בס"ד

הוראות כלליות: על **כל תרגיל** עליכם לכתוב מסמך המציג הסברים על מה שעשיתם, את דרך העבודה שלכם, מדוע בחרתם בדרך שבחרתם. את התוצאות ונתוח של התוצאות והשגיאות (אם ישנן).

את כל הפתרונות עליכם להעלות למודל. הפתרון **חייב לכלול** גם את הקוד עצמו בקובץ של פייתון (ולא קישור) ואת המסמך עם הסברים מפורטים עם טבלאות, תרשימים, פלוטים וכדומה.

**תרגיל 3 ב- IR**

בתרגיל זה אנו נבדוק אוריינטציה סנטימנטלית של כל אחד מהעיתונים.

לכל עיתון ייבדק הסנטימנט בהקשר של פרו-ישראלי או פרו-פלסטיני.

לצורך זה צריכים להתמודד עם האתגרים הבאים: (1) אין לנו דוגמאות מתוייגות למשימה זו, (2) ה- BERT מוגבל ל- 512 טוקנים.

על מנת להתגבר על האתגרים אלו בצורה אוטומאטית, עלינו לבצע את הפעולות הבאות (אם יש לכם רעיונות אחרים אשמח לשמוע):

1. לאסוף מילים בעלי אוריינטציה חיובית כלפי ישראל ואותו דבר כלפי חמאס והפלסטינאים.
2. לעבור על כל הכתבות ולשלוף משפטים בעלי איזכור של מילים **השייכות לקבוצה אחת בלבד** (פרו-ישראלית/פרו-פלסטינית, רשימת מילים מופיעה בנספחים).
3. להריץ סנטימנט על המשפטים הרלוונטיים לשלב 2 (יש לא מעט מסווגי סנטימנט ב- hugginface ).

הואיל ולכל מסווג ישנן שגיאות, נבצע מספר בדיקות אוטומאטיות

* 1. נריץ כל משפט על מספר מסווגים של סנטימנט (מופיעים בנספחים)
  2. נבחר את אלו שכולם (או כמעט כולם הסכימו לתיוג שלהם פרו-ישראלי/פרו-פלסטיני/נייטראלי)
  3. נבנה שלוש קבוצות של משפטים (פרו-ישראלי/פרו-פלסטיני/נייטראלי) בעלי מספר דומה של משפטים.

1. תיוג כתבה מסויימת:
   1. לעבור על כל אחד מהמשפטים בכתבה (השייכים לשלב 2 או העומדים בקריטריונים של שלב 2)
   2. כל משפט צריך לקבל ציון סנטימנטלי (פרו-ישראלי/פרו-פלסטיני/נייטראלי)
   3. לעשות פונקציה המשקללת את הציונים של כל המשפטים בכתבה
   4. בהתאם לפונקציה, להכריע האם הכתבה פרו-ישראלית/פרו-פלסטינית/נייטראלית (ועד כמה היא כזו)
2. לרוץ על כל הכתבות
3. לרוץ על כל העיתונים
4. לתת ציון לכל אחד מהעיתונים לגבי פרו-ישראל ופרו-פלסטינאים לפי פונקציה שפיתחתם או פונקציה קיימת,לבחירתכם; כמובן פונקציה רלוונטית ועניינית והסבירו מדוע בחרתם בפונקציה זו.

לכל משפט בכתבות תשמרו **בקובץ אקסל1**:

* שיוך לעיתון (אפשר לשונית לכל עיתון)
* שיוך לכתבה/מספר כתבה (אינדקס של הכתבה)
* המשפט עצמו
* האם המשפט משוייך לישראל או לפלסטינים (p/I), בדוגמה, "I" (מתייחס לישראל בצורה שלילית)
* ציון שקיבל המשפט (עד כמה פרו-ישראלי/פרו-פלסטיני/נייטראלי) לכל אחד מהמודלים (אם יש ציון מספרי).
* תיוג (ההכרעה הסופית, למשל בדוגמה NEG)
* ציון ממוצע סופי **של הרוב (**למשל בדוגמה 0.935)

דוגמה לקובץ אקסל:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **עיתון** | **מספר כתבה** | **משפט** | **I/P** | **מודל 1 ציון** | **מודל 2 ציון** | **מודל 3 ציון** | **מודל 4 ציון** | **מודל 5 ציון** | **מודל 6 ציון** | **מודל 7 ציון** | **מודל 1 הכרעה** | **מודל 2 הכרעה** | **מודל 3 הכרעה** | **מודל 4 הכרעה** | **מודל 5 הכרעה** | **מודל 6 הכרעה** | **מודל 7 הכרעה** | **majorןty** | **AVG-majorןty-score** |
| A-J | 7 | Children and civilians were murdered in **Israeli** bombings. | I | 0.92 | 0.95 |  |  |  |  |  | NEG | NEG | NUT | NUT | NEG | NEG | NEG | **NEG** | 0.935 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

