**השתמש באלגוריתמים של למידת מכונה כדי לסווג קישור וחיתוך מזמורים**

**סיכום:**

השפה הערבית היא ישות חיה הגדלה ומתפתחת הלכה למעשה ויישום נכון של כל נימוסיה וענפי הדקדוק, הטהרה, הסמנטית והלקסיקלית, ובפרויקט זה אציגאת תרומת הטכנולוגיה לפיתוח השפה הערבית, ובמיוחד את הציור הנכון של הקישור והחיתוך. פרויקט זה נועדבמיוחד לבנות מודל מסווג חכם המקטלג את המילים בערבית המתחילות באות A, ולקטלג את לחישתן לקישור או לחישה חותכת באמצעות שימוש בטכניקות בינה מלאכותית בכלל, ואלגוריתמים של למידת מכונה בפרט על מנת לקבוע קריטריונים מדויקים ונכונים בציורהקישור וחיתוך הלחישות בצורה נכונה, כך שפיתחנו את הטכנולוגיה לתרום לשירותהשפה הערבית. פרויקט זה הסתמך על אוסף של מילים בערבית המתחילות על ידי זמזום, על ידי עיצוב שאלון דיגיטליעםהמשימה של הידור כמה שיותר מילות סטארט-אפ וסיווגן כקישור או חיתוך בהתאם לכללי הדקדוק. האינטרנט נארז על ידי 50 מומחי דקדוק בדרגות שונות, עם 400 מילים סה"כ מסווגות, ולאחר האצה והרחקה של 101 המילים הכפולות, קיבלנו 299 מילים תקפות על הגודל וסוג הנתונים שנאספו ומנגנון הסיווג שאני מקווה להחיל על מודל חוברת העבודה, ובניית החבות הוחלו אלגוריתמים לסיווג בהתאם למדגם המשולב כגון אלגוריתם מכונת התמיכה הווקטורית(SVM) ואלגוריתם Naif BizNB) והאלגוריתםשלהשכן הקרוב KNN)באמצעותשימוש בשפה (sk-learn וספריית פייתון לאחר אימון המודלים של עבודות המשמשות ומדידת הדיוק של אלגוריתמים) בנה כי אלגוריתם מכונת התמיכה הווקטורית SVM השיג את הדיוק הגבוה ביותר של הדגם ב-92%, שהוא גבוה ומספיק כדי לפתור את בעייתהפרויקט.

**אתחול:**

אנו יכולים להגדיר למידת מכונה כמלווה במדע הבינה המלאכותית, המעוניין בעיצוב אלגוריתמים וטכניקות המאפשרים למחשבים ללמוד את עצמם עם אפשרות לפתחאלגוריתמים אלה. אנו יכולים לסקור שני סוגים של למידת מכונה: למידה אינדוקטיבית ותשראתית, למידה אינדוקטיבית מוגדרת כלמידת גישוש ועובדת גם על העיקרון של הסקת כללי נתונים כלליים, השונה מלמידה אינדוקטיבית שבהניתן ללומד את הכללים שהוא צריךליישם. למידת מכונה מבוססת על העיקרון של מערכות למידה מנתונים זמינים, זיהוי דפוסים מתאימים ו קבלת החלטות ללא התערבות אנושית, ומאחר שהמשימה העיקרית של למידת מכונה היא לחלץ מידע מנתונים, למידת מכונה היא שיטה לניתוח נתונים שהופכת את בנייתם של מודלים לניתוח נתונים המכונים מודלים תחזיתיים או טקסונומיים, המאפשרים לחוקרים, מדעני נתונים ואנליסטים לקבל החלטות המאפשרות תוצאות אמינות. בניית מודל ניתוח הנתונים כפופה למספר מנגנונים או שיטות לתוצאות מדויקות, שכן שלב בניית המודל דורש יישום של מספר שלבים מתמטיים או לוגיים המסודרים ומרצף כדי להגיע למודל אנליטי מיומן שיכול לחזות ערכים קרובים יותר לאמת או לסווג נתונים במדויק, קבוצת שלבים זו נקראת האלגוריתם, כפי שמודל סיווג הנתונים נקרא חוברת העבודה. בפרויקט זה עסק SAבשימוש במודלים של אלגוריתמים של למידת מכונה בסיווג שלhumzes קישור וחיתוך ואלגוריתמים אלה יוחלו על מילים בערבית ביחס לשפה הערבית לא קיבל את המזל של לימוד ומחקר כמו שפות אחרות.

**מטרות הפרויקט:**

* פתרון הבעיה של שימוש לא נכון בציור הקישור וחיתוך לחישותבטקסטים בערבית. זהה את אלגוריתמי הסיווג המתאימים ביותר בתוך אלגוריתמי למידת מכונה כדי לקבוע את המיקומים הנכונים לשימוש, לקשר ולחתוךבתחילתהמילה.
* מדוד את האיכות של אלגוריתמי הסיווג המפורסמים ביותר בהבחנה בין הקישור לחתיכות בתחילתהמילה.

**בעיית הפרויקטושאלותיו:**

הבעיה העיקרית של הפרויקט טמונה בשימוש שגוי בקישור שאינו במקומו או חיתוך מזמזמים בעת כתיבת מילים עם האות A בפרט, כאשר רבים טועים בציור או בהשלכת הלחישה, אשר מחליש את כוחה שלהשפה. פרויקט זה תורם להתפתחות השפה הערבית, על ידי בניית מודל סיווג חכם לעורכים ומתרגלים, ולכל משתמשי השפה מהסיווג הנכון של הקישור והחיתוך. מחקר זה מבקש לענות על השאלות הבאות לאור השימוש באלגוריתמים של למידת מכונה המשמשים לסיווג וחיזוי:

* כיצד ניתן להבחין בין שני הקישורים ליצירות בטקסטים בערבית?
* מהם היתרונות של סיווג הקישורים והקטעים בטקסטים בערבית?
* אילו אלגוריתמים של למידת מכונה מתאימים לסווג את הקישור ולחתוך זמזומים בטקסטים בערבית?

**מחקרים קודמים:**

מחקר זה הוא אחד המחקרים החדשים בתחום המחקרים הדקדוקיים והטכניים העוסקים בשימוש באלגוריתמים של למידת מכונה כדי לסווג את הקישור ולחתוך לחישות בטקסטים בערבית, וככל הידוע לי, ובאמצעות המחקר שלי, לא מצאתי מחקרים שנחשפו לנושא זה בעבר בגלל הקושי להשתמש בספריות תוכנה בהתמודדות עם טקסטים בערבית בצדק, אך יש פנים וחפיפה כמעט גדולים של מחקרים אחרים הקשורים לפרויקט לשימוש באלגוריתמים אלה ביישומים דומים ומעניינים, ונהניתי מהם, ועזרתי לי להיכנס לפרויקט שלי, כולל:

**לימודים שכותרתם:**

* חקר דעות במשפטי השוואה בערבית **[1]**

מחקר זה התייחס לבעיית זיהוי מגזין ההשוואה בחקר הדעות המשמשות בטקסט הערבי, והמחקר התמקד בציור דעות ממגזין ההשוואה על ידי הכרת המוצר המועדף על ידי כותב הדעות בהשוואה למוצר אחר אויותר. החוקר ציין כי יש מחקר בתחום זה למשפטים של אנגלית ושפות שלה, אבל עבור משפטים בערבית זהו המחקר הראשון, והמחקר גם השתמש בטכניקה המבוססת על סיווג לשוני וטכניקה אחרת המבוססת על חינוך מכונה.

