**זיהוי התקפות כופר :**

דף הקדמה + דף עבודות קודמות ...

לאחר שקיבלנו רקע על התקפות כופר , הסוגים שקיימים כיום ודרכי פתרון לזיהוי האם הקובץ ננעל.  
כפי שציינתי בהקדמה , ישנם שני סוגים של התקפת כופר:  
נעילת הקובץ , שאין אפשרות לגשת אליו. לרוב זה מלווה בשינוי הסיומת של הקובץ.  
והסוג השני הינו הצפנת המידע בתוכו,   
שהקורא יכול לראות את המידע שהוא צריך.  
אך ללא יכולת להבין את המידע.

**חלק ראשון – קריפטוגרפיה**

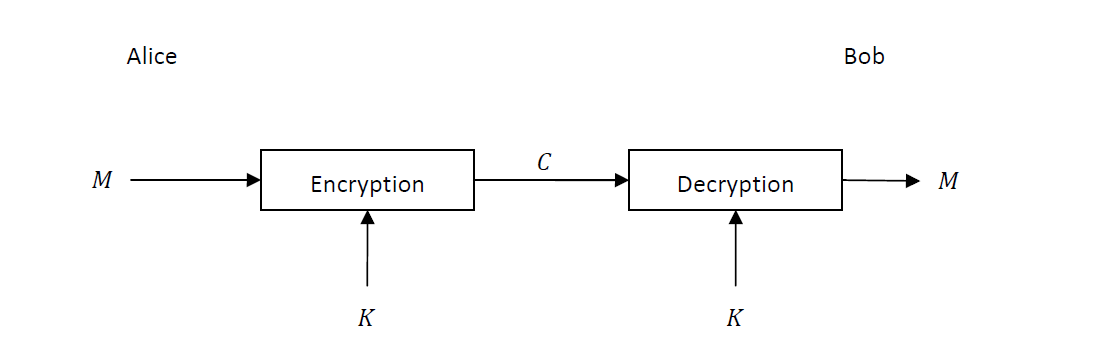
המילה קריפטוגרפיה מורכבת משתי מילים. *קריפטו* שפירושה סתר ו*גרפיה* שפירושה כתיב.

קריפטוגרפיה הוא תחום בן אלפי שנים ובמידה מסוימת הקריפטוגרפיה של היום היא כמו

הקריפטוגרפיה של פעם.  
כמו בכל תחום , ישנם אנשים שמנצלים את הענף לטובה וישנם אנשים שעושים בו שימוש זדוני, שעליהם נדבר במאמר.

במאמר זה נתמקד בתחום ההצפנות בקריפטוגרפיה ונסביר מספר סוגי הצפנות ואת ההבדלים ביניהם,

**הצפנה סימטרית:**  
בעולם ההצפנה הסימטרית ההנחה היא שלשני אנשים שרוצים לתקשר בצורה מוצפנת יש מפתח  
שבעזרתו הם יכולים להצפין הודעות ולפענח הודעות מוצפנות או לאמת הודעות



מספר הצפנות סימטריות נפוצות :

**צופן קיסר**

בתחום ה[קריפטוגרפיה](https://he.wikipedia.org/wiki/%D7%A7%D7%A8%D7%99%D7%A4%D7%98%D7%95%D7%92%D7%A8%D7%A4%D7%99%D7%94), **צופן קיסר**, הידוע גם כ**צופן היסט**, **קוד קיסר** או **היסט קיסר** או **הסטה קיסרית**, הוא אחד הצפנים הפשוטים והידועים בעולם ההצפנה. זהו סוג של [צופן החלפה](https://he.wikipedia.org/wiki/%D7%A6%D7%95%D7%A4%D7%9F_%D7%94%D7%97%D7%9C%D7%A4%D7%94) שבו כל אות בטקסט מוחלפת על ידי אות הנמצאת בהיסט קבוע כלשהו ממנה באלף-בית. למשל אם נקבע את ההיסט להיות 3, האות A תוחלף באות D, האות B תוחלף באות E וכך הלאה. הכינוי קיסר נובע מכך ש[יוליוס קיסר](https://he.wikipedia.org/wiki/%D7%99%D7%95%D7%9C%D7%99%D7%95%D7%A1_%D7%A7%D7%99%D7%A1%D7%A8) נהג להשתמש בצופן על מנת לתקשר עם מפקדיו.

