



אוטומציה של סקר ספרות שיטתי וייצוג האונטולוגיה והסנטימנט כמפת חום - הדגמה בתחום פיתוח תוכנה זריז

איציק דוד
בית הספר למנהל עסקים
אוניברסיטת בר-אילן

רועי גלברד
בית הספר למנהל עסקים
אוניברסיטת בר-אילן

תקציר

סקירת ספרות שיטתית (Systematic Literature Review) הינה אמצעי לזיהוי, הערכה ופרשנות של מגוון המחקרים הרלוונטיים לשאלת מחקר, תחום או תופעה בעלת עניין. כיום, סקירת ספרות שיטתית מבוצעת בעיקר באופן ידני. בהתאם לכך, תהליכי סקירת הספרות דורשים משאבי זמן רבים, מוגבלים בגודלם, ואף חשופים להערכה שגויה ולהטיה. זאת, כאשר מספר הפרסומים וקצב הפרסום הולך וגדל כל העת, כך שהיקף הפרסומים והמאמץ שחוקר נדרש לצורך זה גדלים גם כן. המחקר הנוכחי מהווה ניסיון ראשון במיכון של חלק מתהליכי סקירת הספרות השיטתית באמצעות שילוב של כלים מתחום למידת מכונה (Machine Learning). זאת, באמצעות מתודולוגיה מובנית לאיתור הספרות, חילוץ משפטי הבעת עמדה והערכת הסנטימנט שהם מביעים בהתייחס לרכיבים השונים הנכללים באונטולוגיה של התחום הנחקר. המחקר המוצג מפרט את האופן בו ניתן ליישם למידת מכונה לצורך סקר ספרות שיטתי, וממחיש זאת בתחום של פיתוח תוכנה זריז. כמו כן, המחקר מציע דרך ייחודית להצגת האונטולוגיה של התחום הנחקר, וזאת כמטריצה של רכיבים, דבר המאפשר להציג את הסנטימנט בתמונה מרוכזת כמפת חום. המתודולוגיה המוצעת תורמת גם בהתאמת תחום סקר הספרות השיטתי לדרישות המדע הפתוח, ומאפשרת שחזור הסקירה על ידי עמיתים, היבט שהוא קריטי לאבטחת האובייקטיביות של מחקר מדעי.

מילות מפתח: סקירת ספרות שיטתית, למידת מכונה, ניתוח סנטימנט, אונטולוגיה, מפת חום.

מבוא

במחקרים נוספים, זיהוי נושאי מחקר חדשים, ובחינת השערה מסוימת נתמכת או מופרכת על ידי מידע אמפירי זמין (Gough et al., 2017; Templier & Paré, 2015). כיום, סקירת ספרות שיטתית מבוצעת בעיקר באופן ידני (B. Kitchenham & Charters, 2007; Xiao & Watson, 2019). זאת בעוד מספר הפרסומים וקצב הפרסום הולך וגדל משנה לשנה, ובהתאם לכך המאמץ שחוקרים נדרשים להשקיע במהלך

סקירת ספרות שיטתית (SLR), היא סקירה המקיפה קשת רחבה של פרסומים, במטרה לאתר, להעריך ולאחד את הממצאים הרלוונטיים לתחום הנחקר, במטרה להגיע למסקנה אחידה ואובייקטיבית (Kitchenham, Budgen & Brereton, 2015). סקירת ספרות שיטתית נועדה לתמוך בפעילויות הבאות: סיכום ממצאים קיימים לגבי פרקטיקה או טכנולוגיה, זיהוי פערים במחקר הנוכחי אשר עשויים לסייע בקביעת הצורך

מיכון התהליך תוך שילוב רכיב האונטולוגיה, מאפשר לא רק חסכון משמעותי בזמן והגדלת היקף מקורות הספרות הנסקרים, אלא גם מספק את הדרישה לאובייקטיביות ולאפשרות של שחזור הסקירה על ידי עמיתים, היבט שהוא קריטי כדי להבטיח מדע פתוח ואמין.

ייצוג האונטולוגיה כמטריצה, מאפשר לחוקרים להגדיר באופן שיטתי את מגוון הרכיבים הרלוונטיים לעולם התוכן הנחקר ואת הקשרים בין הרכיבים. ייצוג זה הוא חידוש נוסף לתהליך של סקר ספרות שיטתי, שמאפשר מיקוד בכל אחד מתת המרכיבים בנפרד, תוך אפשרות במקביל להציג תמונה כוללת ואינטגרטיבית. אפשרויות אלו מרחיבות ומעשירות את התוצרים של סקר הספרות, ומוסיפות להם גם את האפשרות להמחשה חזותית באמצעות מפת חום שמקלה את ההצגה והתפיסה של התמונה הכללית על מגוון מרכיביה.

המחקר בנוי משני שלבים: בשלב הראשון הערכת הסנטימנט מבוצעת באופן ידני, ובאופן זה מופק סט נתונים מתויג המהווה את הבסיס לשלב השני (David & Gelbard, 2023). בשלב השני, הערכת הסנטימנט מבוצעת באמצעות למידת מכונה, ורמת הדיוק של הסנטימנט נבדקת אל מול הסנטימנט שנותח באופן ידני. המילונים והאלגוריתמים של למידת המכונה הותאמו למאפיינים של פיתוח תוכנה זריז, כהדגמה לאפשרות התאמתם לכל עולם תוכן באשר הוא.

(et al., 2004). בשנת 2007 פורסם דו"ח המציע קווים מנחים ומקיפים לביצוע סקירת ספרות שיטתית בתחום הנדסת התוכנה.

סקירת ספרות שיטתית כוללת חמישה שלבים (B. Kitchenham & Charters, 2007): הגדרת שאלת המחקר, ביצוע חיפוש לזיהוי מאמרים רלוונטיים, בחירת מאמרים רלוונטיים, הערכת איכות, חילוץ המידע וסינתזה של המידע. תרשים 1 מתאר את תהליך סקירת הספרות והתוצרים שצריכים להתקבל בכל אחד מהשלבים אשר מפורטים בהמשך.

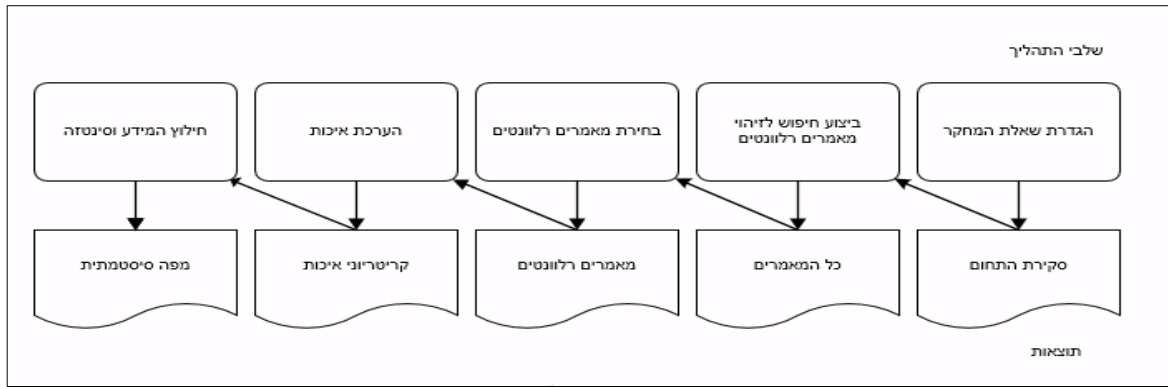
סקירת הספרות (O'Mara-Eves, Alison; Thomas, 2015). תהליכי סקירת הספרות דורשים משאבי זמן רבים (B. Kitchenham & Charters, 2007; Xiao & Watson, 2019) ומתוך שכך, הם מוגבלים בהיקפם ובתדירותם (O'Mara-Eves, Alison; Thomas, 2015). כמו כן, מעצם העובדה שהם מבוצעים באופן ידני, הם חשופים לטעויות בהערכה ולהטיה (Gough et al., 2017; O'Mara-Eves, Alison; Thomas, 2015; Zupic & Čater, 2015). החשיפה לטעויות ולהטיה היא מהותית מאוד לאיכות המחקר (Khan et al., 2001; Templier & Paré, 2015), כיוון שהן יכולות להשפיע על כיוון המחקר, ועל האופן בו החוקר יפרש את הממצאים והמסקנות.