**לימודים שכותרתם:**

* מחקר השוואתי של אלגוריתמים לחקר דעת קהל וניתוח רגשות **ויישומים שלהם[2]**

מחקר זה התייחס לבעיית ריבוי הדעות של לקוחות המאוחסנים במאגרי הנתונים של האינטרנט, אשר העניקו תשומת לב לחיפוש נתונים וניתוח סנטימנט בשנים האחרונות, החוקר ציין כי אנשים הסתמכו על המכונה בסיווג ועיבוד נתונים, שכן הזמינות של כמויות עצומות של צפיות על מוצר אחד מסייעת לחזות את רגשות הלקוחות על ידי ניתוח השקפות המסייעות לא רק להגדיל את הרווחים אלא גם לשפר המוצר המחקר השווה את הטכנולוגיות הקיימות כיום ומשמשות ביישומים מרובים בתחום חקרהדעות. לאחר סקירת המחקרים לעיל, הרעיון של סיווג החמזה התבסס על השימוש במושג חקר בטקסטים ערביים על ידי חוקרים במחקרים הנ"ל בשיטת החיפוש בטקסטים בערבית, עם אלגוריתמי סיווג שונים בשימוש.

**מתודולוגיית פרוייקט:**

* בפרויקט זה נעשה שימוש ברזולוציה דיגיטלית, ומשימת הרזולוציה הוגבלה להידור וסיווג של כמה שיותר מילות סטארט-אפ על ידי מומחי שפה ערבית (עבור שני קישורים או קטעים בהתאם לכללי הדקדוק של הסיווג, הקישור והחתיכות.
* הגישה הסטטיסטית המבוססת על הרזולוציה המתוכננת וקבוצת האימון של המילה שימשה גם בפרויקט זה באמצעות מודל הסיווג, ולפרויקט זה נשתמש באלגוריתמים של למידת מכונה לאחר בחינת תדירות התנאים והמאפיינים (דקדוקים הקשורים למילים שהחלו בלחש וחולצו באמצעות רזולוציה) כדי להסיק את היסודות הבסיסיים שלמודל הסיווג.

**מבנה הפרויקט:**

**הפרויקט מורכב משלושחקירות:**

* המחקר הראשון (המסגרת התיאורטית של הפרויקט): עוסק במושג הבינה המלאכותית, מושג למידת מכונה, אלגוריתמים לסיווג נתונים והקישור והחיתוךלחישות.
* המחקר השני (מאיץ ויישום): עוסק בשלבי היישום של מודלים של עבודות, הכוללים: איסוף תיאורים, ניקוי, מדקטים, קידוד נתונים, זיהוי משתנים עצמאיים ותלויים במדגם הנתונים, כמו גם בניית מודלים שלעבודות.
* המחקר השלישי (תוצאות פלטי הפרויקט): מבחן מודלים של עבודות, מדידת דיוק המודל, תוצאות, תפוקות הפרויקט והיתרונות שלהפרויקט.

**נושא ראשון: המסגרת התיאורטית של הפרויקט:**

**ראשית: מושג הבינההמלאכותית:**

בינה מלאכותית היא ענף של מדעי המחשב, אחד מעמודי התווך של תעשיית הטכנולוגיה של העידן המודרני שלנו, המכונהבינה מלאכותית וניתן להגדירה כיכולת של מכונות ומחשבים לבצע משימות המחקות במידה רבה את המוח האנושי החכם, וניתן לסכם משימות אלה ביכולת לחשוב או ללמוד מניסיונו הקודם, כך שניתן לומר כי בינה מלאכותית שואפת להגיע למערכות שמתנהגות [3] הם לומדים ומבינים כשהם מתנהגים, לומדים ומבינים בני אדם כפי שיש להם**אינטליגנציה.[3]**

**סוגיבינה מלאכותית:**

ניתן לחלק סוגי Ai בהתאם ליכולותיהם לשלושה **סוגים[4]** באופן הבא:

* **בינהמלאכותית מוגבלת:**

זהו סוג שיכול לבצע משימות ספציפיות וברורות כגון יישומי רכב בנהיגה עצמית, תוכנה לזיהוי דיבור, תמונות או שחמט, וסוג זה של בינה מלאכותית הוא הסוג הנפוץ ביותר.

* **בינהמלאכותית כללית:**

זהו מין שיש לו יכולות חשיבה דומות לזו של האדם, שכן הוא גורם למכונה להיות מסוגלת לחשוב בכוחות עצמה ודומה מאוד לחשיבה האנושית, ולמעשה אין יישומים מעשיים למין זה, אך ישנם רק מחקרים הזקוקים למאמץ רב כדי להפוך אותו למציאות, ושיטת הרשתות העצביות היא אחד המודלים של בינה מלאכותית כללית, שכן היא עוסקת בייצור מערכת של רשתות עצביות של המכונה הדומה לזו הכלולה במוח האנושי. **[5]**

* **בינהמלאכותית** ללא הגבלה:

בינה מלאכותית ללא הגבלה היא מהסוג שעשוי לעלות על רמת האינטליגנציה האנושית, ויכולתה לבצע משימות באופן שיכול להיות טוב יותר מזה של בני אדם מיוחדים ובקיאים, ולסוג זה יש מאפיינים נחוצים רבים, כגון למידה, תכנון אוטומטי, היכולת לתקשר וקבלת ההחלטה הנכונה, אך הרעיון של בינה מלאכותית סופר הוא מושג היפותטי שאינו קיים בזמננו.

**בינה מלאכותית יכולה להיות מסווגת גם על פי ארבע הפונקציות השונותהבאות:**

* **מכונות אינטראקטיביות:**

זהו הסוג הפשוט ביותר של בינה מלאכותית מכיוון שאין לו את היכולת ללמוד מניסיון קודם לפתח עסקים עתידיים, ולכן כאן נתקשר עם החוויה הנוכחית כדי לייצר את הדרך הטובה ביותר האפשרית, כגון ציוד עמוק כחול שפותח על ידי יבמ ו- AlphaGo מגוגל.

* **זיכרון מוגבל:**

זיכרון מוגבל יכול לאחסן נתונים היסטוריים קודמים על המערכת הנוכחית לפרק זמן מוגבל, וגישת הנהיגה העצמית היא אחת הדוגמאות היעילות ביותר לדפוס זה, תוך שמירה על המהירות האחרונה של מכוניות אחרות, המרחק בין מכוניות אלה, המהירות המותרת ומידע בסיסי אחר לנהיגה בנתיבי תחבורה. **[6]**

* **תורת ההיגיון:**

סוג זה של בינה מלאכותית פירושו שהמכונה סופגת רגשות אנושיים, מתקשרת ומתקשרת עם בני אדם, ויש לציין כי לא נמצאו יישומים מעשיים עד כה בסוג זה של בינהמלאכותית.

