**מעבדה 6 – כריית נתונים**

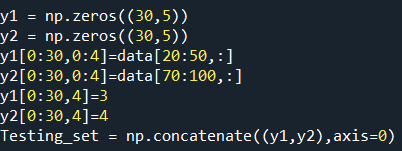
שם מגיש: אילון בן סימון ת"ז: 312162951

שם מגיש: סער ויקטור ת"ז: 312392822

שאלות 6-8:

6. בקטע קוד הבא ניתן לראות את חיבור הנקודות שלא נבחרו ל-training set כלומר הנקודות שייבחרו ל-testing set (30 נקודות מסוג הפרח הראשון ועוד 30 נקודות מסוג הפרח השני).

להלן קטע הקוד המדובר:



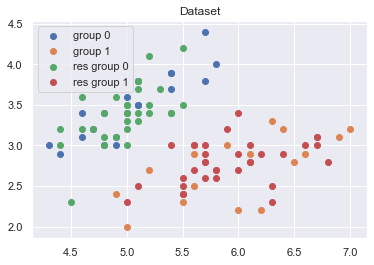
7. להלן התוצאות הגרפיות עבור k=1 (שכן אחד קרוב):

Group 0- קבוצת training set 20 פרחים מסוג 1.

Group 1- קבוצת training set 20 פרחים מסוג 2.

Res Group 0- קבוצת testing set 30 פרחים מסוג 1.

Res Group 1- קבוצת testing set 30 פרחים מסוג 2.



מהגרף ניתן לראות שקבוצת ה- testing set השייכת לסוג הפרח הראשון ("מסוג 1") נמצאת קרוב מאוד לקבוצת ה- training set מאותו סוג פרח ואותו הדבר ניתן לומר על סוג הפרח השני ("מסוג 2").

8. להלן האלגוריתם KNN המחלק את ה-testing dataset לשתי קבוצות לפי k שרירותי:

