International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 1, Jan-Feb-2025, ISSN (Online): 2395-566X

EMO Diary: Daily Diary with Sentiment Analysis

Dr.Kavitha Soppari, Sk Hussain, Gvn Surya, M Pulla Rao
Department of CSE (Artificial Intelligence and Machine Learning), ACE Engineering College, India.

Abstract- The "Daily Diary Writer with Sentiment Analysis" project is a full-stack web application focused on enhancing personal well-being through sentiment analysis. Users can write daily diaries journals, and the system uses natural language processing to analyze the emotions expressed in their entries. This helps users reflect on their emotional patterns over time. Voice input allows for hands-free journaling, making the process more convenient and accessible. The project promotes self-reflection and emotional well-being through detailed sentiment insights.

Index Terms- Daily diary, sentiment analysis, web development, natural language processing, voice input.

I. INTRODUCTION

In an increasingly digital world, self-reflection and emotional awareness have become crucial components of mental well-being. Many individuals turn to journaling as a way to process thoughts and emotions. However, without consistent analysis, it can be challenging to identify emotional patterns and understand personal growth over time. This project, “Daily Diary Writer with Sentiment Analysis,” addresses this need by combining journaling with natural language processing (NLP) for sentiment analysis.

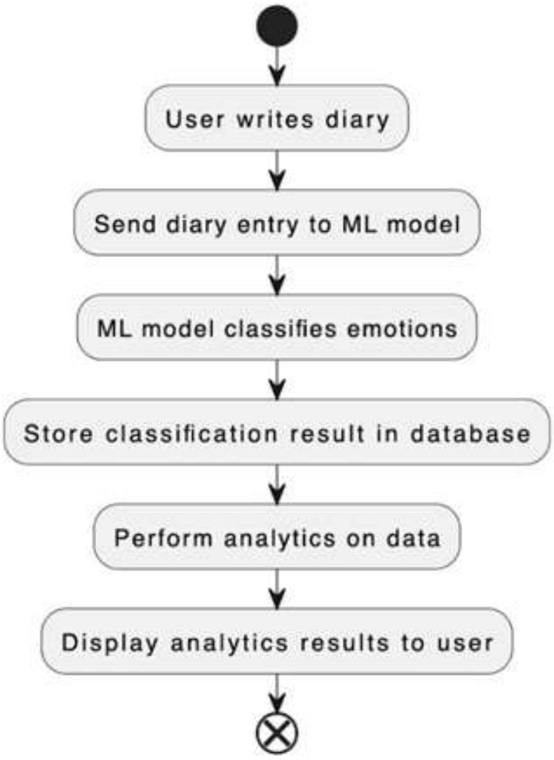

Figure 1: Overall Flow

The application allows users to record daily diary entries, either through text input or voice commands, offering a hands-free, accessible experience. The embedded sentiment analysis model evaluates each entry to detect and analyze emotional states, providing users with meaningful feedback on their emotional patterns. This insight aims to support users in tracking their mental health, recognizing trends, and ultimately fostering a journey of self-reflection and emotional well-being.

Users can use the identified trends to understand and address their emotional challenges.

Workflow Overview: From Diary Entry to Sentiment Insights

The EmoDiary application processes user diary entries through a streamlined, data-driven workflow to deliver meaningful emotional insights:

- **Diary Entry Creation:** Users write daily diary entries, which serve as raw input data reflecting their emotions and thoughts.
- **Emotion Classification:** Each entry is sent to a machine learning model that analyzes the text using sentiment analysis. The model classifies the emotions expressed within the diary entry.
- **Data Storage:** The classified emotions and historical record of each user’s emotional patterns.
- **Analytics Generation:** The stored data undergoes analytical processing to uncover trends, patterns, and insights into the user’s emotional well-being over time.
- **Insightful Visualization:** The analytics results are displayed to the user in a visually engaging format, enabling them to track their emotional journey, identify trends, and gain valuable self-awareness.

International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 1, Jan-Feb-2025, ISSN (Online): 2395-566X

II. PROPERTIES

1. Personalized User Experience
The application should use Firebase Authentication to offer a personalized experience, allowing each user to securely log in and access their personal diary entries, classifications, and analytics.

