

פרויקט חיזוי מגמות בשוק ההון באמצעות DL, ML ו-NLP-

מטרת הפרויקט

בפרויקט זה תבצעו ניתוח נתונים של מחירי מניות וחדשות כלכליות, תבנו מודלים שונים לחיזוי מגמות שוק, ותשוו בין ביצועיהם. הפרויקט יכלול שימוש בטכניקות למידת מכונה (ML) למידה עמוקה (DL) ועיבוד שפה טבעית (NLP) לצורך חיזוי.

שלבי העבודה

1. טעינת הנתונים

- יש להוריד נתונים פיננסיים של מניה אחת לתקופה של שנה לפחות באמצעות yfinance.
 - יש להוריד נתוני חדשות פיננסיות מאותה תקופה ממקור חיצוני כמו Kaggle. למשל (<https://www.kaggle.com/datasets/notlucasp/financial-news-headlines>)
 - יש לוודא ששני מקורות הנתונים מכסים את אותה תקופה וניתנים להשוואה.
- תוצר נדרש:** שתי טבלאות נפרדות מסודרות (למשל ללא שדות ריקים, ללא tuples) – אחת עם נתוני המניה והשנייה עם החדשות הכלכליות.

2. חקר הנתונים (EDA)

- ניתוח מגמות ותבניות בנתוני המניה (ממוצעים, תנודתיות, קורלציות).
 - ניתוח בסיסי של כותרות החדשות (אורך טקסטים, מגמות, מילים נפוצות).
 - יצירת עמודת יעד (Target) בטבלת המניות – האם מחיר המניה עלה, ירד או לא השתנה יחסית ליום הקודם.
- ניתן לחשב זאת למשל באמצעות shift() ולסמן את היעד כ-1 לעליה ואפס ללא שינוי ו-1 לירידה
- תוצר נדרש:** טבלאות וגרפים עם תובנות על נתוני המניה והחדשות.

3. חיזוי מגמות באמצעות מודל קלאסי

- יש לבחור לפחות שני מודלים קלאסיים (כגון Logistic Regression / Random Forest)
- נתוני הקלט יהיו תכונות מספריות בלבד מתוך הנתונים הפיננסיים
- המודל ינסה לחזות האם המניה תעלה, תרד או תישאר זהה ביום הבא.

- יש להשוות ביצועים באמצעות מדדים מתאימים (למשל accuracy, F1-score)

תוצר נדרש: תוצאות השוואת ביצועים בין המודלים הקלאסיים.

4. חיזוי מגמות באמצעות מודל Time Series (LSTM)

- יצירת **חלונות זמן** של מספר ימים אחורה (המשתתפים יצטרכו לקבוע את מספר הימים בעצמם).
- אימון מודל לחיזוי מחיר הסגירה של היום הבא.
- השוואת הביצועים של מודל זה מול המודל הקלאסי.

תוצר נדרש: גרף השוואת מחירי אמת מול מחירים חזויים, כולל ניתוח היתרונות והחסרונות של הגישה.

5. יצירת מדד סנטימנט מכתורות החדשות

- יש להמיר את הכותרות לטקסט מספרי באמצעות **Tokenization + pad_sequences**.
- יש לבנות **מודל LSTM עצמאי** שיקבל כותרת חדשות ויוציא פלט **מדד סנטימנט רציף** בין 0 ל-1 שיהווה מדד סנטימנט עבור כל כותרת
- יש לבחור דרך לייצר **מדד סנטימנט יומי** (שכן בכל יום יש מספר כותרות) מתוך הכותרות של אותו יום (ממוצע, מקסימום/מינימום וכו').
- יש להוסיף **עמודת סנטימנט** לטבלת המניה.

תוצר נדרש: טבלה עם מדד סנטימנט יומי וגרף מגמות הסנטימנט.

6. שילוב הסנטימנט במודלים הקיימים

יש לבדוק האם הוספת מדד הסנטימנט משפרת את החיזוי:

- שילוב **מדד הסנטימנט** במודל הקלאסי (שלב 3).
- שילוב **מדד הסנטימנט** במודל סדרות הזמן (שלב 4).
- ניתוח ההשפעה של הסנטימנט על דיוק החיזוי.

תוצר נדרש: השוואת ביצועי המודלים לפני ואחרי הוספת הסנטימנט.

7. מסקנות והשוואת ביצועים

- השוואה בין כל המודלים שנבנו :

○ **מודל קלאסי ללא NLP**

○ מודל Time Series

○ מודלים עם שילוב הסנטימנט

• דיון על היתרונות והחסרונות של כל גישה.

• הצגת מסקנות מרכזיות מתוך הניתוח.

תוצר נדרש: טבלה מסכמת עם דיוק המודלים, גרפים מסכמים, ומסקנות סופיות.

דרישות ההגשה

• מחברת מסודרת (Jupyter Notebook / Colab) עם **קוד נקי ומוסבר**.

• ויזואליזציות תומכות EDA

• מצגת מסכמת (10-15 שקפים) עם תובנות עיקריות.

• השוואת ביצועים בין המודלים.

בהצלחה