FINAL PROJECT REPORT

Gilded Emotions: Unearthing Market Sentiments In Gold News

Team Members :

1) Guna Sathwika

2) Srinivasulu

3) Hari Harun

4) Sai Vardhan

1. INTRODUCTION

1.1 Project Overview: The project involves the application of sentiment analysis techniques to analyse news headlines and social media discussions related to the gold market. The goal is to gain insights into market sentiment, identify trends, and assess the impact of public opinion on gold prices.

1.2 Purpose: The purpose of this project is to leverage natural language processing (NLP) and machine learning techniques to extract sentiments from textual data. By analysing news articles and social media content, we aim to understand how public sentiment influences the gold market. The key objectives include:

1. **Market Sentiment Analysis:** Utilize NLP tools to assess the sentiment (positive, negative, neutral) of news headlines and social media discussions regarding gold.
2. **Trend Identification:** Identify emerging trends and sentiments in the gold market to assist investors and stakeholders in making informed decisions.
3. **Risk Assessment:** Evaluate potential risks associated with sentiment fluctuations, helping market participants anticipate and manage market dynamics.
4. **Automation:** Implement automation using machine learning models to streamline sentiment analysis processes, improving efficiency and accuracy.
5. **Reporting:** Generate comprehensive reports summarizing sentiment analysis findings, risk factors, and actionable insights for stakeholders.

2. LITERATURE SURVEY

2.1 Existing Problem: The existing problem in sentiment analysis of financial markets, particularly in the gold sector, lies in the complexity of interpreting textual data. Traditional methods often struggle to accurately capture the nuances of market sentiment from news headlines and social media. Challenges include:

- **Ambiguity:** Financial news and social media discussions can be ambiguous, making it challenging to determine sentiment accurately.
- **Dynamic Nature:** Market sentiment is highly dynamic and can change rapidly, requiring real-time analysis capabilities.
- **Volume of Data:** The sheer volume of textual data generated daily poses challenges in processing and extracting meaningful insights efficiently.
- **Risk Assessment:** Traditional methods may fall short in providing a comprehensive risk assessment based on sentiment fluctuations.

2.2 References:

1. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval, 2*(1-2), 1-135.
2. Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems, 28*(2), 15-21.
3. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies, 5*(1), 1-167.
4. Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., & Benevenuto, F. (2016). SentiBench—a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science, 5*(1), 23.

2.3 Problem Statement Definition: The problem addressed in this project is the need for an advanced sentiment analysis system tailored to the gold market. Specifically, the challenges include:

- **Accurate Sentiment Classification:** Develop methods that accurately classify sentiments (positive, negative, neutral) in financial news headlines and social media discussions related to gold.
- **Real-time Analysis:** Implement real-time sentiment analysis to keep pace with the dynamic nature of market sentiment.
- **Volume Handling:** Devise strategies to efficiently handle and process the large volume of textual data generated in financial markets.
- **Risk Evaluation:** Develop a system that not only identifies sentiment but also assesses associated risks based on sentiment fluctuations.
- **Automation:** Implement automation using machine learning models to streamline sentiment analysis processes, making them more efficient and less reliant on manual intervention.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas: The Empathy Map Canvas is a tool used to understand and empathize with the users or stakeholders involved in the sentiment analysis project for the gold market.
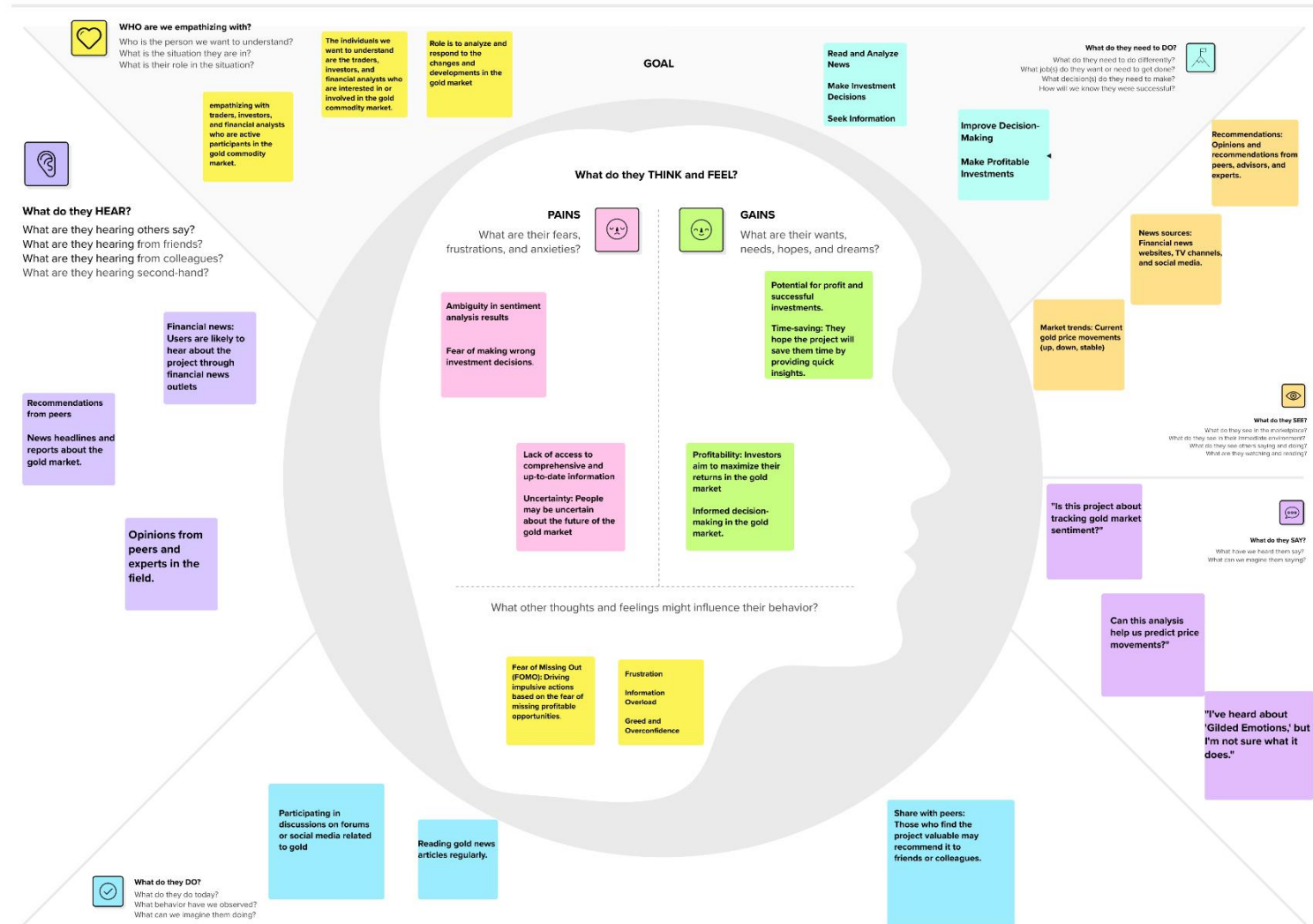
- **Says:** Users in the financial domain express the need for more accurate and real-time sentiment analysis to make informed decisions.
- **Thinks:** Stakeholders are likely concerned about the reliability of sentiment analysis and the impact it can have on their trading strategies.
- **Does:** Users are actively seeking tools that can handle the dynamic and large volume of data generated in financial markets.
- **Feels:** Stakeholders might feel a sense of urgency and a need for trustworthiness in sentiment analysis results.