2. User-Friendly Interface
The platform should provide an intuitive and smooth interface for users to easily write and manage diary and journal entries. Features like text and voice input should be included to enhance the user experience and convenience.

3. Comprehensive Emotion Detection
The sentiment analysis model should use diverse datasets for emotion classification, including emotions, finance, toxic language, mental health, and tweets.

These datasets help capture a broad spectrum of emotions for accurate classification across various contexts.

4. Comparative Study of Algorithms
A thorough comparative study of machine learning algorithms, such as Logistic Regression, Random Forest, SVM, and KNN, should be conducted to identify the best-performing algorithm for each dataset.

The selection should focus on the algorithm that delivers the highest accuracy and reliability for emotion classification.

5. Flask-Based Backend
A Flask-based backend server should be developed to handle sentiment classification requests efficiently, ensuring quick responses. This backend will process diary entries, run sentiment analysis, and manage data flow between the front end and the database.

6. Firebase Database for Data Storage
The Firebase database will be used to store classification results securely, allowing users to retrieve their past entries and classifications. This ensures data consistency and reliable access to past records.

7. Effective Analytics
The application should provide visually engaging and meaningful analytics based on users' sentiment data.

These analytics will allow users to track their emotional patterns over time and gain valuable insights into their mental well-being.

Technique for Model Evaluation

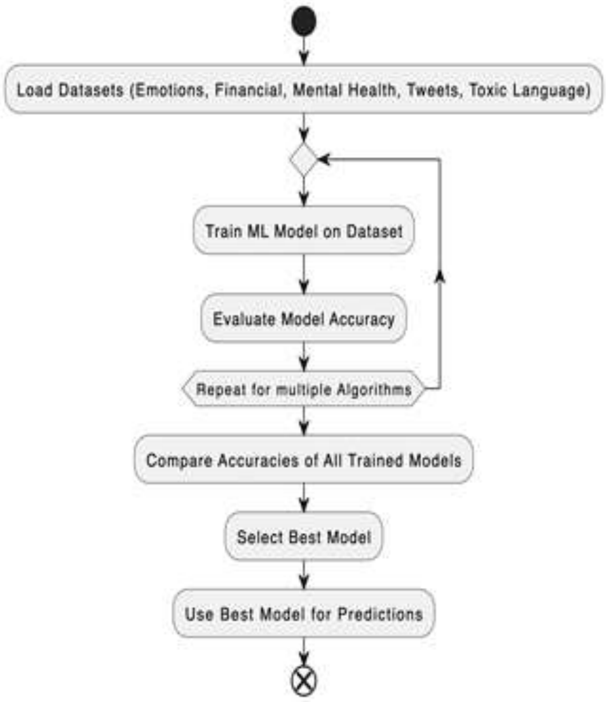

Figure 2: Model Selection Flow

To ensure accurate and reliable sentiment analysis in the "EmoDiary" application, we adopt a comprehensive approach to model evaluation. Since multiple datasets are used, we train several machine learning models on each dataset, experimenting with various algorithms to identify the best-performing model. The following techniques are applied to evaluate and select the optimal model:

Data Splitting

- Each dataset is split into training, validation, and test sets to allow for initial model training and independent performance evaluation.
- The training set is used to fit the model, while the validation set helps fine-tune parameters and avoid overfitting. Finally, the test set provides an unbiased assessment of the model's performance.

TF-IDF Vectorization

- For text data, we employ TF-IDF (Term Frequency-Inverse Document Frequency) vectorization. This technique transforms diary entries into numerical representations based on word frequency, helping the model recognize important terms and their significance.
- TF-IDF helps in weighting more informative words higher than common but less meaningful words, thus improving the model's ability to capture sentiment in textual data.

Evaluation Metrics

- We use multiple evaluation metrics to assess model performance comprehensively, as accuracy alone may not suffice, especially if the data is imbalanced. The primary metrics used include:
- Accuracy: The ratio of correct predictions to the total number of predictions, providing a basic performance measure.
- Precision: The ratio of true positive predictions to the total predicted positives, indicating the relevance of positive predictions.
- Recall: The ratio of true positives to the total actual positives, showing the model's sensitivity to identifying positive emotions.
- F1 Score: The harmonic mean of precision and recall, useful for datasets with class imbalance.
- AUC-ROC (Area Under the Receiver Operating Characteristic Curve): This metric evaluates the model's capacity to distinguish between classes, providing insight into its quality across different thresholds.

