**ANALIZA ȘI PREDICȚIA GRAVITĂȚII** **ACCIDENTELOR RUTIERE**

¹Universitatea Tehnică a Republicii Moldova, Facultatea Calculatoare Informatică și Microelectronică, Departamentul Informatică și Ingineria Sistemelor, Informatica Aplicată

# ABSTRACT

Accidentele rutiere reprezintă o problemă majoră la nivel mondial, având un impact semnificativ asupra sănătății publice și economiei. Înțelegerea cauzelor acestor accidente este esențială pentru dezvoltarea de strategii eficiente de prevenire. În acest studiu, am analizat un set de date care cuprinde diverse aspecte ale accidentelor rutiere, inclusiv momentul zilei, ziua săptămânii, vârsta și sexul șoferului, tipul de vehicul, condițiile de drum și de vreme, tipul de coliziune, numărul de vehicule implicate și multe altele. Acest studiu a identificat perspective și cei mai semnificativi factori care contribuie la severitatea accidentelor rutiere.

Analiza noastră a identificat mai multe factori cheie care contribuie la accidentele rutiere. De exemplu, am observat că majoritatea accidentelor au avut loc în timpul zilei, pe drumuri asfaltate uscate și în condiții meteorologice normale. În plus, majoritatea accidentelor au implicat bărbați cu vârste cuprinse între 18 și 30 de ani.

De asemenea, am constatat că majoritatea accidentelor au rezultat în răniri ușoare, dar au existat și cazuri de răniri grave. Aceste constatări subliniază complexitatea problemei accidentelor rutiere și necesitatea de a aborda această problemă printr-o abordare multidimensională.

Acest set de date este colectat de la departamentele de poliție sub-orașului Addis Abeba pentru munca de cercetare de master. Setul de date a fost pregătit din înregistrările manuale ale accidentelor rutiere din anul 2017-20. Toate informațiile sensibile au fost excluse în timpul codificării datelor și în cele din urmă au 32 de caracteristici și 12316 cazuri de accident. Apoi este preprocesat și pentru identificarea cauzelor majore ale accidentului prin analizarea acestuia folosind diferite metode de regresie precum

Caracteristica țintă este Accident\_severity, care este o variabilă cu mai multe clase. Sarcina este de a clasifica această variabilă pe baza celorlalte 31 de caracteristici pas cu pas, parcurgând sarcina fiecărei zile.

În concluzie, studiul nostru oferă o înțelegere detaliată a factorilor care contribuie la accidentele rutiere. Aceste informații ar putea fi utilizate pentru a dezvolta strategii de prevenire mai eficiente și pentru a reduce impactul accidentelor rutiere asupra sănătății publice și economiei. În continuare, intenționăm să aprofundăm aceste constatări prin analiza detaliată a datelor și prin explorarea altor factori potențiali care ar putea contribui la accidentele rutiere.

# INTRODUCERE

Siguranța rutieră și reducerea accidentelor rutiere au reprezentat o preocupare majoră a planificatorilor de transport, inginerilor și factorilor de decizie. Accidentele rutiere reprezintă o povară grea asupra economiei mondiale din cauza pierderilor semnificative de vieți omenești și a pagubelor materiale. Numărul accidentelor de circulație în Iordania este în creștere, iar consecințele au un impact semnificativ asupra vieții oamenilor și asupra bunăstării lor sociale și emoționale. Este o problemă de comportament complexă datorită multiplicării variabilelor și factorilor care o afectează, cum ar fi comportamentul șoferului, pietonilor, pasagerilor, vehiculului, autostrăzii și caracteristicile infrastructurii. Pentru a dezvolta contramăsuri eficiente, ar trebui efectuată o analiză spațială și temporală folosind R în țările care au asistat la o scădere a numărului de răniți și decese pentru a identifica și a distinge accidentele și apoi a le clasifica în funcție de la momentul apariției, tipul, locația, gravitatea și cauzele acestora.