המחקר הנוכחי מציג מיכון של תהליכי סקירת ספרות שיטתית בתחום של פיתוח תוכנה זריז, באמצעות שילוב של כלי למידת מכונה (Machine Learning). כמו כן, המחקר מציע דרך ייחודית להצגת האונטולוגיה, וזאת כמטריצה של משתנים המייצגת את המרכיבים הנכללים בתחום ואת מגוון הקשרים האפשריים ביניהם. משפטי הבעת העמדה המחולצים במסגרת סקר הספרות משויכים לרכיבי האונטולוגיה אליהם הם מתייחסים, והסנטימנט המובע בהם מסוכם אל כל אחד מהתאים הרלוונטיים שבמטריצה. ייצוג האונטולוגיה כמטריצה, בה מסוכמים ערכי הסנטימנטים הרלוונטיים, מאפשר להציג אותה כמפת חום, שבמבט אחד יכולה לספק את תמונת המצב העולה מסקר הספרות השיטתי.

רקע תיאורטי

א. תחום סקירת הספרות השיטתית (SLR)

סקירת ספרות שיטתית פותחה בתחום הרפואה, מטרתה לאחד את כל הראיות הנוגעות לשאלת מחקר בכדי להגיע למסקנה אחידה באמצעות מתודולוגיה מהימנה, קפדנית וניתנת לשחזור (B. Kitchenham et al., 2015). הצלחת סקירת ספרות שיטתית בתחום הרפואה הובילה לאימוץ השיטה בדיסציפלינות רבות (B. A. Kitchenham)



תרשים 1. שלבי תהליך סקירת ספרות שיטתית

שלב 1 - הגדרת שאלת המחקר

הגדרת שאלת המחקר תשקף את מטרות המחקר ואת שיטת המחקר ומכאן השפעתה על תהליך חיפוש המאמרים הרלוונטיים, תהליך חילוץ המידע מתוך המאמרים, ותהליך סינתזת המידע.

שלב 2 - ביצוע חיפוש לזיהוי מאמרים רלוונטיים

זיהוי כמה שיותר מאמרים רלוונטיים המתייחסים לשאלת המחקר, זאת על ידי שימוש באסטרטגית חיפוש שתמנע הטיה ככל האפשר (Xiao & Watson, 2019). ניתן לפתח מחרוזות חיפוש על ידי זיהוי מילות מפתח ומילים נרדפות להן הנמצאות בגוף שאלת המחקר ושילובם על ידי תנאים בוליאניים של או (OR) ו-גם (AND) (B. A. Kitchenham et al., 2015). חיפוש צריך להתבצע תחילה בבסיסי מידע אלקטרוניים, עם זאת נדרש לבצע חיפוש גם במקורות אחרים, כגון: כתבי עת, כתבי עת של חברות, כנסים וחיפוש כללי באינטרנט.

שלב 3 - בחירת מאמרים רלוונטיים

הערכת רלוונטיות המאמרים מבוצעת על בסיס קריטריונים של הכללה (Inclusion) ואי הכללה (Exclusion), המגדירים מאילו מהמאמרים נתעלם, ואילו מהמאמרים מספקים מידע שיתרום לניתוח שאלת המחקר. יש ליצור פילוט שיבדוק שאכן הקריטריונים שנקבעו מהימיניים ויכולים לסווג מאמרים באופן אפקטיבי. ניתן לשלול שימוש במאמרים על סמך כותרת או תקציר.

שלב 4 - הערכת איכות

בנוסף לקריטריוני הכללה/אי הכללה, יש להעריך את איכות המאמרים הראשוניים שנבחרו, כדי:

לבדוק האם הבדלים באיכות המאמרים מספקים הסבר להבדלים בתוצאות המחקר

לאפשר שימוש במשקולות עבור מאמרים ספציפיים בעת שקלול התוצאות

הנחייה לאופן שבו הממצאים יפורשו ויאפשרו לקבוע את עוצמת המסקנות

למרות שאין הגדרה מוסכמת מהו מחקר "איכותי", ישנם קווים מנחים שמקובלים בתחום (Higgins & Green, 2008; Khan et al., 2001) זופיק וקטר מגדירים זאת כתנאים למזעור ההטיה ולמקסום התוקף של המחקר (Zupic & Čater, 2015).

שלב 5 - חילוץ וסינתזה של המידע

מטרת שלב זה היא לתכנן את תהליך חילוץ המידע מהמאמרים שנבחרו כך שהמידע יתועד באופן מדויק כפי שהוצג במחקר. בסוף התהליך תתבצע סינתזה של המידע על ידי סיכום תוצאות המחקרים. הסינתזה יכולה להיות תיאורית, עם זאת לפעמים ניתן להוסיף סיכום כמותי על ידי שימוש בטכניקות סטטיסטיות של מטא אנליזה (Meta-Analysis) (Xiao & Watson, 2019).

ב. כריית טקסט וניתוח עמדה (Sentiment Analysis)

הכנת הטקסט, לרבות תהליך ניקוי הנתונים, (3)
זיהוי הסנטימנט כלומר זיהוי ושמירת משפטים
 סובייקטיביים הכוללים דעות, אמונות והשקפות.
(4) סיווג הסנטימנט, כחיובי, שלילי או ניטרלי ו-
(5) הצגת התוצאות באופן שיאפשר לחוקר לקבל
 מידע משמעותי באופן נוח ובהיר.

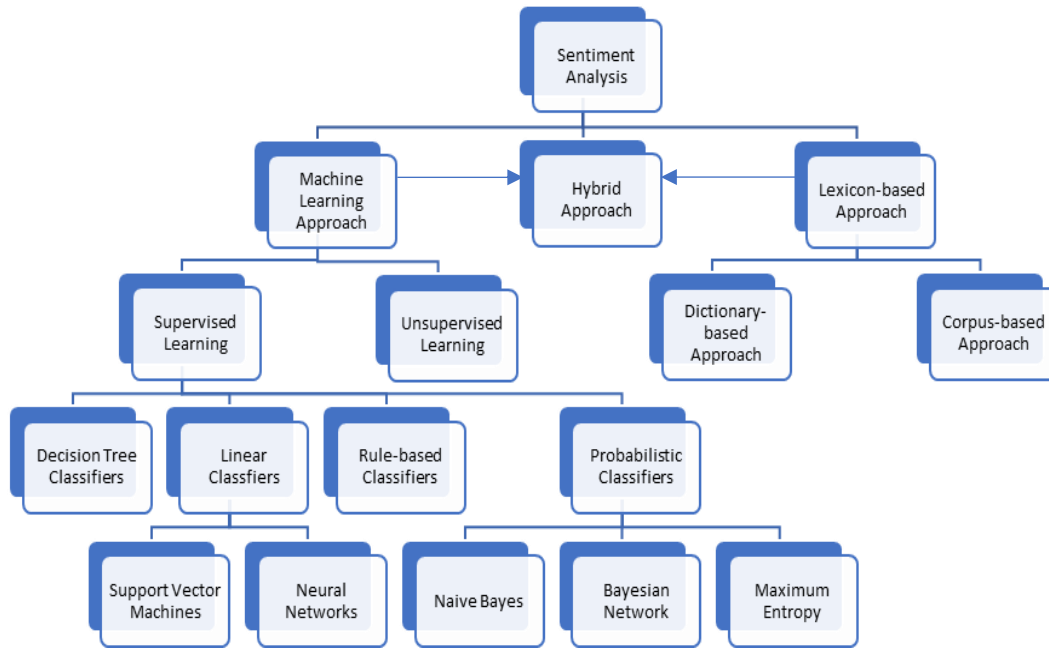
הספרות מגדירה שלוש רמות של סיווג סנטימנט
 (Devika et al., 2016; Medhat et al., 2014).
 (א') סיווג ברמת המסמך – מסווג את חוות הדעת
 המבוטאת במסמך כחיובית או שלילית המבוטאת
 במסמך. היא מתייחסת למסמך כולו כיחידת מידע
 בסיסית הדנה על נושא אחד. (ב') סיווג ברמת
 משפט – מסווג את הסנטימנט המבוטא בכל אחד
 מהמשפטים שזוהו כסובייקטיביים כשלילי, חיובי,
 ניטרלי. (ג') סיווג ברמת היבט - מסווג את
 הסנטימנט ביחס להיבטים ספציפיים של הישות.
 שכן, משתמש יכול לתת דעות שונות על מספר רב
 של היבטים עבור אותה ישות.