**צופן ויז'נר**   
  
**צופן ויז'נר** הוא [צופן החלפה](https://he.wikipedia.org/wiki/%D7%A6%D7%95%D7%A4%D7%9F_%D7%94%D7%97%D7%9C%D7%A4%D7%94) רב־אלפביתי, המחליף כל [אות](https://he.wikipedia.org/wiki/%D7%90%D7%95%D7%AA) במסר באות אחרת מתוך [אלפבית](https://he.wikipedia.org/wiki/%D7%90%D7%9C%D7%A4%D7%91%D7%99%D7%AA) שונה, קרי במפתח שונה.[[1]](https://he.wikipedia.org/wiki/%D7%A6%D7%95%D7%A4%D7%9F_%D7%95%D7%99%D7%96%27%D7%A0%D7%A8#cite_note-1) השימוש במפתח נעשה באופן מחזורי. לאחר שימוש בכל האלפביתים חוזרים לאלפבית הראשון. מיקומה של כל אות במסר המקורי קובע באיזה אלפבית מתוך קבוצת האלפבית של מפתח הצופן להצפינה.[[2]](https://he.wikipedia.org/wiki/%D7%A6%D7%95%D7%A4%D7%9F_%D7%95%D7%99%D7%96%27%D7%A0%D7%A8#cite_note-2) בכל מפתח אלפבית סדר האותיות שונה, כך שכל אות זהה במסר תוצפן לאות אחרת בצופן, על כן לא נשמרת תדירות האותיות שבמסר המקורי, זאת בניגוד ל[צופן חד־אלפביתי](https://he.wikipedia.org/wiki/%D7%A6%D7%95%D7%A4%D7%9F_%D7%97%D7%93-%D7%90%D7%9C%D7%A4%D7%91%D7%99%D7%AA%D7%99) המבצע הזזה או החלפה של כל אותיות המסר במרחק קבוע. ועל כן את צופן ויז'נר לא ניתן לפצח בעזרת [ניתוח תדירויות](https://he.wikipedia.org/wiki/%D7%A0%D7%99%D7%AA%D7%95%D7%97_%D7%AA%D7%93%D7%99%D7%A8%D7%95%D7%99%D7%95%D7%AA).

**צופן אתב"ש**

**כתב אתב"ש**, הנקרא גם "**צופן אתב"ש**", הוא [צופן החלפה](https://he.wikipedia.org/wiki/%D7%A6%D7%95%D7%A4%D7%9F_%D7%94%D7%97%D7%9C%D7%A4%D7%94). בצופן זה מוחלפות האותיות הראשונות בסדר [האלפבית העברי](https://he.wikipedia.org/wiki/%D7%94%D7%90%D7%9C%D7%A4%D7%91%D7%99%D7%AA_%D7%94%D7%A2%D7%91%D7%A8%D7%99) באותיות האחרונות בו. לפי כלל הא-ת ב-ש, [אל"ף](https://he.wikipedia.org/wiki/%D7%90%D7%9C%22%D7%A3) מוחלפת באות [ת"ו](https://he.wikipedia.org/wiki/%D7%AA%22%D7%95), [בי"ת](https://he.wikipedia.org/wiki/%D7%91%D7%99%22%D7%AA) מוחלפת באות [שי"ן](https://he.wikipedia.org/wiki/%D7%A9%D7%99%22%D7%9F), [גימ"ל](https://he.wikipedia.org/wiki/%D7%92%D7%99%D7%9E%22%D7%9C) מוחלפת באות [רי"ש](https://he.wikipedia.org/wiki/%D7%A8%D7%99%22%D7%A9) וכן הלאה. הרעיון הוא לחלק את עשרים ושתיים אותיות האלפבית העברי (לא כולל אותיות סופיות) לשני טורים מקבילים - האחד מתחיל מאל"ף והשני מתחיל מת"ו כבתמונה משמאל:

במקום כל אות במילה אותה מצפינים, כותבים את האות המקבילה בטור ההפוך. לפיכך במקום המילה "שלום" נכתוב "בכפי". הפענוח נעשה בדרך זהה. בצופן החלפה, כאשר ידוע ההיסט, דהיינו הטבלה המתוארת, מלאכת הפענוח קלה מאוד. את ההיסט ניתן להשיג על ידי [ניתוח תדירויות](https://he.wikipedia.org/wiki/%D7%A0%D7%99%D7%AA%D7%95%D7%97_%D7%AA%D7%93%D7%99%D7%A8%D7%95%D7%99%D7%95%D7%AA) מופע של כל אות ואות. לרוב מספיק לפענח אות אחת (בתנאי שהמסר ארוך מספיק), בדר"כ זו בעלת תדירות ההופעה הגבוהה ביותר בשפה, ואז ניתן לחשב את ההיסט.

**צופן פלייפייר**  
בשיטת הצפנה זו מצפינים זוגות של אותיות ([דיגרפים](https://he.wikipedia.org/wiki/%D7%93%D7%99%D7%92%D7%A8%D7%A3)) במקום להצפין אותיות בודדות כמו שנהוג לעשות ב[צופן החלפה](https://he.wikipedia.org/wiki/%D7%A6%D7%95%D7%A4%D7%9F_%D7%94%D7%97%D7%9C%D7%A4%D7%94) פשוט וב[צופן ויז'נר](https://he.wikipedia.org/wiki/%D7%A6%D7%95%D7%A4%D7%9F_%D7%95%D7%99%D7%96%27%D7%A0%D7%A8). שיטת פלייפייר היא לכן משמעותית קשה ל[פריצה](https://he.wikipedia.org/wiki/%D7%A4%D7%99%D7%A6%D7%95%D7%97_%D7%A6%D7%A4%D7%A0%D7%99%D7%9D) מאחר ש[ניתוח תדירות](https://he.wikipedia.org/wiki/%D7%A0%D7%99%D7%AA%D7%95%D7%97_%D7%AA%D7%93%D7%99%D7%A8%D7%95%D7%AA) שמשמש להחלפה פשוטה לא עובד היטב עמו. ניתוח תדירות עדיין יכול להתבצע, אבל על 600 דיגרפים אפשריים במקום על 26 אותיות אפשריות, כלומר ניתוח תדירות הופך למשימה מורכבת בהרבה - ודורש באופן כללי טקסט מוצפן גדול כדי שיוכל להיות שימושי.  
  
**צופן Gronsfeld**

The Gronsfeld cipher is a variant created by Count Gronsfeld (Josse Maximilaan van [Gronsveld](https://en.wikipedia.org/wiki/Gronsveld" \o "Gronsveld) né van Bronckhorst); it is identical to the Vigenère cipher except that it uses just 10 different cipher alphabets, corresponding to the digits 0 to 9). A Gronsfeld key of 0123 is the same as a Vigenere key of ABCD. The Gronsfeld cipher is strengthened because its key is not a word, but it is weakened because it has just 10 cipher alphabets. It is Gronsfeld's cipher that became widely used throughout Germany and Europe, despite its weaknesses.

**הצפנה אסימטרית:**

בעולם האסימטרי לכל ישות יש זוג מפתחות – מפתח ציבורי ומפתח פרטי. המפתח הציבורי משמש

את שאר העולם כדי לשלוח הודעות ולאמת הודעות שהתקבלו ואילו המפתח הפרטי משמש לפענוח

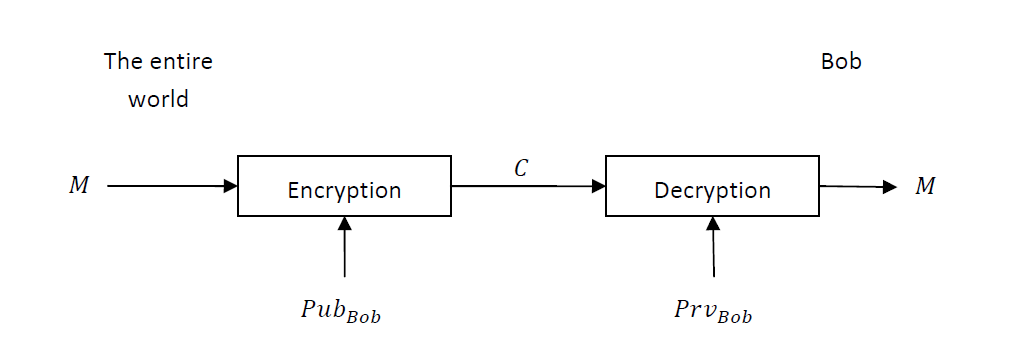
הודעות ולחתימה על הודעות יוצאות.