* **תפיסה עצמית:**

מודעות עצמית היא תחזית עתידית לבינה מלאכותית, ועובדת על עיקרון טכני וסנסציוני מודרני מאוד שבו המכונה יכולה לייצרידע עצמי ותחושות משלה, מה שיהפוך אותה לחכמה יותר מהאדם, והמושג הזה עדיין לא קיים למעשה. **[7]**

**שדות משנה של בינה מלאכותית:**

ל- AI יש תחומי משנה רבים, כגון למידת מכונה, הכוללת מתן אפשרות למחשבים ללמוד ללא תלות בכל ניסיון קודם, כך שמחשבים יכולים לחזות את קבלת ההחלטות הנכונה במהירות, באמצעות פיתוח אלגוריתם המאפשר מצב זה. יש לציין כי מונח זה הוצע לראשונה על ידי ארתור סמואל בשנת 1959 ואנו נתייחס להלן לכמה מהתחומים המפורסמים ביותר של בינה מלאכותית כדלקמן:

* **חיפוש נתונים:**

היא נועדה לחפש ולחקור נתונים ודפוסים ספציפיים במגוון רחב של נתונים באמצעות תוכנה, שכן חברות יכולות להפיק תועלת מחיפוש נתונים בפיתוח הביצועים שלהן, הגדלת המכירות והפחתת עלויות הייצור. **[8]**

* **אחזור מידע ואינטרנט סמנטי:**

הרעיון של אחזור מידע מתייחס לתהליך החיפוש אחר כל סוג של נתונים ומסמכים שעשויים להיות נוכחים באינטרנט באמצעות מושג האינטרנט הסמנטי. שירות האינטרנט הסמנטי ממיר נתוני אינטרנט למסד נתונים גלובלי של מידע מחובר, כך שמכונות יכולות להבין אותו, ולא רק להשתמש בו עבור בני אדם. בדרך זו, המכונה יכולה להזמין כרטיסים באינטרנט, להשתמש במילון מקוון, או דברים אחרים הדורשים בתחילה שימוש ידני כדי להשלים אותם. **[9]**

* **ייצוגידע:**

ייצוג ידע הוא תחום של בינה מלאכותית שמעוניין לגרום למכונות לחשוב ולקבל את ההחלטה הנכונה, שכן הידע שנרכש על ידי המכונה מקובצים ומאוחסנים במסד נתונים המשמש לשיתוף ידע וניהול מרכיביו, ומהווה התייחסות לקבלת החלטות חכמות שיתקבלו על ידי המכונהבעתיד.

* **חשיבה לוגית וחשיבה סבירה:**

חשיבה לוגית בבינה מלאכותית היא צורת חשיבה שונה, משום שהעובדות מוסקות על סמך נתונים זמינים. חשיבה לוגית תואמת את מה שמכונה חשיבת הסתברות, המשתמשת במושגים של הסתברות וחוסר ודאות בידע כדי להתמודד עם כל אי הוודאות העתידית של כל האירועים שעלולים להיות חשודים. **[10]**

* **למידת מכונה:**

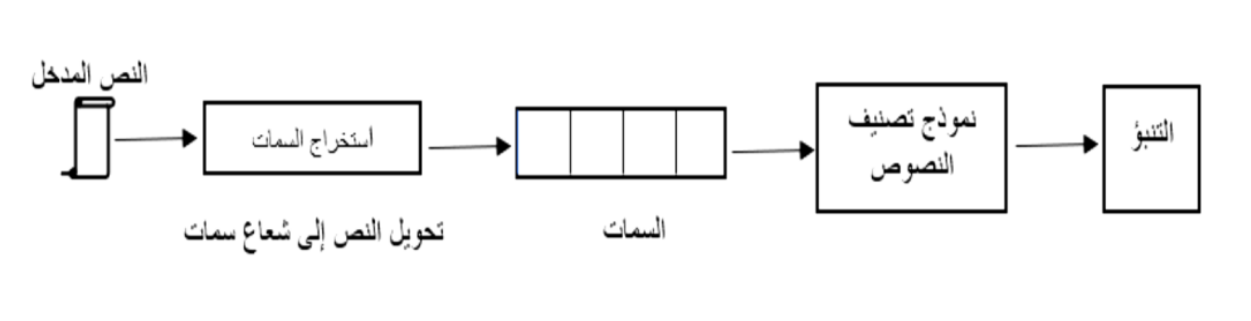
למידת מכונה היא ענף של בינה מלאכותית, הכולל עיצוב ופיתוח של אלגוריתמים וטכניקות המאפשרים למחשבים להחזיק תכונות "למידה". באופן כללי, הלמידה מחולקת לשתי רמות: אינדוקטיבית ואינטרוספקטיבית, שבה הגישה האינטרוספקטיבית שואבת כללים והוראות כלליות מביגדאטה.

**שנית: למידת מכונה:**

המשימה העיקרית של למידת מכונה היא לחלץ מידע בעל ערך מנתונים, ולכן הוא קרוב מאוד לחילוץ נתונים, באמצעות למידת מכונה בניתוח נתונים, דרך לפתח מודלים מורכבים ואלגוריתמים מתאימים לחילוץ נתונים באמצעות תהליכי חיזוי, הנקראים ניתוח חיזוי. מודלים אנליטיים אלה מאפשרים לחוקרים ואנליסטים של נתונים ללמוד החלטות ותוצאות אמינות ויכולים לזהות נתונים מאוחסנים ואתמערכות היחסים שלהם.

ניתן להגדיר מערכות למידת מכונה גם כמערכות המנבאות תחזיות על סמך מה שנתונים קודמים למדו וצריכים להיות מאומנים בדוגמאות רבות של טקסט ותחזיות (סמנים) הצפויים לכל אחד מהם. הנתונים המשמשים להכשרה נקראים ערכת נתוני האימון ונתונים אלה מסווגים מראש עם תכונות וככל שקבוצת האימונים והתכונות שנבחרו מדויקות יותר, כךתחזיות הסיווג, כאשר שיטת למידת מכונה מאומנת, נתוני הדרכה חייבים להיות מומרים למשהו שאתה יכול להבין. המכונה, שבה תכונות מופקות ומומרות לקרניים (ייצוג טקסט לפי מספרים) תסייע לה ללמוד מנתונים קיימים ולבצע תחזיות לגבי טקסטים עתידיים. **[11]**

המודל המאומן יכול לחלץ תכונות מטקסט חדש, לחזות או לסווג טקסטים לפי מאפיינים ספציפיים באמצעות אלגוריתמים לסיווג נתונים כמו באיור (1-1) להלן:



טופס (1-1): מדגים את מנגנון סיווג הטקסט באמצעות אלגוריתמים של סיווג נתונים.