Algorithm Comparison

- We test several machine learning algorithms, including Logistic Regression, Random Forest, Support Vector Machine (SVM), and k-Nearest Neighbors (KNN), on each dataset.
- For each algorithm and dataset, we compute the evaluation metrics listed above. The results are then compared across algorithms to determine the most accurate and reliable model for each dataset.
- After selecting the best-performing algorithm, we conduct hyperparameter tuning to optimize the model's performance further.

Final Model Selection

- The model with the highest performance across multiple datasets and evaluation metrics is selected for deployment in the "EmoDiary" application. This selection aims to balance accuracy, efficiency, and reliability, providing users with meaningful sentiment insights in their diary entries.

By applying these model evaluation techniques, we ensure that the "EmoDiary" application's sentiment analysis is both accurate and dependable, supporting users in tracking their emotional well-being with confidence.

III. LITERATURE ANALYSIS

1. Kristina Machova, 2023, "Detection of Emotion by Text Analysis Using Machine Learning"

In emotion detection from text, several approaches have been explored, ranging from lexicon-based methods to advanced machine learning and deep learning techniques. Here is a summary of the key approaches used in emotion recognition from text, as discussed in the previous section.

Lexicon-based Approach

- Involves using predefined lexicons with words annotated for specific emotions (e.g., happiness, sadness, fear, etc.).
- This method searches for words from the lexicon in the text and computes the probability of each emotion.
- Limitations: Ambiguity in word emotion classification and overlap between emotions make this method less accurate for complex texts.

Machine Learning Approach

- Includes models like Naive Bayes (NB), Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN).
- Naive Bayes uses probabilities for classification, assuming feature independence, but may be outperformed by more complex methods.
- Support Vector Machines (SVM) provide a linear or non-linear classification boundary to maximize margin between classes.
- This method can outperform lexicon-based approaches due to its use of supervised learning, but still may struggle with complex text structures.

Deep Learning Approach

- **Word Embeddings:** Represents words as vectors that capture semantic relationships, enabling deep learning models to understand context.
- **Convolutional Neural Networks (CNNs):** Initially used for image processing but adapted to text through 1D convolution, these models learn local patterns in text.
- **Recurrent Neural Networks (RNNs):** Particularly suitable for sequential data like text, RNNs capture contextual relationships between words. Long Short-Term Memory (LSTM) networks, a special type of RNN, overcome the vanishing gradient problem, enabling the processing of longer sequences.
- The use of Attention Mechanisms in LSTMs helps to focus on critical parts of the sentence, improving performance in emotion detection.

Chatbot Development

- **Rule-based Chatbots:** Use predefined responses based on specific patterns or rules. While effective in certain contexts, these chatbots cannot adapt or learn from new data.
- **Self-learning Chatbots:** These chatbots use machine learning models to improve their performance by learning from interactions. They offer more flexibility but require extensive training data and computational power.

Table 1: Advantages & Disadvantages of Rule based Learning & Self Learning

Chatbot Type	Advantages	Disadvantages
Rule-based	Simple, fast, no need for large datasets	Limited flexibility, cannot adapt to new scenarios
Self-learning	Learns from interactions, adaptable	Requires large datasets, complex to train

2. Shahd Alahdal, 2020, "Diary Mining: Predicting Emotion from Activities, People And Places"

Research on emotion prediction from diaries highlights varied approaches, combining machine learning and natural language processing. Key studies include pre-trained word embeddings with situational factors (activities, people, places) for emotion detection. While this enriched emotional prediction, limitations in handling a broader emotional spectrum were noted due to data sparsity. Another model introduced activity classification through supervised and unsupervised techniques, incorporating people and places to deepen emotional insight, though challenges remained with personalized training data.

