מסמך מלווה לפרויקט כריית נתונים

חלק א' –

סעיף 1 -

ייצור הנתונים שיותאמו לclusters יתבצע באופן הבא:

1) בחירה אקראית של מספר הנקודות שבכל קבוצה כך שסכום הקבוצות יהיה שווה לn

2) K פעמים יתבצעו השלבים הבאים:

1. maxMumberInThisGroup- הגרלת מספר מ 5000 עד 10000 (שהוא ייצג את המספר המקסימלי שיכול להיות בקבוצה הזו)
2. עבור כל מספר שהוגרל בסעיף קודם שהוגרל- יוגרלו dim מספרים מ maxMumberInThisGroup\2 ועד maxMumberInThisGroup וזה בעצם נקודה אחת מסך הנקודות
   1. סעיף זה יתבצע בפונקציית עזר **שמקבלת** את הK הרולוונטי ויוצאת נקודות ואת
   2. חזרה על סעיף 3 C פעמים

הסבר לסעיפים 2,3- סעיפים אלו התבצעו באופן הזה על מנת לאפשר לנקודות פוטנציאליות להיות באותו אשכול

סעיף 2 –

נפעל לפי האלגוריתם אשכול היררכי כל הנקודות מופרדות וכל אחת היא קבוצה בפני עצמה, לכן יהיו בעצם כמה לולאות, שהלולאה הפנימית תחשב מרחקים של כל הנקודות עם כל הנקודות וכל סיום לולאה שזה בעצם סיום החישוב של כל הנקודות עם כל הנקודות נאחד את המינימלים כלומר נלך לפי single link