



The Max Stern
**Yezreel Valley
College**

Business Understanding report

שמות המגישות:

אלמוג יבדייב 305325417

מאי יוסף 318608072

שם המנחה:

יואב זיו



תוכן עניינים

1	שער
2	תוכן עניינים
3	רקע
3	קביעת מטרות עסקיות
4	רקע עסקי
6	יעדים עסקיים וקריטריונים להצלחה
6	מטרות עסקיות
6	קריטריוני הצלחה
7	הערכת מצב
8	מאגר משאבים
8	דרישות, הנחות ומגבלות
9	סיכונים ותוכניות חלופיות
10	טרמינולוגיה
11	עלויות ותועלות
11-12	מטרות מדעי הנתונים וקריטריוני הצלחה
12	תוכנית הפרויקט
14	הערכת כלים ושכניקות



רקע

קביעת מטרות עסקיות

תאונות דרכים הן אחת הסיבות המובילות לפציעות חמורות בעולם, עם השלכות משמעותיות על בריאות הציבור, הכלכלה והחברה. בפרויקט זה אנו מתמקדות בפיתוח מודל מתקדם לחיזוי רמות הפציעה הנגרמות בתאונות דרכים, שמטרתו לתמוך במקבלי החלטות במתן מענה מיידי לאחר תאונה ובשיפור התכנון לטווח הארוך. חיזוי מדויק של רמות הפציעה יסייע לשירותי החירום ולרשויות האכיפה להגיב ביעילות ובדיוק רב יותר, תוך הקצאה חכמה של משאבים והתאמת הפעולות לחומרת האירועים. בנוסף, ניתוח מעמיק של הנתונים יאפשר לזהות דפוסים ארוכי טווח, לספק המלצות לשיפור תשתיות, ולתכנן מדיניות מניעה שתפחית את שיעור התאונות החמורות בעתיד.

הבעיה העסקית:

- כיום, חוסר היכולת לחזות באופן מיידי ומדויק את רמת הפציעה בעקבות תאונה מגביל את יכולתם של שירותי החירום, רשויות האכיפה וחברות הביטוח לקבל החלטות מבוססות. הבעיה משפיעה על:
- שירותי החירום: חוסר היכולת להעריך את רמת הפציעה מגביל את שירותי החירום בהקצאת משאבים וצוותים בהתאם לחומרת האירועים, דבר שעשוי לעכב טיפול במקרים חמורים ולהשפיע על סיכויי ההישרדות של נפגעים.
 - רשויות האכיפה: חוסר מידע מיידי על חומרת התאונה מקשה על ניהול זירת התאונה, כולל סגירת כבישים, ניהול תנועה והכוונת כוחות.
 - רשויות התחבורה: מגבלות בזיהוי אזורים ותנאים שמובילים לתאונות חמורות מונעות תכנון ושדרוג תשתיות בצורה ממוקדת ויעילה.

המטרות העסקיות:

מטרת-על:

פיתוח מודל חיזוי שיספק תחזיות מדויקות על רמת הפציעות בתאונות דרכים. המודל יתבסס על נתונים היסטוריים ויהווה כלי תומך החלטה לשירותי החירום והרשויות התחבורה. בנוסף, המודל יאפשר ניתוח נתונים לזיהוי אזורים ותנאים מסוכנים, ויעניק תובנות לשיפור תשתיות בטיחות ולהפחתת שיעור התאונות החמורות.

Short-Term Goals (מענה מיידי):

1. שיפור המענה של שירותי החירום:
 - תחזיות מדויקות של רמות הפציעה כדי להקצות צוותים וציוד בצורה יעילה.
 - תיעדוף אמבולנסים וטיפול רפואי לתאונות עם פוטנציאל לפציעות חמורות.
2. שיפור ניהול זירת התאונה:
 - מתן מידע בזמן אמת לרשויות האכיפה לצורך סגירת כבישים או ניהול תנועה בצורה מהירה וממוקדת.

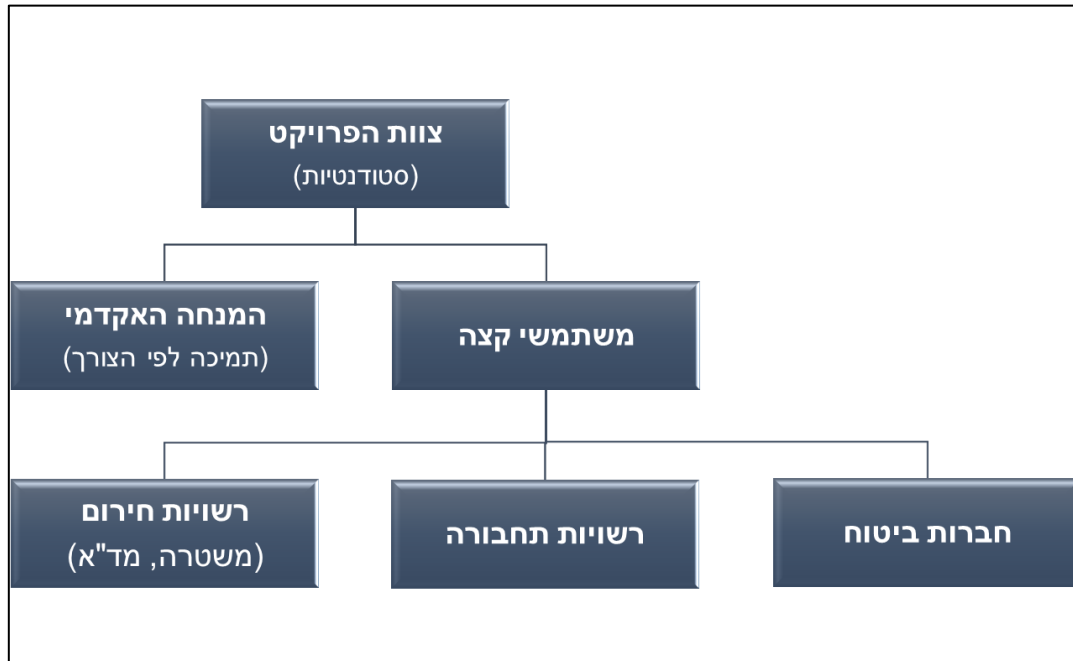
Long-Term Goals (תובנות ותכנון):

1. ניתוח דפוסים לשיפור תשתיות:
 - זיהוי אזורים ותנאים שבהם יש סבירות גבוהה לתאונות חמורות.
 - מתן המלצות לשדרוג תשתיות כמו שיפור תאורה, סימון כבישים והתקנת מעקות בטיחות ועוד.
2. תיעדוף אכיפה באזורים מסוכנים:
 - זיהוי אזורים עם שיעור גבוה של תאונות חמורות לצורך הצבת מצלמות מהירות או הגברת נוכחות משטרתית.