Performance Comparison of Emotion Prediction Models

Table 2: Performance Comparison of Emotion Prediction Models

Model	Accuracy	Precision	Recall
Proposed Model	85.0%	85.0%	82.0%
SVM	83.2%	52.2%	52.2%
Decision Tree	82.4%	46.5%	43.9%
Proposed Model	85.0%	85.0%	82.0%
Naive Bayes	82.3%	45.0%	41.8%

3. Prof. Dr. Friedhelm Schwenker, 2022, "Data Analytics and Machine Learning in Artificial Emotional Intelligence"

The field of Artificial Emotional Intelligence (AEI) enables computers to interpret and respond to human emotions, promoting more natural human-computer interactions. Current AEI technologies leverage data analytics and machine learning, particularly deep learning with multimodal data sources, to recognize emotions based on facial expressions, gestures, voice, and biophysiological signals (e.g., eye movement, ECG, EEG). This interdisciplinary research area focuses on the effectiveness of machine learning methods for multimodal emotional data processing.

Key topics in AEI research include

•Facial Expression Analysis •Emotional Gesture and Pose Classification • Speech-Based Emotion Recognition • Biophysiological Sensor Integration for emotion detection • Multimodal Data Fusion to improve emotion classification accuracy • Sentiment Analysis and emotion recognition in text

This research domain has applications in healthcare, affective computing, and intelligent human-computer interaction, aiming to enhance systems that can assess and respond to emotional states through advanced sensor and signal processing techniques.

Obinna Obeleagu, Yusuf Aleshinloye Abass, Steve A. Adeshina, 2019, "Sentiment Analysis in Student Learning Experience"

This paper explores the use of sentiment analysis (SA) to enhance the objectivity of student feedback and improve learning outcomes. Recognizing the influence of emotions on student experiences, the authors propose combining academic data with sentiment analysis to provide clearer insights into student perceptions. By analyzing both positive and negative sentiments in student feedback, the model helps eliminate biases and offers a more accurate view of learning challenges and successes. The paper highlights the potential of SA to track engagement and emotional reactions, bridging the gap between subjective feedback and objective assessment. Ultimately, this approach offers a data-driven solution to better understand and improve student performance, fostering a more effective learning environment.

The system offers several advantages, particularly in helping teachers efficiently analyze a large number of student diaries. By assessing students' emotional well-being, it provides valuable insights that can improve communication and feedback, ultimately enhancing the learning experience. However, one potential limitation of the study is its focus primarily on students' emotions, which may introduce a bias by overlooking other important factors, such as academic performance or external influences on students' well-being. This narrow focus could affect the overall comprehensiveness of the analysis.

5. Kiran S Raj, Priyanka Kumar, 2021, Automated Human Emotion Recognition and Analysis using Machine Learning

The paper "Automated Human Emotion Recognition and Analysis using Machine Learning" presents a robust system that uses both video-based and text-based analysis to recognize human emotions. The dual approach enhances human-computer interaction by detecting emotions from live video feeds and speech, which is crucial in applications such as customer service and mental health support. By classifying emotions like anger, sadness, happiness, fear, surprise, disgust, and neutrality, the system provides valuable insights into human emotional states. This can be particularly beneficial in contexts such as monitoring emotional well-being, detecting distress in individuals, and identifying hidden sentiments in text, such as in social media messages or scams.

However, the study also highlights some limitations. The accuracy of emotion recognition can be influenced by various factors. In the video-based module, facial expression variations across individuals and environmental factors (such as lighting) can affect the system's ability to accurately classify emotions. Similarly, in the text analysis module, challenges such as speech clarity and the complexity of natural language—particularly sarcasm, ambiguity, and context-dependent expressions—can hinder performance. Despite these challenges, the integration of these two modules—video and text—offers a comprehensive approach to emotion recognition, with significant potential for real-world applications.

In conclusion, while the system presents a promising advancement in human emotion recognition, future improvements in model accuracy and robustness are necessary to mitigate the impact of these influencing factors. The ability to recognize emotions from both visual and textual data can revolutionize fields like customer service, mental health support, and security, but further research is needed to refine the system's effectiveness across diverse scenarios.

Proposed System

The proposed system introduces a novel approach to emotion recognition by leveraging insights derived from diverse datasets, rather than relying on a single or biased data source. This multi-faceted strategy ensures comprehensive emotion analysis, addressing various aspects of human sentiment. Our system incorporates the following datasets to capture a wide range of emotions: Toxic Language Analysis, Mental Health Analysis, Financial Sentiment Analysis, Emotion Analysis, and Tweet-based Emotion Detection. Each dataset is carefully chosen to target specific emotional domains, enabling a holistic understanding of user sentiment.