**שלישית: אלגוריתמים לסיווג נתונים:**

ישנם מספר אלגוריתמים לסיווג נתונים המתאימים בקלות ליישומים לחקר נתוני טקסט לאחר הבדיקה, ואניה קלה להתאמן עם כמויות גדולות וקטנות של נתונים שסופקו, וכאן נסקור את האלגוריתמים המפורסמים ביותר של למידת מכונה לסיווג נתוני טקסט ששימשו בפרויקט זה:

1. **خوارزمية دعم المتجه (SVM-Support Vector Machines):**

המכונהSVM, אלגוריתם זה מבוסס על אלגוריתמי למידה מונחי מכונה המבוססים על ערכת נתונים מוגדרת מראש (קבוצת אימון) בהכשרת האלגוריתם כך שהוא יכול לנתח, לסווג או לזהות כל קבוצה חדשה של נתונים, שפותחה על ידי העולמות ולדימיר Fabank ואלכס שרייבונקיס בשנת 1963 ולאחר מכן פותחה על ידי קורינה קורץ ופאבנק בשנת 1993 ופורסם בשנת 1995. **[12]**

אלגוריתם SVM הוא אחת משיטות הסיווג האוטומטי הפופולריות ביותר המבוססות על יצירת עקומה או רמת שבירה, המפרידה בין דגימות שהוזנו זו לזו, והאלגוריתם מאופיין בשימוש בו בסיווג בעיות עם קטגוריות בינאריות באופן בלעדי. עבור סיווג וזיהוי דפוס, אחת מתכונותיו היא דיוק גבוה בסיווג והיא מוחלת במגוון רחב של תחומים, כולל בחירת קטגוריות טקסט לפי סיווג תמונה. **[13]** אם האלגוריתם יכול למצוא רמת שבירה אחת נמוכה יותר מווקטורים של הנקודה, הסיווג הוא ליניארי, אחרת הסיווג אינו ליניארי. **[14]**

הדיוק של האלגוריתם קובע את יכולתו להפריד בין שני סוגים כך ששתי הנקודות הקרובות ביותר זו לזו רחוקות מלהיות אפשריות, ואנו יכולים לתת שם לרמת ההפרדה עם הקצה או השוליים של ההפרדה. באופן כללי, ככל שהקצה או השוליים של ההפרדה גבוהים יותר, כך השגיאה קטנה יותר בעת הכללה בערכת נתוניםחדשה.

1. **خوارزمية نايف بيز: (Bayes Naive)**

האלגוריתם של Nayef Biz (NB) הוא גם אלגוריתם למידה מונחה מכונה, והוא מסתמך על כללי ההסתברות שגובשו על ידי המדען תומאסביז[15] שבהם ההסתברות מחושבת באמצעות מספר החזרות של ערכים, תדרים ושילובי ערכים בנתונים ידועים מראש (נתוני yip). הוא נחשבלקו הבסיס בסיווג טקסטים ויעיל בתחומים רבים למרות שיש מספר עבודות אחרות עם דיוק גבוה יותר כגון מודלSVM, כמו מודל Bayes Naive מפיץ טקסטים עבור כל פריט באמצעות מודל הסתברות עם הנחות עצמאיות, שיטה זו נפוצה מאוד בתחום סיווג הטקסט, שכן העבודה הבינארית היא אחת השיטות הידועות ביותר של מודל Bayes Naive שהשתמש בייצוג רדיאלי דו-ערך של טקסטים.

מספר שיפורים נעשו למחלקת Biz, כולל התאמת חישוב ההסתברות, הפחתת תכונות וכמה מאפיינים אחרים, וככל שהתיאוריה של Biz מחפשת את האפשרות לאירוע, אירוע נוסף כבר התרחש. האלגוריתם של נאיף בייז הוא גם אחד האלגוריתמים החשובים ביותר של למידת מכונה מכמה סיבות: הקלות של בניית עבודה, וכאן אתה לא צריך להשתמש במה שמכונה הערכת תוכניות עבור משתנים מורכבים שחוזרים על עצמו, וניתן להחיל אותו בקלות על ערכת נתונים גדולה ומטרת האלגוריתם היא לבנות בסיס המאפשר להקצות מבנים עתידיים לפריט מסוים על ידי מתן וקטורים למשתנים המתארים את המבנה הזה, שבאמצעותו המשתמש יכול לעשות סטטיסטיקות רבות להקלות השימוש. **[16]**

1. **خوارزمية الجار الأقرب: K-Nearest Neighbor**

KnN יכול לשמש כעבודת סיווג טקסט פשוטה ויעילה, knn מכיל שני חסרונות: מורכבות חישובית אם הדגימות דומות מאוד, וביצועיהם מושפעים בקלות אם דגימות האימון הן בודדות. ניתן להפחית את המורכבות של KnN באמצעות שלוש שיטות: או על ידי הפחתת מידות הווקטור המייצג את הטקסט, על ידי הפחתת כמות דגימות האימון או על ידי הגבלת מציאת השכנים הקרובים ביותר, כלומר את הערך של K.

KNN td משתמש בסיווג טקסט על-ידי חישוב המרחק שהטקסט בונה וכל הטקסטים בערכת נתוני האימון באמצעות מידה של הבדל או דמיון ביניהם, ולאחר מכן מציאת K הקרוב ביותר הסמוך לכל טקסטי האימון ובחירת מחלקת הטקסט לפריט עם המספר הגדול ביותר של טקסטים בשכנים הקרובים ביותר של טקסטים ואלגוריתמים אחרים שופרה ביותר ממובן אחד. **[17]**

**רביעית: הם לחשו את החיבור ואת החלקים**

בחלק זה של ההיבט התיאורטי, נתייחס לקישור ולחישות חיתוך במילה הראשונה תחתהשמות הבאים:

* **משמעות הזמזום**  **(שפה):**

מילונים רבים של השפה, כולל לשונם של ערביי אבן מנסאר, הצהירו כי הזמזום בשפה פירושו קריצה, קריצה ולחיצה. **[18]**

* **המשמעות של זמזום (במונחים של):**

אל-חמזה במונח בלשנים קיבל מספר הגדרות, ביניהן הגדרתו של אל-אז'ארי בספרו עידון השפה, שם הוא אומר "אני יודע שהלחישה לא סאטירה, אלא כותב פעם אלף ופעם ופעם ו-י' ואלף רכים אין אות, אבל זה חלק מהתקופה שלאחר הפתיחה, והאותיות עשרים ושמונה אותיות עם ו' ואלפא ות', ומתבצעות בלחש עשרים ותשע אותיות ולחישה כאות הנכונה, עם זאת, יש לו מקרים של ריכוך, מחיקה, החלפה והקלה... זו לא אות חלולה, זו טבעת מהפה הרחוק".

**קטעי הזמזום בתחילתהמילה:**

בתחילת המילה, הזמזום מחולקלמדור שלי:

**A-הקישור:**

* **תןלו שם:**

הקישור נקרא על שם שם זה, משום שהוא מגיע להגייה של האות הסטטית הממוקמת בתחילת הנאום, שכן הכלל החשוב בכך הוא שהוא אינו מתחיל עם תושב ואינו מפסיקבתנועה.