לצורך השגת מטרות אלו, ישנם מספר משאבים זמינים לפרויקט. בראש ובראשונה, עומד לרשות הפרויקט מאגר נתונים היסטורי על תאונות המספקים פרטים גיאוגרפיים, דמוגרפיים ומאפיינים שונים של התאונות. הפרויקט יבוצע על ידי צוות של סטודנטיות למדעי הנתונים, בליווי מנחה אקדמי שיספק תמיכה מקצועית. הוא מתבצע על המחשבים האישיים של הסטודנטיות. הכלים בהם נעשה שימוש הם Python וספריות ייעודיות לניתוח נתונים ולבניית מודלים, לצד כלים להצגת תוצאות חזותיות, כמו Tableau או Power BI.

רקע עסקי

מבנה הארגון:



הסברים לתרשים הארגוני:

1. צוות הפרויקט (סטודנטיות)
 - תפקיד ואחריות: אחראיות על פיתוח המודל, איסוף נתונים ותחזיות.
 - תועלת צפויה: פיתוח מודל חיזוי מדויק שיסייע לשירותי החירום, רשויות תחבורה וחברות ביטוח.
2. המנחה האקדמי (תמיכה לפי הצורך)
 - תפקיד ואחריות: מספק ייעוץ מקצועי, תמיכה טכנית ואקדמית לפי הצורך.
 - תועלת צפויה: עזרה בפתרון בעיות טכנולוגיות ובשיקולים אקדמיים לאורך הדרך והכונה מקצועית.
3. משתמשי קצה
 - תפקיד ואחריות: שימוש במודל לחיזוי לשיפור תהליכים בארגון.
 - תועלת צפויה: שיפור המענה המיידי בתאונות דרכים, שיפור תשתיות וייעול תהליכים עסקיים.
4. רשויות חירום (משטרה, מד"א)
 - תפקיד ואחריות: קבלת תחזיות לגבי חומרת הפציעות והקצאת משאבים.
 - תועלת צפויה: שיפור זמני התגובה והקצאת ציוד וצוותים בצורה אופטימלית.
5. רשויות תחבורה
 - תפקיד ואחריות: משתמשים בתובנות המודל לשיפור תשתיות הבטיחות בדרכים.
 - תועלת צפויה: שיפור בטיחות בדרכים, במיוחד באזורים עם תאונות חמורות.
6. חברות ביטוח
 - תפקיד ואחריות: הערכת סיכונים וניהול תביעות ביטוח.
 - תועלת צפויה: התאמת תמחור ביטוח וניהול תביעות בצורה מדויקת יותר.

זיהוי תחום הבעיה:

תחום הבעיה הוא בטיחות בדרכים, ניהול זירות תאונה ומשאבי חירום, עם דגש על חיזוי רמת הפציעות בתאונות דרכים ושיפור תכנון המשאבים. תאונות דרכים מציבות אתגר משמעותי לבטיחות הציבור, לתגובה למצבי חירום ולתכנון משאבים. היעדר היכולת לחזות את חומרת הפציעות באופן מיידי לאחר תאונה מעכב קבלת החלטות יעילה. בעיה זו משפיעה על שירותי החירום בהקצאת משאבים, על רשויות האכיפה בניהול זירות תאונה, ועל רשויות התחבורה בתכנון שיפורי תשתית.

הבהרת הדרישות המקדימות של הפרויקט:

- מוטיבציה: הצורך במענה מדויק ומהיר לתאונות דרכים נובע מהאתגר שבניהול זירות תאונה, הקצאת משאבים וייעול תשתיות. באמצעות תחזיות מדויקות של חומרת הפציעות, ניתן לשפר את זמני התגובה של שירותי החירום, לתכנן משאבים בצורה מיטבית ולייעל תשתיות, ובכך להפחית את כמות התאונות ולשפר את בטיחות הציבור.



- דרישות מקדימות: מסד נתונים היסטורי, כלים טכנולוגיים (Python, ספריות חיזוי), וצוות שמכיר שיטות של למידת מכונה ומדע הנתונים - צוות הפרויקט מורכב משתי סטודנטיות עם ידע בלמידת מכונה ומדע הנתונים, בליווי מנחה אקדמי.
- שימוש קיים במדע הנתונים: זהו פרויקט אקדמי, כך שאין שימוש קודם במדע הנתונים בארגון לצורך זה. הפרויקט ידגים את הפוטנציאל של מדע הנתונים לשיפור תהליכים. במסגרת הפרויקט, מדעי הנתונים משמשים ככלי מרכזי לפיתוח מודל חיזוי. המודל יתבסס על נתונים היסטוריים ממאגרי Data.gov, כולל נתונים גאוגרפיים, דמוגרפיים וקליניים ומדגים את הפוטנציאל של תחום זה לשיפור תהליכי קבלת החלטות, תכנון משאבים וייעול תשתיות.

הפרויקט הוא חלק ממסגרת אקדמית ואושר על ידי המנחה האקדמי כחלק ממטלת הגמר. מכיוון שמדובר בפרויקט אקדמי ולא יישומי בארגון, אין צורך בקידום פנימי של מדעי הנתונים כטכנולוגיה. עם זאת, הפרויקט מדגים את הערך של מדעי הנתונים ככלי משמעותי לשיפור תהליכים בתחום בטיחות בדרכים וניהול משאבים.