To optimize the prediction accuracy for each dataset, we conducted a comparative study of various machine learning algorithms, including Logistic Regression, k-Nearest Neighbors (kNN), Random Forest, Support Vector Machines (SVM), Naive Bayes, and Gradient Boosting. The best-performing model for each dataset was selected based on rigorous evaluation metrics, ensuring robust and reliable predictions. These individual predictions are synthesized to analyze diary entries and provide users with meaningful emotional insights. The system offers detailed visualizations and analytics to help users track their emotional well-being, encouraging self-awareness and promoting mental health. By integrating insights from multiple datasets and fine-tuning models for specific prediction tasks, our system delivers a comprehensive, accurate, and user-focused approach to emotion recognition and analysis.

Emotion Analysis

The Emotion Analysis Phase focuses on classifying emotional expressions in textual data, particularly within Twitter messages. This dataset is structured with two key columns:

- **Text:** Contains Twitter messages as strings.
- **Label:** Represents the primary emotion conveyed, classified into six categories: sadness (0), joy (1), love (2), anger (3), fear (4), and surprise (5).

Each entry comprises a textual message and its corresponding emotional label, offering a comprehensive resource for exploring the nuanced emotional landscape of social media.

Key Features

Textual Data: Provides raw textual content from Twitter, enabling sentiment and emotion analysis in natural language.
Emotion Labels: Captures six fundamental emotions to ensure a broad spectrum of emotional classification.

Use Cases

- **Sentiment Analysis:** Analyze predominant emotions in English Twitter messages.
- **Emotion Classification:** Develop machine learning models to categorize emotions accurately across six classes.
- **Textual Insights:** Investigate linguistic patterns and expressions associated with varying emotional states.

IV. MODEL PERFORMANCE

Based on the results you've provided, the SGD Classifier and Logistic Regression are performing quite well, with the highest accuracy and competitive precision, recall, and F1 scores.we'll proceed with Logistic Regression it for creating the prediction function.

Table 3: Performance Metrics for Emotion Analysis Dataset

Model	Accuracy	Precision	Recall	F1 Score:
Logistic Regression	0.90	0.90	0.90	0.90
Random Forest:	0.86	0.86	0.86	0.86
Naive Bayes:	0.84	0.86	0.84	0.83
KNN	0.53	0.62	0.53	0.55

Financial Sentiment Analysis

The Financial Sentiment Analysis Phase focuses on analyzing financial-related emotions by classifying sentences from financial contexts into sentiment categories. The dataset used in this phase is a combination of two prominent financial datasets: FiQA and Financial PhraseBank. The data has been structured into a single CSV file, making it convenient for sentiment analysis tasks in the financial domain.

Data Set

The dataset consists of two primary columns:

- **Sentence:** Contains financial-related text or sentences, representing statements from financial contexts such as market trends, investment sentiments, and financial reports.

- **Sentiment:** Categorizes the sentiment of each sentence into one of three possible labels:
- Positive, Negative, Neutral Key Features:
- **Financial Context:** Provides real-world, financial-related statements, useful for developing and evaluating sentiment analysis models tailored to finance.
- **Sentiment Labels:** Classifies each sentence into positive, negative, or neutral sentiment, offering a nuanced understanding of the emotional tone in financial discourse.
- Use Cases:
- **Financial Sentiment Analysis:** Analyze the emotional tone of financial documents, market analyses, and news articles.
- **Financial Insights:** Derive insights into market sentiment, investor attitudes, and economic outlooks based on textual analysis.
- **Decision Support:** Use sentiment analysis for decision-making in finance-related tasks like risk assessment, investment analysis, and market forecasting.

