***סיווג קטלניות תאונת דרכים כתלות במאפייני האירוע***

**1 . תיאור מקור מאגר המידע וסביבת העבודה**

תיאור מקור מאגר המידע:

* מאגר המידע שאנו משתמשים נלקח מאתר  kaggle אתר זה מכיל מידע רב על התחומים שונים. <https://www.kaggle.com/datasets>
* מקור הנתונים מתאר את תאונות הדרכים בבריטניה בין השנים 2005-2014
* מאגר נתונים זה מכיל 32 משתנים שקשורים למאפייני התאונות
* הגדרה של תאונת דרכים: תאונת דרכים היא תאונה שבה מעורב כלי רכב אחד או יותר שנועד לנסיעה על דרך סלולה או על דרך עפר.

סביבת עבודה:

* את הנתונים ניתחנו ועיבדנו באמצעות Spark.
* ביצענו למידת מכונה מבוססת איטרציות.
* נבנה מודל בעזרת logistic regression , random forest , Multi Layer Perceptron.
* את בניית המודל ובדיקתו נבצע באמצעות הספרייה MLlib של Spark.
* את האלגוריתמים הרצנו על מחשבי המכללה דרך vlab.

**2. היפותזה**

* במחקר זה נציג היפותזה לחיזוי קטלניות תאונת דרכים בהינתן מאפייני התאונה , אנו נתבסס במחקר על נתונים קיימים במאגר המידע שלנו, כגון:
* סוג הכביש
* מזג האוויר
* מהירות מותרת
* תנאי כביש
* תמרור בצומת
* תנאי תאורה

בעולם בו אנו חיים ישנה מודעות גבוהה לבטיחות בדרכים אך יחד עם זאת לפי ארגון הבריאות העולמי, יותר ממיליון בני אדם נהרגים כל שנה בתאונות דרכים בעולם. בישראל, נפגעים מדי שנה יותר מ-30 אלף איש בתאונות דרכים, מהם מאות הרוגים. מקום המדינה נהרגו בתאונות דרכים יותר אנשים מאשר בכל מלחמות ישראל ביחד. אנחנו מאמינים שאם נוכל לחזות את קטלניות התאונות על בסיס מספר מאפיינים קבועים אפשר יהיה להעלות את המודעות לנושא בצורה ממוקדת ולתת דגש לגורמים הללו ובכך להעלות את היעילות בהתמודדות עם תאונות דרכים.

**3 . תיאור Data set**

הdata set שלנו נלקח מאתר kaggle ומכיל בתוכו 32 מאפיינים של תאונות דרכים ומעל ל1M אירועים שונים של תאונות דרכים. בסיס הנתונים נאסף על תאונות דרכים אשר התרחשו בבריטניה משנת 2005 עד 2014.

לאור כך בחרנו לא להתחשב במאפיינים גאוגרפיים שרלוונטיים רק לבריטניה.

המאפיינים שבחרנו למחקר שלנו :

* **Road type**: type of the road in which the accident happened, categorical coding (1=Roundabout, 2=One-way street, 3=Dual carriageway, 6= Single carriageway, 7=Slip road, 12= One-way street/Slip road).
* **Speed limit:** the max speed limit on road, continual decimal variable
* **First road class:** class of the road (1= Motorway, 2=A(M), 3=A ,4=A ,4=B, 5=C, 6= Unclassified).
* **Second road class:** class of second road (0= Not at junction or within 20 meters, 1= Motorway, 2=A (M), 3=A ,4=A ,4=B, 5=C, 6= Unclassified).
* **Junction control:** sign post at the accident location (0= Not at junction or within 20 meters ,1= Authorized person, 2=Auto traffic signal, 3=Stop sign ,4=Give way or uncontrolled, -1=Data missing or out of range).
* **Weather condition:** weather condition at accident location (1= Fine no high winds,2= Raining no high winds, 3=Snowing no high winds ,4=Fine + high winds ,5=Raining + high winds ,6=Snowing + high winds, 7=Fog or mist, 8= Other, -1 = Data missing or out of range)
* **Road surface condition:** Road surface condition at accident location (1= Dry 2= Wet or damp, 3=Snow ,4= Frost or ice ,5= Flood over 3cm. deep ,6= Oil or diesel, 7= Mud, -1 = Data missing or out of range)
* **Special condition at sight :** categorical coding (0-none , 1= Auto traffic signal – out , 2= Auto signal part defective,3= Road sign or marking defective or obscured, 4= Roadworks , 5= Road surface defective, 6= Oil or diesel ,7=mud ,-1= data missing )
* **Light condition** : (1= Daylight , 4=Darkness - lights lit , 5=Darkness - lights lit ,6=Darkness - no lighting , 7=Darkness - lighting unknown , -1=Data missing or out of range).
* **Junction detail :** (0= Not at junction or within 20 meters ,1= Roundabout, 2=Mini-roundabout traffic signal,3=T or staggered junction ,5=Slip road ,6=Crossroads , 7=More than 4 arms (not roundabout) ,8=Private drive or entrance , 9=Other junction, -1=Data missing or out of range).

