

קורס פרויקט גמר - הצעת לפרויקט

שם: אירה ויטנזון

מסלול ב: she codes; Data analysis

נושא הרעיון:

מה מאפיין פוסט פופולרי ב Reddit? ניתוח סטטיסטי של פוסטים בפורום מסוים (subreddit) וזיואלחציה של הניתוח ע"י תרשימים שונים כגון היסטוגרמה, תרשים פאי, פיזור ועוד. המידע מורכב מהטקסט של הפוסט, התגובות שהוא קיבל, המשתמש שפרסם אותו, השעה ו metadata נוסף. כשלב מתקדם המתבסס על הניתוח, גזירת מאפיינים של פוסטים ומימוש בפיתוח של מודל המנבא כמה upvotes פוסט קיבל בנקודת זמן מסוימת. השוואה של מדדי דיוק של מודלים שונים לחיזוי ה upvotes והסקת מסקנות לגבי השאלה - מה מאפיין פוסט פופולרי?

מדוע בחרת ברעיון:

עניין אותי מה מאפיין פוסטים שמשומנים כ hot/top, האם יש תנאים מסוימים שעזרו להם להגיע לפופולריות.

זה יכול לעזור למי שרוצה לפרסם פוסט במטרה מסוימת להשיג את התוצאות שהיא/הוא מחפשת/ת

שפת הפרויקט:

שפת פיתוח, ספריות פיתוח המיועדות לניתוח נתונים ובניית מודלים

האם יש ידע מקדים? אם כן, פרטי:

קורסי מבור אקדמיים nlp, machine learning, data mining

קורס Data analysis במסגרת She codes