תיאור הפתרון הנוכחי:

כיום, קיימים מספר פתרונות המתמקדים בהפחתת תאונות דרכים ובשיפור הבטיחות בדרכים, אך אין פתרון מקיף לחיזוי מיידי של רמת הפגיעה בעקבות תאונה. להלן מספר פתרונות קיימים, יתרונותיהם וחסרונותיהם:

- **מערכות בטיחות אקטיביות ברכב:** מערכות כגון התרעה מפני התנגשות ובלימה אוטומטית נועדו למנוע תאונות או להפחית את חומרתן. מחקרים הראו כי מערכות אלו מפחיתות את מספר התאונות, אך לא מספקות מידע על רמת הפגיעה לאחר התרחשות התאונה.
 - **יתרונות:** מפחיתות את הסיכון לתאונות ומגבירות את בטיחות הנוסעים.
 - **חסרונות:** אינן מספקות מידע לאחר התרחשות התאונה, כגון הערכת רמת הפגיעה, ולכן לא תומכות בקבלת החלטות מיידיות של שירותי החירום.
- **שימוש בנתוני אפליקציות ניווט:** בארה"ב, משרד התחבורה בחן שימוש בנתוני אפליקציות כמו Waze לפיתוח מודלים לחיזוי סיכוני תחבורה ותאונות דרכים. מודלים אלו מתמקדים בזיהוי נקודות סיכון בכבישים, אך אינם מספקים חיזוי מיידי של רמת הפגיעה לאחר תאונה.
 - **יתרונות:** מאפשרים זיהוי אזורים מסוכנים ותכנון תשתיות משופר.
 - **חסרונות:** אינם מספקים תחזיות בזמן אמת על חומרת הפגיעה בתאונות ספציפיות, ולכן מוגבלים בתמיכתם בשירותי החירום.
- **מערכות למניעת הסחות דעת של נהגים:** חברות כמו סייברוואן פיתחו מערכות המזהות שימוש בטלפון נייד במהלך הנהיגה וחוסמות אפליקציות מסיחות דעת. מערכות אלו מפחיתות את הסיכון לתאונות הנגרמות מהסחת דעת, אך אינן מספקות מידע לאחר התרחשות התאונה.
 - **יתרונות:** מפחיתות את הסיכון לתאונות הנגרמות מהסחות דעת.
 - **חסרונות:** אינן מספקות מידע על רמת הפגיעה לאחר תאונה, ולכן לא תומכות בקבלת החלטות של שירותי החירום.
- **מערכות בטיחות ברכב:** חברות כמו MDGo הישראלית פיתחו מערכות המדווחות על חומרת פגיעות לאחר התרחשות התאונה, אך אינן מספקות תחזיות מיידיות בזמן אמת.

רמת הקבלה בארגונים:

הפתרונות הנוכחיים מתמקדים בעיקר במניעת תאונות ושיפור הבטיחות בדרכים, וזוכים לקבלה רחבה בקרב יצרני רכב, רשויות תחבורה ונהגים. עם זאת, היעדר פתרון המספק חיזוי מיידי של רמת הפגיעה לאחר תאונה מותיר פער בתמיכה בקבלת החלטות של שירותי החירום, רשויות האכיפה וחברות הביטוח.

סיכום:

הפתרונות הקיימים מתמקדים במניעת תאונות ובהפחתת חומרתן, אך אינם מספקים חיזוי מיידי של רמת הפגיעה לאחר התרחשות תאונה. היעדר מידע זה מגביל את יכולתם של שירותי החירום להקצות משאבים בצורה אופטימלית ולספק טיפול מהיר ויעיל לנפגעים. לכן, יש צורך בפיתוח פתרון חדשני שיספק תחזיות בזמן אמת על רמת הפגיעה, וישפר את התגובה המיידית והאסטרטגית של הגורמים המעורבים.



יעדים עסקיים וקריטריונים להצלחה

מטרות עסקיות

מטרת הפרויקט היא לפתח מערכת מבוססת מדעי נתונים לחיזוי רמת הפציעות בתאונות דרכים. המערכת נועדה לתת מענה לשירותי החירום ורשויות האכיפה, על ידי מתן תחזיות מדויקות ומהירות לאחר תאונה, שיסייעו בקבלת החלטות מבוססות ובניהול משאבים יעיל.

תיאור הבעיה לפתרון באמצעות מדעי הנתונים:

הבעיה העסקית בפרויקט היא חוסר היכולת לחזות באופן מיידי ומדויק את רמת הפציעה בעקבות תאונה, דבר שמגביל את היכולת של שירותי החירום ורשויות התחבורה לקבל החלטות מהירות ומבוססות.

שאלות עסקיות:

- כיצד ניתן לחזות את רמת הפציעה הצפויה בתאונה דרכים בהתבסס על נתונים היסטוריים, כדי לאפשר שיפור בתגובה של שירותי החירום?
- איך ניתן להעניק תחזיות מדויקות שיאפשרו לרשויות התחבורה לשפר את תכנון התשתיות באזורים מועדים לתאונות חמורות?
- כיצד ניתן להשתמש בנתוני תאונות היסטוריים כדי לחזות את רמת הפציעה הנגרמת בתאונה?
- איך ניתן לשפר את הקצאת המשאבים של שירותי החירום באמצעות תחזיות מדויקות?
- אילו דפוסים ניתן לזהות בנתוני תאונות כדי לזהות אזורים ותנאים בעלי סיכון גבוה?
- איך ניתן למנף את ניתוח הנתונים לזיהוי שיפורי תשתיות קריטיים במיקומים ספציפיים?

דרישות עסקיות נוספות:

- להבטיח שלא ייגרם עיכוב בשירותי החירום בתגובה לתאונות, ושהמשאבים יוקצו בצורה מדויקת בזמן אמת.
- להימנע מהשלכות של חיזוי שגוי, שיכולות להוביל לניהול לא נכון של המשאבים או למענה לקוי בשטח.

היתרונות הצפויים במונחים עסקיים:

- חיזוי מדויק של רמות הפציעה צפוי לשפר את זמני התגובה של שירותי החירום ב-20%.
- שיפור בתכנון התשתיות ושיפור אכיפת החוקים באזורים בעלי סיכון גבוה יכול להפחית את שיעור התאונות החמורות ב-15%.

קריטריוני הצלחה

קריטריוני ההצלחה לפרויקט מתמקדים במדדים אובייקטיביים וסובייקטיביים, על מנת להבטיח שהמערכת תשיג את מטרותיה בשיפור בטיחות בדרכים, הקצאת משאבים וניהול תגובת חירום. קריטריונים אלו מתואמים עם המטרות העסקיות של הפרויקט ושואפים לספק תוצאות מדירות ומשמעותיות.

