

פרויקט מחקר במדעי הנתונים-רשתות חברתיות

פרויקט זה מהווה פרויקט סיכום לחומר הנלמד בקורסים השונים במדעי הנתונים. במהלך המסלול נלמדו קורסים שנתנו כלים שונים לטיפול ועבודה עם נתונים, מתוך מטרה ללמוד דפוסים כאלו ואחרים באופן מיועד ובלתי מיועד מתוך אוסף הנתונים. בעזרת דפוסים אלו ניתן להציג מסקנות מחקריות ולבנות מודלים של למידת מכונה.

נושא המחקר או האפליקציה שתממשו צריך להיות סביב נתונים ורעיונות שקשורים ברשתות חברתיות, עליכם לבחור נושא מדויק עם יכולת איסוף נתונים מספקת לעבודה מחקרית. המחקר יכול להתרכז ברשת אחת (רק פייסבוק) או לבצע השוואה בין רשתות שונות (השוואה בפוסטים טוויטר מול פייסבוק)

לדוגמא

- תפוצת פוסטים בהיבט הזמן והמרחב בטוויטר בהשוואה למגיפות קודמות (עופות, חזירים וכו') ובהתייחסות לגלים השונים.
- זיהוי פוסטים קונספירטיביים.
- דירוג וסיווג הרגש בפוסטים ככל שמתקרבים לארוע מרכזי כלשהו
- ניתוח אושיות רשת בתחום מסוים
- כלי המזהה פייק ניוז בהקשר נושא כלשהו
- זיהוי אלימות בפוסטים/תמונות/ווידאו
- זיהוי מצוקה
- דיפלומטיה בראי הרשתות החברתיות

לפרויקט זה מספר חלקים

1. מציאת רעיון ונתונים מתאימים למחקר (כאמור חייב להיות קשור או מתוך הרשתות החברתיות).
2. ניתוח סטטיסטי והיכרות עם הנתונים
3. שלבי עיבוד מקדים של הנתונים
4. בניית מודל
5. הערכת המודל (חזרה לשלב 3 לניסוי נוסף)
6. סיכום ודיון בתוצאות

יהיה עליכם להציג מסמך מסכם (מומלץ ב Jupyter notebook) המציג את השלבים השונים בבניית המערכת, על פי השלבים שנלמדו בתחום במהלך התואר. יש לתת תשומת לב מיוחדת למדדי ההערכה והביצוע של המודלים שלכם - כולל התייחסות השוואתית למדדים מקובלים בתחום.

שלבים ומסמכי התקדמות :

1. יש להגיש מסמך (מומלץ ב Jupyter notebook) המתאר את תחום העבודה הנבחר, מאגר הנתונים שנבחר לפרויקט ושלבים של עיבוד מקדים שבוצע עליו. תיאור המאגר אמור להכיל: כמות דוגמאות, וניתוח סטטיסטי של הנתונים (לדוגמא בטקסט נרצה לדעת פרטים כמו - כמות מילים במסמך, כמות תווים, מילים נפוצות בשפה, מילים נפוצות בטקסט, תווים מיוחדים וכו') תיאור העיבוד המקדים אמור להכיל את כל השלבים שבוצעו, כלים שהופעלו בכל שלב, ופורמט

הנתונים לאחר העיבוד המקדים (מודל הצגת נתונים - לדוגמא ייצוג ווקטורי).

2. יש לבנות את השלבים הבאים של הפרויקט ולהגיש דו"ח המתאר:
אופי המודל

מודל הייצוג שנבנה על הנתונים שנבחרו (למשל: רלציוני, VSM, גרף וכד')

מודל למידה שנבחר (למשל: SVM, K-MEANS, KNN, NB וכד')

יש לנמק את בחירתכם,

יש לציין את הכלים שהשתמשתם בהם, ולהראות צילומי מסך מהפעלת תכנית שבניתם בדו"ח.

3. יש לבצע את השלבים האחרונים של הפרויקט בתוכנה:

לבצע ניסויים לצורך הערכת איכות המודל שבניתם, לבנות דו"ח סופי (כולל תיאור של הניסויים ותוצאות)

ומצגת קצרה (בסביבות של 10 שקפים) עם תיאור הבעיה, שיטה שנבחרה, תוצאות והדגמת

התכנה. ניתן גם להכין סרטון קצר ל-2 דקות.

בהצלחה !