans. In acest scop vom folosi la antrenare și datele obținute pe senzorii din vecinătatea senzorului ținta pentru predicție.

**Descrierea tehnica a studiilor efectuate**

Modelele combinatoriale au fost mai puțin utilizate în studiile anterioare pentru prognozarea fluxurilor de trafic. Acestea vor fi folosite pentru cercetare aici. Arborii de decizie sunt printre cele mai populare metode folosite atât pentru clasificare, cât și pentru regresie. Modelul intuitiv al unui arbore de decizie permite interpretarea ușoară și robustă a rezultatelor. Modelul forestier aleator este un instrument puternic în învățarea automată și a fost folosit într-o mare varietate de aplicații. Este un model care combină mai mulți arbori de decizie într-un proces de învățare în ansamblu. Ideea principală din spatele metodelor de învățare în ansamblu este de a construi un model mai bun de predicție prin combinarea unui grup de modele care oferă o predicție mai slabă. Un alt model combinatorial care dă rezultate foarte bune în multe aplicații este modelul GB (Gradient Boosting), care combină de asemenea mai mulți arbori de decizie într-un proces de învățare lentă. Merita menționat și faptul că, atât modelul forestier aleatoriu, cât și modelul GB sunt robuste la zgomote și valori aberante (care sunt des întâlnite în datele despre fluxul de trafic obținute de la senzorii de trafic).

Cel puțin următorii algoritmi de învățare automată vor fi utilizați pentru experimente cu

setul de date furnizat de Caltrans Performance Measurement System [1]:

- arbori de decizie;

- modelul forestier aleator;

- modelul GB (Gradient Boosting).

Este bine cunoscut (Schapire et al. [8]) faptul că metodele de votare sunt utile pentru îmbunătățirea performanțelor algoritmilor de învățare automată. Ca exemplu, regula de vot Borda [9] este utilizată recent cu succes de Zhou și Qiu [10] pentru îmbunătățirea predicțiilor algoritmului forestier aleatoriu. Prin urmare, intenționăm să încercăm experimental ca să îmbunătățim performanța anumitor algoritmi de învățare automată folosind idei din teoria voturilor, care sunt descrise de exemplu în Bruns et al. [11] sau în Bruns și Ichim [12].

**Bibliografie**

[1] European Commission. Disponibil la <https://ec.europa.eu/>.

[2] INRIX, 2020 Global Traffic Scorecard. Disponibil la <https://inrix.com/scorecard/>.

[3] Caltrans Performance Measurement System. Disponibil la [https://pems.dot.ca.gov/.](https://pems.dot.ca.gov/)

[8] R. Schapire, Y. Freund, P. Bartlett, W. Lee, Boosting the margin: a new explanation for the effectiveness of voting methods. The Annals of Statistics 26 (1998), 1651 – 1686.

[9] J. de Borda, Memoire sur les elections au scrutiny, Histoire de'Academie Royale Des Sci. 102 (1781), 657 – 665.

[10] Y. Zhou, G. Qiu, Random Forest for Label Ranking. Expert Systems with Applications 112 (2018), 99 – 109.

[11] W. Bruns, B. Ichim, C. Söger, Computations of volumes and Ehrhart series in four candidates elections, Annals of Operations Research 280 (2019), 241 – 265.

[12] W. Bruns, B. Ichim, Polytope volume by descent in the face lattice and applications in social choice, Mathematical Programming Computation (2020), available online at [https://link.springer.com/](https://link.springer.com/article/10.1007/s12532-020-00198-z).