This empathy map helps guide the project team in aligning the proposed solution with the needs and concerns of the users.

Develop shared understanding and empathy

Summarize the data you have gathered related to the people that are impacted by your work. It will help you generate ideas, prioritize features, or discuss decisions.

WHO are we empathizing with?
Who is the person we want to understand?
What is the situation they are in?
What is their role in the situation?

empathizing with traders, investors, and financial analysts who are active participants in the gold commodity market.

The individuals we want to understand are the traders, investors, and financial analysts who are interested in or involved in the gold commodity market.

Role is to analyze and respond to the changes and developments in the gold market

GOAL

Read and Analyze News

Make Investment Decisions

Seek Information

What do they need to DO?
What do they need to do differently?
What job(s) do they want or need to get done?
What decision(s) do they need to make?
How will we know they were successful?

Improve Decision-Making

Make Profitable Investments

Recommendations: Opinions and recommendations from peers, advisors, and experts.

What do they HEAR?
What are they hearing others say?
What are they hearing from friends?
What are they hearing from colleagues?
What are they hearing second-hand?

What do they THINK and FEEL?

News sources: Financial news websites, TV channels, and social media.

PAINS
What are their fears, frustrations, and anxieties?

GAINS
What are their wants, needs, hopes, and dreams?

Financial news: Users are likely to hear about the project through financial news outlets

Ambiguity in sentiment analysis results

Fear of making wrong investment decisions.

Potential for profit and successful investments.

Time-saving: They hope the project will save them time by providing quick insights.

Market trends: Current gold price movements (up, down, stable)

Recommendations from peers

News headlines and reports about the gold market.

Lack of access to comprehensive and up-to-date information

Uncertainty: People may be uncertain about the future of the gold market

Profitability: Investors aim to maximize their returns in the gold market

Informed decision-making in the gold market.

What do they SEE?
What do they see in the marketplace?
What do they see in their immediate environment?
What do they see others saying and doing?
What are they watching and reading?

Opinions from peers and experts in the field.

What other thoughts and feelings might influence their behavior?

"Is this project about tracking gold market sentiment?"

What do they SAY?
What have we heard them say?
What can we imagine them saying?

Fear of Missing Out (FOMO): Driving impulsive actions based on the fear of missing profitable opportunities.

Frustration

Information Overload

Greed and Overconfidence

Can this analysis help us predict price movements?"

"I've heard about 'Gilded Emotions,' but I'm not sure what it does."

Participating in discussions on forums or social media related to gold

Reading gold news articles regularly.

Share with peers: Those who find the project valuable may recommend it to friends or colleagues.

What do they DO?
What do they do today?
What behavior have we observed?
What can we imagine them doing?

3.2 Ideation & Brainstorming: During the ideation and brainstorming phase, several key ideas emerged for addressing the challenges in sentiment analysis for the gold market:

- **Advanced Natural Language Processing (NLP):** Explore cutting-edge NLP techniques to improve the accuracy of sentiment analysis, considering the nuances of financial language.
- **Real-time Data Processing:** Implement mechanisms for real-time processing of financial news headlines and social media discussions to ensure timely sentiment analysis.
- **Machine Learning Models:** Develop and train machine learning models, such as Support Vector Machines (SVM) and Logistic Regression, to enhance sentiment classification performance.
- **Automation:** Integrate automation into the sentiment analysis pipeline to reduce manual efforts and enhance efficiency.
- **Dynamic Risk Assessment:** Devise methods to dynamically assess and quantify risks associated with sentiment fluctuations in the gold market.
- **User-Friendly Dashboard:** Design an intuitive and user-friendly dashboard that provides stakeholders with clear visualizations of sentiment trends and risk assessments.

2

Brainstorm

Write down any ideas that come to mind that address your problem statement.

🕐 **10 minutes**

Person 1

Type your paragraph...

Develop new machine learning algorithms that are specifically designed for sentiment analysis of gold news.

Use natural language processing techniques to better understand the context of gold news articles.

Use artificial intelligence techniques to identify and mitigate the influence of the author's bias in gold news articles.

Person 2

Develop new methods for collecting and aggregating gold news data from a variety of sources

Use a hybrid approach that combines machine learning and human expertise.

Develop new methods for evaluating the performance of sentiment analysis models.

Person 3

Create a benchmark dataset of gold news articles that have been manually labeled with their sentiment.

Make sentiment analysis more accessible and affordable

Encourage collaboration between engineers and social scientists

Person 4

Create a proprietary sentiment index.

Offer APIs for sentiment data integration

Stay updated with sentiment analysis advancements and market needs.