אופן ההגשה וכתיבת הדו"ח על התרגיל:

עליכם לכתוב מסמך readme שיכיל בראשו את שמות הסטודנטים, ת.ז., ומספר קבוצת התרגיל. המסמך יציג הסברים על מה שעשיתם, את דרך העבודה שלכם, מדוע בחרתם בדרך שבחרתם. האם היה קושי לזהות את הסנטימנט של משפט או של כתבה? מדוע דרך זו או דרך אחרת להתמודד עם המשימה טובה/לא—טובה? האם גיליתם "התנהגויות מוזרות" של עיתונים/כתבות מסויימות? נניח שאתם עורכים של העיתון, מה הייתם אומרים לכתבים הכותבים את הכתבות?

את המסקנות שלכם לגבי כל אחד מהעיתונים. לכל עיתון, תרשים עוגה של אחוז הכתבות לכל קטיגוריה.

את התוצאות ונתוח של התוצאות **והשגיאות** (אם ישנן) ותובנות שהגעתם אליהן.

**קובץ אקסל2** המכיל 4 לשוניות, לשונית לכל עיתון; לכל אינדקס של כל כתבה, הרלוונטית ללשונית, את הציון של כל המשפטים המופיעים בכתבה; את הציון של הכתבה ואת ההכרעה/התיוג הסופי של הכתבה.

את כל הפתרונות עליכם להעלות למודל. הפתרון **חייב לכלול** את (1) הקוד בקובץ של פייתון (ולא קישור) ואת (2) מסמך ה- readme עם הסברים מפורטים עם טבלאות, תרשימים, פלוטים וכדומה, ואת (3) קובצי האקסל.

בנוסף, עליכם להכניס לספריה את קובץ ה- readme ואת קובצי האקסל, שם הספריה יהיה שמות הסטודנטים שעשו את התרגיל ולכווץ בזיפ את כל הספריה; וגם את זה להעלות למודל.

**נספחים:**

מילים פוטנציאליות לישראל/פלסטין (**אתם רשאים להוריד מילים או להוסיף מילים, ברור לי שחלק מהמילים עדיף שלא יופיעו**).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Israeli words** | |  | **Palestine words** | |
| Cabinet | cabinet |  | Abbas | abbas |
| Colonizers | colonizers |  | Displaced | displaced |
| Government | government |  | Freedom fighters | freedom fighters |
| Homeland | homeland |  | Gaza | gaza |
| Humanitarian Aid | humanitarian aid |  | Gazans | gazans |
| IDF | idf |  | Hamas | hamas |
| Iron Dome | iron dome |  | Hassan Nasrallah | hassan nasrallah |
| Israel | israel |  | Hezbollah | hezbollah |
| Israeli | israeli |  | Houthis | houthis |
| Jerusalem | jerusalem |  | Humanitarian Crisis | humanitarian crisis |
| Jewish | jewish |  | Intifada | intifada |
| Knesset | knesset |  | Iran | iran |
| Mossad | mossad |  | Muhammad Sinuar | muhammad sinuar |
| Netanyahu | netanyahu |  | Naim Qassem | naim qassem |
| Occupation | occupation |  | Nakba | nakba |
| Occupied Territories | occupied territories |  | Nukhba | nukhba |
| Occupiers | occupiers |  | Oppressed | oppressed |
| Parliament | parliament |  | Organization | organization |
| Settlers | settlers |  | Palestine | palestine |
| Tel Aviv | tel aviv |  | Palestinians | palestinians |
| Tel-Aviv | tel-aviv |  | PLO | plo |
| West Bank | west bank |  | Refugees | refugees |
| West-Bank | west-bank |  | Resistance | resistance |
| Zionism | zionism |  | Resisters | resisters |
| Zionist entity | zionist entity |  | Sinuar | sinuar |
| Zionist regime | zionist regime |  | Terrorists | terrorists |
| Zionist State | zionist state |  | Tyrants | tyrants |
|  |  |  | Victims | victims |

מודלים של סנטימנט (אתם מוזמנים למצוא אחרים/נוספים):

1. <https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest>
2. <https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment>
3. <https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment>
4. <https://huggingface.co/siebert/sentiment-roberta-large-english>
5. <https://huggingface.co/lxyuan/distilbert-base-multilingual-cased-sentiments-student>
6. <https://huggingface.co/finiteautomata/bertweet-base-sentiment-analysis>
7. <https://huggingface.co/j-hartmann/sentiment-roberta-large-english-3-classes>