סיווג הסנטימנט מבוצע במגוון טכניקות, תרשים 3
 מציג חלוקה לשלוש גישות שונות: גישת למידת
 מכונה (Machine Learning Approach), גישה
 מבוססת לקסיקון (Lexicon-Based Approach),
 וגישה היברידית (Medhat et al., 2014; Pradhan
 et al., 2016).

טכניקות של כריית טקסט מאפשרות לאחזר באופן
 אוטומטי ידע ממאגרי מידע המכילים נתונים
 טקסטואליים לא מובנים בקנה מידה גדול, ובאופן
 שניתן לשחזר אותו שוב ושוב (Debertoli et al.,
 2016). כריית טקסט מבוצע במגוון רחב של
 טכניקות, החל מספירה כמותית של מופעי מילים,
 ועד להבנה של התכנים הנידונים בטקסט
 (Kwartler, 2017; Silge & Robinson, 2017).
 טכניקות אלו מאפשרות לחוקרים להתגבר על
 המגבלות הקיימות בניתוח של טקסט רחב היקף
 ולהניב תובנות שאחרת לא היה ניתן להסיק
 (Debertoli et al., 2016). המחקר בנושא עיבוד
 שפה טבעית (NLP) החל משנת 1950, והתמקד
 בנושאים של תרגום, אחזור מידע, סיכום, מענה על
 שאלות, זיהוי וסיווג נושאים (Topic Modeling),
 וניתוח רגש-עמדה (Sentiment Analysis).

ניתוח סנטימנט עוסק בניתוח דעות, הערכות,
 רגשות, עמדות של אנשים כלפי אובייקטים שונים
 (כגון: מוצרים, ארגונים, אנשים, אירועים),
 וסיווגם כחיובי/שלילי/ או ניטרלי (Jongeling et
 al., 2015). ניתוח הסנטימנט נפוץ מאוד בייחוד
 ברשתות חברתיות, אך יש לו השפעה רבה גם על
 תחומים אחרים, כגון: מדעי הניהול, מדעי
 המדינה, מדעי החברה, שכן הם מושפעים
 מדעותיהם של אנשים (Pang & Lee, 2008).

תהליך ניתוח סנטימנט כולל חמישה שלבים
 (D'Andrea et al., 2015): **(1) איסוף המידע, (2)**



תרשים 2. טכניקות לסיווג סנטימנט

(Liu et al., 2012), (al., 2017) ביקורות למסעדות (Pham & Le, 2018), וביקורת של מוצרי צריכה (Jeyapriya & Kanimozhi, 2015; Severyn et al., 2016).

בניגוד ללמידה המונחית, למידה בלתי מונחית (Unsupervised learning) אינה זקוקה לסט אימון וניתן להשתמש בה לכל מידע ללא צורך בנתונים מסווגים מראש (Medhat et al., 2014). היא מאפשרת לאתר דפוסים וקשרים בטקסט ומיישמת את הידע על נתונים חדשים, על ידי שימוש בשתי שיטות עקרוניות: Clustering ו-Topic Modeling (Kwartler, 2017). זה וקופראפו (De & Koppurapu, 2013) השתמשו ב-clustering על מנת לקבץ רעיונות שהועלו בפורטל Ideas. האנג ושות' (Huang et al., 2017) ביצעו ניתוח topic modeling באמצעות שיטת LDA עבור הודעות שפורסמו בבלטפורמה לפרסום תכנים קצרים (microblogging). פו ושות' (Phu et al., 2017) ניתחו בטכניקה הזו ביקורות גולשים לסרטים שפורסמו ב-Facebook.

שימוש בלמידת מכונה אינו דורש מילון ייעודי, כפי שנדרש בטכניקת Lexicon-Based Approach,

בלמידת המכונה יש שתי גישות חלופיות: גישה מונחית וגישה לא מונחית. למידה מונחית (Supervised learning) מתייחסת לסיווג שהוגדר מראש בסט נתונים שמשמש לאימון המכונה (Training set). יתרון השיטה הוא ביכולת להתאים וליצור מודלים והקשרים ספציפיים. מנגד, החיסרון העיקרי הוא הקושי בסיווג מידע עם מאפיינים שונים מאלו ששימשו לאימון (Medhat et al., 2014). השיטות הפופולאריות ביותר ללמידה מונחית הן: Naïve Bayes (NB), Bayesian Networks (BN), Maximum Entropy (ME), Stochastic Gradient Descent (SGM), Support Vector Machine (SVM) ו-Neural Network (NN). טריפתי ושות' (Tripathy et al., 2016) בחנו תוצאות של ניתוח סנטימנט עבור האלגוריתמים SGD, ME, SVM ו-NB, באמצעות חילוץ מאפייני N-gram מביקורות של סרטים ממאגר IMDb. בעוד שהאלגוריתמים השונים סיפקו תוצאות דיי דומות, הם מצאו שככל שה-N עבור שיטת n-gram גדול יותר, כך הדיוק של האלגוריתמים קטן יותר. תוצאות דומות התקבלו במחקרים בתחומים אחרים, כגון ניתוח תכנים ב-Twitter (Vilares et al., 2016).

למצוא מילים עם אוריינטציה לתחום ולהקשר ספציפי על ידי איתור תבניות תחביריות המופיעות ברשימה של מילים המשמשות כזרעים (seeds) שנבדקים על קורפוס רחב תוך שימוש ביחסים כגון: "and", "or", "but", "either", המאפשרים לזהות מילים נוספות עם אותה אוריינטציה. לדוגמא, במשפט "this book is good and attractive", מכיוון שהמילה "good" בעלת משמעות חיובית השימוש באילוץ התחבירי "and" בעצם נותן את אותה האוריינטציה גם למילה "attractive". באופן זה, לטכניקת הקורפוס יש יתרון באיתור מילים עם אוריינטציה לתחום או הקשר ספציפיים (Hemmatian & Sohrabi, 2019; Medhat et al., 2014).

הטכניקה ההיברידי (Hybrid Approach) משלבת למידת מכונה בגישות המילונים. טבלה-1 מציגה סיכום של מאפייני השיטות שתוארו בפרק זה.

ומאפשר יכולות דיוק גבוהות גם על טקסט מתחומים שלא נכללו בסט האימון (Devika et al., 2016). בניגוד לכך, גישת Lexicon-Based מתבססת על מילון המפרט את מילות הרגש-עמדה, כגון: "טוב", "יפה", "מדעים", "רע", "מפחיד" וכד'. קוטביות (polarity) של מילה מתארת את התחושה וההערכה שמילה מסוימת מבטאת (Hemmatian & Sohrabi, 2019). הקוטביות של מילה יכולה להיות משוקללת כמספר המבטא יחס או הסתברות של רמת הקוטביות החיובית ביחס לשלילית (Hemmatian & Sohrabi, 2019). את המילון ניתן ליצור באופן ידני, משימה שדורשת זמן רב, או באופן אוטומטי, על ידי שימוש במספר מילים כזרעים (seeds) העוברים תהליך של הרחבה.