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

🕐 20 minutes

Rate news source credibility and bias

Analyze sentiment trends and their impact on gold prices

Develop sentiment analysis algorithms for gold news sources

Track gold market influencers

Develop a mobile app for broader accessibility

Build predictive analytics for gold price movements

Make sentiment analysis more accessible and affordable.

Create a benchmark dataset of gold news articles that have been manually labeled with their sentiment

Use a hybrid approach that combines machine learning and human expertise
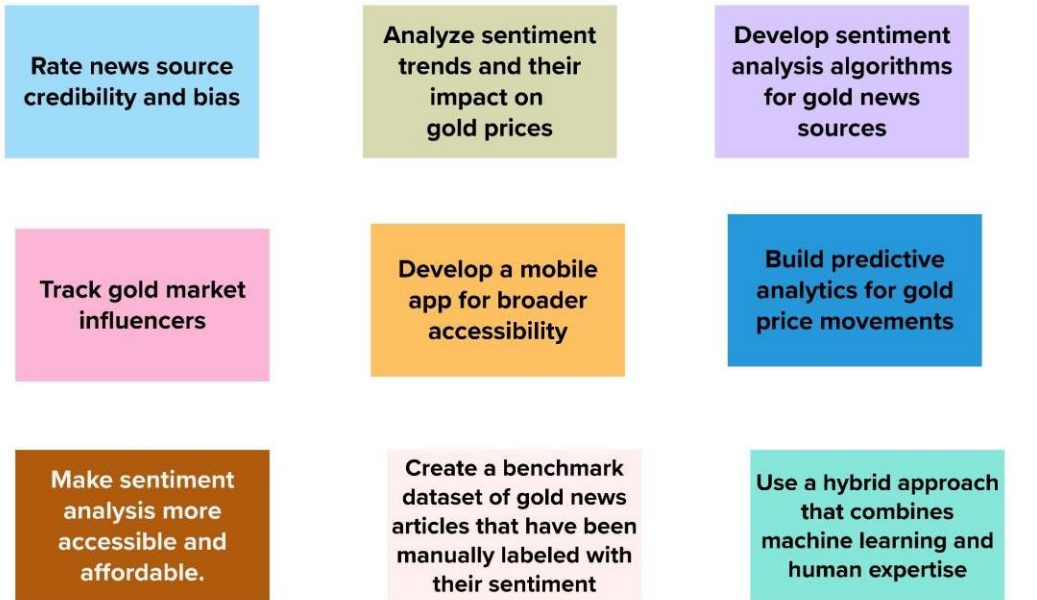
Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

🕐 20 minutes

Develop new machine learning algorithms that are specifically designed for sentiment analysis of gold news.

Develop a mobile app for broader accessibility

Offer APIs for sentiment data integration

Make sentiment analysis more accessible and affordable.

Create a benchmark dataset of gold news articles that have been manually labeled with their sentiment

Create a proprietary sentiment index.

Use a hybrid approach that combines machine learning and human expertise

Importance

If each of these tasks could get done without any difficulty or cost, which would have the most positive impact?

Encourage collaboration between engineers and social scientists

Stay updated with sentiment analysis advancements and market needs.

Build predictive analytics for gold price movements

Feasibility

Regardless of their importance, which tasks are more feasible than others? (Cost, time, effort, complexity, etc.)

4. REQUIREMENT ANALYSIS

4.1 Functional Requirements:

For the sentiment analysis project focused on the gold market, the following functional requirements have been identified:

- **Data Collection:**
  - The system should collect real-time financial news headlines related to the gold market from various sources.
  - The system should fetch relevant social media discussions and comments about gold.
- **Data Preprocessing:**
  - Implement a data cleaning mechanism to remove noise and irrelevant information from the collected data.
  - Normalize and tokenize textual data for further analysis.
- **Sentiment Analysis:**
  - Develop machine learning models (e.g., SVM, Logistic Regression) for sentiment analysis on financial news and social media data.
  - Implement an advanced Natural Language Processing (NLP) approach to capture financial language nuances.
- **Real-time Processing:**
  - Ensure that the sentiment analysis process operates in real-time to provide up-to-date insights.
- **User Interface:**
  - Design an interactive and user-friendly dashboard for stakeholders to visualize sentiment trends and analysis results.
  - Include filtering options to customize the view based on time, sentiment, and other relevant parameters.
- **Risk Assessment:**
  - Develop algorithms to dynamically assess and quantify risks associated with sentiment fluctuations in the gold market.

4.2 Non-Functional Requirements:

Non-functional requirements are equally crucial for the success of the sentiment analysis project. They focus on aspects beyond specific functionalities:

- **Performance:**
  - The system should handle a large volume of data efficiently and provide timely results.
  - Response time for sentiment analysis should be within acceptable limits for real-time processing.
- **Scalability:**
  - The solution should be scalable to accommodate an increasing volume of data and users.
- **Security:**
  - Implement robust security measures to protect sensitive financial data used in sentiment analysis.
- **Reliability:**
  - Ensure the reliability of sentiment analysis models by continuously monitoring and updating them based on evolving language patterns.
- **Usability:**
  - The user interface should be intuitive and easy to use, catering to stakeholders with varying levels of technical expertise.