2.1 עיבוד בסיס הנתונים

על מנת שבסיס הנתונים יהיה כזה שניתן יהיה לקבל ממנו מודלים ומסקנות סטטיסטיות נדרש לבצע הכנה מקדימה , אנחנו ביצענו את העיבודים הבאים:

* מחיקת שורות שיש בהם מאפיינים ריקים (/NULLמחרוזת ריקה)
* מחיקת שורות שבהם מופיע קידוד -1 מאחר וזה קידוד לנתון ריק
* הפיכת המשתנה התלוי (חומרת התאונה ) לבינארי כלומר תאונה חמורה או קלה
* חלוקת הנתונים לאימוןTraining) ) ולבדיקה (Test) ביחס של 0.8 0.2

**3 . תיאור האלגוריתמים**

* 1. Logistic regression

רגרסיה לוגיסטית היא [מודל סטטיסטי](https://he.wikipedia.org/wiki/%D7%9E%D7%95%D7%93%D7%9C_%D7%A1%D7%98%D7%98%D7%99%D7%A1%D7%98%D7%99) המתאר קשר אפשרי בין [משתנה איכותי/קטגורי](https://he.wikipedia.org/wiki/%D7%9E%D7%A9%D7%AA%D7%A0%D7%94_%D7%90%D7%99%D7%9B%D7%95%D7%AA%D7%99), המכונה "המשתנה המוסבר", ובין משתנים אחרים המכונים "משתנים מסבירים". המשתנים המסבירים יכולים להיות איכותיים או כמותיים. המודל מאפשר לאמוד את מידת ההשפעה של שינוי בערכו כל אחד מהמשתנים המסבירים על ערכו של המשתנה המוסבר.

**פונקציית הסיגמואיד**

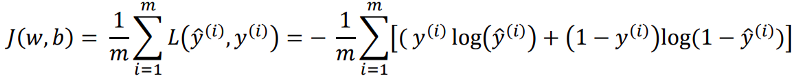
* פונקציה לוגיסטית (פונקציה בעלת עקומה בצורת S), אשר באמצעותה מחושב ערך החיזוי. הפונקציה מניבה פלט בין 0 ל-1 עבור משתנה תלוי שהוא בינארי.
* אם נקבל תוצאה גדולה או שווה ל-0.5 נקבע 1 עבור המשתנה, אחרת 0.
* פונקציית הסיגמואיד מקבלת כקלט וקטור משקלים w (Weight) מוכפל במשתנים הבלתי תלויים x, ואל תוצאת הכפל מחברים את הסטייה b.

**פונקציית ההפסד**

* נרצה למצוא את הערך הטוב ביותר עבור המשקלים באמצעות מזעור פונקציית ההפסד למינימום.
* תחילה, נבחר ערך רנדומלי עבור המשקלים וכדי למדוד את טיב האלגוריתם נחשב את פונקציית ההפסד המוגדרת כך:
* 
* כאשרy ̂ הוא וקטור תוצאת פונקציית הסיגמואיד (החיזוי), y הוא וקטור המשתנים התלויים של סט האימון.

**פונקציית העלות**

* מטרת פונקציית ההפסד היא למזער את השגיאה בין ערך החיזוי לבין הערך האמיתי וכך להגיע לפתרון אופטימלי.
* כדי לקבל תוצאה שימושית יש לחשב את הממוצע של פונקציית ההפסד עבור כל סט האימון.
* פונקציה זו נקראת פונקציית עלות

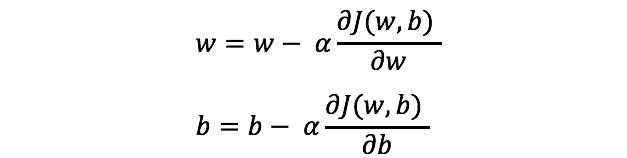


כאשרy ̂ הוא וקטור תוצאת פונקציית הסיגמואיד (החיזוי), y הוא וקטור המשתנים התלויים של סט האימון. m היא כמות הנתונים.

