Am folosit codul din rezolvarea laboratorului 5, căruia i-am schimbat citirea și afișarea.

După deschiderea fișierelor pentru citire, am salvat în liste doar informațiile de care am nevoie, adică listele cu \_labels conțin doar valorile 0 și 1, iar listele cu \_samples au pe fiecare “linie” o listă formată doar din cuvintele frazei criptate.

Exemplu:

test\_samples[0] = ['k>.h', 'j:TW@', "'g", 'cWUX', '}xDd', 'fzsFU%', 'zq|=p}', '<p#o', '#fEw', 'ApziUd', 'gjjAh', "&rk<'", '@he=', 'Akefe@m', 'jkjA', 'HrWDpi', 'hg', '@mof@m', 'n(E&hj', 'fZ}', '@:f}', 'm}h@(']

train\_lables = ['1', '1', '1', '1', '0', '1', '0', '0', '0', '0', '1', '0', '1', '1', '0', '1', '1', '1', '1', '1', '1', '0', '0', '0', '1', '0', '0', '0', '0', '0', '0', '1', … ,’1’]

In continuare am definit vocabularul folosind metodele din Bag\_of\_words și am format matricele train/test/validation\_features pe care le-am normalizat folosind tipul “l2”, salvând normalizările în scaled\_train/test/validation\_data. Apoi am definit modelul (folosid svm din biblioteca sklearn, cu parametru de penalitate pentru eroare C=100 și kernel linear), l-am antrenat și am salvat predicția label-urilor în predicted\_labels\_svm.

In final, am calculat și afișat F1-score-ul și am format fișierul txt de export, ce conține id-urile din test\_samples și label-urile din predicted\_labels\_svm, desparțite printr-o virgulă. Am folosit Google Colab pentru a rula codul.