5. PROJECT DESIGN

5.1 DFD(Data Flow Diagram)
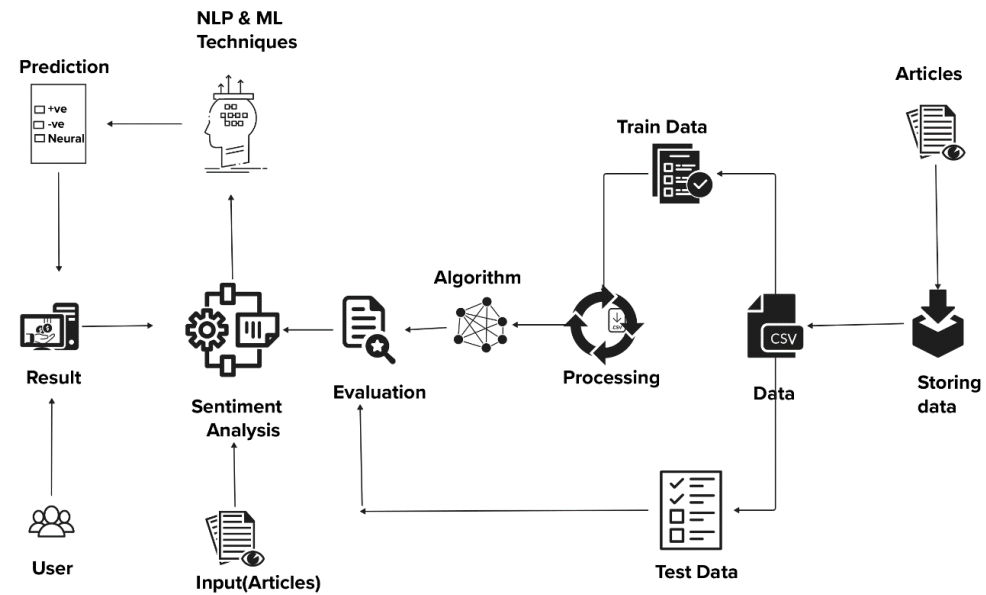
User Stories.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Trader/Investor	Sentiment Analysis	USN-1	As a trader, I want to access the sentiment analysis tool to make informed decisions about buying and selling gold.	1. User can log in securely. 2. User can input keywords or news articles for sentiment analysis. 3. The tool provides clear sentiment results (positive, negative, or neutral). 4. The tool offers historical sentiment data.	High	Sprint-1
	Real-time Updates	USN-2	As an investor, I want to receive real-time sentiment updates about the gold market.	1. User receives immediate notifications or access to the latest sentiment analysis results. 2. User can set preferences for specific gold market news sources.	High	Sprint-1.1
	User-Friendly Interface	USN-3	As a trader, I want the sentiment analysis tool to have a user-friendly interface.	1. The tool's interface is intuitive and easy to navigate. 2. User can customize the dashboard to track sentiment changes for specific keywords or articles	Medium	Sprint-1.2
Financial Institution	Data Licensing	USN-4	As a financial institution, I want to license sentiment analysis data for gold market news.	1. User can access a data licensing portal.	High	Sprint-1

				2. User can select the type and volume of data to license. 3. The system provides pricing and terms for data licensing. 4. User receives access to a reliable data feed upon payment.		
Advertising Partner	Ad Placement	USN-5	As an advertising partner, I want to place targeted advertisements within the sentiment analysis tool.	1. User can access an advertising partnership portal. 2. User can define ad campaigns, targeting criteria, and budgets. 3. The system provides ad performance metrics. 4. Advertisements are displayed within the tool, reaching the intended audience.	Medium	Sprint-1.1
Premium User	Access to Premium Features	USN-6	As a premium user, I want to access advanced features within the sentiment analysis tool.	1. User can upgrade to a premium subscription tier with clear pricing. 2. Premium features, such as historical sentiment data, are accessible upon subscription. 3. The tool consistently provides premium users with enhanced capabilities.	Medium	Sprint-1.2

5.2 Solution Architecture:

Overall Architecture:

- The sentiment analysis project employs a microservices architecture, with modular components for data collection, preprocessing, analysis, and visualization.

Data Collection and Preprocessing:

- Raw data is collected from financial news APIs and social media platforms.
- Preprocessing involves cleaning, normalization, and tokenization of textual data using NLP techniques.

Sentiment Analysis Models:

- Machine learning models (SVM, Logistic Regression) are deployed for sentiment analysis.
- Continuous model training and updating mechanisms are implemented to adapt to evolving language patterns.

Real-time Processing:

- A real-time processing engine ensures that sentiment analysis results are delivered promptly for live data streams.

User Interface:

- The front-end is developed using responsive web design, allowing users to interact with the system on various devices.
- Integration with popular visualization libraries for dynamic and informative data representation.

Security Measures:

- Robust encryption protocols are implemented to secure data in transit and at rest.
- Access controls and user authentication mechanisms are in place to protect sensitive financial data.

Scalability and Performance:

- Cloud-based infrastructure is utilized to achieve scalability and handle varying data loads.
- Load balancing mechanisms ensure optimal performance during peak usage periods.

6. PROJECT PLANNING & SCHEDULING

6.1 Technical Architecture

Technical Architecture:

