***Fake News***

מבוא

על מנת להתמודד עם בעיית הFake News נצטרך קודם להבין מהו Fake News.

חדשות מזויפות הן המייצגות מידע כוזב במסווה של חדשות אותנטיות.

יוצרי ה Fake News מפרסמים באמצעות מספר פלטפורמות בדרך כלל חשאיות במטרה להטעות את צרכני התוכן ולגרום לדבר להתפרסם בצורה ויראלית דרך ערוצי תקשורת שונים.

הפלטפורמות בהם מתפרסמים Fake News:

* Client- Server: אתרי חדשות, רשתות חברתיות, video-sharing website
* Peer to Peer: WhatsApp, Telegram

מטרות המפרסמים:

* כאוס.
* פגיעה באדם\ אישות מסוימת.
* פגיעה במותג מסוים.

הקשיים בהתמודדות עם Fake News

* השגת החדשות. לעיתיים החדשות מפורסמות בקבוצות סגורות ומשם מתפשטות.
* ניתוח החדשות והחלטה האם הוא אמתי. עקב העבודה שרבים מן האנשים אשר קוראים Fake News אכן עלולים להתקשות בלהבין האם מדובר בחדשות אמיתות או מזויפות.
* זיהוי המקור הראשוני של מפרסם החדשות.

מטרת המערכת:

יכולת לזהות חדשות מזויפות בשלב מוקדם ככל הניתן.

תכנון המערכת

במערכת יהיו מספר שלבים:

1. בניית מודל שבעזרתו נוכל לחזות לגבי נכונות ידיעה.
2. הרצה מחזורית של:
   1. ניטור ידיעות חדשות.   
      (ננסה לפתור ע"י יצירת משתמשים וניסיון להצטרף לקבוצות סגורות ובצורה אוטומטית למשוך מהם ידיעות)
   2. חיזוי נכונות ידיעה.  
      את בעיית ניתוח הידיעה והחלטה האם אמתית אני סבורה שבאמצעות מערכת מבוססת על אלגוריתם מעולם ה Machine Learning נוכל להתגבר על בעיה זו לאורך זמן.  
      המערכת תהייה semi supervised learning ובמקרה שלא תצליח לחזות היא תיעזר במומחה אשר יעשה בדיקה בעצמו ויזין למערכת מידע אמין בלבד.

למערכת מספר מצבים:

* מצב למידה: במצב זה המערכת בונה מודל בעזרת מידע קיים.
* מצב ניבוי: במצב זה המערכת מקבלת News ומחליטה לגביו האם Fake or Real

המודל אותו תייצר המערכת יהיה עפ"י עקרון "חוכמת ההמונים".  
המודל יהיה מורכב ממספר מודלים קטנים שייבנו בעזרת אלגוריתמים ממשפחות ML שונות.  
כל אחד מהמודלים הקטנים יוכל לקבל News ולספק ניבוי מספר בין 0-1.  
הניבוי של המערכת יהיה ממוצע משוקלל כאשר לכל מודל קטן יהיה משקל משלו   
(Acc /TotalAcc).   
כאשר הניבוי לגבי News יהיה 0.5 המערכת תבקש את עזרת המומחה.

בשלב זה (אבטיפוס) המאפיינים שהתייחס אליהם ב News הם:

* **כותרת**: האם הכותרת כתובה באותיות גדולות בלבד או מכילה סימני קריאה?
* **תאריך פרסום הכתבה**: שנה, חודש ומספר יום
* כמות הכתבות האמיתות שפרסם אותו הכתב בעבר
* **כמות שגיאות הכתיב בכתבה**
* **המיקום הגיאוגרפי בו צולמו תמונת מהכתבה**
* **האם שמו של האתר הוא וריאציה של שם של אתר מוכר**
* לבדוק את **פרטי צור הקשר של האתר** (כתובת אמתית, אימייל,..)
* **התאריך בו צולמה תמונות הכתבה לעומת תאריך פרסום הכתבה**
  1. ידוע גורמים שונים על מנת למונע מהידיעה להתפשט.  
     (לידע גורמים אינו בעיה ובאמצעות API נוכל לספק מידע זה לכל מי שחפץ בו או לפנות לשירותים של מערכות אחרות.)
  2. קבלת פידבק ישיר או עקיף. (כרגע נסתפק בעקיף, למערכת יהיה API אל מספר מערכות מרכזיות ואמיניות שאם הידיעה פורסמה שם כנראה שהיא נכונה ואם לא אז כנראה שמזויפת)
  3. שיפור המודל בהתאם לפידבק.