שם: אביעד טוויג

ת.ז: 319090171

**שיחזור התוצאות על סט הבדיקה**

יש להוריד את המודל [מהקישור הזה](https://drive.google.com/file/d/1FFY7WDN_cna4tBVcRdHUPLcE8yXC56kz/view?usp=sharing).

יש לשים את כל קבצי הפייתון, את הקובץ של המודל ואת הקובץ של סט הבדיקה באותה תיקייה.

יש להריץ את הקובץ ב-main וללחוץ על 1 ואז אנטר (אם משהו מהקבצים הנ"ל לא נמצא שם תיווצר הודעת שגיאה והתוכנית תיפול).

לבסוף ייווצר קובץ csv שנקרא: predict\_on\_set\_by\_char\_by\_most.csv

הקובץ שנוצר, נוצר על ידי החיזוי של סט הבדיקה על המודל ועם הסתמכות על כך שכל מילה מכילה פונט בודד. ובהתאם "הולך על פי הרוב" של החיזוי של הפונט במילה עבור כל תו, במידה ואין רוב החיזוי שיופיע הוא בהתאם לחיזוי של התו הבודד.

**הערות נוספות על הקוד**

החלוקה של קבצי הפייתון:

1. train.py – מכיל 4 מתודות:
2. הוספת רעש מלח פלפל לתמונה.
3. יצירת תמונות ולייבלים מתוך data set (תוך כדי שימוש במתודת רעש).
4. בניית סט האימון והוולדיזציה לפי שם הקובץ (תוך כדי שימוש במתודה הקודמת)
5. אחסון התמונות והלייבלים בתוך קובץ h5.
6. myModel.py – מכיל 4 מתודות:
7. יצירת הסטים המתאימים של האימון והוולדיזציה עבור המודל (תוך כדי שימוש בקבצי ה-h5 שנבנו ב-train.py).
8. יצירת סט מתאים של הבדיקה רק של תמונות.
9. בניית המודל ושמירה של המודל הטוב ביותר.
10. חיזוי של המודל על סט הבדיקה (תוך כדי שמירה החיזוי כקובץ h5).
11. results.py – מכיל 4 מתודות:
12. מתודה ראשית שיוצרות 4 רשימות (החיזוי עבור כל אות, רשימה של אותיות, רשימה של מילים ובמידה ולסט יש פונט רשימה של הפונט האמיתי).
13. בניית הקובץ csv רק לפי החיזוי של האות עצמה.
14. בניית הקובץ csv לפי החיזוי של האות עצמה ולקחת בהתחשבות את זה שלכל מילה יש פונט אחד.
15. בניית הקובץ csv לפי החיזוי עבור המילים.
16. main.py – מכיל מחלקה אחד, 5 מתודות ועוד מלל באזור הגלובלי:
17. המחלקה היא של צבעים ושל הדגשות עבור ההדפסות.
18. המתודה הראשונה משתמשת במתודות של train.py ליצירת הסט אימון ווולדיזציה.
19. המתודה השנייה משתמשת במתודות של myModel.py לבניית המודל מסט האימון והוולדיזציה.
20. המתודה השלישית משתמשת במתודות של myModel.py לבניית סט הבדיקה והחיזוי על סט הבדיקה.
21. המתודה הרביעית משתמשת במתודות של results.py ליצירת קבצי ה-csv בהתאם לדרישה.
22. המתודה החמישית היא לצורך שאלת שאלות של המשתמש במידה והוא רוצה לשחזר את התהליך.
23. המלל מחוץ למתודות הוא לצורך שאלת שאלות בסיסית של המשתמש אם הוא רק רוצה לחזות וליצור קובץ csv או שהוא רוצה לשחזר את כל התהליך.

ניתן לשחזר את כל התהליך שביצעתי באופן הבא (מספיק שרק קבצי האימון וקובץ הבדיקה יהיו בתיקייה של קבצי הפייתון).

הכוונה בתהליך הוא: יצירת סט התמונות של האימון (אם יש גם קובץ validation אז גם סט תמונות שלו) -> בניית המודל בהתאם למודל שהגדרתי, ביחד עם סט התמונות (וסט הוולדיזציה אם קיים) -> בניית סט התמונות של הבדיקה -> חיזוי על סט הבדיקה -> יצירת קובץ csv.

יש להריץ את ה-main וללחוץ על 2 ואז אנטר, ואז יש לבחור את מה שבדיוק רוצים לבדוק כלומר איזה שלב בתהליך. בכל בחירה יהיו שאלות שמוודאות שכל הקבצים הנדרשים עבור כל שלב קיימים.

בנוסף, ישנה אופציה לקבל קבצי csv שונים בהתאם לחיזוי שנוצר:

1. קובץ csv כפי שציינתי לעיל.
2. קובץ csv שמכיל את החיזויים עבור כל תו, מבלי להתייחס לכך שמילה מכילה פונט בודד.
3. קובץ csv שמכיל את החיזויים עבור כל מילה (אם יש רוב של פונט אחד אז הוא החיזוי, אם במילה אחת יש תיקו על המקום הראשון בין שניים אז כל אחד מהפונטים שבתיקו מקבל 0.5, אם יש תיקו בין שלושתם אז כל אחד מקבל 0.333.

אם בסט הבדיקה יש פונטים אז קיימת אופציה ליצור את אותם קבצי csv עם עוד שלוש עמודות שמכילות את הפונט האמיתי של אותו תו (1 עבור הפונט, 0 עבור השאר).