קריטריונים אובייקטיביים:

הקריטריונים האובייקטיביים הם קריטריונים כמותיים שניתן למדוד בצורה ברורה, וכל מטרה עסקית תורגמה לקריטריון הצלחה מדיד:

1. שיפור המענה של שירותי החירום:
 - קריטריון הצלחה: היכולת לחזות את רמות הפציעה ב-85% דיוק ולהקצות משאבים בצורה מדויקת, מה שיביא לשיפור של 20% לפחות בזמני התגובה של שירותי החירום.
2. תיעדוף אמבולנסים וטיפול רפואי לתאונות חמורות:
 - קריטריון הצלחה: הציפייה היא שהמודל יוכל להבחין בצורה מדויקת בתאונות חמורות ולהגביר את תיעדוף המשאבים והאמבולנסים עבור תאונות אלו, ולהגביר את זמני התגובה ב-15%.
3. שיפור ניהול זירת התאונה:
 - קריטריון הצלחה: מתן מידע בזמן אמת שיביא לצמצום של 20% לפחות בשעות סגירת הכבישים וייעול תהליך ניהול התנועה.



4. ניתוח דפוסים לשיפור תשתיות:
- קריטריון הצלחה: שיפור היכולת לזהות אזורים עם סיכון גבוה לתאונות חמורות ולתרום לשדרוג תשתיות בטיחותיות באזורים אלו, עם צפייה לירידה של 10%-15% בתאונות חמורות באזורים שזוהו כבעייתיים.
5. תיעדוף אכיפה באזורים מסוכנים:
- קריטריון הצלחה: הצבה של מצלמות מהירות ושיפור הנוכחות המשטרתית באזורים המסוכנים יוביל לירידה של 10% לפחות בתאונות חמורות באזורים אלו.

קריטריונים סובייקטיביים:

קריטריונים אלו מתמקדים במדדים איכותיים שלא ניתן למדוד באופן ישיר, אך הם משפיעים על ההצלחה הכוללת של הפרויקט:

1. שביעות רצון המשתמשים:
 - קריטריון הצלחה: הצלחה תוגדר כאשר גורמים בשירותי החירום ורשויות התחבורה, יביעו שביעות רצון גבוהה מהיכולת של המודל לשפר את תהליך קבלת ההחלטות בשטח, כולל יכולתו לתמוך בהחלטות, בתכנון משאבים ותיעדוף תשתיות.
2. שיפור שיתוף הפעולה:
 - קריטריון הצלחה: רשויות חירום ורשויות התחבורה ידווחו על שיפור בשיתוף הפעולה ביניהן כתוצאה מהשימוש במודל.
3. תרומה למחקר ופיתוח בתחום בטיחות בדרכים:
 - קריטריון הצלחה: המודל יקבל הכרה אקדמית כחדשני ויוצג בכנסים או פרסומים מקצועיים בתחום.

יישור מדדים סובייקטיביים:

כדי להבטיח שהקריטריונים הסובייקטיביים יימדדו באופן אחיד, יש לזהות את הגורמים שיעריכו את הצלחת המודל:

- מפקדי שירותי החירום (בתי חולים ומשטרה): ימדדו את תרומת המודל לשיפור זמני התגובה ותכנון המשאבים
- מנהלי רשויות התחבורה: יבדקו אם התחזיות עזרו לזהות אזורים מסוכנים ולתכנן שיפורי תשתיות.

הערכת המצב

הפרויקט מבוסס על מסד נתונים היסטורי המכיל כ-189,333 רשומות של תאונות דרכים. הנתונים כוללים מגוון משתנים, כגון פרטי התאונה, מידע גיאוגרפי, נתונים דמוגרפיים ומאפייני הסביבה, שמאפשרים לבצע ניתוח מעמיק ולבנות מודל חיזוי מדויק.

צוות הפרויקט כולל שתי סטודנטיות למדעי הנתונים, האחראיות על כל שלבי פיתוח המודל, כולל ניתוח נתונים, עיבוד מקדים ובניית מודל חיזוי. בנוסף, מנחה אקדמי המספק תמיכה מקצועית במקרה של אתגרים או צורך בהכוונה.

גורמי סיכון:

1. חוסר איזון בנתונים: שיעור נמוך יחסית של תאונות חמורות עשוי להוות אתגר בעת בניית המודל.
2. איכות הנתונים: קיימת אפשרות של נתונים חסרים או נתונים שאינם מדויקים.
3. מורכבות המודל: קיימת סבירות שהמודל הראשוני לא יעמוד ברמת הדיוק הנדרשת.
4. מגבלות זמן ומשאבים: תיתכן מגבלה של זמן לביצוע כל שלבי הפרויקט או מגבלות משאבים טכניים.

תוכנית חלופית:

1. שימוש בטכניקות כמו SMOTE או משקלים מותאמים במהלך אימון המודל.
2. ביצוע עיבוד מקדים מדוקדק לניקוי ושיפור איכות הנתונים. הכולל בדיקות עקביות, מילוי נתונים חסרים, או הסרת רשומות לא תקינות.
3. ביצוע ניתוח מקדים לבחירת המודל המתאים ביותר. אם המודל הראשוני לא ישיג את רמת הדיוק הרצויה, נעבור למודלים מורכבים יותר.
4. חלוקת העבודה לשלבים ברורים ותיעדוף משימות קריטיות כדי לעמוד בזמנים.

מאגר משאבים

משאבי חומרה:

כיום, הפרויקט מתבצע באמצעות מחשבים אישיים של חברי צוות הפרויקט (שתי הסטודנטיות), עם יכולות עיבוד בסיסיות. המחשבים מצוידים בכלים הדרושים, כגון Python וספריות מתאימות, ומאפשרים עיבוד הנתונים באופן מקומי. במידה ויידרש כוח עיבוד חזק יותר, קיימת אפשרות להשתמש במשאבי מחשב ענן (כגון Google Colab) המספקים תמיכה בעיבוד נתונים ועומדים בדרישות הטכניות לניתוח נתונים מסיבי ובבניית מודלים מבוססי Machine Learning.

מקורות נתונים ומאגרי ידע:

- מקורות נתונים זמינים: מסד נתונים הכולל כ-189,333 רשומות תאונות דרכים, המכיל נתונים בפורמט CSV על משתנים גיאוגרפיים, דמוגרפיים, סביבתיים ועוד. נתונים אלו זמינים לשימוש בפרויקט ללא צורך בגישה למסדי נתונים חיים.
- פורמט ושמירה: הנתונים נשמרים בפורמט CSV על גבי מחשבים אישיים וניתנים לעיבוד באמצעות ספריות Python.
- שימוש בנתונים חיצוניים: נכון לעכשיו, אין כוונה לרכוש נתונים חיצוניים נוספים.
- נושאי אבטחת מידע: לא זוהו בעיות אבטחת מידע, שכן הנתונים הם ציבוריים ונגישים למטרות אקדמיות בלבד.

