

כיוון 2 -

שלב 1:

- מתוך קבוצת הTEST (על אחד המודלים)  
נוציא את ה6 הרשומות של NVIDIA ו-TESLA ונשאר עם קבוצת TEST של 178 רשומות שבה לכל חברה יש יותר מ5 רשומות  
- לכל חברה מתוך 11 החברות שנשארו, בוחרים רנדומלית 5 רשומות המשוייכות לכל חברה (company). אנחנו נקבל בסה"כ 55 רשומות שנבחרו להיות המיצגות של המשרות: 5 משרות לכל חברה  
יתר 123 הרשומות שנשארו הן יהיו קבוצת המועמדים הפונים למשרות  
- הקבוצה הזאת מגדירה את רשומת המשרות והמועמדים לכל שמונת המודלים

שלב 2:

עבור כל אחת מהאפשרויות של וריאציה של גיל מגדר ופונק' מרחק - ועבור כל משרה, נחשב את המרחק בין כל אחד המועמדים למשרה הזאת (123 מרחקים) ואוציא בצורה מדורגת - קטן לגדול.  
בקובץ התוצאות: עבור משרה מסוימת יתקבלו ה11 מועמדים הקרובים ביותר.  
נקבל 8 קבצי תוצאות אשר כל קובץ מכיל 55 שורות

תוצאה:

מבנה רשומת התוצאה:

רשומת המשרה | 11 הרשומות הקרובות ביותר | המרחק שלהן

טיפול בערכים חסרים:

איפה נמצא המימוש לערכים חסרים במידע הנומרי והקטגורי

מדדים:

Inter Cluster Distances V

WCSS חדש V

1. ביצוע precision & recall לפי כל חברה, (לכל קובץ תוצאות בנפרד) ע"פ הדגשים שכתבנו ביום ראשון
2. Accuracy - לפי הגדרה של TRUE X לפי 1, 3, 5 X = או לפי 11 | 13 X = בכיוון 2 או 1 בהתאמה
3. nDCG -  
i - הוא המיקום בדירוג החברות (i - נע בין 1 - 11 | 13 בכיוון 2 או 1 בהתאמה,  
rel - היא 0 אם זה שונה מהתווית (label) ו- 1 אם זה זהה לתווית  
p - מספר הדירוגים = 11 | 13

הDCG תהיה הציון הסופי של שאילתה שהוא קטן מאחד או אחד (ציון > 1)

התוצאה הסופית של כל השאילתות תהיה ממוצע של ציוני ה- DCG