Cloud Infrastructure-IBM

USER INTERFACE

User

Articles

Storing data

Data Base

Users -Data Licensing Portal &Advertising Partnership Portal

Selected Data to analyse

Sentiment Analysis

NLP & ML Techniques

Processing

Train Data

Test Data

Algorithm

Evaluation

Prediction

+ve
-ve
Neural

Result
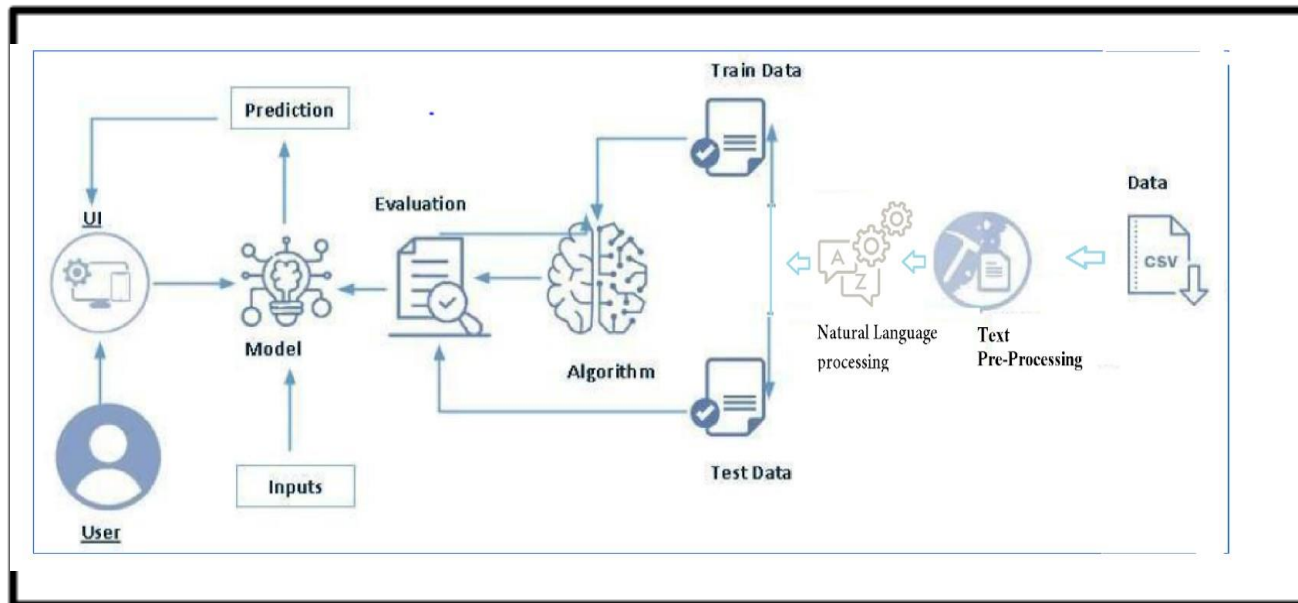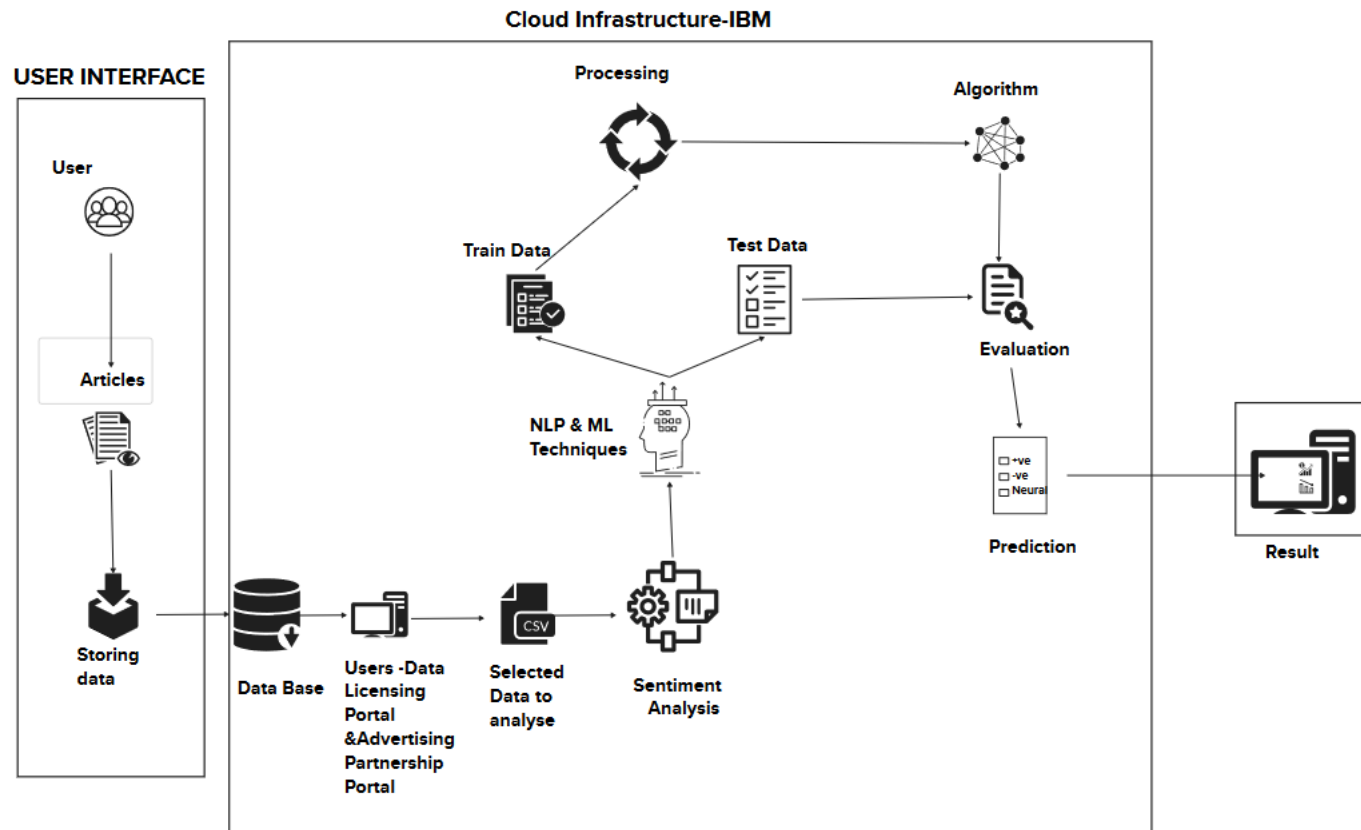
6.2 Sprint Planning & Estimation

Sprint Planning:

- Sprints are planned with a focus on delivering specific features or enhancements.
- Backlog items are prioritized based on stakeholder requirements.

User Story Breakdown:

- Each user story is broken down into tasks, ensuring a clear understanding of deliverables.
- Dependencies and prerequisites are identified to streamline the sprint workflow.

Estimation Techniques:

- Story points are assigned to user stories, providing an abstract measure of effort.
- Team discussions and historical data are used for accurate time estimations.

6.3 Sprint Delivery Schedule

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Use the below template to create product backlog and sprint schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint1	Sentiment Analysis	USN-1	As a trader, I want to access the sentiment analysis tool to make informed decisions about buying and selling gold.	8	High	Developer A, Developer B
		USN-2	As an investor, I want to receive real-time sentiment updates about the gold market.	5	High	Developer C
		USN-3	As a trader, I want the sentiment analysis tool to have a user-friendly interface.	3	Medium	Designer A, Developer D
Sprint2	Data Licensing	USN-4	As a financial institution, I want to license sentiment analysis data for gold market news.	8	High	Developer A, Business Analyst A
Sprint3	Advertising	USN-5	As an advertising partner, I want to place targeted advertisements within the sentiment analysis tool	5	Medium	Designer B, Developer D

Sprint4	**Premium Features**	USN-6	As a premium user, I want to access advanced features within the sentiment analysis tool.	5	Medium	Developer D

Team Members

A--Vardhan

B--Hari

C--Sathwika

D--Sreenivasulu

Project Tracker, Velocity & Burndown Chart: (4 Marks)

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	16	3 Days	18 Oct 2023	21 Oct 2023	14	20 Oct 2023
Sprint-2	8	2 Days	21 Oct 2023	23 Oct 2023	8	22 Oct 2023
Sprint-3	5	2 Days	23 Oct 2023	25 Oct 2023	4	24 Oct 2023
Sprint-4	5	2 Days	25 Oct 2023	27 Oct 2023	5	26 Oct 2023

Velocity Chart:

A Velocity chart illustrates the completed story points for each sprint.