Înțelegerea factorilor sub care oamenii sunt răniți în principal în accidente de mașină ar putea îmbunătăți standardele generale de siguranță. Variabilele care influențează nivelul de rănire prelungită a pasagerilor într-un accident de vehicul încorporează atribute ale individului, elemente naturale, condițiile de pe autostradă la momentul incidentului și calitățile intrinseci ale vehiculului însuși. În această cercetare, a fost utilizată o colectare considerabilă de informații despre accidente, alături de tehnici de învățare, pentru a dezvălui conexiunile dintre numărul de accidente în comparație cu diferitele niveluri de severitate a rănilor și factorii legați de accident.

Caracteristica țintă este Accident\_severity, care este o variabilă cu mai multe clase. Sarcina este de a clasifica această variabilă pe baza celorlalte 31 de caracteristici pas cu pas, parcurgând sarcina fiecărei zile.

**DATE ȘI STATISTICI SUMARE**

**Setul de date**

Studiul nostru s-a bazat pe setul de date "Road Accidents Severity Dataset", disponibil pe platforma Kaggle[1]. Acest set de date cuprinde informații variate despre diferite situații de accident și circumstanțele accidentelor, precum și gravitatea acestora dupa anumiți factori cum ar fi sexul conducătorului, vârsta, timpul zilei, calitatea drumului etc.

***Preprocesarea datelor***

Pentru preprocesarea datelor am studiat cu ajutorul summary() structura datasetului și am filtrat datele nule și cele care nu sunt necesare cum ar fi timpul accidentului.

***Analiza exploratorie***

Analiza exploratorie a fost efectuată pentru a evidenția relațiile dintre variabile și pentru a identifica tendințe semnificative. Cu ajutorul pachetelor ggplot2 și tidyverse, am creat vizualizări informative, cum ar fi bar-plot-uri și pie ploturi care ilustrează coeficienții repartizarea accidentelor pe sex, vârstă și servicii.

***Modelarea liniară***

Setul de date a fost împărțit într-un set de antrenare și unul de testare folosind funcția sample.split. Apoi, am construit modelul liniar cu ajutorul funcției lm.

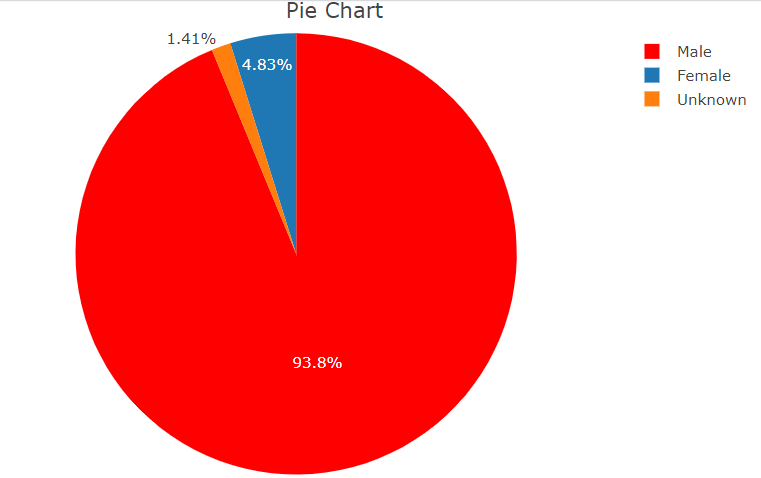
***Evaluarea performanței modelelor***

La sfârșit am ilustrat în formă de diagrame rezultatul comparării rezultatului corect cu cel prezis de model împreună cu loguri pentru rezultatele numerice

**REZULTATE**

Figura 1 reprezintă un Pie plot care arată repartiția accidentelor pe sex, astfel se observă că cel mai mare procentaj de accidente rutiere s-au produs cu participarea sexului masculin – 93.8%,

apoi urmează sexul Femenin cu 4.83%, iar la urmă 1.41% rămase nu este cunoscut sexul.

****

**Figura 1.** Repatizarea pe sex al accidentelor

Figura 2 reprezintă un grafic vertical care afișează datele despre victimele accidentelor în funcție de tipul suprafeței drumului și de gravitatea accidentului.