נמצא את הערכים של הפרמטרים w ו-b כדי למזער את פונקציית העלות

**Gradient Descent**

* אלגוריתם אופטימיזציה למציאת מינימום של פונקציה.
* משתמשים בו על מנת למזער את פונקציית העלות.
* במקרה של פונקציית העלות נצטרך להגדיל או להקטין את המשקלים w ואת הסטייה b.



* הנגזרות החלקיות לפי w ו-b מייצגות את ההשפעה של שינוי בw וב-b על פונקציית העלות.
* ע"י מציאת השיפוע והפיכתו לשלילי כיוון ההתקדמות בכל איטרציה יהיה שלילי לכיוון המינימום.
* = learning rate: כמה גדול הצעד בכל איטרציה.
* ככל שהצעד קטן יותר, זמן מציאת המינימום מתארך, וככל שהוא גדול יותר, גדל הסיכוי לפספס את המינימום.

תוצאות ההרצה שלנו:

* 1. Multi Layer Perceptron
* אלגוריתם multi layer perceptron הוא סוג של רשת עצבים מלאכותית, רשת צמתים (נוירונים), המשמש לסיווג בינארי
* האלגוריתם מורכב לפחות משלושה שכבות של צמתים, שכבת הקלט, שכבה נסתרת (שלא ניתנת לצפייה) ושכבת פלט, כל צומת בשכבה אחת מחוברת לצומת בשכבה הבאה עם משקולת
* רשת עצבים מלאכותית מקבלת קלט כווקטור
* פרט לצמתי הקלט, כל הצמתים האחרים (המייצגים פונקציה) ממפים כניסות לפלטים על ידי שילוב לינארי של הכניסות עם משקולות הצומת והbias, ויישום פונקציית הפעלה לא לינארית (סיגמואיד בspark) המגדירה את הפלט בהינתן קלט.
* לצורך אימון, האלגוריתם עושה שימוש בטכניקת למידה הנקראת backpropagation, מודל מתמטי חישובי שפותח בהשראת תהליכים מוחיים או קוגניטיביים.
* הלמידה מתרחשת בנוירון על ידי שינוי המשקולות (בדרך כלל מייצגות את הקישוריות בין נוירונים בתוך הרשת) זאת על סמך כמות השגיאה בפלט לעומת התוצאה הצפויה.
* הטעויות מחלחלות בחזרה ברשת ומתבצעים כוונוני משקלות בעזרת האלגוריתםGradient descent, המנסה להקטינן.
* במידה ותיקוני המשקולות נעשים בשיעור קטן, גדל הסיכוי להתכנסות של רשת העצבים לטעות מינימלית, יתכן שתהיה התכנסות למינימום מקומי, אבל בפועל השיטה נותנת תוצאות טובות וזאת משום שניתן להפעילה מספר פעמים עם משקולות התחלתיות שונות.

תוצאות ההרצה שלנו:

3.3. Random forest

האלגוריתם של יערות אקראיים בונה מקבץ של עצים רבים

כמו כן, האלגוריתם מגריל תצפיות (במקום להשתמש בכל התצפיות הוא משתמש במדגם שלהן), לצורך בנייתו של עץ.

אלגוריתם זה יכול לצמצם את ההשפעות של קורלציה בין משתנים, וכמו כן, הוא נותן הזדמנות למשתנים מסבירים שונים לבוא לידי ביטוי, אפילו אם הם לא בעלי העוצמה החזקה ביותר.

שלבי האלגוריתם:

1. ראשית האלגוריתם ייצור bootstrap dataset, מתוך הטבלה המרכזית שלנו יבחר שורות אקראיות (ניתן לבחור שורות יותר מפעם אחת)

2. על טבלת זאת האלגוריתם יבנה עץ החלטה, בשלב הראשוני בחירת עמודות אקראית, חישוב מדד הgini והחלטה מי יהיה ,rootלאחר מכן בכל שלב יבחרו באופן אקראי עמודות שונות יחושב שוב מדד gini עד בניית עץ ההחלטה .