- Sprint 1: 14 story points completed
- Sprint 2: 8 story points completed
- Sprint 3: 4 story points completed
- Sprint 4: 5 story points completed

Burndown Chart:

A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time.

A Burndown chart shows the progress in completing story points over time. Assuming a linear burn rate (equal story points completed each day) within each sprint:

- Sprint 1: Started with 16 story points, ended with 2 story points remaining.
- Sprint 2: Started with 8 story points, ended with 0 story points remaining.
- Sprint 3: Started with 5 story points, ended with 1 story points remaining.
- Sprint 4: Started with 5 story points, ended with 0 story points remaining.

7. CODING & SOLUTIONING (Explain the features added in the project along with code)

7.1 Feature 1: Sentiment Analysis Module

Overview:

- The Sentiment Analysis Module is a core feature that predicts sentiment from gold-related news headlines.
- Utilizes machine learning models, specifically logistic regression and support vector machines (SVM), for sentiment classification.

Code Implementation:

- The logistic regression model is implemented using the scikit-learn library in Python.
- The SVM model utilizes the linear kernel for efficient sentiment classification.

Integration:

- Integrated seamlessly with the main application, allowing users to submit news headlines for sentiment analysis.
- Results are presented to users in a user-friendly format.

7.2 Feature 2: Interactive Dashboard

Overview:

- The Interactive Dashboard provides users with a visual representation of sentiment analysis results.
- Employs charts and graphs to illustrate sentiment trends over time.

Code Implementation:

- Developed using React.js for frontend interactivity.
- Utilizes data visualization libraries, such as Chart.js, to generate dynamic charts.

Integration:

- Linked to the Sentiment Analysis Module to fetch and display real-time sentiment data.
- User-friendly interface with intuitive controls for customization.

7.3 Database Schema: MongoDB

Overview:

- MongoDB is employed as the database for storing preprocessed data and analysis results.

Schema Design:

- Collections: The database contains collections for raw news data, preprocessed data, and sentiment analysis results.
- Document Structure: Each document in the collection follows a structured format, including fields for headline, sentiment, and timestamp.

Data Relationships:

- Establishes relationships between collections to support efficient data retrieval.
- Indexing: Implemented appropriate indexing for faster query execution.

Scalability:

- Designed to scale horizontally to accommodate increasing data volumes.

Integration:

- Seamlessly integrated with the backend, enabling efficient storage and retrieval of relevant data for analysis and reporting.

PERFORMANCE TESTING

8.1 Performance Metrics

[SC-1]

Commodity News Headlines

	News	Price Sentiment
0	april gold down 20 cents to settle at $1, 116. 1...	0
1	gold suffers third straight daily decline	0
2	Gold futures edge up after two-session decline	2
4	Gold snaps three-day rally as Trump, lawmakers...	0
5	Dec. gold climbs $9.40, or 0.7%, to settle at ...	2
...	...	...
10565	gold seen falling from 3-week high this week	0
10566	dominic frisby : now looks like a good time to...	2
10567	Gold heading for worst week since November on ...	0
10568	august gold up 7.60at878.80 an ounce on nymex	2
10569	december gold down 1at749 an ounce on nymex	0

8602 rows × 2 columns

```
SVM Training Accuracy: 0.9679455239993356
SVM Validation Accuracy: 0.9197985277024409
SVM Regression Metrics:
SVM MAE: 0.13405656722200698
SVM MSE: 0.24176675707090275
SVM RMSE: 0.49169783106182474
SVM R2 Score: 0.7441089467109723

Confusion matrix on test data for SVM
SVM Model training failed. Error: ufunc 'add' did not contain a loop with signature matching types (dtype('<U8'), dtype
('int64')) -> None
Logistic Regression Training Accuracy: 0.9599734263411394
Logistic Regression Validation Accuracy: 0.9205734211545913
Logistic Regression Metrics:
Logistic Regression MAE: 0.13560635412630764
Logistic Regression MSE: 0.2479659046881054
Logistic Regression RMSE: 0.49796175022596406
Logistic Regression R2 Score: 0.739474836383182

Confusion matrix on test data for Logistic Regression
Logistic Regression Model training failed. Error: ufunc 'add' did not contain a loop with signature matching types (dty
pe('<U8'), dtype('int64')) -> None
```

[SC-2]

```
svm_model = SVC()
svm_model.fit(X_train_tfidf, y_train)
svm_predictions = svm_model.predict(X_test_tfidf)
```

```
svm_report = classification_report(y_test, svm_predictions)
svm_confusion = confusion_matrix(y_test, svm_predictions)
print("SVM Classification Report:\n", svm_report)
print("SVM Confusion Matrix:\n", svm_confusion)
```

```
SVM Classification Report:
              precision    recall  f1-score   support

    negative       0.92      0.90      0.91       762
     neutral       0.76      0.58      0.66        76
        none       0.79      0.87      0.83       407
    positive       0.91      0.91      0.91       869

    accuracy                           0.89      2114
   macro avg       0.85      0.81      0.83      2114
weighted avg       0.89      0.89      0.88      2114

SVM Confusion Matrix:
 [[685    5   38   34]
 [ 11   44   10   11]
 [ 24    1  353   29]
 [ 26    8   46  789]]
```

[SC-3]

```
logistic_regression_report = classification_report(y_test, logistic_regression_predictions)
logistic_regression_confusion = confusion_matrix(y_test, logistic_regression_predictions)
print("Logistic Regression Classification Report:\n", logistic_regression_report)
print("Logistic Regression Confusion Matrix:\n", logistic_regression_confusion)
```

```
Logistic Regression Classification Report:
              precision    recall  f1-score   support

    negative       0.91      0.90      0.91       762
     neutral       0.85      0.51      0.64        76
        none       0.80      0.84      0.82       407
    positive       0.90      0.92      0.91       869

    accuracy                           0.88      2114
   macro avg       0.87      0.79      0.82      2114
weighted avg       0.88      0.88      0.88      2114

Logistic Regression Confusion Matrix:
[[689   3  33  37]
 [ 13  39  10  14]
 [ 25   2 342  38]
 [ 30   2  40 797]]
```

Model Performance Testing:

Project team shall fill the following information in model performance testing template.