משאבי כוח אדם:

המנחה האקדמי מספק תמיכה מקצועית והכוונה בפרויקט במידת הצורך. עם זאת, אין מעורבות של מומחים נוספים מבחינה עסקית או טכנולוגית. בנוסף, אין צורך בתמיכה טכנית מצד מנהלי מסדי נתונים או צוות טכני אחר, שכן הפרויקט מנוהל במלואו על ידי שתי הסטודנטיות, אשר משתמשות בכלים סטנדרטיים לעיבוד נתונים ומבצעות את כל שלבי העבודה בעצמן.

דרישות, הנחות ומגבלות

קביעת דרישות:

הנתונים המשמשים לפרויקט נלקחו מאתר Data.gov האמריקאי, הכוללים מידע מגוון על תאונות דרכים. הנתונים ציבוריים, ואין מגבלות משפטיות או מגבלות אבטחת מידע על השימוש בהם או על תוצאות הפרויקט. הפרויקט מבוצע במסגרת אקדמית בלבד ואינו מיועד לשימוש מסחרי, ולכן אין מגבלות חוקיות על פרסום תוצאותיו במתכונת אקדמית.

הפרויקט מתבצע במסגרת פרויקט גמר אקדמי, ולכן הוא כפוף ללוחות זמנים מוגדרים מראש. כל הגורמים המעורבים, כולל הסטודנטיות והמנחה האקדמי, פועלים בתיאום כדי לעמוד בדרישות לוחות הזמנים ולהשלים את כל שלבי הפרויקט במועד הנקבע.

תוצאות הפרויקט יוצגו באמצעות דוחות, גרפים ומודלים מפורטים, המיועדים לצרכים אקדמיים בלבד. התוצאות לא יפורסמו באינטרנט ולא ישולבו במערכות חיצוניות או בסיסי נתונים, אלא ישמשו להצגה במסגרת הלימודים בלבד.

הבהרת הנחות:

גורמים כלכליים שעשויים להשפיע על הפרויקט:

בפרויקט זה, אין גורמים כלכליים שעשויים להשפיע עליו, כגון עלויות ייעוץ או מוצרים מתחרים, מאחר שמדובר בפרויקט אקדמי שאינו תלוי בתקציבים חיצוניים, שכן הכלים הטכנולוגיים הנדרשים כמו python ומאגר הנתונים מ data.gov הם חינמיים ונגישים.

הנחות לגבי איכות הנתונים:

הנתונים שנלקחו מ Data.gov הם ציבוריים ומבוססים, אך ייתכן שיש בהם אי-דיוקים, ערכים חסרים או נתונים שאינם עקביים. ההנחה היא שהנתונים איכותיים דיים לפיתוח המודל, תוך שימוש בטכניקות לניקוי ועיבוד מקדים של הנתונים כדי לטפל בבעיות כמו ערכים חסרים או חריגות.

ציפיות נותן החסות/צוות הניהול מתוצאות הפרויקט:

המנחה האקדמי והצוות הניהולי מצפים להבין הן את תהליך העבודה והמודל עצמו, כולל ההיגיון מאחורי בחירת האלגוריתמים והמשתנים, והן את התוצאות הסופיות. בנוסף, הם מעוניינים בהצגה ברורה של הממצאים, עם דגש על גרפים וטבלאות ויזואליים, הממחישים את המסקנות ואת הערך שהמודל מספק.



אימות ומגבלות:

הגישה לנתונים בפרויקט אינה מוגבלת ואינה דורשת סיסמאות או הרשאות מיוחדות, שכן הנתונים הם ציבוריים וזמינים דרך הפלטפורמה של Data.gov. מבחינת הגבלות חוקיות, הנתונים נלקחו ממקור ציבורי זה, ולכן לא צפויות מגבלות משפטיות על השימוש בהם או על פרסום תוצאות הפרויקט. מבחינה כלכלית, הפרויקט אינו דורש תקציב נוסף, שכן כל המשאבים הנדרשים: חומרה, תוכנות ונתונים זמינים ללא עלות נוספת. כל הדרישות הכלכליות מכוסות במסגרת האקדמית, ואין מגבלות כלכליות שצפויות לעכב את התקדמות הפרויקט.

סיכונים ותוכניות חלופיות

במהלך הפרויקט עשויים להתעורר סיכונים שיש לקחת בחשבון מראש. להלן פירוט הסיכונים המרכזיים ודרכי ההתמודדות האפשריות עימם:

ליו"ר (Scheduling)

- סיכון: התהליך עשוי לקחת זמן רב יותר מהמתוכנן בשל עיבוד כמויות גדולות של נתונים או התמודדות עם תקלות טכניות.
- תוכנית חלופית: יצירת תכנית עבודה גמישה הכוללת זמני גיבוי למשימות קריטיות ותיעדוף משימות חיוניות בפרויקט.

תקציב (Financial)

- סיכון: ייתכן שיהיה צורך בתוכנות או משאבים נוספים שלא הוערכו מראש.
- תוכנית חלופית: שימוש בתוכנות, כלים ומאגרי נתונים חינוניים (כגון Python ו- Tableau Public) הבטחת זמינות משאבים במסגרת התמיכה האקדמית.

איכות הנתונים (Data Quality)

- סיכון: ייתכן שהנתונים יכילו ערכים חסרים, טעויות או חוסר עקביות.
- תוכנית חלופית: שימוש בטכניקות עיבוד מקדים, כמו ניקוי נתונים, מילוי ערכים חסרים, ובדיקת עקביות הנתונים. במידת הצורך, ייתכן חיפוש נתונים חיצוניים משלימים.

תוצאות (Results)

- סיכון: התוצאות הראשוניות של המודל עשויות לא להיות מדויקות או לא לספק תובנות משמעותיות, דבר שיצריך התאמות נוספות ועבודה חוזרת.
- תוכנית חלופית: שיפור המודל על ידי כוונן היפר-פרמטרים, בחינת מודלים חלופיים מתקדמים יותר, ושילוב מדדים נוספים לשיפור איכות החיזוי.

חוסר איזון בנתונים (Data Imbalance)

- סיכון: שיעור נמוך של תאונות חמורות עלול לפגוע בדיוק המודל.
- תוכנית חלופית: שימוש בטכניקות לאיזון נתונים כמו SMOTE, התאמת משקלים או בחירת מדדים שמדגישים דיוק בקבוצות קטנות.