3. כעת חוזרים שוב ושוב על התהליך ובונים עצים שונים, בצורה כזאת מקבלים מגוון עצים שונים

4. לבסוף נריץ את הנתונים על כל העצים ונקבל תוצאה של כן/לא עבור כל עץ

5. לבסוף התוצר המתקבל הוא ממוצע החיזויים על פני כלל העצים

תוצאות ההרצה שלנו:

**4. מסקנות**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Logistic regression | Random forest | Multi Layer Perceptron |
| מספר תאונות בבניית המודל | 1.2M | 1.2M | 1.2M |
| זמן ריצה(בשניות) | 38.8 | 44.36 | 96 |
| אחוז דיוק | 87.371 | 87.4% | 85.584% |

על בסיס התוצאות שקיבלנו משלושת האלגוריתמים אפשר לראות שהמשתנים הבלתי תלויים שנבחרו (מאפייני התאונה) אכן מאפשרים חיזוי מדויק למדי של חומרת התאונה(המשתנה התלוי).   
בנוסף ניתן לראות ששלושת האלגוריתמים מנבאים בדיוק דומה וגבוה את חומרת התאונה (מה שמעיד על התאמה טובה של האלגוריתמים לבסיס הנתונים).

כמו כן מביצועי האלגוריתמים עולה שזמן הריצה של R.F נמוך בכ 13% מהבא אחריו , אמנם עבור בסיס הנתונים שלנו זה פער די זניח אך עבור בסיסי נתונים גדולים יותר הפער הזה יכול להיות משמעותי מאד.

**5 . תאונות דרכים - ההיבט האנושי.**

תאונת דרכים היא תאונה שבה מעורב כלי רכב אחד או יותר שנועד לנסיעה על דרך סלולה או על דרך עפר.

על פי נתוני ארגון הבריאות העולמי, בכל שנה 1.2 מיליון אנשים נהרגים וכ-20 עד 30 מיליון נפצעים בתאונות דרכים ברחבי העולם. העלות הכלכלית העולמית השנתית של תאונות הדרכים הוערכה בשנת 2003 ב-518 מיליארד דולר בשנה, ועלותן למשק הישראלי הוערכה בשנת 2019 בכ-17 מיליארד שקל בשנה.

תאונות דרכים בישראל

שיעור ההרוגים בתאונות דרכים ב[ישראל](https://he.wikipedia.org/wiki/%D7%99%D7%A9%D7%A8%D7%90%D7%9C) נמצא בירידה מתמדת מאז אמצע שנות ה-70 אז עמד שיעור ההרוגים על כעשרים הרוגים ל-100,000. בשנת 2016 עמד הנתון על 4.3 הרוגים ל-100,000 תושבים. (בשנת 2009 עמד על כ-5.2 הרוגים בשנה ל-100,000 תושבים). בהשוואה בינלאומית, ישראל היא אחת המדינות בעלות שיעור ההרוגים בתאונות דרכים לנפש הנמוך ביותר בעולם. אם משווים את מספר ההרוגים למספר כלי הרכב, המצב בישראל פחות מזהיר - ישראל היא בעלת שיעור דומה של הרוגים למספר כלי רכב לזה של [ארצות הברית](https://he.wikipedia.org/wiki/%D7%90%D7%A8%D7%A6%D7%95%D7%AA_%D7%94%D7%91%D7%A8%D7%99%D7%AA), כ-15 הרוגים ל-100,000 כלי רכב, שיעור הגבוה מרוב מדינות [מערב אירופה](https://he.wikipedia.org/wiki/%D7%9E%D7%A2%D7%A8%D7%91_%D7%90%D7%99%D7%A8%D7%95%D7%A4%D7%94). שיעור ההרוגים לנסועה הוא כ-8.8 הרוגים למיליארד ק"מ.[[7]](https://he.wikipedia.org/wiki/%D7%AA%D7%90%D7%95%D7%A0%D7%AA_%D7%93%D7%A8%D7%9B%D7%99%D7%9D#cite_note-%D7%9E%D7%92%D7%9E%D7%95%D7%AA-7)

הירידה בשיעור ההרוגים לנפש בתאונות דרכים החל מאמצע שנות ה-90 משמעותית יותר מהגידול באוכלוסייה, כך שהמספר המוחלט של הרוגים בתאונות דרכים בישראל ירד משיא של כ-700 הרוגים (כ-550 בתחומי הקו הירוק) ב-1995 למינימום שנרשם בשנת 2009, בה נהרגו 370 בני אדם. עם זאת, חלק מהירידה נובעת מכך שבעקבות הסכמי אוסלו חדלה ישראל לכלול במספר ההרוגים את ההרוגים הלא ישראליים בתחומי [יהודה ושומרון](https://he.wikipedia.org/wiki/%D7%99%D7%94%D7%95%D7%93%D7%94_%D7%95%D7%A9%D7%95%D7%9E%D7%A8%D7%95%D7%9F) ו[חבל עזה](https://he.wikipedia.org/wiki/%D7%97%D7%91%D7%9C_%D7%A2%D7%96%D7%94). באותה תקופה נרשמה גם ירידה ממוצעת של כ-15,000 במספר השנתי של הנפגעים בתאונות דרכים, וירידה של כ-10,000 במספר תאונות הדרכים השנתי בכבישי הארץ.