Axul x este etichetat cu “Tipul\_suprafeței\_drumului” cu categorii: Drumuri de asfalt, Drumuri de pământ, Drumuri de asfalt cu unele deteriorări, Drumuri de pietriș, și Altele.

Axul y este etichetat cu “Numărul\_de\_victime” variind de la 0 la 8.

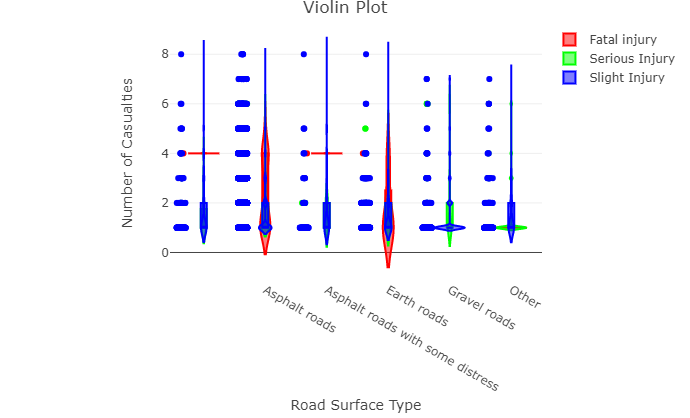
Există trei culori care reprezintă gravitatea accidentului: roșu pentru rănire ușoară; verde pentru rănire gravă; galben pentru rănire fatală, așa cum este indicat în legenda intitulată “Gravitatea\_accidentului”.

Fiecare tip de suprafață a drumului are bare care indică numărul de victime pe nivelul de gravitate a accidentului.

Drumurile de asfalt au numere mai mari în toate categoriile, în special în rănirile ușoare.

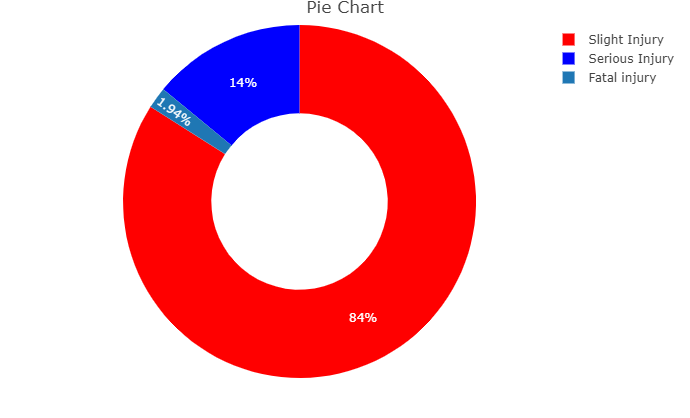
Drumurile de pământ au mai puține cazuri, dar un număr notabil de răniri grave.

Drumurile de pietriș arată, de asemenea, un număr semnificativ de răniri ușoare.



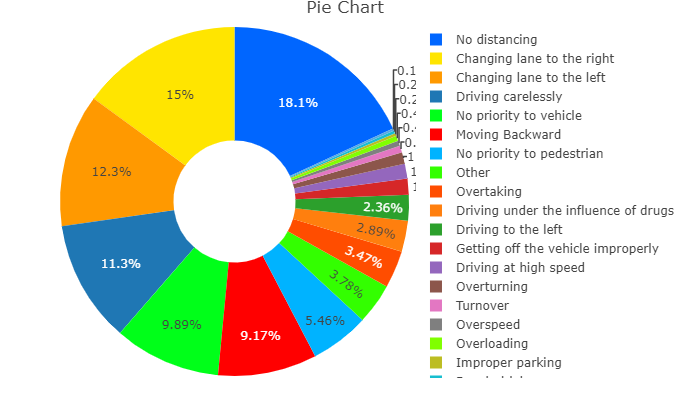
**Figura 2.** Impactul tipului de suprafață a drumului asupra gravității și numărului de victime în accidente

În figura 3, se observă că în accidentele rutiere cel mai des sunt accidente slabe fără sau cu traume slabe. Acest lucru se poate explica prin existența a multor accidente în orașe la viteze mici sau în ambuteiaje, când accidentul cauzează foarte puține daune. Urmează accidentele serioase, după care persoanelor li se acordă ajutor medical – 14%. În cel mai rău caz sunt decesele cu o cotă de aproape 2%.