* **הגדרה:**

זוהי הלחישה הקבועה בהתחלה, ונופלת במקרה של הקשר, וbn מאליק הצהיר: "אם זה נקרא מה הוא קודם קישור, הלחישה תיחתך אם זה בהעברה ממעשה כגון: (ראיתי שאני יודע)" מצוין אם אדם נקרא **על שמו[19]**.

* **תנועתהקישור:**

המקור בקישור הוא, שהוא שבור כמו: ללכת, למעט אלף הגדרות שפתוחות, כמו גם אלף אנשי ימין כמו שאתה אומר: "איימן אללה על ידי פתיחתהאלף." **[20]**

* **מיקומי רכזת:**

לקישור שלוש עמדותכדלקמן:

**שמות:**

במילה הראשונה, חמזה אל-וסל ממוקם בעשרה שמות בשפה (שם, בן, בת, שתיים, אישה, בן, אם אללה, א)ומספר שמות אלה מוזכרים בקוראן.

**B-פעולות:**

הקישור במעשים המתחילים הוא שתימערכות:

* מעשה העבר: החוליה בעבר של החמישייה היא כמו: להיפתח, כמו גם את העבר של המשושים כגון:סליחה.
* עשה זאת: הקישור בעשותו כן נמצאבמקומות הבאים:
  + סדר המערכה המשולשת הוא כמו: ניצחון, < ניצחון.
  + סדר המעשה בן חמש השנים הוא כמו: ניצחון עלהאמת.
  + סדר מופע שש המפלגות הוא כמו: סלח עלאשמתך.

**C-מכתבים:**

אין קשר באותיות למעט באות אחת, האינדוקציה (ה) או מה שמכונה כאב שמש, כגון השמש, כמו גם בסלאם ירחי כגון: **ירח[21]**

**B/ זמזוםהחיתוך:**

* **תןלו שם:**

ההומזה של החתיכות נקראת על שם זה כי זה חותך ושומר את הקודמים של האותיות, כגון האמרה שלך: שיעור - לימוד, הלחישה לחתוך את הדאל מן הלחישה שקיבלה את זה, כמו גם נאסר - אנסר, הלחישה הזמינה הבא, שהוא האות של n עבור הלחישה שקיבלהאת זה.

* **חיתוךתנוחות** זמזום:

ליצירות יש שלוש עמדותכדלקמן:

**שמות:**

היצירות מופיעות בכל השמות בערבית כמו אחמד, אמג'ד, אכרם, אסד,כמו גםבששת השמות, כל מה שמתחיל בלחש הם קטעים כגון: "אחיך-אבא", למעט 10 השמות שלחשתם היא חוליה.

**B-פעולות:**

החלקים ממוקמים בפעליםכדלקמן:

* + העבר של המעשה המשולש הוא כמו: לקחת לאכול.
  + עבר הרביעייה כגון: אכרם - הושג
  + סדר הפעולה הארבעה הוא כמו:אכרם.
  + מקור הרביע הוא כמו: אזהרהמאזעקה.

**C-מכתבים:**

כל האותיות בערבית הזדעזעו מפיסת זמזום, למעט המכירות (N-N-N-SO-22 ל) **[22]**

* **ההבדל הוא הבן שלי, הלחישה שלי היא החיבור**והחתיכות:

על מנת להבחין ולהבחין בין הקישור לבין החלקים, יש צורך לדעת את הדברים הבאים:

* החלקים באים דוממים אוזזים.
* הקישור שמגיע רקזז.
* לחישת החלקים קבועה בתחילת הנאום ובסופו, כגון: הוא שאל באמצע וקרא בסוף.
* הקישור נופל תוך כדי דיבור תוך הוכחת תחילה בלבד.
* הקישור הוא רק מוגזם, בעוד החלקים הם מקוריים או **עודפים.[23]**

**נושא 2: יישום ויישום**

**אתחול:**

הנתונים של היום מתנפחים מיום ליום ומקורותיהם רבים, והדבר מוביל לחשיפת נתונים אלה לבעיות רבות המפחיתות את איכות הנתונים כגון המספר הגדול של הנתונים שאבדו וחוסר העקביות של הנתונים, ולכן בפרויקט זה חילקתי את שלבי היישום של מודלים של עבודות נתונים לשישה שלבים, החל משלב עיצוב הרזולוציה ואיסוף הנתונים וכלה בשלב מדידת הדיוק של מודלים של עבודות, כמו באיור הבא (1.2):

****

**טופס מס' (1-2): שלבי יישום מודלים של עבודות**

במחקר זה נתייחס לארבעת השלבים הראשונים של יישום מודלים של עבודות ושני השלבים הנותרים יטופלו בשלישי.

* **שלב עיצוב הרזולוציה ותיאור הנתונים.**

המילים בערבית שהחלו בלחש נאספו על ידי עיצוב שאלון דיגיטלי (נספח א') ועוצבו באמצעות מודלים של גוגל ושמם (פתרון סיווג מילים טירונות עם קישור או חיתוך), משימת ההחלטה הוגבלה להרכבת כמה שיותר מילות סטארט-אפ על ידי מומחים בערבית סווגו על ידי מומחים כקישור או חיתוך בהתאם לכללי הדקדוק הידועים, והשאלון פורסם בכתובת האינטרנט הבאה:

[**https://docs.google.com/forms/d/e/1FAIpQLSdtwnflWQ7hoTdVbKfZUHDVix4fmXPE3grXMiQGIDem9Dsc0Q/viewform**](https://docs.google.com/forms/d/e/1FAIpQLSdtwnflWQ7hoTdVbKfZUHDVix4fmXPE3grXMiQGIDem9Dsc0Q/viewform)

הוא מולא על ידי 50 מומחי דקדוק בדרגות שונות במטרה להשיג מדגם הומוגני, ומספר המילים הכולל שסווג היה 400מילים. המילים הנגזרות מהטקסטים לדוגמה מסווגות למילים החל מקישור המוזכר בפרויקט זה על ידי Wasl ומילים המתחילות בלחש קטע המוזכר בפרויקט זה.

משתנים בלתי תלויים (מאפיינים) חולקו לשלושה מאפיינים בהתאם לכללי הדקדוק המכונים מאוחר יותר בטבלה (1.2).

תבנית הקובץ של ערכת הנתונים נבחרה כ- csv והכילה מאפיינים (משתנים בלתי תלויים):

* Variable :(diacritic) הוא משתנה דיגיטלי עם אורך של (1) מספר המציין את תנועת האות A בתחילת המילה (חריץ, שבר, חבילה) המיוצגת על-ידי ערכים דיגיטליים (1,2,3) בהתאמה.
* משתנה :(ספירה) הוא משתנה דיגיטלי המתייחס למספר אותיות המילה כמילה בערכים המכילים שניים, שלושה או ארבעה תווים ... וכו ', הם מיוצגים דיגיטלית בערכים (10... 4 או 3,2)בהתאמה.
* המשתנה :(שם עצם) הוא משתנה דיגיטלי המציין כי המילה (לא שם, שם רגיל, של עשרת השמות, מתוך ששת השמות, שם מחובר, שם פעיל) יוצגה על ידי מספרים (6, ... 3, 2, 1, 0, בהתאמה.
* משתנה :(פועל) הוא משתנה דיגיטלי המציין כי המילה (לא יעילה, מעשה עבר, פעולת פעולה, פעולת סדר) מיוצגת דיגיטלית

(0,1, 2,3) בהתאמה.