מאז הקמת מדינת ישראל נהרגו יותר משלושים אלף בני אדם בתאונות דרכים, יותר מבכל מלחמות ישראל ופיגועי הטרור גם יחד.

בשנת 2016 נהרגו בישראל 373 בני אדם בתאונות דרכים, עלייה של 4.7% לעומת השנה הקודמת. ב-2016 תאונות דרכים היו הגורם ל-0.85% ממקרי המוות בישראל, אשר מתפלגים בקירוב כך: כ-1% ממקרי המוות של גברים וכ-0.5% ממקרי המוות של נשים.

סוג כלי הרכב המעורבים בתאונות

[משאיות](https://he.wikipedia.org/wiki/%D7%9E%D7%A9%D7%90%D7%99%D7%AA) מעורבות בכ-17% מהתאונות הקטלניות בעוד הן מהוות רק 2.3% מכלי הרכב בישראל. [אוטובוסים](https://he.wikipedia.org/wiki/%D7%90%D7%95%D7%98%D7%95%D7%91%D7%95%D7%A1) מעורבים בתאונות קטלניות פי 9.6 יותר מאשר כלי רכב פרטיים. הסיבות העיקריות לכך הן כמות הנסיעות המרובה (לעומת כלי רכב פרטיים) וה[תנע](https://he.wikipedia.org/wiki/%D7%AA%D7%A0%D7%A2) הגדול של כלי הרכב הכבדים. כ-40% מתאונות הדרכים מתרחשות ב[ערים](https://he.wikipedia.org/wiki/%D7%A2%D7%99%D7%A8).

מספרם של רוכבי ה[אופנוע](https://he.wikipedia.org/wiki/%D7%90%D7%95%D7%A4%D7%A0%D7%95%D7%A2) שנהרגים בתאונות דרכים גבוה יחסית לאחוז האופנועים בכלל אוכלוסיית כלי הרכב ויחסית לנסועה שלהם. כ-10% מכלל ההרוגים בתאונות הם רוכבי אופנוע, כאשר האופנועים מהווים רק ארבעה אחוזים מכלל כלי הרכב, ואחוז הנסועה שלהם הוא רק אחוז אחד מכלל כלי הרכב. מספר התאונות בישראל ללא יהודה ושומרון בהן היו מעורבים אופנועים בשנים 1990–2008 נע בין 2,300 ל-4,500 בשנה (בקירוב).

סיבות עיקריות לתאונות בישראל

על פי נתוני [משטרת ישראל](https://he.wikipedia.org/wiki/%D7%9E%D7%A9%D7%98%D7%A8%D7%AA_%D7%99%D7%A9%D7%A8%D7%90%D7%9C), הסיבות העיקריות לתאונות הקשות הן: אי-ציות ל[רמזור](https://he.wikipedia.org/wiki/%D7%A8%D7%9E%D7%96%D7%95%D7%A8), סטייה מנתיב (בעיקר ניסיונות [עקיפה](https://he.wikipedia.org/w/index.php?title=%D7%A2%D7%A7%D7%99%D7%A4%D7%94_(%D7%A0%D7%94%D7%99%D7%92%D7%94)&action=edit&redlink=1) על קו לבן רצוף), אי-מתן [זכות קדימה](https://he.wikipedia.org/wiki/%D7%96%D7%9B%D7%95%D7%AA_%D7%A7%D7%93%D7%99%D7%9E%D7%94_(%D7%AA%D7%97%D7%91%D7%95%D7%A8%D7%94)), אי-[שמירת מרחק](https://he.wikipedia.org/wiki/%D7%A9%D7%9E%D7%99%D7%A8%D7%AA_%D7%9E%D7%A8%D7%97%D7%A7), מהירות מופרזת ביחס לתנאי הדרך (להבדיל מ[מהירות שאינה חוקית](https://he.wikipedia.org/w/index.php?title=%D7%94%D7%92%D7%91%D7%9C%D7%AA_%D7%9E%D7%94%D7%99%D7%A8%D7%95%D7%AA_%D7%9E%D7%95%D7%AA%D7%A8%D7%AA&action=edit&redlink=1) ואי-מתן זכות קדימה להולכי רגל. כשליש מההרוגים בתאונות הם הולכי רגל. חלק ניכר מהולכי הרגל שנהרגו, נפגעו בעת שחצו [מעבר חצייה](https://he.wikipedia.org/wiki/%D7%9E%D7%A2%D7%91%D7%A8_%D7%97%D7%A6%D7%99%D7%99%D7%94) מסומן.