זה פותר את הבעייתיות של מספר המפתחות במקרה של קריפטוגרפיה סימטרית. כאן לאדם יש רק

שני מפתחות וכל העולם יכול לשלוח לו הודעות מוצפנות והוא יכול לשלוח הודעות חתומות וכל אחד

יכול לוודא את החתימה.

מצד שני, עדיין לא ברור איך מתבצעת הפצת המפתחות ציבוריים.



כל מי שרוצה לשלוח הודעה לבוב יכול להשתמש במפתח הציבורי של בוב וכדי לפענח בוב צריך

להשתמש במפתח הפרטי שלו. ההנחה היא כמובן שקשה מאוד לפענח הודעה מוצפנת בלי המפתח

הפרטי ושקשה מאוד לגלות את המפתח הפרטי מתוך ידיעת המפתח הציבורי

.

**הצפנת** **RSA**

**RSA** היא מערכת [הצפנת](https://he.wikipedia.org/wiki/%D7%A7%D7%A8%D7%99%D7%A4%D7%98%D7%95%D7%92%D7%A8%D7%A4%D7%99%D7%94) [מפתח ציבורי](https://he.wikipedia.org/wiki/%D7%9E%D7%A4%D7%AA%D7%97_%D7%A6%D7%99%D7%91%D7%95%D7%A8%D7%99) [דטרמיניסטית](https://he.wikipedia.org/wiki/%D7%90%D7%9C%D7%92%D7%95%D7%A8%D7%99%D7%AA%D7%9D_%D7%93%D7%98%D7%A8%D7%9E%D7%99%D7%A0%D7%99%D7%A1%D7%98%D7%99) מעשית הראשונה שהומצאה והיא עדיין בשימוש נרחב במערכות [אבטחת מידע](https://he.wikipedia.org/wiki/%D7%90%D7%91%D7%98%D7%97%D7%AA_%D7%9E%D7%99%D7%93%D7%A2) מודרניות, [תקשורת מחשבים](https://he.wikipedia.org/wiki/%D7%AA%D7%A7%D7%A9%D7%95%D7%A8%D7%AA_%D7%9E%D7%97%D7%A9%D7%91%D7%99%D7%9D) ו[מסחר אלקטרוני](https://he.wikipedia.org/wiki/%D7%9E%D7%A1%D7%97%D7%A8_%D7%90%D7%9C%D7%A7%D7%98%D7%A8%D7%95%D7%A0%D7%99). ב־RSA, כבכל מערכת מפתח ציבורי, מפתח ההצפנה אינו סודי והוא שונה ממפתח הפענוח שנשמר בסוד, על כן היא נקראת **אסימטרית**. האסימטריה ב־RSA נובעת מהקושי המעשי שב[פירוק לגורמים](https://he.wikipedia.org/wiki/%D7%A4%D7%99%D7%A8%D7%95%D7%A7_%D7%9C%D7%92%D7%95%D7%A8%D7%9E%D7%99%D7%9D_%D7%A9%D7%9C_%D7%9E%D7%A1%D7%A4%D7%A8_%D7%A9%D7%9C%D7%9D) של מספר פריק שהוא כפולה של שני [ראשוניים](https://he.wikipedia.org/wiki/%D7%9E%D7%A1%D7%A4%D7%A8_%D7%A8%D7%90%D7%A9%D7%95%D7%A0%D7%99) גדולים, שהיא [בעיה פתוחה](https://he.wikipedia.org/wiki/%D7%91%D7%A2%D7%99%D7%94_%D7%A4%D7%AA%D7%95%D7%97%D7%94_%D7%91%D7%9E%D7%AA%D7%9E%D7%98%D7%99%D7%A7%D7%94) ב[תורת המספרים](https://he.wikipedia.org/wiki/%D7%AA%D7%95%D7%A8%D7%AA_%D7%94%D7%9E%D7%A1%D7%A4%D7%A8%D7%99%D7%9D).