ניתן לסווג את טכניקות המילונים לשתי שיטות עיקריות dictionary-based approach ו-corpora-based approach (Hemmatian & Sohrabi, 2019; Medhat et al., 2014).

Dictionary-Based Approach מבוססת על זיהוי ידני של מערך מילים והגדרת הקוטביות שלהם. לאחר מכן, האלגוריתם מרחיב את המערך על ידי חיפוש מילים נרדפות או מילים בעלות משמעות הפוכה בטזאורוסים ומילונים מקוונים (Mohammad et al., 2009), אך מגבלת השיטה נעוצה בכך שאינה יכולה למצוא מילים עם אוריינטציה ייחודית לתחום ספציפי. לדוגמא, מילון שמשמש לסיווג ביקורות של סרטים או מוצרים לא יתאים לסווג טקסט הקשור להנדסת תוכנה ועלול לשגות בסיווג הסנטימנט (Jongeling et al., 2015). ראתין ושות' (Rathan et al., 2018) השתמשו בשיטה זו לזיהוי וסיווג סנטימנט, על בסיס ציורים ואימוזים בטוויטר, עבור תכונות של טלפונים סלולריים. בעבודתם הם מציגים את החשיבות בתיקון טעויות והתאמת המילונים בעבודה על טקסט מטוויטר ביחס לטקסט עיתונאי או מקצועי.

לעומת זאת, טכניקת ה-Corpora-based approach מתגברת על חיסרון זה ומאפשרת

חסרונות	יתרונות	דיוק	יעילות	טכניקות
למידת מכונה (Machine learning)				
יכולת לנתח מספר רב של קטגוריות תלוי באופן משמעותי בסט האימון	אפקטיביות בגילוי סובייקטיביות	גבוה מאוד	איטי	למידה מפקחת Supervised
דורש מומחה בתחום עם ידע לשוני בעל עלויות גבוהות	עמיד בפני "רעש" שקיים במידע			
זמן לאימון האלגוריתם לוקח זמן, בייחוד כאשר מדובר במידע בעל קטגוריות מרובות	לא נדרש התערבות של מומחה	בינוני	מהיר	למידה לא מפקחת Unsupervised
היכולת שלו לנתח מספר קטגוריות עדיין טרם הוכח	יעיל וישים על תחומים שונים			
Lexicon-based Approach				
לא ניתן לסווג מילים בתחום תוכן שאינו נמצא במילון	אינו דורש סט אימון	נמוך	מהיר מאוד	Dictionary-based
גישה נוחה וקלה למילים במילון ולאוריינטציה שלהם	קיים קושי עבור טקסטים עם תלות סמנטית מספק תוצאות טובות לתחומי תוכן פחות נפוצים			
בעל יכולת למצוא מילים עם הביצועים משתנים בהתאם להיקף של אוריינטציה לתחום תוכן ספציפי	מספק תוצאות טובות יותר כאשר תחומי התוכן שונים	נמוך	מהיר מאוד	Corpus-based
קיים קושי לכסות את כל המילים כאשר מדובר בתוכן טקסטואלי גדול	חייב טכניקה משלימה, לא ניתן להשתמש רק בו			

טבלה 1. השוואה בין הטכניקות לסיווג סנטימנט

ג. למידת מכונה בסקירת ספרות שיטתית

את אסטרטגיית החיפוש, הם לא סיפקו ראיות אמפיריות אשר תומכות במסקנתם. וואלס ושותי (Wallace et al., 2010) צמצמו כמעט מחצית מהמחקרים שדרשו בדיקה ידנית באמצעות שימוש בגישה היברידית אשר שילבה מסווג SVM לסיווג מחקרים בתחום הביורפואי. ניסוי (Felizardo et al., 2011) שהשווה בין הטכניקה הידנית ולמידת מכונה מצא שהקבוצה שהשתמשה בלמידת מכונה סיננה מאמרים באופן מהיר יותר תוך שרמת הדיוק בבחירת מאמרי היסוד (Primary Studies) הייתה זהה.

סקירת ספרות שיטתית אודות שימוש בלמידת מכונה לזיהוי Primary Studies, סקרה 44 מאמרי סקירה שפורסמו בין השנים 2006 ל 2014, טענה שעדיין קשה להסיק מסקנות נחרצות אודות היעילות של למידת המכונה, אך ייתכן שזה נבע מתחומי המחקר המגוונים שנכללו בסקירה (O'Mara-Eves, Alison; Thomas, 2015). בניגוד

נכון להיום יישום פרקטיקות של למידת מכונה במסגרת סקירת ספרות שיטתית בוצע בעיקר בתחום הרפואה והתמקד בשלב בחירת מאמרים רלוונטיים לסקירה. החלוצים בתחום (Aphinyanaphongs et al., 2005) השתמשו

בטכניקות סיווג מסוג naïve Bayes ו-Support Vector Machines (SVM) לצורך בחירת מאמרים רלוונטיים לסקירה במאגר ACP journal club. הם השוו את התוצאות שהתקבלו באמצעות טכניקות אלו לבין התוצאות שהתקבלו באמצעות clinical query filter שהינה גישה מסורתית לבחירת מאמרים בתחום הרפואה. הם סיכמו שהשימוש בטכניקות למידת מכונה אפקטיבי יותר מהניתוח הידני. באופן דומה, אנאנידו ושותפיו, (Ananiadou et al., 2009) השתמשו במסווג מסוג SVM לבחירת מאמרים רלוונטיים בתחום מדעי החברה. למרות שהכותבים טענו שטכניקות למידת מכונה משפרות

בתהליך בחירת מאמרים רלוונטיים לסקירה. בנוסף אחד המחקרים מצא שהמסווגים Naïve Bayes ו-SVM עולים ביעילותם על טכניקת KNN. בנוסף המסווג Naïve Bayes הפיק תוצאות עם השיעור הנמוך ביותר של טעויות (False Positive). מחקר נוסף בנושא, שנעשה לאחרונה, תומך גם כן בממצאים חיוביים אלו (Popoff et al., 2020).

הסנטימנט נבדקת אל מול הסנטימנט שנתח באופן ידני.

כמו כן, המחקר מציע דרך ייחודית להצגת האונטולוגיה, וזאת כמטריצה של משתנים המייצגת את המרכיבים הנכללים בתחום ואת מגוון הקשרים האפשריים ביניהם. משפטי הבעת העמדה המחולצים במסגרת סקר הספרות משויכים לרכיבי האונטולוגיה אליהם הם מתייחסים, והסנטימנט המובע בהם מסוכם אל כל אחד מהתאים הרלוונטיים שבמטריצה.

איתור וסינון המאמרים בוצעו באמצעות שאילתות SQL. מאמרי היסוד עברו תהליך של ניתוח ידני והמידע אודותם, כולל משפטי הבעת העמדה, שיוכם לרכיבי האונטולוגיה, וסיווגם כשליליים או חיוביים, נשמר בבסיס נתונים ייעודי ששימש את השלב השני של המחקר בו תהליך הניתוח בוצע באמצעות טכניקות של למידת מכונה.

