חלק א׳ – נתוני ביטוח רכבים

# הגדרת הבעיה

## תיאור כללי של עולם התוכן הנחקר

עולם התוכן של המחקר עוסק בהתנהגות נהגים ושימוש במידע הקיים עליהם לצורך בניית מודל חיזוי. במחקר זה נרצה לחזות האם אדם יממש את ביטוח הרכב שלו במהלך השנה הבאה בהינתן מאפיינים עליו.

תחום זה של ניתוח נתונים וחיזוי מבוסס התנהגות נהגים הוא בעל חשיבות רבה בשל השפעותיו הכלכליות והחברתיות, הן עבור חברות הביטוח והן עבור הציבור. הבנת גורמים המשפיעים על מימוש ביטוח עשויה לסייע בחיזוי טוב יותר של סיכונים, בתמחור מדויק של פוליסות ביטוח, ואף בקידום אמצעים לשיפור הבטיחות בדרכים.

סט המידע שבידינו כולל מדגם של 8,000 אנשים, עם נתונים דמוגרפיים כגון רקע וגיל, ונתוני נהיגה הכוללים את מספר הפרות התנועה שביצעו, מספר דוחות על מהירות, סוגי הרכבים שלהם, ושנות הייצור שלהם.

בהתבסס על מידע זה, תחום המחקר עוסק במציאת קשרים ותובנות בין מאפייני הנהגים והתנהגותם לבין סיכויי מימוש הביטוח. בעבר, מחקרים בתחום זה השתמשו במודלים סטטיסטיים ובאלגוריתמים של למידת מכונה כדי לנתח דפוסים ולחזות התנהגויות, תוך שימוש בעיבוד מקדים למניעת רעש נתונים והתמקדות בתכונות מרכזיות כמו הרגלי נהיגה ותדירות עבירות.

מסקנות ממחקרי עבר מצביעות על כך שמודלים מתקדמים, כמו רשתות עצביות, עשויים לספק דיוק גבוה יותר בזיהוי קשרים מורכבים בין משתנים, במיוחד במקרים שבהם יש נתונים גדולים ומגוונים. כמו כן, היכולת לחזות תנאים קיצוניים, כגון נהגים בעלי סיכון גבוה במיוחד למימוש הביטוח, נותרה אתגר מרכזי המחקר.

מטרת המחקר הנוכחי היא לבחון האם ניתן לבנות מודל חיזוי על בסיס נתוני המדגם שבידינו, כזה שיוכל לספק לחברות הביטוח כלי יעיל לניהול סיכונים ותכנון תמחור מדויק.

## הגדרת שאלת המחקר

האם ניתן לבנות מודל חיזוי מספק בעזרת נתוני המדגם כדי לחזות האם לקוח יממש את ביטוח הרכב שלו בשנה הבאה?

# הבנת הנתונים

## תיעוד מקורות הנתונים ומשמעותם

< טבלה שמסבירה עבור כל משתנה מה הסיפור שלו >

< איך קיבלנו כל נתון >

## הסתברויות אפריוריות וקשרים בין מאפיינים

< מלא שיט >

איכות הנתונים

< טיפול במשתנים חסרים >

< נתונים לא הגיוניים >

הכנת הנתונים

< השמטת מאפיינים רועשים / חסרי חשיבות / תצפיות בעלות חוסר רב > + הסברים

< קטגוריזציה למשתנים רציפים >

< יצירת מאפיינים חדשים ממשתנים קיימים – אינטראקציה / פונקציות על משתנים >