**חלק שני – מכונות למידה**

**למידת מכונה** (Machine Learning; לעיתים מכונה גם **למידה חישובית**) היא תת-תחום ב[מדעי המחשב](https://he.wikipedia.org/wiki/%D7%9E%D7%93%D7%A2%D7%99_%D7%94%D7%9E%D7%97%D7%A9%D7%91) וב[בינה מלאכותית](https://he.wikipedia.org/wiki/%D7%91%D7%99%D7%A0%D7%94_%D7%9E%D7%9C%D7%90%D7%9B%D7%95%D7%AA%D7%99%D7%AA) המשיק לתחומי ה[סטטיסטיקה](https://he.wikipedia.org/wiki/%D7%A1%D7%98%D7%98%D7%99%D7%A1%D7%98%D7%99%D7%A7%D7%94) וה[אופטימיזציה](https://he.wikipedia.org/wiki/%D7%90%D7%95%D7%A4%D7%98%D7%99%D7%9E%D7%99%D7%96%D7%A6%D7%99%D7%94_(%D7%9E%D7%AA%D7%9E%D7%98%D7%99%D7%A7%D7%94)). התחום עוסק בפיתוח [אלגוריתמים](https://he.wikipedia.org/wiki/%D7%90%D7%9C%D7%92%D7%95%D7%A8%D7%99%D7%AA%D7%9D) המיועדים לאפשר ל[מחשב](https://he.wikipedia.org/wiki/%D7%9E%D7%97%D7%A9%D7%91) ללמוד מתוך דוגמאות, ופועל במגוון משימות חישוביות בהן התכנות הקלאסי אינו אפשרי. אין לבלבל בין תחום זה, שבו המחשב הוא הלומד, ובין [למידה ממוחשבת](https://he.wikipedia.org/wiki/%D7%9C%D7%9E%D7%99%D7%93%D7%94_%D7%9E%D7%9E%D7%95%D7%97%D7%A9%D7%91%D7%AA), שבה המחשב משמש כעזר למידה על ידי הרצת [לומדה](https://he.wikipedia.org/wiki/%D7%9C%D7%95%D7%9E%D7%93%D7%94) או בדרך אחרת

To solve the employment problem (or similar problems), we build a machine learning model (called a classifier) that, when given a new instance (a user), predicts a label (whether she is employed or not).

Train / Test Split

* In order to train our classifier, we must have some labeled data (for example, we must know for some users whether they are employed or not).
* In order to evaluate the performance of our classifier, we split our data into train and test sets, train our classifier on the train set, and then test it on the test set.

Recall and Precision

|  |  |  |
| --- | --- | --- |
| **Confusion Matrix** | **Classified as Positive** | **Classified as Negative** |
| **Really Positive** | True Positive | False Negative |
| **Really Negative** | False Positive | True Negative |

* Accuracy?
  + Trues / All
  + (True Positive + True Negative) / (True Positive + True Negative + False Negative + False Positive)
* Recall
  + What fraction of positives did we actually find?
  + True Positive / Really Positive
  + True Positive / (True Positive + False Negative)
* Precision:
  + If we say positive, how precise are we?
  + True Positive / Classified as Positive
  + True Positive / (True Positive + False Positive)
* A combination of precision and recall:
* 2 (Precision \* Recall) / (Precision + Recall)
* E.g.:
  + Precision = 1, recall = 0,
    - F-Measure = 0
  + Precision = recall = 0.5
    - F-Measure = 0.5
  + Precision = recall = x
    - F-Measure = x

**מכונת למידה SVM**

**מכונת תמך וקטורי** (באנגלית **Support Vector Machine**, לרוב נכתב ונהגה כראשי-תבות **SVM**) היא טכניקה של [למידה מונחית](https://he.wikipedia.org/wiki/%D7%9C%D7%9E%D7%99%D7%93%D7%94_%D7%9E%D7%95%D7%A0%D7%97%D7%99%D7%AA) (supervised learning) המשמשת לניתוח נתונים ל[סיווג](https://he.wikipedia.org/wiki/%D7%A1%D7%99%D7%95%D7%95%D7%92_(%D7%A1%D7%98%D7%98%D7%99%D7%A1%D7%98%D7%99%D7%A7%D7%94)) ול[רגרסיה](https://he.wikipedia.org/wiki/%D7%A8%D7%92%D7%A8%D7%A1%D7%99%D7%94_(%D7%90%D7%A0%D7%9C%D7%99%D7%96%D7%94)).