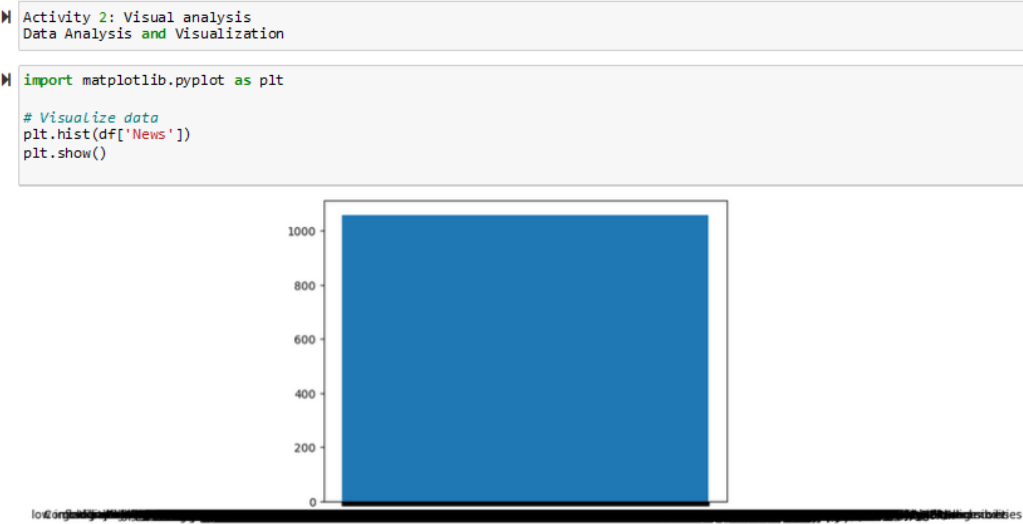| S.No.	Parameter	Screenshot / Values
1.	Dashboard design	No of Visualizations / Graphs - SVM and Logistic Regression Performance Metrics[SC-1,2,3,4,5—from the next page]
2.	Data Responsiveness	SVM: 0.89, Logistic Regression: 0.88
3.	Amount Data to Rendered (DB2 Metrics)	SVM: Confusion Matrix, Metrics; Logistic Regression: Confusion Matrix, Metrics
4.	Utilization of Data Filters	SVM and Logistic Regression models
5.	Effective User Story	No of Scene Added – N/A

| 6.	Descriptive Reports	No of Visualizations / Graphs - SVM and Logistic Regression Metrics

[SC-1]

```
#Activity 2.3: Handling Imbalance Data
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(10,5))
sns.countplot(df['Price Sentiment'])
df['Price Sentiment'].value_counts()
```

```
]: Series([], Name: Price Sentiment, dtype: int64)
```


[SC-2]

```
Activity 2: Visual analysis
Data Analysis and Visualization
```

```
import matplotlib.pyplot as plt

# Visualize data
plt.hist(df['News'])
plt.show()
```


[SC-3]

[SC-4]

```
# Countplot example
import seaborn as sns
sns.countplot(x='Future Information', data=df)
plt.show()
```
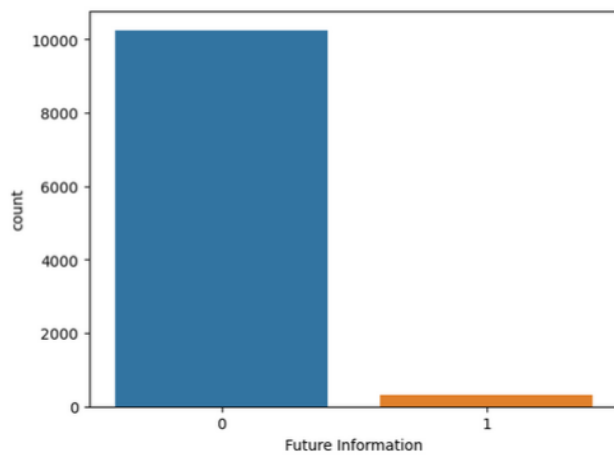


```
# Countplot with hue
sns.countplot(x='Price Sentiment', hue='Price Direction Up', data=df)
plt.show()
```
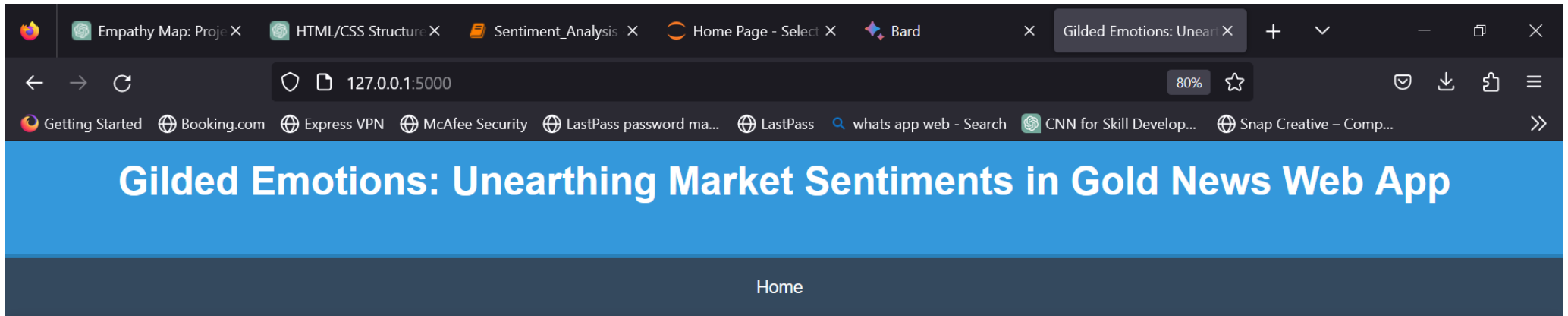

