הצגה עצמית -

שלום, אני גל לוי ואני ליאל כהן סטודנטים שנה ב למדעי המחשב וכעת נציג לכם את פרויקט הסיום שלנו בקורס מדעי הנתונים.

שאלת המחקר –

שאלת המחקר שלנו היא - **מה הקשר בין המאפיינים של מחשב לבין מחיר השוק שלו, ובאיזו מידה נוכל לחזות את המחיר על סמך המאפיינים הללו?**

כסטודנטים ככלל וכסטודנטים למדעי המחשב בעלי עניין רב בשוק המחשבים, הפרויקט שלנו מתמקד בפיתוח מודל למידת מכונה שיכול לחזות את מחיר השוק של מחשבים על סמך המאפיינים שלהם.

על ידי ניתוח תכונות כמו CPU, זיכרון ומפרטים רלוונטיים אחרים, שאפנו לחשוף את הדפוסים והקשרים הבסיסיים המשפיעים על תמחור המחשב.

על ידי איסוף מערך נתונים מקיף של רישומי מחשב והמחירים המתאימים להם, יישמנו אלגוריתמי למידת מכונה כדי ליצור מודל חיזוי.

במטרה שמודל זה יספק תובנות חשובות לסטודנטים, קונים, מוכרים וקמעונאים ויעזור להם להעריך את מחירי המחשבים בצורה מדויקת יותר ולקבל החלטות מושכלות בשוק המחשבים הדינמי.

מקורות להרכשת המידע –

ביצענו הרכשה משני אתרים

האתר הראשון ZAP – אתר ZAP הוא אתר המתמחה במכירה וקניית מוצרי טכנולוגיה, בעיקר מחשבים ומוצרים קשורים לעולם המחשוב והסלולר.

אתר ZAP מהווה מקור מועיל עבור אלו שמחפשים לרכוש מוצרי טכנולוגיה באיכות

ולכן הרכשנו מתוכו את המאפיינים השונים של מחשבים ניידים.

האתר השני בו השתמשנו הוא - cpubenchmark.

cpubenchmark הוא אתר פופולרי המספק ציוני מדידה למעבדים שונים

האתר משתמש בתוכנת ההשוואה של PassMark הנחשבת לאמינה ומספקת תוצאות עקביות להשוואת ביצועי CPU .

הרכשת ZAP -

כדי להרכיש את הנתונים מאתר ZAP השתמשנו בספריית סלניום בפייתון.

האתר בנוי מעמודים המכילים אייטמים של מחשבים ולכן

חילצנו את כל האייטמים מכל עמוד. כדי לקבל את כל המאפיינים ראינו כי הם נמצאים תחת **פרטים נוספים**

כך למעשה חילצנו מכל אייטם את כל המאפיינים שלו.

יצרנו Class לביצוע ההרכשה המקבל URL ודרייבר. המחלקה מכילה פונקציונליות של

Get\_laptops אשר באמצעותה קיבלנו את כל האייטמים של המחשבים מכל העמודים.

Get\_details אשר באמצעותה קיבלנו את כל המאפיינים של כל מחשב.

לאחר מכן ביצענו הרכשה לאתר benchmark –

גם כאן השתמשנו בספריית סלניום וחילצנו את המידע על ידי ה ID של הטבלה.

ניתן לראות את פונקציית crawl\_cpu אשר מבצעת זאת.

לאחר שקיבלנו שני DF המכילים את המידע ביצענו crosscheck בין המעבדים שהכרשנו מZAP לציונם מאתר cpubenchmark והוספנו את זה כעמודה המייצגת פרמטר בDF המקורי.

ניקוי נתונים –

לאחר יצירת ה dataset שרצינו ביצענו ניקוי נתונים והפכנו את הפרמטרים להיות ב TYPE המתאים.

למשל מחיר המלווה לייצוג ₪ - הורדנו ובנוסף המרנו ערכים מ string ל int או float.

בנוסף, אתר ZAP מציג טווח מחירים ולכן בחרנו לחשב את המחיר הממוצע שהוא המשתנה המושפע.

לאחר מכן, ביצענו מחיקה של חריגים והורדה של דופליקציות וערכים חסרים.

ויזואליזציה –

ניתן לראות בגרף הראשון את התפלגות המאפיינים השונים שבחרנו, מתוכם נתמקד בכמה דברים מעניינים.

ניתן לראות את היצרניות אשר נבדקו בפרויקט שלנו – נשים לב כי כמות המחשבים הגדולה ביותר הייתה של לנובו ולאחר מכן של אסוס דל וHP.

המחיר הגבוה ביותר של מחשבים היה של מייקרוסופט ואילו של שיאומי הנמוכים ביותר.

השתמשנו במפת חום על מנת להבין את הקשרים בין המשתנים השונים שבדקנו.

נשים לב כי RAM CAPACITY בעל ההשפעה הרבה ביותר על המחיר הממוצע ולאחר מכן storage capacity ו cpu mark

בנוסף ניתן לראות כי לאורך השנים חלו עליות וירידות במחירים וכי החל מ2022 חלה עליה משמעותית במרבית היצרניות.

למידת מכונה –

מרבית הנתונים שלנו הם נתונים מספריים נומריים ובינאריים ולכן השתמשנו תחילה ברגרסיה ליניארית

קיבלנו כי הקשר בין המשתנים אותם בדקנו עומד על 70 אחוז וניתן לראות שזה די לינארי.

אך לא הסתפקנו בלמידת מכונה אחת ולכן ביצענו בדיקות נוספות

השתמשנו בKNN, עצי החלטה ויער רנדומלי

ניתן לראות כי התוצאות השתפרו וכי יער רנדומלי הביא את החיזוי הטוב ביותר ועומד על 87 אחוזים.

לסיכום –

בהתבסס על הממצאים שלנו, הפרמטרים המרכזיים שיש להם השפעה משמעותית על מחיר המחשב הם CPU, קיבולת RAM ויכולת אחסון.

בנוסף, גם מערכת ההפעלה ומועד הכניסה ל-ZAP משפיעים על המחיר, אם כי במידה פחותה.

נציין כי ככל שתאריך הכניסה לאתר מאוחר יותר, המחיר נוטה לעלות.

יש לציין שגם לגורמים כמו משקל, התאמה לגיימינג ותכונות אבטחה יש השפעה על המחיר, אך במידה פחותה.

תודה רבה.