מטרה - בניית מודל למידת מכונה שיכול ללמוד מהמידות של האירוסים הללו שהמין שלהם ידוע, כך שיוכל לחזות את המינים לקשתית חדשה.

החומרים הבאים שימשו לבניית הדגמים:

1.“Introduction to Machine Learning with Python: A Guide for Data Scientists”.

Book by Andreas C. Müller and Sarah Guido

2.אתר https://www. scikit-learn.org/

3. אתר https://www.python.org/

בחרתי בשני מודלים (שני אלגוריתמים):

k-nearest neighbors and logitic regression

במערך הנתונים מופיעים שלושה סוגי האיריסים והמאפיינים של כל פרח. מספר פרחים 150

כדי לטעון את Iris dataset אני משתמש בפונקציה load\_iris של מודול dataset של ספריית scikit-learn

from sklearn.datasets import load\_iris

iris\_dataset = load\_iris()

iris\_dataset מכיל מפתחות וערכים:

Keys of iris\_dataset:

dict\_keys([‘data’, ‘target’, ‘frame’, ‘target\_names’, ‘DESCR’, ‘feature\_names’, ‘filename’])

מפתח DESCR הוא תיאר של מערך הנתונים

מפתח target\_names הוא סוגי פרחים:

Target names: [‘setosa’ ‘versicolor’ ‘virginica’]

מפתח feature\_names הם השמות של מאפייני הפרחים

feature names: [‘sepal length (cm)’, ‘sepal width (cm)’, ‘petal length (cm)’, ‘petal width (cm)’]

מפתח data הוא מדידות כמותיות של פרחים - אורך עלי הכותרת, רוחב עלי הכותרת, אורך עלי הכותרת ורוחב עלי הכותרת

Data:

[[5.1 3.5 1.4 0.2]

[4.9 3. 1.4 0.2]

[4.7 3.2 1.3 0.2]

[4.6 3.1 1.5 0.2]

[5. 3.6 1.4 0.2]]

…………………….

מפתח target הוא סוגים של פרחים

Target:

[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2

2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2

2 2]

שמות של פרחים

0 – setosa, 1 – versicolor, 2 – virginica

בהתבסס על iris\_dataset, עליך לבנות מודל החוזה סוגים של איריס לקבוצת מדידות חדשה.

צריך לבנות מודל אשר ינבא סוגים של פרחים לקבוצת מדידות חדשה

מחלקים את הנתונים הראשוניים לשני חלקים:

החלק הראשונה לבניית המודל training set

החלק השני הוא להערכת (בדיקת) המודל test set

השתמש בפונקציה train\_test\_split לחלוקת הנתונים הראשונים

פונקציה זו מערבבת את מערך הנתונים ומחלקת אותו לשני חלקים לא שווים training set –75%, test set – 25%:

from sklearn.model\_selection import train\_test\_split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(iris\_dataset[‘data’],iris\_dataset[‘target’],random\_state=0)

מערכים של training set

X\_train, y\_train

מערכים של test set

X\_test, y\_test

בניית מודל בשיטה k-nearest neighbors

שימוש בכיתה KNeighborsClassifier ליצירת אובייקט knn

מספר שכנים = 1(פרמטר n\_neighbors)

from sklearn.neighbors import KNeighborsClassifier

knn=KNeighborsClassifier(n\_neighbors=1)

מתבצעת השיטה FIT של אובייקט KNN לבנות מודל

knn.fit(X\_train,y\_train)

שיטת SCORE של אובייקט KNN משמשת לחישוב נכונות המודל

accuracy = knn.score(X\_test,y\_test)

נכונות המודל לחבילת הבדיקות = 0.97

זוהי תחזית גבוהה. לכן מודל זה אמין בשימוש

בניית מודל בשיטה logitic regression

שימוש בכיתה LogisticRegression ליצירת אובייקט logreg

הפוך לכוח הסדרה = 1(פרמטר C)

from sklearn.linear\_model import LogisticRegression

logreg=LogisticRegression(C=1)

מתבצעת השיטה FIT של אובייקט logreg לבנות מודל

logreg.fit(X\_train, y\_train)

שיטת SCORE של אובייקט logreg משמשת לחישוב נכונות המודל

accuracy = knn.score(X\_test,y\_test)

נכונות המודל לחבילת הבדיקות = 1.00

זוהי תחזית גבוהה. לכן מודל זה אמין בשימוש