למסקנותיהם, שני מחקרים אמפיריים נוספים (Bekhuis & Demner-Fushman, 2012; García Adeva et al., 2014) שהתמקדו בתחום הרפואה, ערכו השוואה בין מסווגי סנטימנט שונים, k-Nearest Neighbors (KNN), naïve Bayes ו-SVM. שני המחקרים האלו הגיעו למסקנה שלמידת מכונה יכולה לתמוך באופן טוב ויעיל

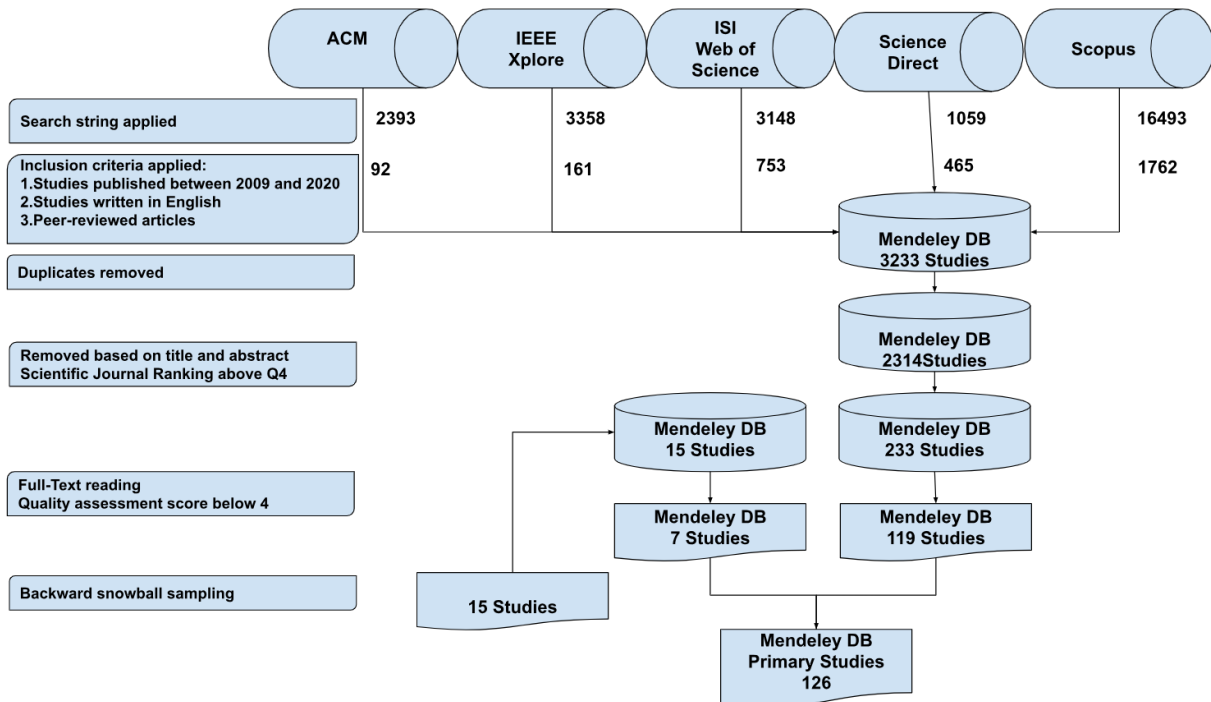
שיטת המחקר

המחקר הנוכחי מציג אוטומציה של סקר ספרות שיטתי וייצוג האונטולוגיה והסנטימנט כמפת חום, תוך הדגמת הדבר בתחום פיתוח תוכנה זריז. לפיכך, המחקר בנוי משני שלבים: בשלב הראשון הערכת הסנטימנט מבוצעת באופן ידני, ובאופן זה מופק סט נתונים מתויג המהווה את הבסיס לשלב השני, במסגרתו הערכת הסנטימנט מבוצע באמצעות למידת מכונה, כשהמילונים והאלגוריתמים ללמידת המכונה עוברים התאמה למאפיינים של פיתוח תוכנה זריז, ורמת הדיוק של

א. SLR - איתור מחקרי יסוד בתחום

פיתוח תוכנה זריז

המחקר הנוכחי מתמקד בשאלת ההתאמה של מתודולוגית פיתוח זריז להיבטים שונים של פרויקט פיתוח תוכנה. תרשים 3 מציג את תהליך סקירת הספרות השיטתית שנעשתה בתחום פיתוח תוכנה זריז. התהליך החל באיסוף של 264,451 מאמרים, שאותרו במאגרי המידע המובילים בתחום, ובהפעלת מנגנוני סינון שהוליכו לצמצום האוסף לכדי סט של 26 מאמרי יסוד (Primary studies).



תרשים 3. תהליך איסוף וסינון המאמרים במסגרת ה-SLR

ב. ייצוג האונטולוגיה כמטריצה דו-ממדית

מרכיבי עולם התוכן של פיתוח תוכנה זרזו והקשרים ביניהם, הוגדרו כמטריצה דו-ממדית המשלבת בציר ה-X את תחומי הידע בניהול פרויקטים (knowledge areas) כפי שהם מוגדרים ב-PMBOK (Project Management Body of Knowledge), ובציר ה-Y מציגה את שלבי מחזור החיים של מערכת ממוחשבת (SDLC Software Development Life Cycle).

ג. ניתוח ידני של הסנטימנט

המשפטים שאותרו ב-126 מאמרי היסוד, והסנטימנט המבוטא בהם שויכו לתאים הרלוונטיים במטריצה, כאשר כל משפט יכול להתייחס למספר תאים. ערכי הסנטימנט המתייחסים לכל תא סוכמו לערך מספרי, והמטריצה המספרית שהתקבלה תורגמה למפת החום (כמוצג בתרשים 4). התרגום המספרי בוצע באופן הבא:

- S מתאר את המאמרים שנבחרו כרלוונטיים לסקירת הספרות (בדרך כלל נקראים

(Development Life Cycle), כפי שמקובל בתחום הנדסת התוכנה. באופן זה התקבלה מטריצה שמכילה 16 עמודות של תחומי ידע ו-5 שורות של שלבים במחזור החיים, המסתכמים ב-80 "תאים" שכל אחד מהם מייצג שילוב אפשרי של עולם ידע בהקשר לשלב ספציפי במחזור החיים. ייצוג זה של האונטולוגיה מאפשר הצגה יעילה וקומפקטית הן של הרכיבים השונים והן של הקשרים שביניהם. גודלה של המטריצה ממחיש את מורכבותו של התחום ומורכבות הניתוח שיידרש כדי לתאר את הסנטימנט ביחס לכל אחד מהתאים במטריצה.

$$S = \{S1, S2, \dots, \text{כלומר } ("Primary Studies" S126)\}$$

- P מתאר את השלבים במחזור חיים של מערכת אשר נכללו באונטולוגיה שהוגדרה, כלומר P = {Architecture and Design, Development, Quality Assurance, Maintenance, SDLC General}
- C מתאר מאפייני פרויקט שנגזרו מה-PMBOK ונכללו באונטולוגיה שהוגדרה, כלומר C = {Quality, Complexity, Requirements, Schedule, Budget, Risk, Productivity, Communication, Distribution, Decision-

- שהוגדרו באונטולוגיה התוצאה כדלקמן:
 $POL(s,p,c) = \sum \text{Sentiment Score}$
2. נרמול ערכי $POL(s,p,c)$ לערכים 1,0 ו -1 על בסיס התנאים הבאים:
 - i. עבור $POL(s,p,c) > 0$ הגדר $POL(s,p,c) = 1$
 - ii. עבור $POL(s,p,c) < 0$ הגדר $POL(s,p,c) = -1$
 - iii. עבור $POL(s,p,c) = 0$ הגדר $POL(s,p,c) = 0$
 3. חישוב אגרגטיבי של ציון הסנטימנט עבור כל תא באופן הבא:

$$\text{Sum of Sentiment Score } (p,c) = \sum_{s=1}^{126} POL(s, p, c)$$
 5. עבור כל אחד מהתאים (p,c) הוגדר הערך CS (Cell Score) באופן הבא:
 6. אם $\text{Number of studies } (p,c) < 5$ הגדר כ 'Lack of evidence' אחרת חשב את הערך CS (Cell Score) באופן הבא:

$$CS = \frac{\text{Sum of Sentiment Score}(p,c)}{\text{Number of studies}(p,c)}$$
 8. עבור כל $CS(p,c)$ הוגדר סיווג וצבע כפי שמוצג בטבלה 2.