* המשתנה:(שם תואר) הוא משתנה דיגיטלי המציין שהמילה (לא שם תואר, מאפיין) מיוצגת דיגיטלית על-ידי מספרים (0,1) בהתאמה.
* האות היא משתנה דיגיטלי המצייןשהמילה (לא אות, תו) מיוצגת במספרים באופן דיגיטלי (0,1) בהתאמה.
* המשתנה (the)הוא משתנה דיגיטלי המצייןשהמילה (שאינה ידועה בשם Al, המכונה Al) יוצגה על-ידי מספרים באופן דיגיטלי (0,1) בהתאמה.
* קובץ Dataset נקרא מילים בערבית,טבלה (2-1) המציגה חלק מתיאור נתוני הפרופיל (Words\_Arabic. csv) להלן:

**תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי**

* **שלב סיווג נתונים**

אין ספק, כאשר איכות הנתונים נמוכה, זה ישפיע באופן בלתי נמנע על תוצאות הניתוח. בפרויקט זה נעשהשימוש במספר שיטותלניקוי נתונים בטקסטים שנאספו, ושלב ניקוי הנתונים כלל את השלבים הבאים:

* התמודדות עם נתונים שאבדו
* מחקנתונים כפולים.

לאחר ניקוי נתונים בטקסטים שנאספו, לאחר האצה ואי-הכללה של מילים כפולות וכן השלמת נתונים חסרים, מספר המילים שלא נכללו הפך ל- 101, קיבלנו 299 מילים חוקיות להחלה על טופס חוברת העבודה.

תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי

טבלה (2.2) מסבירה את הכללים לאיתור הקישור ולחתיכה של זמזום.

הנתונים בטבלה (1-2) מקודדים וערכים מילוליים מומרים לערכים דיגיטליים כך שאלגוריתמים יכולים לטפל בהם לפילוח זמנים 2-2, מה שמסביר את הכללים לאיתור הקישור ולחיתוך הלחישות והנתונים לאחר קידודו כמו בטבלה.

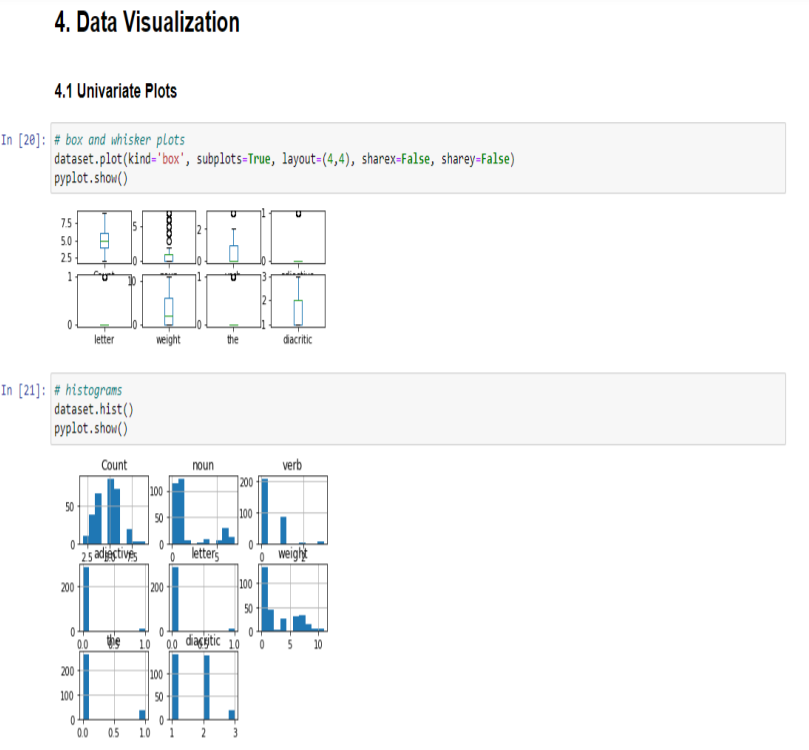
(3.2) הדברים הבאים:

תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי

קידוד נתונים בטבלה (2-3) בקובץ ערכת הנתונים(.csv בערבית)

לאחר השלמת תהליך קידוד הנתונים, הנתונים מיוצגים באמצעות Python וייבוא ספריות (sklearn-scipy-numpy) באמצעות העורך (Jupyter) כמו באיור (2-2) כדלקמן:



צורה (2-2 :)צג נתונים עם יומן באמצעותמחברת Jupyter)

* **שלב הבנייה וההכשרה של מודלים שלעבודות:**

הנתונים המיוצגים חולקו לנתוני הדרכה (67%) ולנתונים ניסיוניים (33%) מסך הנתונים (299) שנרשמו כהכנה לבניית מודל אימוני הנתונים באמצעות אלגוריתמי סיווג נבחרים (אלגוריתם תמיכה וקטורית SVM - אלגוריתם Naif Biz NB - אלגוריתם השכן הקרוב ביותר (KNN) שנבחר ביחס לגודל המדגם וערכי הנתונים עם האלגוריתמים הנ"ל והמודל נבנה באמצעות פונקציות ספרייתLearn-Skכמו באיור (2-3)a להלן:

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

**טופס (2-3): בניית מודלים לסיווג באמצעות עורך (Jupyter)**

**שלישית: תוצאות ותוצאות הפרויקט**

**ראשית: בדיקת מודלים של עבודות**

לאחר השלמת שלב הבנייה של מודלים של עבודות באמצעות אלגוריתמים ,(NB, SVM KNN) עברנו לשלב הבדיקה של מודלים אלה, כך שנוכל למדוד את איכותם מאוחר יותר, המודלים נבדקו על ידי ביצוע תחזיות על המודלים לאחר אימון באמצעות פונקציות הספרייה (sklearn),איור (3-1) מראה את המבחן של מודלים עבודות באמצעות פונקציית החיזוי כדלקמן:

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

טופס (3-1) מציג את המבחן של מודלים של עבודות באמצעות פונקציית חיזוי

**שנית: מדידת הדיוק של מודלים של עבודות**

הערכת התוצאות של מודלים לחקר נתונים היא שלב חשוב המאפשר לנו להגדיר את המודל היעיל ביותר, האפקטיביות של המודל נמדדת על ידי דיוק התוכנית במקום ואופי הנתונים המשמשים לבניית מודלים ממלא תפקיד מרכזי ביעילותם, ויש מסגרות סטטיסטיות רבות הבוחנות מודלים לסיווג, החשובות שבהן הן:

* **חישוב דיוק ממוצע של דיוק ממוצע**

האם הממוצע המתמטי של יחסי הדיוק החזויים הנכונים עבור כל קטגוריה המסופקת על ידי המודל למספר הסיווגים בפועל של קטגוריה זו בערכת נתוני הבדיקה, והדיוק הממוצע של התחזיות עבור האלגוריתמים המשמשים כפי שמוצג בטבלה (3.1) אשר מראה כי אלגוריתם SVM)יש את היחס הגבוה ביותר בעת חישוב הדיוק הממוצע שבו הוא הגיע לדיוק שלה (%94) בעוד האלגוריתם שלי (NB) ו