כנהוג בתחום זה, דוגמאות האימון מיוצגות כווקטורים ב[מרחב ליניארי](https://he.wikipedia.org/wiki/%D7%9E%D7%A8%D7%97%D7%91_%D7%9C%D7%99%D7%A0%D7%99%D7%90%D7%A8%D7%99). עבור בעיות סיווג, בשלב האימון מתאימים מסווג שמפריד נכון ככל האפשר בין דוגמאות אימון חיוביות ושליליות. המסווג שנוצר ב-SVM הוא [המפריד הליניארי](https://he.wikipedia.org/w/index.php?title=%D7%9E%D7%A4%D7%A8%D7%99%D7%93_%D7%9C%D7%99%D7%A0%D7%99%D7%90%D7%A8%D7%99&action=edit&redlink=1) אשר יוצר מרווח גדול ככל האפשר בינו לבין הדוגמאות הקרובות לו ביותר בשתי הקטגוריות. כאשר מוצגת נקודה חדשה, האלגוריתם יזהה האם היא ממוקמת בתוך הקו המגדיר את הקבוצה, או מחוצה לו.

**מכונת למידה KNN**

**אלגוריתם השכן הקרוב** או ***k*-Nearest Neighbors algorithm** (או בקיצור **k-NN**) הוא אלגוריתם חסר [פרמטרים](https://he.wikipedia.org/wiki/%D7%A4%D7%A8%D7%9E%D7%98%D7%A8_%D7%A1%D7%98%D7%98%D7%99%D7%A1%D7%98%D7%99) ל[סיווג](https://he.wikipedia.org/wiki/%D7%A1%D7%99%D7%95%D7%95%D7%92_(%D7%A1%D7%98%D7%98%D7%99%D7%A1%D7%98%D7%99%D7%A7%D7%94)) ול[רגרסיה מקומית](https://he.wikipedia.org/wiki/%D7%A8%D7%92%D7%A8%D7%A1%D7%99%D7%94_%D7%9E%D7%A7%D7%95%D7%9E%D7%99%D7%AA).[[1]](https://he.wikipedia.org/wiki/%D7%90%D7%9C%D7%92%D7%95%D7%A8%D7%99%D7%AA%D7%9D_%D7%A9%D7%9B%D7%9F_%D7%A7%D7%A8%D7%95%D7%91#cite_note-1) בשני המקרים הקלט תלוי ב-k התצפיות הקרובות במרחב התכונות. k-NN יכול לשמש לסיווג או לרגרסיה:

* **k-NN לסיווג** – בהינתן קלט של דוגמה חדשה, האלגוריתם משייכה לקבוצה. הדוגמה משויכת למחלקה הנפוצה ביותר בקרב k השכנים הקרובים (כאשר k מוגדר כמספר חיובי שלם, בדרך כלל מספר קטן). אם k=1 האובייקט משויך למחלקה של השכן הבודד הקרוב ביותר.
* **k-NN לרגרסיה** – בהינתן דוגמה חדשה, האלגוריתם מחזיר ערך מאפיין לדוגמה. ערך זה הוא ממוצע ערכים של ערכי k השכנים הקרובים ביותר.

k-NN הוא אלגוריתם לימוד מבוסס מופעים, או [למידה עצלה](https://he.wikipedia.org/w/index.php?title=%D7%9C%D7%9E%D7%99%D7%93%D7%94_%D7%A2%D7%A6%D7%9C%D7%94&action=edit&redlink=1), שבו הפונקציה מקורבת באופן מקומי בלבד וכל החישובים נדחים עד סיווגה. אלגוריתם k-NN הוא מבין האלגוריתמים הפשוטים ביותר בתחום [למידת המכונה](https://he.wikipedia.org/wiki/%D7%9C%D7%9E%D7%99%D7%93%D7%AA_%D7%9E%D7%9B%D7%95%D7%A0%D7%94).