על פי מחקרי [הרשות לבטיחות בדרכים](https://he.wikipedia.org/wiki/%D7%94%D7%A8%D7%A9%D7%95%D7%AA_%D7%94%D7%9C%D7%90%D7%95%D7%9E%D7%99%D7%AA_%D7%9C%D7%91%D7%98%D7%99%D7%97%D7%95%D7%AA_%D7%91%D7%93%D7%A8%D7%9B%D7%99%D7%9D), תקופות של מתח ולחץ, לדוגמה [פיגועים](https://he.wikipedia.org/wiki/%D7%98%D7%A8%D7%95%D7%A8), מעלות את מספר תאונות הדרכים.

כסיבה לשיעור הגבוה (לעומת הממוצע במערב אירופה) של תאונות דרכים בישראל ביחס לכמות הרכבים והנסועה, אחד ההסברים הוא שבישראל שיעור הנהגים הצעירים גבוה יחסית, ונהגים אלו בסיכון גבוה יותר להיות מעורבים בתאונות דרכים (במדינות OECD, נהגים צעירים מעורבים בתאונות דרכים בשיעור של קרוב ל-300% משעורם בקרב הנהגים). הסבר נוסף רואה קשר סיבתי במס הגבוה על כלי רכב, שגורם לכך ששיעור כלי הרכב הישנים הנעים בכביש גבוה, ורכבים ישנים נמצאים בסיכון גבוה יותר להיות מעורבים בתאונות דרכים.

הולכי רגל ותאונות דרכים

בעשור האחרון הולכי רגל מהווים באופן עקבי כשליש מההרוגים בתאונות דרכים בישראל. במדינות ה- OECD, אחוז הולכי הרגל מתוך סך ההרוגים בתאונות דרכים הוא כ-20% בממוצע. בשנת 2016 נהרגו 377 איש בתאונות דרכים, מהם 104 הולכי רגל (28%). בשנה זו נפצעו קשה 1,723 איש, 31% מהם הולכי רגל. בשנת 2015 נהרגו 356 בני אדם, כשליש מתוכם (32%, שהם 114 איש) היו הולכי רגל.

מקרב הולכי הרגל שנהרגו או נפצעו קשה בדרכים עירוניות (3,398 איש המהווים 87%):

* 30% נפגעו במעברי חצייה (23% במעבר לא מרומזר ו-7% במעבר מרומזר) שאמורים לשמש, כביכול, מקומות בטיחותיים לחציית הולכי הרגל.
* 34% נפגעו בעת חצייה שלא במעבר חצייה
* % 33 נפגעו לא בשעת חצייה (לדוגמא, נפגעו על הכביש או על המדרכה; אם כי ברוב המקרים מיקום הפגיעה איננו ידוע).

פילוח התאונות החמורות מסוג פגיעה בהולך רגל, לפי סוג הרכב המעורב :

* 60% מהתאונות החמורות מסוג פגיעה בהולך רגל אירעו כתוצאה ממעורבות כלי רכב פרטיים. ב- 9% מהתאונות מסוג זה היו מעורבים אוטובוסים.
* באשר לנפגעים לנסועה , בשנת 2015 אוטובוסים היו מעורבים פי 5 יותר מרכב פרטי בתאונות הולכי רגל חמורות לכל מיליארד קילומטר נסועה.