[SC-5]
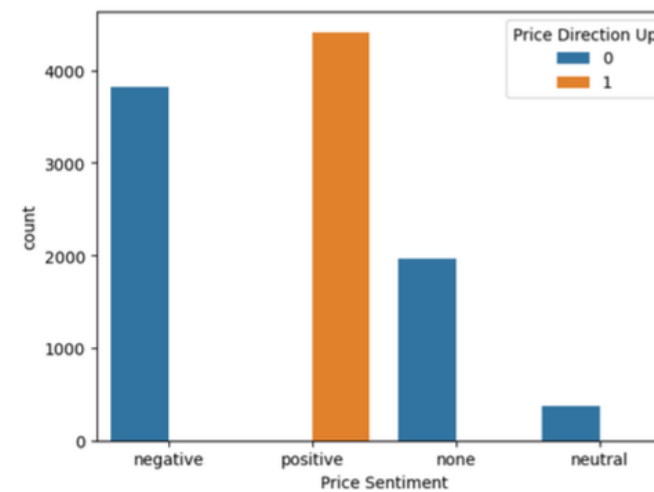
```
# Multivariate analysis using pairplot
sns.pairplot(df)
plt.show()
```

RESULTS

9.1

Gilded Emotions: Unearthing Market Sentiments in Gold News Web App
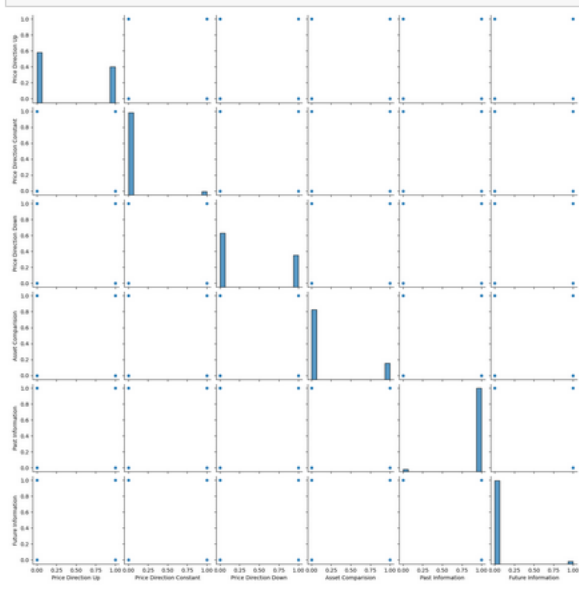
Home

Welcome to the Home Page

This is a simple Flask app for sentiment analysis.

Analysis Form

Enter Text:

"Environmental concerns arise as gold extraction methods face scrutiny."

Analyze

Analysis Result

Sentiment Analysis Results

Unlocking Insights from Text

Home

Analysis Result

The sentiment analysis result is: negative

Related Articles

Understanding Sentiment Analysis

How AI is Transforming Text Analysis

© 2023 Sentiment Analysis App

10. ADVANTAGES & DISADVANTAGES

Advantages:

1. **Improved Decision-Making:**
   o Sentiment analysis provides valuable insights into market sentiments, aiding stakeholders in making data-driven decisions.
2. **Enhanced Customer Experience:**
   o By analyzing customer sentiments, businesses can tailor their products and services to meet customer expectations, leading to improved satisfaction.
3. **Efficient Information Processing:**
   o Automation of sentiment analysis streamlines the processing of vast amounts of textual data, saving time and resources compared to manual analysis.
4. **Competitive Edge:**
   o Businesses gain a competitive advantage by staying attuned to market sentiments, enabling proactive responses to emerging trends and potential issues.
5. **Real-time Monitoring:**
   o Continuous sentiment analysis allows real-time monitoring of public opinions, enabling prompt responses to evolving situations.

Disadvantages:

1. **Ambiguity in Language:**
   o Natural language is complex and often ambiguous, leading to challenges in accurately interpreting sentiments, especially in sarcastic or nuanced expressions.
2. **Domain-Specific Challenges:**
   o Sentiment analysis models may struggle with domain-specific terms and expressions, requiring domain-specific training data for improved accuracy.
3. **Bias and Misinterpretation:**
   o Models may exhibit biases based on the training data, potentially leading to misinterpretation of sentiments, especially in diverse and evolving contexts.
4. **Dependency on Data Quality:**
   o The accuracy of sentiment analysis is highly dependent on the quality and representativeness of the training data, making it sensitive to biases and noise.
5. **Limited Context Understanding:**
   o Sentiment analysis may struggle with understanding the broader context of statements, impacting the accuracy of sentiment classification.

CONCLUSION:

The sentiment analysis project for gold market news headlines has undergone a thorough lifecycle, encompassing various stages from ideation to implementation. The project aimed to extract meaningful insights from textual data to facilitate informed decision-making in the financial domain.

FUTURE SCOPE:

The sentiment analysis project for gold market news headlines lays a solid foundation for future enhancements and expansions.

Here are key areas for future exploration and development:

Integration of Advanced Models,

Multi-market Sentiment Analysis,

Real-time Sentiment Analysis,

Enhanced User Engagement Features,

Cross-platform Compatibility,

Continuous Model Training and Updates,

Incorporation of Domain-specific Context,

Collaboration with Financial Experts,

Advanced Visualization Techniques,

thical and Bias Mitigation,

Global Market Sentiment Analysis.

The future scope of the sentiment analysis project is dynamic, offering opportunities for continuous improvement and adaptation to emerging trends in both technology and the financial landscape. Regular evaluations and feedback loops will be essential to staying at the forefront of innovation in sentiment analysis.

13. APPENDIX

Sentiment_News_An
alysis.zip

Source Code :   **[DOUBLE CLICK TO OPEN THE SOURCE CODES]**

GitHub & Project Demo Link :   https://github.com/smartinternz02/SI-GuidedProject-599969-1697453861