- making, Low-Skills, High-Skills, Small-Medium Scale, Large Scale}.
- (P,C) מגדיר מיקום של תא באונטולוגיה שהוגדרה
 - $POL(s,p,c)$ מגדיר את סכום תוצאת הסנטימנטים עבור מאמר S ותא (p,c) .
- סיכום אגרגטיבי של מספר המאמרים הנבדלים עבור כל תא במטריצה חושב באופן הבא:
- $$\text{Number of studies } (p,c) = |\{S1,S2,\dots,S126\}|$$
- ציון הסנטימנט עבור כל תא במטריצה היה חשוף להטיה, מכיוון שכל אחד מהמחקרים שזוהה כרלוונטי לסקירה (S) מציג מספר שונה של סנטימנטים, כך שמחקרים עם מספר גבוה יותר של משפטי הבעת עמדה עלולים היו להשפיע יותר על הציון הסופי. לפיכך, הוגדר תהליך שמטרתו להבטיח שכל אחד מהמחקרים שנכללו בסקירה (S) ישפיעו באופן שווה על הציון הסופי:
1. חישוב סכום ציון הסנטימנט עבור כל אחד מ-126 המאמרים ב $\{S\}$ לכל אחד מהתאים

#	Condition	Classification type	Heat Map color
1	$\text{Number of studies}_{(p,c)} \leq 5$	Lack of evidence (L)	Orange
2	$-1 \leq CS < -0.6$	Strong negative consensus (SN)	Red
3	$-0.6 \leq CS < -0.3$	Weak negative consensus (WN)	Pink
4	$-0.3 \leq CS \leq 0.3$	Disagreement (D)	Yellow
5	$0.3 < CS \leq 0.6$	Weak positive consensus (WP)	Light Green
6	$0.6 < CS \leq 1$	Strong positive consensus (SP)	Dark Green

טבלה 2. אופן סיווג התא והגדרת צבע בהתאמה לערך ה CS

Life Cycle Phases (SDLC)	Knowledge Areas (PMBOK)															
	Quality	Budget	Schedule	Requirements	Productivity	Decision making	Risk	Communication	Distribution	Complexity		Scale		Skills		General
										Low	High	Small-Medium	Large	Low	High	
Architecture & Design	WN	D	WN	WN	WN	WN	L	D	SN	L	SN	L	SN	SN	L	SN
Development	WP	L	SN	SP	D	L	L	WP	L	L	L	L	SN	SN	SP	L
Quality Assurance	WP	L	SN	D	D	L	L	WP	L	L	L	L	WN	L	L	L
Maintenance	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
SDLC General	SP	D	D	D	WP	WN	WN	WP	WN	L	WN	L	SN	SN	SP	L

SP Strong Positive consensus
 SN Strong Negative consensus
 D Disagreement
WP Weak Positive consensus
 WN Weak Negative consensus
 L Lack of evidence

תרשים 4. מפת החום של הסנטימנט אודות התאמת הגישה הזריזה לפיתוח תוכנה

פסקה שאוחזרו עבור שישה תחומי ידע ספציפיים מה PMBOK. הלקסיקון עודכן תוך כדי בחינת התוצאות לצורך התאמת הלקסיקון לתחום הפיתוח הזריז, לרבות הוספה והסרה של מונחים, שינוי עוצמת הקוטביות, שינוי סוג valance shifter (לדוגמה, negators, amplifiers, de-amplifiers, and adversative conjunctions). הקוטביות עבור כל אחד מהסנטימנטים נורמלה לערכים של 1 ו-10. המתארים סנטימנט חיובי או שלילי בהתאמה. חישוב אגרטיבי של ציון הסנטימנט עבור כל אחד מתחומי הידע בוצע בדומה לניתוח הסנטימנט הידני המוצג בסעיף 3.

תוצאות המחקר

כאמור, המידע הכולל את משפטי הבעת העמדה, שיוכם לרכיבי האונטולוגיה, וסיווגם כשליליים או חיוביים, נשמר בבסיס נתונים ייעודי ושימש כסט נתונים מתויג לצורך הערכת יכולת הביצוע של ניתוח הסנטימנט באמצעות למידת מכונה. בשלב ראשון ניתוח הסנטימנט הוגבל לשישה תחומי ידע ספציפיים מאלו המוגדרים ב-PMBOK, להבדיל מהגדרת האונטולוגיה כמטריצה דו-ממדית בעלת 80 תאים, במטרה לפשט את פעולת ההתאמה של המילונים ושל האלגוריתמים. טבלה 3 מציגה את שיוך המאמרים לששת תחומי הידע ואת חישוב הסנטימנט המצטבר.

ד. ניתוח הסנטימנט באמצעות למידת מכונה

תהליך ניתוח הסנטימנט בוצע באמצעות פלטפורמת R וגובש על בסיס ההנחיות ודוגמאות של מקרה-בוחר (use-cases) שהוצגו בספר "Mining in Practice With R" (Kwartler, 2017). לצורך סיווג הסנטימנט כשלילי או חיובי נבחרה חבילת תוכנה עבור פלטפורמת R בשם "SentimentR", מסווג מבוסס לקסיקון, עבור כל אחד מהמונחים שנמצא בלקסיקון מוגדר מספר רציף בין 5- ל-5 אשר מתאר עד כמה הסנטימנט שלילי או חיובי ומכונה בספרות בשם "קוטביות" ומכאן והלאה נשתמש במונח הזה. "SentimentR" מאפשרת לחשב את הקוטביות עבור משפט יחיד או פסקה ואינו מצריך מידע לאימון.

"SentimentR" נבחרה לנוכח המלצתו של נלדי (Naldi, 2019) שבחר מאפיינים, יתרונות, חסרונות של מגוון חבילות תוכנה לניתוח סנטימנט באמצעות פלטפורמת R ביניהם: SentimentR, sentiment analysis ו-syuzhet, Rsentiment. בסיס הבחינה הומלץ להשתמש בחבילה "SentimetR" מכיוון שנותנת את המענה הטוב ביותר למילים הופכיות (negators) (Naldi, 2019).

בדומה לניתוח הסנטימנט באופן ידני שתואר בסעיף הקודם, הקוטביות חושבה עבור משפט או

PMBOK knowledge areas	Number of mapped PSs	Percentage of mapped PSs	Aggregated sentiment scores
(1) Scope	49	21%	-14
(2) Schedule	33	14%	-9
(3) Cost	25	11%	-2
(4) Quality	53	22%	8
(5) Communication	72	30%	7
(6) Risk	5	2%	-3
Sum	237		

טבלה 3. המאמרים העיקריים שנבחנו בכל אחד מתחומי הידע והסנטימנט האגרסיבי

לאור הצלחת מקרה הבוחן הני"ל, הורחבו התאמות המילונים והאלגוריתמים במטרה לנסות ולשחזר את מפת החום שהתקבלה בניתוח הידני שהתייחס לכל מרכיבי האונטולוגיה (80 תאים). תרשים 5

מציג את מפת החום שהתקבלה בהפעלת האלגוריתמים לניתוח הסנטימנט, ניתן לראות את הדמיון הרב אל התוצאות שהתקבלו בניתוח ידני (תרשים 4).

Life Cycle Phases (SDLC)	Knowledge Areas (PMBOK)															
	Quality	Budget	Schedule	Requirements	Productivity	Decision making	Risk	Communication	Distribution	Complexity		Scale		Skills		General
										Low	High	Small-Medium	Large	Low	High	
Architecture & Design	WN	D	WN	WN	D	WN	L	D	WN	L	WN	L	SN	SN	L	WN
Development	D	L	WN	SP	D	L	L	WP	L	L	L	L	SN	SN	SP	L
Quality Assurance	WP	L	SP	D	D	L	L	WP	L	L	L	L	WN	L	L	L
Maintenance	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
SDLC General	SP	D	D	D	D	D	WN	D	WN	L	WN	L	WN	WN	SP	L

SP Strong Positive consensus
 SN Strong Negative consensus
 D Disagreement
WP Weak Positive consensus
 WN Weak Negative consensus
 L Lack of evidence

תרשים 5. מפת החום של הסנטימנט כפי שהתקבלה באמצעות למידת מכונה

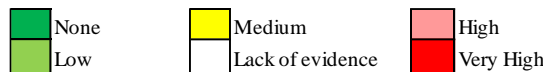
טבלה 4 מציגה Confusion Matrix של התאמת סנטימנט המכונה לסנטימנט הידני, ואת ערכי ה- Recall וה- Precision.