היפגעות הולכי רגל בני 65+:

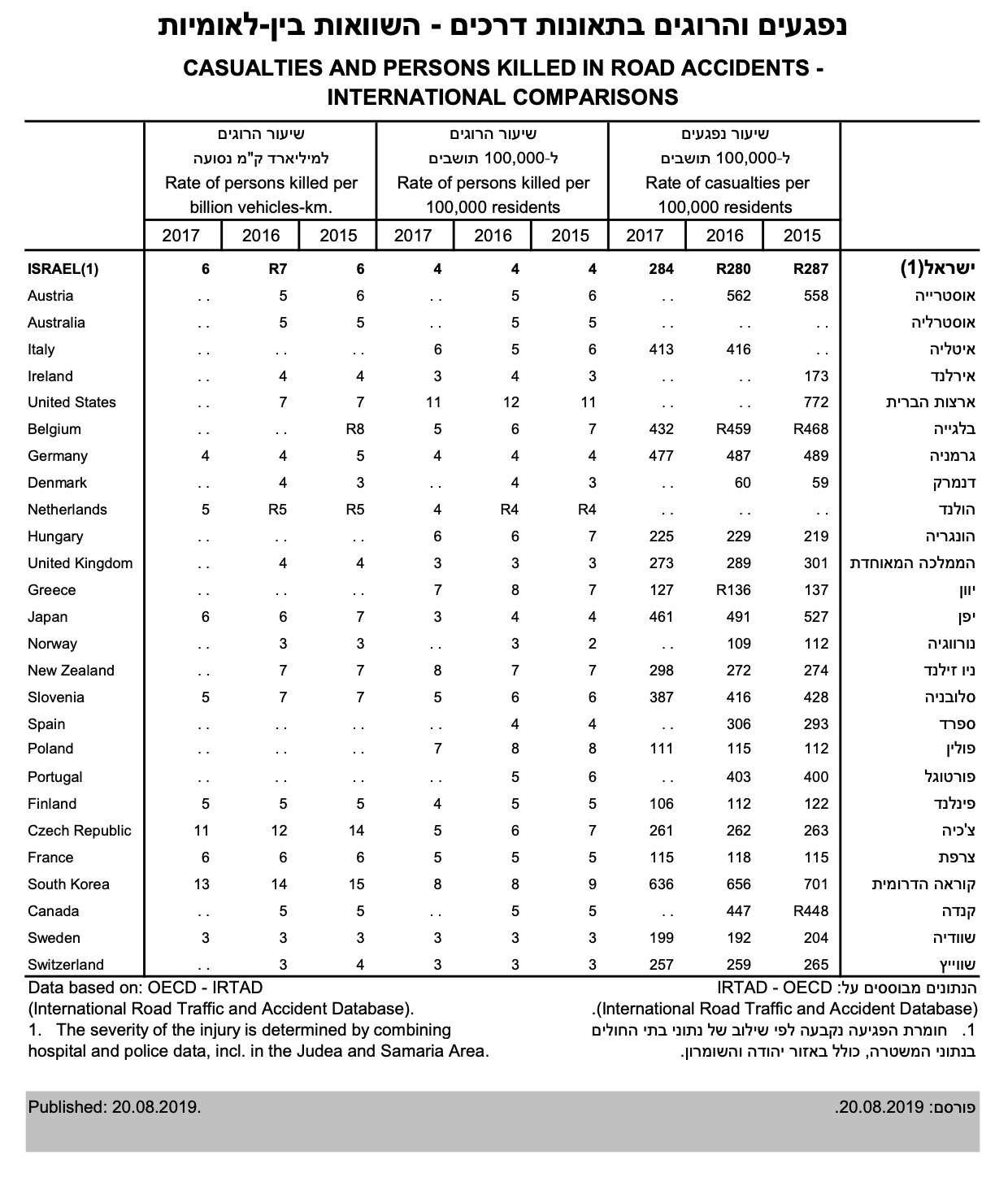
* + הולכי רגל קשישים (בני 65+) הם הקבוצה הפגיעה ביותר מקרב הולכי הרגל. בין השנים 2010 – 2015 נהרגו או נפצעו קשה 1,210 קשישים המהווים 31% מסך הולכי הרגל שנפגעו באותה תקופה, פי 3 משיעורם באוכלוסייה (בני 65+ מהווים כ-10% מאוכלוסיית ישראל).
* כ-60% מהולכי הרגל הקשישים שנפגעו הם בני 75 ומעלה.  95% מהולכי הרגל הקשישים נפגעו בדרכים עירוניות (שלושת רבעי מהם בערים הגדולות והבינוניות – ירושלים, תל-אביב, חיפה, חולון ואחרות).  בבחינת מיקום התאונה, 30% מהקשישים נפגעו במעברי חצייה לא מרומזרים; 30% בעת חצייה שלא במעבר חצייה; ו-30% לא בשעת חצייה (בדומה לכלל הולכי רגל שנפגעו בתאונות דרכים חמורות).

עבירות נהגים ו/או הולכי רגל בתאונות חמורות מסוג פגיעה בהולך רגל (שנת 2015):

עבירות הנהגים ו/או הולכי רגל בתאונות מסוג זה מתפלגות כדלקמן:

* ב-67% מהתאונות החמורות מסוג פגיעה בהולך רגל שאירעו בשנת 2015 נרשמו עבירות לנהגים המעורבים. העבירה השכיחה בתאונות מסוג זה היא אי-ציות נהגים לחובת מתן זכות קדימה (כ- 60% מעבירות הנהגים).
* ב-30% מהתאונות החמורות נרשמו עבירות של הולכי רגל. העבירה שנרשמה בשכיחות גבוהה היא התפרצות הולך הרגל לכביש (33% מעבירות הולכי הרגל).
* 3% לא ידוע (מקרים בהם לא ניתן היה להכריע מהי העבירה שגרמה לתאונה).

נפגעים והרוגים בתאונות דרכים השוואות בינלאומיות :



**6. הפרק הפילוסופי**

כפי שציינו בפרקים הקודמים ובעיקר בפרק ההיבט האנושי בעיית תאונות הדרכים הינה בעיה גדולה ורוחבית בכל המדינות (מפותחות ומתפתחות כאחד).

אין ספק שאם נדע לחזות במדויק כתלות במאפיינים מסוימים את חומרת תאונות הדרכים , המדינה והאזרחים יוכלו לתת דגש על המאפיינים הללו ובכך להוריד בצורה ניכרת את כמות הנפגעים מתאונות אלו.

אחד היתרונות הגדולים בשימוש באלגוריתמי סיווג כפי שעשינו במחקר שלנו הוא שלאחר תהליך לימוד ובחינה של המכונה ניתן להגיד מה דיוק החיזוי שמתקבל על סמך המאפיינים שנבחרו ובכך לזהות האם למאפיינים הללו ישנה השפעה על חומרת התאונה (במידה ולא הייתה המודל על בסיסם לא היה מצליח לחזות את חומרת התאונות ).

כמובן שיתרון נוסף הוא יתרון האוטומטיות והיעילות , בעזרת האלגוריתמים והתשתית שבחרנו הצלחנו בעשרות שניות בודדות לבנות מודל שיאפשר חיזוי של חומרת התאונה כתלות במאפיינים קבועים וזאת עבור מיליון ומאתיים אלף תאונות שונות!

החסרונות של החיזוי בעזרת האלגוריתמים שהשתמשנו בהם והסקת מסקנות מהם זה שישנה חשיבות רבה לבסיס הנתונים הנבחר , למדנו שחשוב להשקיע זמן ומאמץ בהתאמת בסיס הנתונים לעבודה עם האלגוריתמים הנבחרים וכמובן חשוב שבסיס הנתונים יהיה רחב ומייצג ככל הניתן.

המסקנות שלנו מהעבודה הן שניתוח נתוני עתק בשילוב אלגוריתמי סיווג הינם כלים שיכולים לעזור מאד ללמוד על גורמי סכנה בתאונות דרכים ולהציף הזדמנויות חדשות וכיווני הסתכלות חדשניים במאבק זה.

לדעתנו בעתיד ניתן יהיה לשלב טכנולוגיות מידע (טלפון , WAZE, מחשבי הרכב וכו) עם אלגוריתמים מהסוג שעבדנו איתם על מנת לחשב בזמן אמת את מידת הסכנה של הנהג במידה ותתרחש תאונת דרכים (טכנולוגיות המידע יוכלו לספק את הנתונים הנדרשים בגזרת המאפיינים ועל בסיסים האלגוריתמים המדוברים יוכלו לחזות את מידת הנזק של התאונה עוד לפני התרחשותה).

**7. ביבליוגרפיה**

* <https://www.gov.il/BlobFolder/reports/injury_to_pedestrians_in_israel/he/injury_to_pedestrians_in_israel.pdf>
* <https://www.cbs.gov.il/he/publications/DocLib/2019/19.ShnatonTransportandRoadSafety/st19_21.pdf>
* <https://he.wikipedia.org/wiki/%D7%AA%D7%90%D7%95%D7%A0%D7%AA_%D7%93%D7%A8%D7%9B%D7%99%D7%9D>
* <https://www.kaggle.com/>
* Risk and protection factors in fatal accidents ; Emmanuelle Dupont 1a , Heike Martensena , Eleonora Papadimitrioub , and George Yannisb