

### Task1

1. בהינתן טקסט או קטע מתוכו, נרצה לזהות את כותב הטקסט.  
(בעזרת פתרון לבעיה הזאת נוכל לזהות גניבה ספרותית או העתקה.)
  2. נלך לכיוון של Machine learning, כלומר היינו רוצים להשתמש בסט דוגמאות אימון על מנת ללמוד את המאפיינים של כותב ספציפי ולהשתמש בהם כדי להכריע מיהו הכותב עבור טקסט חדש.  
אלגוריתמים אפשריים לשימוש: SVM, Decision Tree.  
בעזרת אלגוריתמים אלו נוכל לייצר קלאסיפיקציה מתאימה לפי הכותב.  
נוכל להפעיל את האלגוריתם ליצירת ה- decision tree תוך שימוש במאפיינים שונים ולהשוות בן התוצאות השונות.
  3. כדי לאפיין כותב ספציפי, נשתמש במאפיינים הנמצאים בטקסטים שהוא כתב, לדוגמא: שימוש בשפה עשירה (כמה השפה שלו מגוונת בטקסט), היחס בין מספר המילים הקצרות לעומת מספר המילים הארוכות וכדומה.
  4. את המאפיינים האלו נוכל להציג בצורה וקטורית ובעזרתם להפעיל את האלגוריתמים שהצגנו.
4. ניקח מאגר של טקסטים ונחלק אותם לשני חלקים, סט אימון וסט טסט. נפעיל את האלגוריתמים על סט האימון ואז נבחן אותו על ידי סט הטסט.

### חברי הקבוצה:

רועי שתיוי – roi1255 308192822

דנית ישעיהו – danity 312434269

נגה שטרן – nogastern 302388715

ורדית ארקש – vardit 305134140