פרוייקט DS סופי – חיזוי שער מניות תוך שימוש בעיבוד טקסט, מודלי TS, ורשתות ניורונים.

**משתנה נצפה:**

- האם המניה תעלה/תרד מחר?

- מה יהיה ערך המניה מחר ?

**משתנים מסבירים**

- נתוני עבר של אותה מניה

- נתוני עבר של מניות אחרות

- **כתבות – ניתוח סנטימנט**

- ציוצים – ניתוח סנטימנט

- חגים, ימים בשבוע, עונתיות

-

**עיבוד נתונים**

- נעשה חיזוי סנטימנט לכתבות ולציוצים

- מבחנים סטטיסטיים לקביעת סיבתיות גרנג'ר, ADF, סטטיונריות, והתאמת הנתונים למודלים

- נרמולים וכו'

**מודלים**

- מודל פייסבוק (GEM)

- מודל VAR אם נראה שזה מתאים

- מודל STEM שמקבל (1) את נתוני הבסיס (2)את תוצאות המודלים האחרים ומכין תוצאה משלו.

-

1. לאסוף הרבה מקורות מידע שיכולים להשפיע על מניות ואפשר לאסוף אותם בזמן אמת (כתבות/ציוצים/מניות אחרות עם לאג)

2. לבנות מודל LSTM שחוזה שני דברים: (א)מה יהיה המחיר מחר (ב)האם השוק יעלה מחר ביותר מ 0.5%, ומשתמש בפיצ'רים משלב 1

3. אולי לבנות מודל VAR או ARIMA, שני שיטות של כלכלנים שפחות מערבות ML

4. לשפר את 2 על הvalid

5. להכין "סימולציה", שלוקחת רק את המקרים בהם המודל היה בטוח ביותר מ 90% שמחר השוק יעלה ביותר מ 0.5%, ולבדוק מה היה קורה אם היינו בכל יום כזה קונים מניות ב 1000$ ומוכרים אותם ביום למחרת (לפי השערי סגירה).

**מקורות מידע**

כתבות:

<https://www.kaggle.com/datasets/snapcrack/all-the-news?select=articles3.csv>

<https://www.kaggle.com/datasets/rmisra/news-category-dataset>

<https://components.one/datasets>

**מניות**:

<https://www.kaggle.com/datasets/jimschacko/tesla-inc-last-5-years-stock-historical-data>

<https://finance.yahoo.com/quote/GOOG/history?period1=1433548800&period2=1591833600&interval=1d&filter=history&frequency=1d&guccounter=1&guce_referrer=aHR0cHM6Ly90b3dhcmRzZGF0YXNjaWVuY2UuY29tL3RpbWUtc2VyaWVzLWZvcmVjYXN0aW5nLXByZWRpY3Rpbmctc3RvY2stcHJpY2VzLXVzaW5nLWZhY2Vib29rcy1wcm9waGV0LW1vZGVsLTllZTE2NTcxMzJiNQ&guce_referrer_sig=AQAAAHE_6l6gCKYhddDIU8ntCG3Pn4KV6NMPIDTi6K1CYlYHVoc2iGySMjLHpSehcWnACgLg_nC-5Cis8MXMrU1LuNnIGVlcxJJAPlhnfSYt0hMkjQm-LtynbqQmMY8vsfUKDIEdiwuAs_5TI8e5uGdO8s53YdWnefqkDamPb5avyNir>

**מודלים**

מודל GEM עם המנגנון של פייסבוק:

<https://taufik-azri.medium.com/forecasting-through-economic-uncertainty-multivariable-time-series-analysis-with-var-and-prophet-e6b801962acb>

בין פייסבוק לVAR:

<https://thingsolver.com/forecasting-var-prophet/>

<https://stats.stackexchange.com/questions/481112/what-are-the-underlying-statistical-differences-between-the-vector-autoregressio>

חיזוי עם LSTM

<https://towardsdatascience.com/lstm-time-series-forecasting-predicting-stock-prices-using-an-lstm-model-6223e9644a2f>

מודל ARIMA

<https://towardsdatascience.com/time-series-forecasting-predicting-stock-prices-using-an-arima-model-2e3b3080bd70>

מודל עם הסקה בייסיאנית

<https://projecteuclid.org/journals/annals-of-applied-statistics/volume-16/issue-1/Inference-in-Bayesian-additive-vector-autoregressive-tree-models/10.1214/21-AOAS1488.short>