		Predicted			
		Positive	Negative	Total	Recall
Actual	Positive	457	58	515	88.74%
	Negative	69	357	426	83.80%
	Total	526	415	941	86.50%
	Precision	86.88%	86.02%	86.50%	-

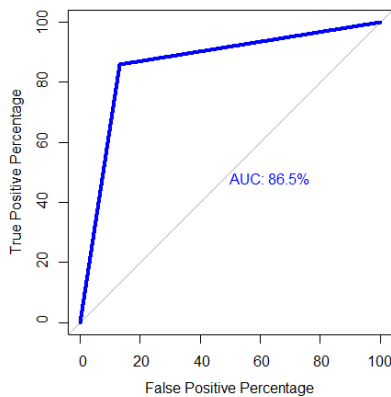
טבלה 4. Confusion Matrix של סיווג הסנטימנט ע"י המכונה

תרשים 6 ממחיש את מידת הדימיון-מרחק בין התוצאות שהתקבלו בניתוח הידני (תרשים 4) ובין אלו שהתקבלו בניתוח הממוחשב (תרשים 5), תוך הדגשה ברקע אדום של התא היחיד בו התקבלה הערכה שונה באופן מהותי.

Life Cycle Phases (SDLC)	Knowledge Areas (PMBOK)															
	Quality	Budget	Schedule	Requirements	Productivity	Decision making	Risk	Communication	Distribution	Complexity		Scale		Skills		General
										Low	High	Small-Medium	Large	Low	High	
Architecture & Design	0	0	0	0.06	0.25	0	L	0.15	0.25	L	0.3	L	0.07	0.1	L	0.4
Development	0.31	L	0.2	0	0	L	L	0.26	L	L	L	L	0	0.29	0.38	L
Quality Assurance	0.06	L	1.60	0.29	0.06	L	L	0	L	L	L	L	0	L	L	L
Maintenance	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
SDLC General	0	0.17	0.05	0	0	0.25	0.29	0.04	0.12	L	0	L	0.2	0.07	0.19	L



תרשים 6. דימיון-מרחק בין התוצאות שהתקבלו בניתוח הידני ובניתוח הממוחשב



תרשים 7. גרף ה-AUC של המסווג

ניתן לראות כי רק בתא אחד, המייצג שילוב של הרכיבים Quality assurance ו-Schedule, הסנטימנט של הניתוח הממוחשב שונה מהותית מהסנטימנט הידני. חשוב לציין שניתוח הסנטימנט בתא הנ"ל התבסס על מספר מצומצם של מאמרים (חמישה) דבר המעלה את הצורך להבטיח מספר מינימלי של משפטי הערכה כדי שניתן יהיה לקבל תוצאה אמינה בהערכתם על ידי המכונה.

תרשים 7 מציג את ערך ה-AUC של המסווג. כמו מדדי ה-Precision וה-Recall, גם ערך ה-AUC שהתקבל משקף סיווג באיכות טובה.

סיכום ומסקנות

המחקר הנוכחי מציג תרומה בשני תחומים. האחד הוא תחום פרויקטי פיתוח תוכנה, והשני הוא במיכון של תהליכי סקירת ספרות שיטתית והתאמת מילונים ואלגוריתמים של למידת מכונה לצורך הערכת סנטימנט בעולם תוכן ספציפי באשר הוא.

בתחום פיתוח פרויקטי תוכנה המחקר מציג סקר ספרות שיטתי הבודח את ההתאמה של מתודולוגית הפיתוח הזריז למאפיינים השונים של פרויקט פיתוח תוכנה. המחקר מחדש הן באופן בו הוא מציג את האונטולוגיה הרלוונטית לתחום – כמטריצה דו-ממדית, המשלבת בציר ה-X את תחומי הידע בניהול פרויקטים (knowledge areas) כפי שהם מוגדרים ב-PMBOK, ובציר ה-Y מציגה את שלבי מחזור החיים של מערכת

את הדרישה לאובייקטיביות ולאפשרות שחזור הסקירה על ידי עמיתים, ולהבטחת מדע פתוח.

בנקודה הזו הקורא יכול לשאול האם מה שבוצע במסגרת מחקר זה יכול היה להתבצע באמצעות ChatGPT? והתשובה לשאלה זו היא שלילית לחלוטין. ל-ChatGPT יש יכולת גבוהה מאוד לתאר ולנסח דברים, אך אין לו שתי יכולות עקרוניות: אין לו את היכולת לבסס את התיאורים על מקורות ספציפיים, כל שכן על משפטים ספציפיים בתוך אותם מקורות, ואם נתעקש אזי אחת משתיים, או שהוא יודה בכך שאין לו את היכולת הזו, או שהוא ימציא-ידמיון מקורות סתמיים ויגיד שהם הבסיס לדבריו, מה שמכונה הלוצינציה. היכולת הנוספת שאין ל-ChatGPT, שנובעת מהמגבלה הקודמת, זו היכולת ליצור אגרגציה כמותית של מקורות. כלומר להציג דעה מגוונת במתכונת כגון: עמדתם של א ו-ב היא X, ואילו עמדתם של ג ו-ד היא Y.

בניגוד למגבלות אלו, המחקר הנוכחי מציג תמונה המאפשרת עקיבות ושקיפות מלאה אודות כל אחד מהשלבים הנכללים בו, ובכל שלב להצביע על מילה / משפט עמדה / המקור הספרותי מהם נובע כל ערך וכל הערכה.

למרות מגבלה זו של ChatGPT הרי שתחום זה מתפתח בימים אלו באופן מואץ, ואנו נערכים למחקר המשך שישלב את יכולותיו של ChatGPT במסגרת של תהליכי הגדרת האונטולוגיה, וזיהוי והערכת הסנטימנטים, באופן שינסה להבטיח אמינות על ידי מניעה של הלוצינציות.

ממוחשבת, כפי שמקובל בתחום הנדסת התוכנה. סקר הספרות השיטתי מציג לא רק את התחומים בפיתוח תוכנה בהם הספרות מציינת שהמתודולוגיה הזריזה מתאימה ותורמת להם, אלא גם את התחומים בהם היא אינה מתאימה, וכן את התחומים בהם אין מספיק עדויות בספרות וראוי שהקהילה בתחום תחקור ותבחן.

בנוסף, המחקר מציג דרך חזותית חדשה להצגת התמונה העולה מסקר הספרות השיטתי, וזה באמצעות מפת חום המציגה את האגרגציה הכמותית של הסנטימנט העולה מהסיכום של כלל המקורות, תוך אפשרות להתייחס הן אל התמונה הכללית המתייחסת אל התחום כולו, והן אל התמונה הפרטנית בכל אחד ממרכיבי האונטולוגיה השונים.

בתחום של מיכון תהליכי סקירת ספרות שיטתית והתאמת מילונים ואלגוריתמים של למידת מכונה לצורך הערכת סנטימנט בתחום ידע מסוים, בהקשר זה המחקר הציג את תוצאות סקר הספרות הידני, בתחום פיתוח התוכנה הזריז, כמקרה בוחן המדגים את האפשרות ליישם תהליכי למידת מכונה לצורך עריכת סקר ספרות שיטתי והערכת הסנטימנט של כל אחד מהמרכיבים הנכללים באונטולוגיה של עולם התוכן שנבחר. תוצאות המסווג שפותח מעיד על יכולת גבוהה המותנית במספר מינימלי של מקורות שהוא נחוץ כדי להבטיח שהערכת הסנטימנט של המכונה לא יהיה דומה להערכת הסנטימנט שהיה מתקבל בנייתו ידני של הספרות. תובנה זו מאפשרת לחוקרים לא רק חסכון משמעותי בזמן והסרת כל מגבלה לגבי היקף מקורות הספרות הנסקרים, אלא גם מספק

