המרת פרויקט קורס במבוא למדעי הנתונים לפיתוח מודל כחלק מפרויקט גמר next step

הפרויקט שלנו מבוסס על 2 מודלים עיקריים.

1. מודל ״דרך קריירה״- מודל שמקבל מהמשתמש את העבודה שהוא רוצה להגיע אליה, ואת הפרטים הנוכחיים עליו (כישורים, לימודים, ניסיון). על ידי ניתוח של קו״ח של משתמשים שהגיעו לעבודה הזאת, המודל ימצא מכנה משותף וימליץ למשתמש לעקוב אחרי מסלול זה בכדי להגיע לעבודת החלומות שלו
2. מודל ״עבודה עכשיות״ – המודל הזה מוצא למשתמש משרות שרלוונטיות למצב הנוכחי שלו.

בהמרה של פרויקט זה אנחנו נפתח את מודל העבודה העכשווי.

**חלק a**

שלב ב - Exploration + Data preprocessing

בחלק זה במקור יש צורך להסיר מהנתונים בתי מלון לא רלוונטים ולהמיר דאטא ordinal למספרים בעלי יחס סדר.

בהמרה שלנו אנחנו נוציא מהדאטא סטים שלנו מידע רלוונטי, וננקה את הדאטא שלא רלוונטי.

שלב ג - תחזית מחירי החדרים

במקור בחלק זה יש לממש פתרון לחזות את מחירי חדרי המלון על פי הנתונים שאספתם.

בהמרה שלנו נממש פתרון שמביא למשתמש את אחוזי ההתאמה שלו למשרה ספציפית. נממש זאת בעזרת אלגוריתמי רגרסיה.

שלב ה - ביצוע Reverse Engineering לאלגוריתמי המיון של האתרים

בשלב זה נבצע את המיון של המודל

(לא היה לי מושג איך לנסח את זה עם יותר מילים מזה)

**חלק b**

בכדי להביא לידי ביטוי בפרויקט גם רגרסיה, גם קלאסיפיקציה וגם קלאסטרינג, מימוש הרגרסיה והקלאסטרים יתבטא כך:

נבצע קלאסטרינג על מאגר הגוב טייטלים ונקבל קבוצות יותר כלליות (לדוגמא רופא) ולאחר מכן בקבוצה שהכי מתאימה למצבו הנוכחי של המשתמש נבצע רגרסיה על משרות ספציפיות בהתאם לגוב טייטל זה, על ידי שימוש בapi של לינקדין (לדוגמא נוירולוג באסף הרופא, רופא משפחה בראשון לציון...) ולסווג לכל משרה את אחוז ההתאמה שלה לנתונים של המשתמש.

רעיונות נוספים:

* רגרסיה
  + להראות עבור המשרה שרלוונטית למשתמש עכשיו כמה זמן כנראה הוא יעבוד במשרה זאת, בהתבסס על קורות החיים של אנשים שעבדו גם במשרה דומה.
  + לחזות את התאמת המשרה ללקוח ולתת תשובה באחוזים
* קלאסטרינג-
  + לסווג קו״ח לפי מילות מפתח
  + לעשות קלאסטרינג למידע ולהציג למשתמש מגוון משרות מאותה הקבוצה (פשוט בלי רגרסיה כמו ברעיון שבחרנו)