משאבים טכנולוגיים (Technological Resources)

- סיכון: מגבלות טכנולוגיות עלולות להאט את עיבוד הנתונים או את אימון המודל.
- תוכנית חלופית: ייעול התהליכים בעזרת כלים טכנולוגיים מתקדמים ושימוש בכלים חיצוניים. שיתוף פעולה עם המנחה האקדמי לצורך פתרון בעיות טכנולוגיות.



טרמינולוגיה

כדי להבטיח שכל המעורבים בפרויקט מבינים זה את זה בצורה ברורה ואחידה, נדרש להגדיר מילון מונחים הכולל מושגים טכניים ועסקיים הקשורים לפרויקט. מילון זה יסייע בצמצום בלבול, יקל על תהליך העבודה, ויבטיח שכל הצוות משתמש באותם מונחים באופן עקבי.

מילון מונחים:

מונחים עסקיים:

1. **Injury Severity (חומרת פציעה):** מדד המגדיר את רמת הפציעה שנגרמה בתאונת דרכים. (No Apparent Injury ללא פציעה נראית לעין, Possible Injury פציעה אפשרית, Suspected Minor Injury חשד לפציעה קלה, Suspected Serious Injury חשד לפציעה חמורה, Fatal Injury פציעה קטלנית).
2. **High-Risk Areas (אזורים בסיכון גבוה):** אזורים שבהם קיימת סבירות גבוהה לתאונות חמורות, הדורשים התערבות תשתיתית או הגברת אכיפה.
3. **Public Data (נתונים ציבוריים):** נתונים פתוחים ממקורות כמו Data.gov.
4. **Emergency Services (שירותי חירום):** ארגונים כמו משטרה, אמבולנסים, המשתמשים בתחזיות המודל לשיפור זמן התגובה והקצאת המשאבים.
5. **Success Metrics (מדדי הצלחה):** קריטריונים ברורים המודדים את הצלחת המודל והפרויקט, כגון דיוק התחזיות, הפחתת זמני תגובה ושיפור בתכנון תשתיות.

מונחים טכניים במדעי הנתונים:

6. **Target Variable (משתנה מטרה):** המשתנה שעליו המודל מבצע תחזיות, במקרה זה - חומרת הפציעה.
7. **Predictive Modeling (מודלים חיזויים):** שימוש בנתוני עבר כדי לחזות אירועים עתידיים ולספק תובנות מעשיות.
8. **Machine Learning (למידת מכונה):** תחום מדעי הנתונים שמתמקד בפיתוח אלגוריתמים שמסוגלים ללמוד ולחזות מתוך נתונים היסטוריים.
9. **Data Cleaning (ניקוי נתונים):** זיהוי ותיקון נתונים חסרים, כפולים או שגויים לפני בניית המודל.
10. **Data Preprocessing (עיבוד מקדים):** פעולות שנעשות על הנתונים לפני הכנסתם למודל, כמו נרמול ערכים, קידוד קטגוריות או מילוי ערכים חסרים.
11. **Hyperparameter Tuning (כוונון היפר-פרמטרים):** תהליך של כונון הפרמטרים של מודל החיזוי כדי לשפר את ביצועיו.
12. **Class Imbalance (חוסר איזון בנתונים):** מצב שבו קטגוריה אחת מיוצגת באופן נמוך לעומת קטגוריות אחרות, מה שעלול להשפיע על דיוק המודל.
13. **SMOTE:** שיטה לאיזון נתונים על ידי יצירת דוגמאות סינתטיות, המשמשת בעיקר לטיפול במקרים של חוסר איזון בין קטגוריות בנתוני המודל.
14. **Performance Metrics (מדדי ביצוע):** מדדים המשמשים להערכת איכות המודל, כולל דיוק (Accuracy), רגישות (Recall) ודיוק ספציפי (Precision).
15. **Accuracy (דיוק):** מדד המשווה בין התחזיות הנכונות של המודל לבין התחזיות השגויות מתוך כלל התחזיות.
16. **Precision/Recall (דיוק/זיהוי):** מדדים להערכת איכות התחזית.
17. **Feature Engineering (הנדסת תכונות):** יצירה או התאמה של משתנים לשיפור דיוק המודל.
18. **Random Forest / XGBoost:** אלגוריתמים ללמידת מכונה עבור חיזוי מדויק.
19. **Regression (רגרסיה):** שיטה לחיזוי רמות פציעה על בסיס משתנים אחרים.



כלים טכנולוגיים:

1. **Python**: שפה לניתוח נתונים ובניית מודלים.
2. **Google Colab**: פלטפורמה לעיבוד בענן ותמיכה במודלים גדולים.
3. **Tableau / Matplotlib**: כלים להצגת ממצאים גרפיים.
4. **Heatmap (מפת חום)**: ייצוג חזותי של נתונים באמצעות צבעים שונים, המאפשר זיהוי אזורים מסוכנים לפי תאונות וחומרת פציעות ותבניות גיאוגרפיות.

עלויות ותועלות

הנתונים לפרויקט נלקחו ממאגר Data.gov, שהוא מקור ציבורי, ולכן אין עלויות ישירות הקשורות לאיסוף או שימוש בנתונים. עיבוד הנתונים יתבצע באמצעות Python, תוך שימוש בכלים נוספים ותוכנות חינוכיים. מכיוון שהפרויקט מבוצע על מחשבים אישיים של הסטודנטיות, העלויות התפעוליות הן נמוכות מאוד, ללא צורך במשאבים טכנולוגיים מתקדמים.

העבודה על הפרויקט דורשת השקעת זמן משמעותית, המוערכת בהרבה שעות עבודה של צוות הסטודנטיות. שעות אלו מוקדשות לניקוי הנתונים, פיתוח המודל, ניתוח התוצאות והכנת המצגות הסופיות. זמן העבודה מהווה עלות עקיפה חשובה שיש לקחת בחשבון כחלק מהערכת המשאבים הנדרשים לפרויקט. אין עלויות הקשורות לפריסת התוצאות, מאחר שהן מיועדות להצגה במסגרת פרויקט הגמר בלבד, ללא שילוב במערכות חיצוניות או פרסום ציבורי.

מבחינת תועלות, המודל יספק תחזיות מדויקות של רמות הפגיעה בתאונות דרכים, מה שיאפשר את יכולת ההקצאה והתגובה של שירותי החירום, רשויות התחבורה וחברות הביטוח. בנוסף, ניתוח הנתונים צפוי לזהות דפוסים הקשורים לתאונות חמורות, לאתר מוקדי סיכון גיאוגרפיים ולספק המלצות לשיפור תשתיות.

