הערות של רונן מההצגה- 12.04

* **זמני השקעה** – יומי זה מהיר מידי, גם שבועי זה קצר מידי. אך רבעוני ארוך מידי – חודשי זה בדכ הה-happy medium
* **מודלים**- LSTM זה טוב בדכ לניבוי של מחירים. כדאי להתחיל עם זה, זה הימור טוב
* איך לעבוד עם CNBC: בעיקר לנסות לעבוד על הSENTIMENT. לנסות לקבל ציון משוקלל מFINBERT על כל כתבה.
  + פינברט לכל משפט זה יחסית פשטני. מה שמעניין זה רק **לזהות משפטים שיש ניבוי על כיוון השוק**
  + לבנות קלאספייר שיזהה משפטים כאלה, (מדברים על כיוון השוק, כיוון הכלכלה - רמת המאקרו). ואז עליהם להריץ FINBERT ואז זה כבר יהיה יותרת מדויק. ואז אם נגיד אנחנו רואים סדרה של ימים שמאוד שלילי, יודעים שכדאי לצאת מהשוק.
  + נצטרך לייצר TRAINING DATA של מפשטים שאנשים אומרים שהשוק הולך לעולות/ לרדת. אפשר לעשות חיפוש I PREDICT MARKET WILL GO DOWN וכן'. לנסות לזהות כמה שיותר משפטים כאלה. נגיד 50 שליליים ו-50 חיוביים ואז זה יהיה ה TRAINING DATA DATA . (אפשר גם לבדוק את האופציה של להשתמש GOOGLE TRENDS)
  + את הקלאסיפייר יש לבנות באמצעות FINBERT, לא בייס, LOG REG או אחרים, בגלל שברט תופס את הסמנטיקה, יעבוד יותר טוב. (ספציפית FINBERT).
* ב-**LABELS** – אין תשובה קלה להאם צריך להיות מולטיקלאס או בינארי. צריך לעשות ניסיונות. אולי רק עולה/ יורד, אולי להוסיף מחלקה שלישית של "כלום". BERT מאפשר הסתברויות, אך צריך להבין מה הTHRESHOLD שלנו.
* גם אפשרי להעשות ניבוי לאחור. (הצעה של איתן). אבל אז יכול להיות בעיה של overfitting , כי לא נתפוס את הפספוסים- איפה שניבו עליה אבל לא היה בסוף , לדוג'. אפשר להתחיל ככה ואז להרחיב את הPATTERN ולקוות שמצא עוד כאלה שמתאימים
* Semantic sentence similarity - יעזור לנו למצוא משפטים עם משמעות דומה. ככה אפשר להרחיב מדוג' אחת להרבה דוגמאות נוספות. כמה לינקים רלוונטים שרונן שלח (אך אפשר בהחלט לחפש עוד):
  + <https://datascience.stackexchange.com/questions/23969/sentence-similarity-prediction>
  + <https://github.com/topics/paraphrase-detection>
* כדאי לעבוד במקביל על האסטרטגיות (חלקות עבודה), וגם כדאי מאוד לעבוד ב-GITHUB, יעזור לעבוד בצורה מסודרת