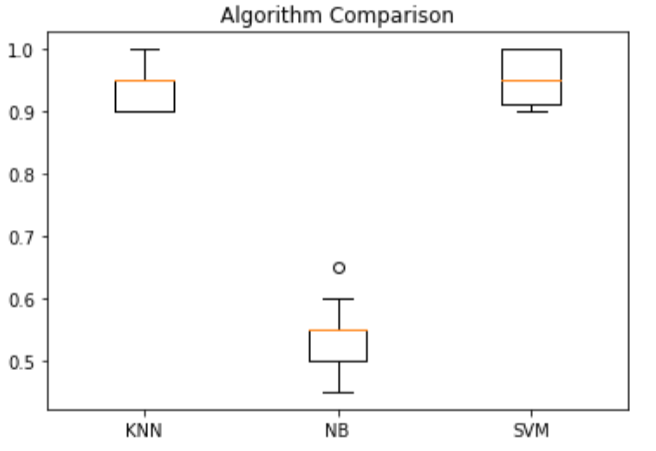
(KNN) דיוק ממוצע (54%) ו -(85%) בהתאמה:

תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי

**טבלה (1-3): דיוק ממוצע של אלגוריתם**

איור (3-2) מציג גרף של הדיוק הממוצע של אלגוריתמים (SVM), NB ו- KNN.



**צורה (2-3): גרף דיוק ממוצע של אלגוריתם**

* **חישוב דיוק כולל של טוטה דיוק**

היחס בין סך התחזיות הנכונות שסופקו על ידי הטופס לסך הסיווגים בפועל בערכת נתוני הבדיקה, הדיוק הכולל של האלגוריתמים המשמשים כפי שמוצג בטבלה 3.2)

תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי

**טבלה (2-3) דיוק כולל של אלגוריתמים**

מהטבלה לעיל, לאלגוריתם של knn יש דיוק כולל של 91% בהשוואה לאלגוריתם SVM, שהיה בעל שיעור הדיוק הכולל הגבוה ביותר של 92%, בעוד שלאלגוריתם NB היה שיעור הדיוק הכולל הנמוך ביותר של 78%.

* **חיפוש מטריצת בלבול של** מטריצה

מטריצת הבלבול מציגה את מספר האירועים הצפויים כראוי ואת מספר האירועים הצפויים בטעות בקבוצת הבדיקה עבור כל פריט בהשוואה למספר המקרים בפועל של פריטים אלה. הדירוג של מטריצה זו הוא n × n כאשר n הוא מספר הפריטים בעמודת משתנה היעד) משתנה הצאצא, ומטריצת הבלבול של האלגוריתמים המשמשים לסיווג משתנה התוצאה נמצאה לשתי קטגוריות (קישור wasl ורעד לחתוך (Gtaa כמו באיור (3-3) כדלקמן:

**תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי**

**איור (3-3): מטריצת הבלבול של אלגוריתמים(SVM, NB, KNN)המשמשת לסווג את הקישור ולחיתוך הזמזומים**

**שלישי: תוצאות**

השתמשנו באלגוריתם של מכונת תמיכה וקטור (SVM )נאיף ח'וורזמיה bes (נ.ב) ואלגוריתם. שכן הקרוב ביותר (KNN )מיון לחשת חיבור וחיתוך (ב תחילת המילה.) והוא להתקיים בחר באפשרות זו אלגוריתמים בהתבסס על מספר גורמים, הגודל המתאים ביותר קבוצה נתונים עם זה אלגוריתמים והרבה. מאפיינים מאומץ על זה ב תהליך הסיווג, להתקיים עבודה מודל הדרכה לכל אלו אלגוריתמים זה התבהר מ במהלך חישוב הדיוק של המודל עבור כל אלגוריתם החורג מאלגוריתם מכונת התמיכה וקטור (SVM ) שם קיבלתי את שיעור הדיוק הגבוה ביותר של הדגם הגיע %92, שהוא גבוה ומספיק. لפתרון בעיית החיפוש בסיסיוזה היה. בסיסי ב גישה אלגוריתםבאחוז גבוה זה נמצא המספר הגדול של מאפיינים להשתמש ב מיון (בוצע השתמש 8 משתנים עצמאית) מה שתרם ב יצר מודל הדירוג מציאותי יותר..

**רביעית: תפוקות מחקר והיתרונות שלמחקר:**

מטרת הפרויקט שהושלם הייתה לבנות מודל חכם המקטלג את הקישור והחיתוך (בתחילת המילה) באמצעות אלגוריתמים לסיווג נתונים על מנת לקבוע קריטריונים מדויקים ונכונים בכתיבה מדויקת של טקסטים בערבית כדי לתרום ולסייע בהתאמת הטכנולוגיה למדידת איכות אלגוריתמי הסיווג המפורסמים ביותר בהבחנה כדי לשרת את השפה הערבית, כמו גם מטרת המחקר בין הקשר לחיתוך לחישות בתחילת המילה, ומודל אינטליגנטי זה שתוכנן יכול לתרום לפיתוח אלגוריתמי הסיווג המפורסמים ביותר בשירות השפה הערבית, כמו גם למטרה של מחקר בין שני הלחישות לחיתוך בתחילת המילה, ומודל אינטליגנטי זה שתוכנן יכול לתרום לפיתוח אלגוריתמי הסיווג המפורסמים ביותר בשירות השפה הערבית. ערביתכדלקמן:

* המודל החדש יכול לשמש כדי לסקור את המחקר המדעי ולהבטיח כי הלחישה נמשכת במילים כראוי, תורם לשלומם של אלמנטים של מחקר מדעי.
* המודל החדש יכול לשמש כדי לסקור מאמרי חדשות שפורסמו ברשתות החברתיות השונות על ידי התקנתו בסעיף תוספים לדפדפן כדי לוודא כי הלחישה נמשכת כראוי בכל המילים של המאמרים שפורסמו, אשר תורם להתפתחותהתקשורת הערבית החדשה.
* המודל החדש יכול לשמש עם יישומי סיור ונחשב לחלק אותנטי של מערכות הפעלה ניידות כדי לוודא כי פקודות מתורגמות נכתבות בערבית והתחילו בלחשכראוי.
* ניתן להתאים את הדגם החדש למערכות תרגום המשמשות בכנסים, מטוסים ורכבות, ומסייעות להציג את הזמזום כראוי.

בשורה התחתונה, אנו יכולים להשתמש בדגם החדש ככלי תוכנה שניתן לשלב עם כל התוכנות והמערכות הטכניות המציגות טקסט קריא או כתוב בערבית באמצעות מערכות תוכנה כדי לסווג את הלחישה במילים המתחילות בקישור אובגזירה.