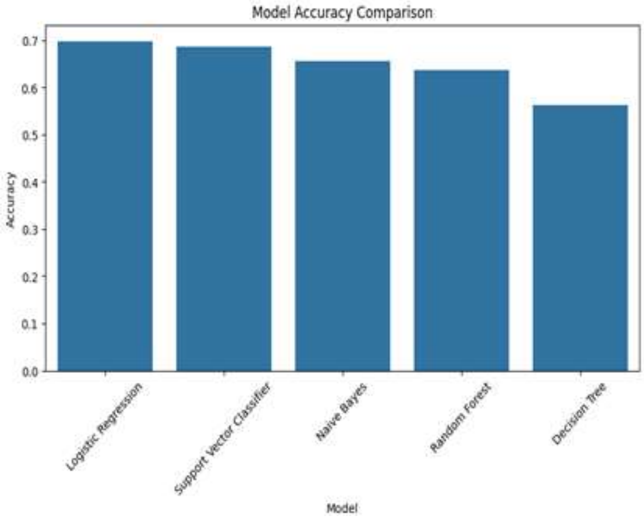

Figure 3: Model Accuracy Comparison

We will be utilizing Logistic Regression for this task, as it outperforms other models in terms of accuracy and efficiency, making it the most suitable choice for our analysis.

V. MENTAL HEALTH ANALYSIS

The mental health analysis phase leverages machine learning models to predict various mental health statuses based on user-provided textual statements. The dataset, which contains 53,043 entries, includes statements labeled with one of the following mental health categories: Normal, Depression, Suicidal, Anxiety, Bipolar, Stress, and Personality Disorder.

The main objective of this phase is to use these textual entries to classify the mental health status of the individual. For effective classification, we utilized multiple machine learning algorithms, including Logistic Regression, Random Forest, Support Vector Machine (SVM), Naive Bayes, K-Nearest Neighbors (KNN), and XGBoost. These models were selected and evaluated based on their ability to classify the mental health statuses accurately.

The preprocessing steps involved cleaning the data, handling missing values, and transforming the text into numerical features using the TF-IDF vectorization technique. The text was preprocessed to remove any irrelevant or duplicate entries, and the data was then split into training and testing sets.

After applying the different models, a detailed performance evaluation was conducted using accuracy, precision, recall, and F1-score metrics. The results showed that SVM and XGBoost provided the best performance, with XGBoost slightly outperforming SVM in terms of recall and F1-score. The final model choice was made based on these performance metrics, ensuring that the selected model provides the most accurate and reliable predictions for mental health classification.

The insights and predictions made by these models help in understanding the emotional and psychological well-being of individuals, offering a deeper analysis of their mental health status. This analysis plays a vital role in identifying potential mental health issues and providing timely interventions.

Table 4: Performance Metrics for Mental Health Analysis Dataset

Model	Accuracy	Precision	Recall	F1 Score:
Logistic Regression	0.75	0.78	0.64	0.68
Random Forest:	0.70	0.82	0.55	0.61
Naive Bayes:	0.64	0.76	0.47	0.52
SVM	0.75	0.79	0.66	0.71
KNN	0.33	0.64	0.17	0.13

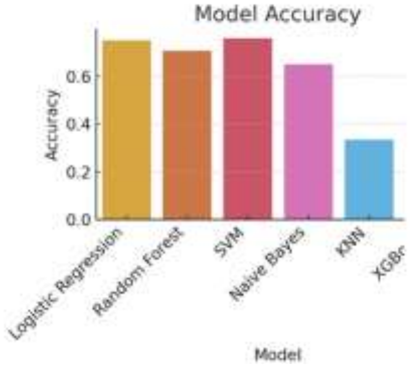

Figure 4: Visualization of Accuracies

VI. TOXIC LANGUAGE ANALYSIS

The Toxic Language Analysis phase aims to identify and reduce toxic language in user diary entries. This phase uses a dataset with two attributes: text (representing the content of the diary entry) and label (indicating whether the text is toxic or non-toxic). The goal is to classify whether a given diary entry contains toxic language, which is crucial in fostering a safe and supportive environment for users. By analyzing the toxicity of language, the system can help individuals be more mindful of their word choices, promoting healthier communication.

For this task, various machine learning models were applied and evaluated based on their performance in detecting toxic language. The models used in the study include Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes, and Gradient Boosting. These models were assessed using four key metrics: Accuracy, Precision, Recall, and F1 Score.

The dataset was preprocessed by cleaning the text data and transforming it into numerical features using the TF-IDF vectorization method. Each model was trained on this vectorized data and tested for its ability to correctly classify toxic versus non-toxic language.