רשימה ביבליוגרפית

- Ananiadou, S., Rea, B., Okazaki, N., Procter, R., & Thomas, J. (2009). Supporting systematic reviews using text mining. *Social Science Computer Review*, 27(4), 509–523. <https://doi.org/10.1177/0894439309332293>
- Aphinyanaphongs, Y., Tsamardinos, I., Statnikov, A., Hardin, D., & Aliferis, C. F. (2005). Text Categorization Models for High-Quality Article Retrieval in Internal Medicine. *Journal of the American Medical Informatics Association*, 12(2), 207–216. <https://doi.org/10.1197/JAMIA.M1641>
- Bekhuis, T., & Demner-Fushman, D. (2012). Screening nonrandomized studies for medical systematic reviews: A comparative study of classifiers. *Artificial Intelligence in Medicine*, 55(3), 197–207. <https://doi.org/10.1016/j.artmed.2012.05.002>
- D'Andrea, A., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, Tools and Applications for Sentiment Analysis Implementation. *International Journal of Computer Applications*, 125(3), 26–33. <https://doi.org/10.5120/ijca2015905866>
- De, A., & Kopparapu, S. K. (2013). Unsupervised clustering technique to harness ideas from an Ideas Portal. *Proceedings of the 2013 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2013*, 1563–1568. <https://doi.org/10.1109/ICACCI.2013.6637413>
- Debortoli, S., Müller, O., Junglas, I., & vom Brocke, J. (2016). Text mining for information systems researchers: An annotated topic modeling tutorial. *Communications of the Association for Information Systems*, 39(1), 110–135. <https://doi.org/10.17705/1cais.03907>
- Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment Analysis: A Comparative Study on Different Approaches. *Procedia Computer Science*, 87, 44–49. <https://doi.org/10.1016/j.procs.2016.05.124>
- Felizardo, K. R., Salleh, N., Martins, R. M., Mendes, E., Macdonell, S. G., & Maldonado, J. C. (2011). Using visual text mining to support the study selection activity in systematic literature reviews. *International Symposium on Empirical Software Engineering and Measurement*, 77–86. <https://doi.org/10.1109/esem.2011.16>
- García Adeva, J. J., Pikatza Atxa, J. M., Ubeda Carrillo, M., & Ansuategi Zengotitabengoa, E. (2014). Automatic text classification to support systematic reviews in medicine. *Expert Systems with Applications*, 41(4 PART 1), 1498–1508. <https://doi.org/10.1016/j.eswa.2013.08.047>
- Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews*. Sage.
- Hemmatian, F., & Sohrabi, M. K. (2019). A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review*, 52(3), 1495–1545. <https://doi.org/10.1007/s10462-017-9599-6>
- Higgins, J., & Green, S. (2008). Cochrane Handbook for Systematic Reviews of Interventions. In *Cochrane Database of Systematic Reviews*. <http://onlinelibrary.wiley.com/doi/10.1002/9780470712184.fmatter/summary>
- Huang, F., Zhang, S., Zhang, J., & Yu, G. (2017). Multimodal learning for topic sentiment analysis in microblogging. *Neurocomputing*, 253, 144–153. <https://doi.org/10.1016/j.neucom.2016.10.086>
- Jeyapriya, A., & Kanimozhi, S. (2015). Extracting Aspects and Mining Opinions in Product Reviews using Supervised Learning Algorithm A.Jeyapriya. *IEEE*, 548–552.
- Jongeling, R., Datta, S., & Serebrenik, A. (2015). Choosing your weapons: On sentiment analysis tools for software engineering research. *2015 IEEE 31st International Conference on*

- Software Maintenance and Evolution, ICSME 2015 - Proceedings*, 531–535. <https://doi.org/10.1109/ICSM.2015.7332508>
- Khan, K. S., Ter Riet, G., Glanville, J., Sowden, A. J., Kleijnen, J., & others. (2001). *Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews* (Issue 4 (2n)). NHS Centre for Reviews and Dissemination.
- Kitchenham, B. A., Budgen, D., & Pearl Brereton. (2015). *Evidence-based software engineering and systematic reviews*. CRC press.
- Kitchenham, B. A., Dybå, T., & Jørgensen, M. (2004). Evidence-based software engineering. *Proceedings - International Conference on Software Engineering*, 26, 273–281. <https://doi.org/10.1109/icse.2004.1317449>
- Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering* (Issue January).
- Kwartler, T. (2017). Text Mining in Practice with R. In *Text Mining in Practice with R*. WILEY. <https://doi.org/10.1002/9781119282105>
- Liu, J., Seneff, S., & Zue, V. (2012). Harvesting and summarizing user-generated content for advanced speech-based HCI. *IEEE Journal on Selected Topics in Signal Processing*, 6(8), 982–992. <https://doi.org/10.1109/JSTSP.2012.2229690>
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- Mohammad, S., Dunne, C., & Dorr, B. (2009). Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, August*, 599–608. <https://doi.org/10.3115/1699571.1699591>
- Naldi, M. (2019). *A review of sentiment computation methods with R packages*. 1–11. <http://arxiv.org/abs/1901.08319>
- O'Mara-Eves, Alison; Thomas, J. M. M. A. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic Reviews*. <https://doi.org/10.1145/2945078.2945085>
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1--2), 1–135.
- Pham, D. H., & Le, A. C. (2018). Learning multiple layers of knowledge representation for aspect based sentiment analysis. *Data and Knowledge Engineering*, 114(January 2017), 26–39. <https://doi.org/10.1016/j.datak.2017.06.001>
- Phu, V. N., Dat, N. D., Ngoc Tran, V. T., Ngoc Chau, V. T., & Nguyen, T. A. (2017). Fuzzy C-means for english sentiment classification in a distributed system. *Applied Intelligence*, 46(3), 717–738. <https://doi.org/10.1007/s10489-016-0858-z>
- Popoff, E., Besada, M., Jansen, J. P., Cope, S., & Kanters, S. (2020). Aligning text mining and machine learning algorithms with best practices for study selection in systematic literature reviews. *Systematic Reviews*, 9(1), 1–12. <https://doi.org/10.1186/s13643-020-01520-5>
- Pradhan, V. M., Vala, J., & Balani, P. (2016). A survey on sentiment analysis algorithms for opinion mining. *International Journal of Computer Applications*, 133(9), 7–11.
- Rathan, M., Hulipalled, V. R., Venugopal, K. R., & Patnaik, L. M. (2018). Consumer insight mining: Aspect based Twitter opinion mining of mobile phone reviews. *Applied Soft Computing Journal*, 68, 765–773. <https://doi.org/10.1016/j.asoc.2017.07.056>

- Severyn, A., Moschitti, A., Uryupina, O., Plank, B., & Filippova, K. (2016). Multi-lingual opinion mining on YouTube. *Information Processing and Management*, 52(1), 46–60. <https://doi.org/10.1016/j.ipm.2015.03.002>
- Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. “O’Reilly Media, Inc.”
- Templier, M., & Paré, G. (2015). A framework for guiding and evaluating literature reviews. *Communications of the Association for Information Systems*, 37(September 2015), 112–137. <https://doi.org/10.17705/1cais.03706>
- Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*, 57(March), 117–126. <https://doi.org/10.1016/j.eswa.2016.03.028>
- Vilares, D., Alonso, M. A., & Gómez-Rodríguez, C. (2017). Supervised sentiment analysis in multilingual environments. *Information Processing and Management*, 53(3), 595–607. <https://doi.org/10.1016/j.ipm.2017.01.004>
- Wallace, B. C., Trikalinos, T. A., Lau, J., Brodley, C., & Schmid, C. H. (2010). Semi-automated screening of biomedical citations for systematic reviews. *BMC Bioinformatics*, 11. <https://doi.org/10.1186/1471-2105-11-55>
- Xiao, Y., & Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. *Journal of Planning Education and Research*, 39(1), 93–112. <https://doi.org/10.1177/0739456X17723971>
- Zupic, I., & Čater, T. (2015). Bibliometric Methods in Management and Organization. *Organizational Research Methods*, 18(3), 429–472. <https://doi.org/10.1177/1094428114562629>