הצעה למודל ML:

1. מספר הפיצ'רים יהיה כמספר המילונים.
2. עבור כל שורה ב-Data Set (שמייצגת הודעה שנכתבה בצ'אט) – ננקה את ההודעה לשלושה חלקים תחביריים: נושא, נשוא ומושא.
3. נשערך את ערכי הפיצ'רים ע"פ שיוך של כל חלק תחבירי למילון שמייצג את הפיצ'ר המתאים.
   לדוגמה: עבור הפיצ'ר שמייצג את המילון עבור מילים עם הקשר מתמטי – אם הנושא או הנשוא או המושא שייכים למילון הנ"ל, נסמן ב-1 את הפיצ'ר. יש עוד לוגיקות שאפשר לממש בהקשר הזה.

המצב כרגע: יש שלושה מילונים. בהינתן משפט אנו מבצעים חלוקה תחבירית של המשפט. שולחים כל אחד מהחלקים (נושא, נשוא ומושא) לפונקציה שבודקת לאיזה רגע קריטי המילה נקשרת ע"י בדיקה באיזה מילון המילה קיימת.
בהינתן תיוג של כל המילים במשפט, נחליט על תיוג המשפט באופן הבא:

1. אם לפחות אחת מהמילים תויגה כ-DS, נחזיר DS.
2. אם אף אחת מהמילים לא תויגה כ-MAT וגם TC, נחזיר NMD.
3. נחזיר את המקסימום מבין MAT ו-TC.

נשאר לבצע:

1. לנקות את המילים מתחיליות, רבים (ה' הידיעה וכד'). לדאוג שמכל מילה תחזור הצורה המקורית שלה.
2. לחשוב איך לשלב למידת מכונה בנתונים הקיימים שהוצאנו.
3. לכתוב טסטים.
4. לבדוק ביצועים.
5. לחבר בין החלק של יקיר ואחיעד לשלנו.