לבסוף, הערכת העלויות והתועלות מראה כי עלויות הפרויקט הן נמוכות יחסית, בעוד שהתועלות הצפויות הן משמעותיות ביותר. יחס זה מדגיש את הערך האקדמי של הפרויקט ואת הפוטנציאל הממשי שלו ליצור כלי חיזוי אפקטיבי שמספק ערך אמיתי לשיפור הבטיחות בדרכים ותהליכי קבלת החלטות.

מטרות מדעי הנתונים וקריטריוני הצלחה

סוג הבעיה במדע הנתונים:

פרויקט זה מתמקד בבעיה של סיווג (Classification), שבה המטרה היא לחזות את רמת הפגיעה (Injury Severity) בתאונות דרכים. המודל ייעד רמות פגיעה לקטגוריות מוגדרות מראש, כולל "ללא פגיעה", "פגיעה אפשרית", "פגיעה קלה", "פגיעה חמורה", ו"פגיעה קטלנית".

מטרות טכניות:

היעד המרכזי הוא פיתוח מודל סיווג מדויק ואמין, המבוסס על אלגוריתמים מתקדמים כמו Random Forest ו-XGBoost, שיוכל לבצע תחזיות בזמן אמת בהתבסס על נתונים היסטוריים.

- התחזיות צריכות להתבצע תוך פחות מ-10 שניות מרגע קבלת הנתונים, על מנת לאפשר לשירותי החירום להיערך ולהגיב באופן מיידי.
- המודל ישתמש במאפיינים גיאוגרפיים, דמוגרפיים, סביבתיים ועוד. כדי לשפר את דיוק התחזיות ולזהות דפוסים חשובים בנתונים.
- בנוסף, יפותחו Heatmaps המבוססים על נתונים גיאוגרפיים (Longitude ו-Latitude), שיאפשרו זיהוי אזורים גיאוגרפיים בעלי סיכון גבוה לתאונות חמורות. כלי זה יסייע לרשויות התחבורה בקבלת החלטות מושכלות לשיפור תשתיות ולמניעת תאונות בעתיד.



תחזיות קצרות טווח:

המודל יספק תחזיות מיידיות (Real-Time) על חומרת הפציעות בעקבות תאונות דרכים, עם תוקף שימוש עד 24 שעות מרגע התרחשות התאונה. תחזיות אלו יישמשו את שירותי החירום לתגובה מיידית, כגון הקצאת אמבולנסים וסגירת כבישים.

תחזיות לטווח ארוך:

המודל יבצע ניתוחים חודשיים של דפוסי תאונות וזיהוי אזורים מסוכנים. תובנות אלו יאפשרו לרשויות התחבורה לתכנן שיפורי תשתיות לתעדף משאבים בהתאם לתחזיות של סיכון אזורי בטווח של שלושה עד שישה חודשים קדימה.

תוצאות רצויות- מדדים מספריים להצלחת המודל:

- **דיוק (Accuracy):** המודל שואף להגיע לדיוק של לפחות 85% בתחזיות רמת הפציעה.
- **Recall ו-Precision:** הדגש יושם על זיהוי מדויק של פציעות חמורות וקטלניות, עם יעד של לפחות 80% בכל אחד מהמדדים.
- **זמן תגובה:** תחזיות המודל יופקו תוך פרק זמן של 10 שניות מרגע קבלת הנתונים.
- **שיפור תשתיות ואכיפה:** תחזיות המודל יובילו לירידה של 15% בתאונות חמורות באזורים שזוהו כבעייתיים.

מדדים סובייקטיביים והגורם המודד הצלחה:

- **שביעות רצון המשתמשים:** המודל ייחשב מוצלח אם גורמים מקצועיים, כמו נציגי שירותי החירום ורשויות התחבורה, ידווחו על שיפור בתהליך קבלת ההחלטות והקצאת המשאבים.
- **שיפור שיתוף הפעולה:** המודל ייבחן גם לפי יכולתו לשפר את התקשורת ושיתוף הפעולה בין רשויות החירום והתחבורה.
- **תרומה למחקר ופיתוח:** המודל יוערך על פי תרומתו למחקר אקדמי, תוך קבלת הכרה כחדשני ופרסומו בכנסים או מאמרים מקצועיים בתחום בטיחות בדרכים.

גורמי הערכה:

- **משתמשים ישירים:** נציגי הרשויות המשתמשות במודל.
- **גורמים אקדמיים:** מנחה הפרויקט וצוות אקדמי להערכת התרומה המחקרית.

פריסת המודל:

המודל ייפרס על פלטפורמה מתאימה (למשל, שרת ענן או מערכת מקומית) שתאפשר קבלת נתונים בזמן אמת והצגת תוצאות דרך דוחות גרפיים וכלים כמו Tableau או Power BI. התוצאות יישמשו גם למענה מידי לשירותי חירום וגם לתכנון אסטרטגי עבור רשויות התחבורה וחברות הביטוח.

תכנית הפרויקט

שלב 1: איסוף וניקוי נתונים (4-6 שבועות)

תיאור: איסוף נתונים ממקורות ציבוריים כמו Data.gov, ניקוי ערכים חסרים וזיהוי אי-עקביות. בחינת התפלגות המשתנים והסרת נתונים חריגים או לא תקינים.

מטרות:

- יצירת מאגר נתונים איכותי ונקי.
- הכנת הנתונים לשלב הניתוח והמודלים.

שלב 2: ניתוח ראשוני של הנתונים (4-6 שבועות)

תיאור: ביצוע ניתוח מעמיק לזיהוי מגמות, דפוסי ראשוניים וקשרים בין משתנים. בדיקות סטטיסטיות, התפלגות קטגוריות, וזיהוי משתנים חשובים לפיתוח המודל.

מטרות:

- הבנת הנתונים לעומק.
- זיהוי משתנים חשובים לשלב המודלים.



שלב 3: איזון נתונים ועיבוד מקדים (2-4 שבועות)

תיאור: איזון עמודת המטרה בסט האימון בלבד, באמצעות טכניקות כמו SMOTE או עדכון משקלים, התאמות נוספות, כמו נרמול ערכים (Scaling) וקידוד קטגוריות (Encoding) הכנת הנתונים הסופי לשלב המודל.