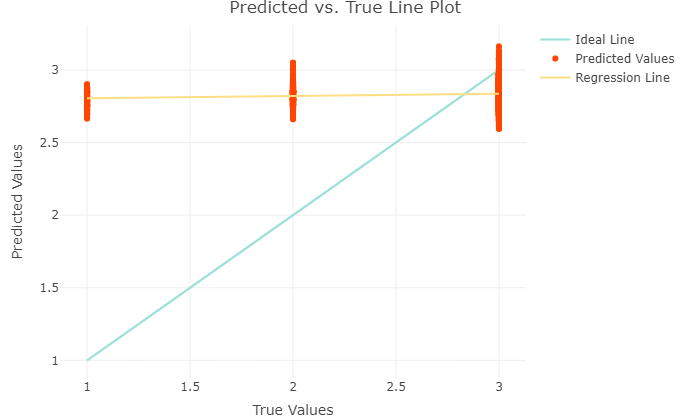


**Figura 3.** Repartizarea accidentelor pe nivel de traumatisme

În figura 4 observăm repartizarea accidentelor în dependență de cauzele provocarii acestora. Cel mai mare procent are cauza de nerespectare a distanței de mers – 18.1%, urmată de schimbarea benzii spre dreapta – 15% și pe stânga – 12.3%. Un mare procent la fel îi aparține condusului fără grijă – 11.3%. Neașteptarea priorității are 9.89 %, iar mersul cu spatele d- 9.17%.



**Figura 4.** Repartizarea accidentelor pe cauze



**Figura 5.** Rezultatul modelului

**DISCUȚII**

Accidentele rutiere au devenit o problemă majoră în timpurile noastre, având un impact semnificativ asupra sănătății publice și a economiei. Înțelegerea factorilor care contribuie la accidente poate ajuta la dezvoltarea de strategii eficiente de prevenire.

În acest studiu, am analizat un set de date care include diverse caracteristici legate de accidente rutiere, cum ar fi ora zilei, ziua săptămânii, banda de vârstă a șoferului, sexul șoferului, nivelul de educație, relația dintre șofer și vehicul, experiența de conducere, tipul de vehicul, proprietarul vehiculului, anul de serviciu al vehiculului, defectul vehiculului, zona în care a avut loc accidentul, benzile sau medianele, aliniamentul drumului, tipurile de intersecții, tipul suprafeței drumului, condițiile suprafeței drumului, condițiile de lumină, condițiile meteorologice, tipul de coliziune, numărul de vehicule implicate, numărul de victime, mișcarea vehiculului, clasa victimei, sexul victimei, banda de vârstă a victimei, severitatea victimei, munca victimei, starea de sănătate a victimei, mișcarea pietonului, cauza accidentului și severitatea accidentului.

Conform studiului nostru, cei cinci predictorii principali pentru severitatea accidentelor sunt ora zilei, ziua săptămânii, banda de vârstă a șoferului, sexul șoferului și nivelul de educație. Aceste rezultate pot fi utilizate pentru a dezvolta strategii de prevenire a accidentelor mai eficiente și pentru a îmbunătăți siguranța rutieră. Cu toate acestea, este important de menționat că valabilitatea acestor modele pentru diferite locații, populații cu comportamente diferite de conducere, stiluri de viață, rase și patrimoniu genetic rămâne încă necunoscută. În plus, performanța și valabilitatea acestor predicții pot varia considerabil în funcție de contextul specific. Prin urmare, este esențială continuarea cercetării în acest domeniu.

**References**

Setul de date [Sursă electronică <https://www.kaggle.com/datasets/kanuriviveknag/road-accidents-severity-dataset?select=RTA+Dataset.csv> ]

RDocumentation, resursă web pentru documentație privitor la RStudio , ‘RDocumentation’ [Accesat:04.12.23], disponibil: <https://www.rdocumentation.org/>