**המלצות:**

1. החל אלגוריתמים המתאימים להתמודדות עם מטרות עם קטגוריות סיווג מרובות באמצעות אלגוריתמים יצירת אינדקס מלבד אלה לאלתר מטרות עם קטגוריות בינאריותבלבד.
2. פתח ספריות sklearn לתמיכה מלאה בערבית.
3. פתח את המודל כדי לשמש כיצירה עבורטקסטים קוליים בערבית.

**שוליים:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

1. ד"ר עלאא מצטפא אל-היליס: בחינת דעות בערבית משפטים השוואתיים, כתב העת הערבי הבינלאומי לאינפורמטיקה, כרך 2, גיליון 4, 2013.
2. ד"ר ראנה זוהיר עבדול גאני אל עוביידי, ד"ר גידה עבדול עזיז אל-טאליב: מחקר השוואתי של אלגוריתמים לחקר דעות וניתוח רגשות ויישומים שלהם, כתב העת רפדיין למדעי המחשב ונפח המתמטיקה (12,) גיליון II, 2018.
3. אמל כאזם מירא,ת'אר יאר יאסים קטעה. 2019.) יישומיםשל בינה מלאכותית בחינוך מנקודת המבט שלההוראה באוניברסיטה.
4. אמל כאזם מירה וב עריכה: יאסים קטעה. )2019.) יישומים של בינה מלאכותית בחינוך מנקודת המבט שלההוראה באוניברסיטה.
5. ד"ר .M מוסטפא עובייד (2018). ניתוח מתקדם וחקר נתונים, בית המחשבה הערבי, מהדורה ראשונה, כרך 1,קהיר.
6. בורקוב, א. (2019). ספר למידת המכונה בן מאה העמודים (כרך א'). קנדה: אנדרי בורקוב.
7. פלורין, ז. (2011). כריית נתונים: מושגים, מודלים וטכניקות. ספרינגר-ורלג. ברלין היידלברג.
8. איאן, ה. וו ואייבה, פ' (2005). כריית נתונים: כלים וטכניקות מעשיות של למידת מכונה, מהדורה שנייה. אלסבייר סן פרנסיסקו: ארה"ב
9. ג'יאווי, ח', מישלין, ק' וג'יאן פ. (2012). כריית נתונים: מושגים וטכניקות, מהדורה שלישית. אלסבייר בע"מ: ארה"ב.
10. Kalyani, G. ו Jaya, א Lakshmi הערכת ביצועים של טכניקות סיווג שונות עבור גילוי חדירה. כתב העת להנדסת מחשבים (IOSRJCE).
11. Kalyani, G. ו Jaya, א Lakshmi הערכת ביצועים של טכניקות סיווג שונות עבור גילוי חדירה. כתב העת להנדסת מחשבים (IOSRJCE).
12. מייקל, ג'יי.איי.B וגורדון, ס.ל. (2004). טכניקות כריית נתונים לניהול שיווק, מכירות וקשרי לקוחות, מהדורה שנייה. ויילי הוצאה לאור, Inc. אינדיאנפוליס, אינדיאנה: ארה"ב.
13. אמל כאזם מירה וב עריכה: יאסים קטעה. (2019.) יישומים של בינה מלאכותית בחינוך מנקודת מבטם שלמורים באוניברסיטה.
14. ד"ר .M מוסטפא עובייד (2018). ניתוח מתקדם וחקר נתונים, בית המחשבה הערבי, מהדורה ראשונה, כרך 1,קהיר.
15. הסטטיסטיקאי האנגלי תומאס(חי בתקופה זו) 1761-1701לספירה(.. . ניסח מקרה מיוחד של התיאוריה המפורסמת הנושאת אתשמו, משפט ' בייס) למרות שזה לא פורסם בחייו אבל פורסם לאחר מותו על ידי ריצ'רד פרייס.
16. ד"ר.M מוסטפא עובייד, (2018) ניתוח מתקדם וחקר נתונים, בית המחשבה הערבי, מהדורה ראשונה, כרך 1,קהיר.
17. אמל כאזם מירה ובעריכה: יאסים קטעה. ) 2019.) יישומיםשל בינה מלאכותית בחינוך מנקודת המבט שלההוראה באוניברסיטה.
18. מוחמד בן כרם בן עלי, אבו אל-פדל, ג'מאל אלדין, בנו של אל-אנסארי האפריקאי. )1993.) 18 לשון הערבים. דאר סאדר - מהדורת ביירות: שלישי (1/17).
19. מוחמד בן עבדאללה, בנו של מאלכ אלטי אלג'יאני, אבו עבדאללה, ג'מאל אלדין. חקירה: עבד אל-מוני אחמד חרדי. (ד"ט). המרכז למחקר מדעי ולהתחדשות אוניברסיטת אום אלקורה בפקולטה למורשת אסלאמית של המהדורה של מכה: ראשון (3/1466). (
20. אבו אל-פתאח עוסמאן בן ג'וני אל-מוסולי, תחקיר: פאז פארס. (DT) לזרוח בערבית. בית ספרי התרבות. כווית-

(226-225).

1. מוחמד רפיק מומאן אלשבקי. (2015) פניני זהב בהסבר על הקדמת האי. עזה- מהדורה פלסטינית: ראשון (77).
2. עבד אל-חל מוחמד אלנקרטס, תחקיר: מוחמד ח'ליל ראס. (2003) מקיף בערבית. בית הספר הלאומי - מהדורת לבנון: ראשון (178-162).
3. טאהר בן מוחמד זועבי אלביריני מאלג'יריה, שיח'. (2019) . בית הספרים המדעיים - ביירות -לבנון.

**מקורות והפניות**

ראשון: מקורות ואזכורים ערביים

תמונה שמכילה טקסט, מסמך, צילום מסך

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, מסמך, צילום מסך

התיאור נוצר באופן אוטומטי

שנית: מקורות זרים והפניות

[13] בורקוב, א. (2019). ספר למידת המכונה בן מאה העמודים (כרך א'). קנדה: אנדרי בורקוב. [14] פלורין, ז. (2011). כריית נתונים: מושגים, מודלים וטכניקות. ספרינגר-ורלג. ברלין היידלברג.

[15] איאן, ה. וו ואייבה, פ' (2005). כריית נתונים: כלים וטכניקות מעשיות של למידת מכונה, מהדורה שנייה. אלסבייר בע"מ סן פרנסיסקו: ארה"ב.

[16] ג'יאוויי, ח', מישלן, ק. וג'יאן פ. (2012). כריית נתונים: מושגים וטכניקות, מהדורה שלישית. אלסבייר בע"מ: ארה"ב.

[17] קאליאני, ג' וג'איה, א. לקשמי הערכת ביצועים של טכניקות סיווג שונות לזיהוי חדירות. כתב העת להנדסת מחשבים (IOSRJCE).

[18] מייקל, ג'יי.איי.B וגורדון, ס.ל. (2004). טכניקות כריית נתונים לניהול שיווק, מכירות וקשרי לקוחות, מהדורה שנייה. ויילי הוצאה לאור, Inc. אינדיאנפוליס, אינדיאנה: ארה"ב.

**תוספת**  **(A)**

סיווג מאוית N של מילים טירון עם קישור או לחתוך

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

