Great Learning 10 YEARS

← Go Back to Introduction to Natural Language Processing

Course Content

Problem Statement - Stock Market News Sentiment Analysis and Summarization

Submission type	:	File Upload
Due Date	:	Sep 28, 4:30 AM MST
Total Marks	:	60
Available from	:	Sep 05, 7:30 AM

Description

Business Context

The prices of the stocks of companies listed under a global exchange are influenced by a variety of factors, with the company's financial performance, innovations and collaborations, and market sentiment being factors that play a significant role. News and media reports can rapidly affect investor perceptions and, consequently, stock prices in the highly competitive financial industry. With the sheer volume of news and opinions from a wide variety of sources, investors and financial analysts often struggle to stay updated and accurately interpret its impact on the market. As a result, investment firms need sophisticated tools to analyze market sentiment and integrate this information into their investment strategies.

Objective

With an ever-rising number of news articles and opinions, an investment startup aims to leverage artificial intelligence to address the challenge of interpreting stock-related news and its impact on stock prices. They have collected historical daily news for a specific company listed under NASDAQ, along with data on its daily stock price and trade volumes.

As a member of the Data Science and AI team in the startup, you have been tasked with analyzing the data, developing an AI-driven sentiment analysis system that will automatically process and analyze news articles to gauge market sentiment, and summarizing the news at a weekly level to enhance the accuracy of their stock price predictions and optimize investment strategies. This will empower their financial analysts with actionable insights, leading to more informed investment decisions and improved client outcomes.

Data Dictionary

* Date: The date the news was released
* News: The content of news articles that could potentially affect the company's stock price
* Open: The stock price (in $) at the beginning of the day
* High: The highest stock price (in $) reached during the day
* Low: The lowest stock price (in $) reached during the day
* Close: The adjusted stock price (in $) at the end of the day
* Volume: The number of shares traded during the day
* Label: The sentiment polarity of the news content
    * 1: Positive
    * 0: Neutral
    * -1: Negative

Important Note

- Please set the runtime to T4-GPU in Google Colab. Please follow the below instructions to the runtime to T4-GPU
    - Click on "Runtime" in the menu bar
    - Select "Change runtime type" from the dropdown menu
    - In the "Hardware accelerator" section, choose "GPU"
    - You may see multiple GPU options; choose "GPU" if you specifically want a T4 GPU
    - After selecting the GPU option, click on the "SAVE" button

Submission Guidelines

1. There are two ways to work on this project:

i. Full-code way: The full code way is to write the solution code from scratch and only submit a final Jupyter notebook with all the insights and observations.

ii. Low-code way: The low-code way is to use an existing solution notebook template to build the solution and then submit a business presentation with insights and recommendations.

The primary purpose of providing these two options is to allow learners to opt for the approach that aligns with their individual learning aspirations and outcomes. The below table elaborates on these two options.

Submission type	Who should choose	What is the same across the two	What is different across the two	Final submission file [IMP]	Submission Format
Full-code	Learners who aspire to be in hands-on coding roles in the future focussed on building solution codes from scratch	Perform exploratory data analysis to identify insights and recommendations for the problem	Focus on code writing: 10-20% grading on the quality of the final code submitted	Solution notebook from the full-code template submitted in .html format	.html
Low-code	Learners who aspire to be in managerial roles in the future-focussed on solution review, interpretation, recommendations, and communicating with business		Focus on business presentation: 10-20% grading on the quality of the final business presentation submitted	Business presentation in .pdf format with problem definition, insights, and recommendations	.pdf

Please follow the below steps to complete the assessment. **Kindly note that if you submit a presentation along with the notebook, ONLY the presentation will be evaluated.** Please make sure that all the sections mentioned in the rubric have been covered in your submission.

i. Full-code version

- Download the full-code version of the learner notebook.
- Follow the instructions provided in the notebook to complete the project.
- Clearly write down insights and recommendations for the business problems in the comments.
- Submit only the solution notebook prepared from the learner notebook [format: .html]

ii. Low-code version

- Download the low-code version of the learner notebook.
- Follow the instructions provided in the notebook to complete the project.
- Prepare a business presentation with insights and recommendations to the business problem.
- Submit only the presentation [format: .pdf]

2. Any assignment found copied/plagiarized with other submissions will not be graded and awarded zero marks.

3. Please ensure timely submission as any submission post-deadline will not be accepted for evaluation.

4. Submission will not be evaluated if

- it is submitted post-deadline, or,
- more than 1 file is submitted.

Best Practices for Full-code submissions

- The final notebook should be well-documented, with inline comments explaining the functionality of code and markdown cells containing comments on the observations and insights.
- The notebook should be run from start to finish in a sequential manner before submission.
- It is important to remove all warnings and errors before submission.
- The notebook should be submitted as an HTML file (.html) and NOT as a notebook file (.ipynb).
- Please refer to the FAQ page for common project-related queries.

Best Practices for Low-code submissions

- The presentation should be made keeping in mind that the audience will be the Data Science lead of a company.
- The key points in the presentation should be the following:
  - Business Overview of the problem and solution approach
  - Key findings and insights which can drive business decisions
  - Business recommendations
  - Focus on explaining the key takeaways in an easy-to-understand manner.
  - The inclusion of the potential benefits of implementing the solution will give you the edge.
- Copying and pasting from the notebook is not a good idea, and it is better to avoid showing codes unless they are the focal point of your presentation.
- The presentation should be submitted as a PDF file (.pdf) and NOT as a .pptx file.
- Please refer to the FAQ page for common project-related queries.

Happy Learning!

Scoring guide (Rubric) - Stock Market News Sentiment Analysis and Summarization - Project Rubric

Criteria	Points
Exploratory Data Analysis - Problem definition - Univariate analysis - Bivariate analysis - Use appropriate visualizations to identify the patterns and insights - Key meaningful observations on individual variables and the relationship between variables	8
Data Preprocessing - Split the target variable and predictors - Split the data into train, validation, and test sets	2
Word Embeddings - Using Word2Vec - Using GloVe - Using Sentence Transformer	10
Sentiment Analysis - Comment on which metric to use and why - Build a ML model with each of the following embedding techniques - Word2Vec - GloVe - Sentence Transformer - Perform hyperparameter tuning for the ML model for each of the following embedding techniques - Word2Vec - GloVe - Sentence Transformer - Comment on model performance across different metrics - Choose the best model from the ones built with proper reasoning - Check the performance of the final model on the test set Note: The ML model to use is at the discretion of the learner.	16
Content Summarization - Group the data at a week-level - Load the large language model from Hugging Face - Create a function to define the model parameters and generate a response - Define the instruction for the task (the task is to identify the top three positive and negative events from the week that are likely to impact the stock price) - Apply the response generation function to get an output from the model - Create a DataFrame containing the necessary fields from the model's output in a structured manner	12
Actionable Insights and Recommendations - Share your observations and insights from the analysis conducted - Provide recommendations for the business	4
Presentation/Notebook - Overall Quality - Structure and flow - Crispness - Visual appeal - All key insights and recommendations covered? OR - Structure and flow - Well commented code - All key insights and recommendations covered?	8
Points	60

Submit Assignment

Upload file

Add comments

Submit Assignment