**חלק ג:**

**שאלה 5.1:**

Accuracy is = 93.81%

**שאלה 6:**

1. מטרת הגיזום היא להחליש את ה overfitting ע"י הקטנת העץ. הפעולה של הגיזום מחליפה תת העץ בעלה, בגיזום מוקדם נעשית הפעולה לפני פיתוח תת העץ.
2. חלק רטוב
3. 2. הגרף שקיבלנו:

Chart, line chart

Description automatically generated

3. עבור גיזום גדול מ 5 קיבלנו ירידה ב validation accuracy, לכן נבחר פרמטר הגיזום להיות 5. עבור גיזום עם פרמטר גיזום 5 קיבלנו שה test accuracy היא 95.92% וזה גבוהה מערך שקיבלנו 93.81%.

**חלק ד:**

**שאלה 7:**

הכללה של מסווג השכן הקרוב, אובייקט ייבחן ע"י K השכנים הקרובים.

יתרונות:

1. פשוט למימוש
2. אלגוריתם KNN לומד ומסווג רק כאשר מתקבלת בקשה לסיווג, לכן אפשר להוסיף data חדשה ולא לפגוע ב accuracy של האלגוריתם.

חסרונות:

1. אלגוריתם KNN לומד ומסווג רק כאשר מתקבלת בקשה לסיווג, לכן הוא אלגוריתם עצלן. אלגוריתם זה יתקשה לתפקד כאשר מדובר על כמויות גדולות של אימון.
2. האלגוריתם לא עובד עם **high dimensions,** קשה לאלגוריתם לחשב את המרחק.

**שאלה 8:**

1. היא קבוצת החזקה:
2. שהוא שווה ל , הקבוצה S היא בגודל D.