

## השפעת תכונות מטה-דאטה על זיהוי אירועים בטוויטר

סאמר אעמר

בהנחיית: ד"ר דוד בודוף, ד"ר אוסי מוקרין

טוויטר היא רשת חברתית מובילה אשר מחברת בין מאות מיליונים של משתמשים פעילים מכל העולם. למרות שרוב הדיונים והדיווחים על טוויטר הם על נושאים שאינם מעניינים לחדשות, אנשים גם נוטים לדווח ולעדכן, בזמן אמת, על אירועים חדשותיים (כמו רעידת אדמה, תסונאמי, אירועים מקומיים למיניהם). הרבה עבודות מחקר בוצעו כבר כדי לבחון את היכולת לזהות דיווחים שקשורים לאירועים חדשותיים בטוויטר והוצעו מספר גישות שונות כדי לפתור את הבעיה. חלקן הגדול של מחקרים התייחסו לתכנים הטקסטואליים של הציוצים והציעו אלגוריתמים תוך כדי שימוש ב-NLP, ספירת מלים, מופעים של מילות מפתח שקשורות לאירועים, מופעים של מילות צמודות, ועוד. חלק אחר מהגישות הציעו להתמקד בתכונות מטה-דאטה בלבד שלא קשורות לתוכן של הציוצים. כגון, תכונות גיאוגרפיות, אינטראקציות בין משתמשים, תכונות רשת או תכונות של הפרופילים של משתמשים.

בעבודה זו אנו מעלים את השאלות הבאות: מה מאפיין התנהגות משתמשים ברשת כשדנים באירועים חדשותיים בטוויטר? ואיזה תכונות ייחודיות מאפיינות דיונים אלו בהשוואה לכל דבר אחר? האם יש סוגים שונים של אירועים חדשותיים ואיזה תכונות מאפיינות אותם ומבדילות ביניהם? בעבודה זו אנו נותנים מענה על שאלות אלו על ידי ניתוח של מסד נתונים של 30 מיליון ציוצים אמיתיים. נגדיר מספר של תכונות מטה-דאטה של הציוצים שלא קשורות לטקסט ונשתמש ב machine learning כדי לבחון את יכולתם לזהות ולאפיין אירועים חדשותיים בטוויטר. דבר אשר יעמיק את הבנתנו להתנהגות המשתמשים של טוויטר כאשר הם דנים באירועים חדשותיים.