מטרות:

- התמודדות עם חוסר איזון בקטגוריות.
- הבטחת יציבות המודל והתאמתו לנתונים.

שלב 4: פיתוח מודל ראשוני (6-8 שבועות)

תיאור: בניית מודל ראשוני באמצעות אלגוריתמים כמו Random Forest ו-XGBoost, ובחינת ביצועי המודל הראשוני.

מטרות:

- פיתוח גרסה בסיסית של המודל.
- בחירת האלגוריתם המתאים ביותר להמשך.

שלב 5: הערכת המודל ושיפורים (4-6 שבועות)

תיאור: הערכת ביצועי המודל באמצעות Accuracy, Precision, Recall, ו-F1 Score. כוונת היפר-פרמטרים, שיפור המודל ואיטרציות חוזרות. בנוסף, ניסיונות עם מודלים שונים וגרסאות חיזוי שונות על מנת לשפר ביצועים.

מטרות:

- שיפור ביצועי המודל.
- זיהוי מגבלות ושיפור התחזיות.

שלב 6: פיתוח Heatmap ולוחות מחוונים (4-6 שבועות)

תיאור: יצירת Heatmaps אינטראקטיביים ולוחות מחוונים להצגת תחזיות ותובנות באופן חזותי.

מטרות:

- להקל על תהליכי קבלת החלטות ושיפור תשתיות.

שלב 7: סיכום והצגת תוצאות (2-4 שבועות)

תיאור: הכנת מסמכים, מצגת מסכמת ודוחות להצגת תוצאות הפרויקט במסגרת האקדמית.

מטרות:

- הצגת הפרויקט בצורה מקצועית וממצה.

נקודות החלטות קריטיות לתוכנית:

תחנות ביקורת ונקודות החלטה:

- אישור ניקוי ואיכות הנתונים לאחר שלב עיבוד הנתונים הראשוני.
- בחירת המודל הראשוני על סמך ביצועיו.
- סקירת Heatmaps ולוחות מחוונים ואישורם בהתאם לדרישות.

שלבים חוזרים:

- איזון והערכת המודל: איטרציות חוזרות לשיפור ביצועי המודל לאחר כל הערכה.
- עדכון ויזואליזציות: בחינת הנתונים המוצגים ושיפורם לפי פידבק.

משאבים ומאמץ נדרש:

- צוות: שתי סטודנטיות עם ליווי מנחה אקדמי.
- כלים: Python, Google Colab, Tableau (או כלים דומים להצגת תובנות).
- נתונים: נתוני תאונות דרכים ציבוריים ממאגר Data.gov.



הערכת כלים וטכניקות

במסגרת פרויקט זה, נבחרו כלים ושיטות עבודה המספקים מענה מקיף למטרות המיידיות והאסטרטגיות של חיזוי רמת פגיעה בתאונות דרכים וזיהוי אזורים מסוכנים. הכלים והטכניקות מאפשרים גם חיזוי בזמן אמת של רמות הפגיעה, לשיפור התגובה של שירותי החירום, וגם ניתוח נתונים ארוך טווח לתמיכה בהחלטות אסטרטגיות של רשויות התחבורה והבטיחות.

כלים מרכזיים: שפת התכנות Python נבחרה בזכות היכולות הרחבות שלה לכל שלבי תהליך מדעי הנתונים. באמצעות ספריות כמו Pandas ו-Numpy ניתן לבצע עיבוד מקדים וניקוי נתונים, בעוד ש Scikit-learn ו-XGBoost מאפשרות פיתוח והערכת מודלים חיזויים, תוך שימוש באלגוריתמים מתקדמים כגון Random Forest ו-XGBoost. ספריות Matplotlib ו-Seaborn משמשות ליצירת גרפים ו-Heatmaps להצגת תובנות ויזואליות. ספריית Imbalanced-learn מספקת פתרונות לטיפול בחוסר איזון בנתונים באמצעות טכניקות כמו SMOTE.

הפלטפורמה Google Colab מספקת סביבת עבודה בענן המאפשרת עיבוד מסיבי ותמיכה בחומרה מתקדמת, דבר המייעל את תהליך אימון המודלים. לוחות מחוונים אינטראקטיביים מפותחים באמצעות Tableau או Power BI להצגת נתונים ו-Heatmaps שיסייעו בתיעודף תשתיות ואכיפה.

שיטות וטכניקות עיקריות: בשלב העיבוד המקדים מתבצע ניקוי נתונים, איזון קטגוריות המטרה באמצעות SMOTE, עדכון משקלים, לצד נרמול ערכים, לשיפור יציבות המודל. בשלב הניתוח הראשוני מתבצעות בדיקות מעמיקות לזיהוי דפוסים, מגמות וקשרים בין משתנים כמו מהירות מותרת, סוג התאונה (חזיתית/צדית), מזג אוויר ועוד. ניתוח זה יניח את היסודות לפיתוח Heatmaps בשלב מתקדם יותר, שיסייעו בזיהוי אזורים מסוכנים ותיעודף משאבים.

לצורך פיתוח מודלים, נעשה שימוש באלגוריתמים מתקדמים כמו Random Forest ו-XGBoost, המספקים תחזיות מדויקות. בשלב ההערכה נעשה שימוש במדדים כמו Accuracy, Precision, Recall ו-F1 Score, לצד כוונן היפר-פרמטרים ושיפורים איטרטיביים.

מענה לצרכים עסקיים:

- **מענה מיידי:** תחזיות בזמן אמת מספקות לשירותי החירום ולמשטרה מידע על רמת הפגיעה, לייעול הקצאת משאבים וניהול זירת תאונה, כולל הקצאת אמבולנסים, ניהול תנועה, וסגירת כבישים.
- **מענה אסטרטגי-ארוך טווח:** Heatmaps מזהים אזורים מסוכנים על פי כמות תאונות ורמות פגיעה, ומאפשרים לרשויות התחבורה לתעדף שיפורי תשתיות כגון הצבת מצלמות תנועה, שיפור תאורה, והגברת שיטור.

סיכום: כלים וטכניקות אלו נבחרו כדי לענות על הצרכים המידיים וארוכי הטווח של הפרויקט, תוך שימוש בפתרונות מבוססי נתונים המשלבים חיזוי בזמן אמת ותובנות אסטרטגיות לשיפור הבטיחות בדרכים. הבחירה בכלים אלו מבטיחה פתרון יעיל וממוקד לייעול משאבים ולשיפור תשתיות באופן חכם ומבוסס נתונים.