שקלול תרומתם של השכנים יכול להיות שימושי גם במקרה של סיווג וגם במקרה של רגרסיה, כך שמשקל השכנים הקרובים תורם יותר לממוצע מהשכנים הרחוקים יותר. לדוגמה שיטת [שקלול](https://he.wikipedia.org/w/index.php?title=%D7%A9%D7%A7%D7%9C%D7%95%D7%9C&action=edit&redlink=1) נפוצה מורכבת כך שנותנים לכל שכן משקל של ‎1⁄d, כאשר d הוא המרחק לאותו שכן.[[2]](https://he.wikipedia.org/wiki/%D7%90%D7%9C%D7%92%D7%95%D7%A8%D7%99%D7%AA%D7%9D_%D7%A9%D7%9B%D7%9F_%D7%A7%D7%A8%D7%95%D7%91#cite_note-2)

השכנים נלקחים מתוך סדרת אובייקטים של מחלקה (עבור k-NN לסיווג) או אפיון הערך (עבור k-NN לרגרסיה) ידועים. חיסרון בולט של האלגוריתם הוא רגישותו למבנה המקומי של הנתונים.

**חלק שלישי – DATASET**

בשביל ללמד את המכונה כמו שצריך, אנו צריכים מאגר מידע גדול כדי ללמד את המכונה.  
אז בכדי להשיג את המערכת זיהוי הטובה ביותר אימנו את המכונה שלנו על מאות אלפי קבצים.  
המידע תמיד היה מחולק לקבצים לא מוצפנים וקבצים מוצפנים.

חשוב לציין שכל המידע הלא מוצפן היה המון ספרים וקבצים טקסט אמיתיים

בנוסף , כל המידע המוצפן היה מוצפן באחוזים שונים בכדי ללמד את המכונה כמו שצריך ולהצליח להגיע לפיתרון אונליין. שתוכל להציל את הקובץ במהירות המירבית  
בכדי להגיע לתוצאות מיריביות במכונה, כל המידע חולק ל3 סוגים.  
1. רק תווים של אותיות באנגלית  
2. רק תווים ומספרים.  
3. כל סוגי התווים באסקי.

**חלק רביעי – הפתרון שלי**

לאחר שהבנו את מעט מתחום הקריפטוגריה ומעט מתחום הלמידת מכונה.  
אני יכול להסביר לכם את שיטת הפיתרון שלי.

לאחר חקירה מעמיקה של מאות אלפי קבצי טקסט, ראיתי לא מעט דפוסים שחזרו על עצמם בההצפנות.  
רובם ככולם של ההצפנות מגיעות במקרה אחד למילה שאין אפשרות בכלל לבטא ובמקרה השני למילה שלא קיימת במילון.

הפיתרון האידיאלי זה לעבור על כל המילים בטקסט, ואם זה מילה שקיימת במילון לשמור את זה כפיצר.  
פתרון זה מומלץ לעבודה אופליין על הקבצים. אך בכדי לבדוק את כל המילים במאות אלפי קבצים צריך לא מעט שבועות שהמכונה תעבוד כדי להגיע לפלט.  
כדי למצוא את האיזון בין תשובות כמה שיותר מדוייקות , אפילו עם הצפנה חלקית לבין מהירות הפיתרון.  
לאחר המון ניסויים שביצעתי, מצאתי את השילוב בין יעילות ונכונות.  
שילבתי מספר פיצרים למכונת למידה שהם :  
1. בדיקת השורות בכל קובץ טקסט  
2. דגימת 2 מילים משורה ובדיקה האם הם מילים באנגלית.  
3. בדיקה האם המבנה של המילה הוא תקין מבחינת מבנה המילה.  
4. האם זה מספר.  
5. האם זה מילה החשודה כמילה מוצפנת מהניסיון והקריאה על המון סוגי הצפנות.

את הפתרונות שלי חילקתי למספר גרפים שונים בכדי