**הצעת פרויקט: חיזוי האם ישארו דירות ריקות לאחר ההגרלה**

מגישים:

אורי ביטון-213868797

דניאל גולדשמיד-209328244

data.gov.ilמערך הנתונים שבחרנו נמצא באתר:

מערך המידע נותן מידע מפורט על הגרלות דיור במסגרת תוכניות ממשלתיות כמו "מחיר למשתכן". מערך הנתונים כולל נתונים על מספר הנרשמים והזוכים בהגרלה, פרטי הפרויקט, מחירי הדירות, סוגי זכאות, ועוד. מידע זה מאפשר לנו לנתח ולחזות תהליכים הקשורים להגרלות

מוטיבציה:

הפרויקט מתמקד בשאלה האם ישארו דירות ריקות בתום ההגרלה. הנושא חשוב כי הוא יכול לחשוף דפוסים של חוסר התאמה בין היצע לביקוש, כמו עודף דירות בפרויקטים מסוימים או חסמים לנרשמים. הבנה כזו יכולה לסייע למקבלי החלטות לשפר את תהליך הקצאת הדירות ולמקסם את יעילותו בפרויקטים עתידיים. בנוסף, הנושא רלוונטי לציבור הרחב, שכן הוא קשור לשיפור הנגישות לדיור

מתודולוגיה:

נשתמש בטכניקות למידת מכונה לעיבוד וניתוח המידע:

1) עיבוד נתונים מקדים

2) טיפול בערכים חסרים או שגויים

3) ניתוחי ויזואליזציה שיכולים לספק לנו תובנות משמעותיות

4) בדיכת כדאיות יצירת עמודות חדשות בעזרת שילובים של עמודות קיימות

חיזוי מונחה: (Supervised Learning)

נבנה מודל לחיזוי על סמך נתוני העבר, תוך שימוש באלגוריתמים של קלאסיפייאר על מנת לחזות האם יישארו דירות ריקות לאחר הגרלה או לא תוך בדיקה של מגוון מודלים ובחירה של המודל הטוב ביותר.

ולאחר מכן בניית כלי לבדיקת חשיבות הפיצ'רים על החיזוי

ניתוח בלתי-מונחה: (Unsupervised Learning)

זיהוי קבוצות בעלי מאפיינים דומים היכולות להציג לנו תובנות חשובות כבעלי עניין

ניסויים מתוכננים:

בניית מודל לחיזוי דירות ריקות:

חיזוי משתנה המטרה: האם נותרו דירות ריקות לאחר ההגרלה.

שילוב מאפיינים רלוונטיים כמו מספר הדירות, כמות הנרשמים לפי קבוצות זכאות, מחירי הדירות, ומצב הפרויקט.

חלוקה לסט אימון וסט בדיקה וביצוע ניתוח feature importance לזיהוי הגורמים המרכזיים שמשפיעים על התוצאה.

חקירת דפוסים בלתי-מונחים:

יצירת קבוצות פרויקטים על בסיס נתונים כמו מיקום גיאוגרפי, מצב הפרויקט, וסוגי הנרשמים, במטרה לזהות דפוסים שאינם ברורים ממבט ראשוני.

הערכת ביצועים:

בדיקת דיוק החיזוי של המודל, עם שאיפה לדיוק של לפחות 80%

זיהוי גורמים מרכזיים שמובילים לדירות שנותרות ריקות לאחר ההגרלה

תוצאה רצויה:

מודל מדויק המסוגל לחזות אם ישארו דירות ריקות, יחד עם הבנת הגורמים לכך

תובנות שיכולות לתרום לשיפור תהליך ההגרלות בעתיד