Table 5: Performance Metrics for Toxic Language Analysis Dataset

Model	Accuracy	Precision	Recall	F1 Score:
Logistic Regression	0.91	0.92	0.90	0.91
Random Forest:	0.89	0.87	0.90	0.89
Naive Bayes:	0.88	0.84	0.95	0.89
SVM	0.91	0.92	0.91	0.91
KNN	0.59	0.55	0.97	0.70

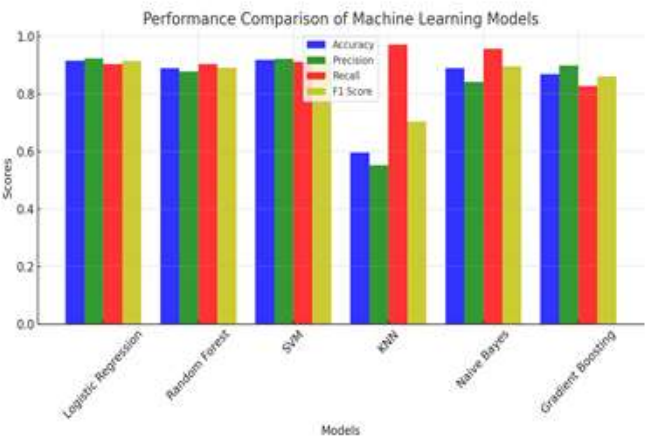

Figure 5: Performance Comparison of Machine Learning Models

Conclusion

Based on the results, Support Vector Machine (SVM) outperformed the other models, achieving the highest accuracy (0.9180) and a balanced performance across precision, recall, and F1 score. Logistic Regression also performed well, with a very close overall score. K-Nearest Neighbors (KNN) had a notably low F1 score, indicating a poor balance between precision and recall, despite having a high recall. Naive Bayes performed well with a high recall but lower precision, while Random Forest and Gradient Boosting were also effective but did not match the top performers.

Tweet Emotion Analysis

The Tweet Emotion Analysis phase leverages a dataset consisting of 40,000 entries of tweet data. Each entry contains a tweet_id, the sentiment label, and the content of the tweet. The goal of this phase is to categorize the emotions expressed in the tweets into various sentiment categories. This analysis helps in understanding the emotional tone of user-generated content on social media platforms, which can be insightful for monitoring public sentiment and improving user well-being.

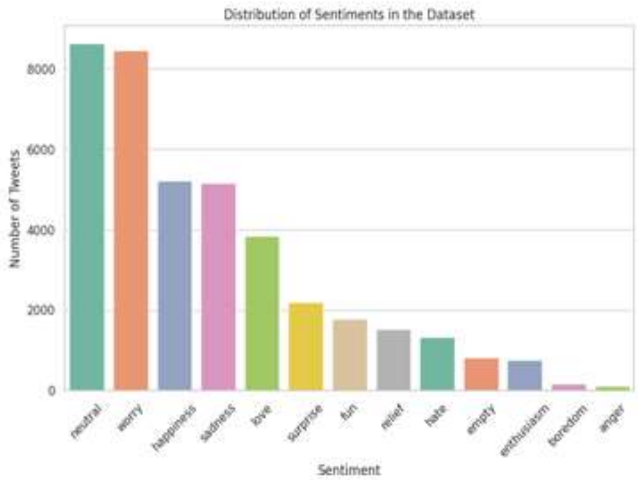

Figure 6: Distribution of Sentiments in the Dataset

Dataset Overview

The dataset consists of the following three columns:
tweet_id: A unique identifier for each tweet. sentiment: The emotional label assigned to each tweet. The sentiment labels include a wide range of emotions such as neutral, worry, happiness, sadness, love, surprise, fun, relief, hate, empty, enthusiasm, boredom, and anger.

Content: The text content of the tweet, which is analyzed to predict the associated sentiment.
Sentiment Distribution

The distribution of sentiments across the dataset is as follows:

- Neutral: 8,638 tweets

- Worry: 8,459 tweets
- Happiness: 5,209 tweets
- Sadness: 5,165 tweets
- Love: 3,842 tweets
- Surprise: 2,187 tweets
- Fun: 1,776 tweets
- Relief: 1,526 tweets
- Hate: 1,323 tweets
- Empty: 827 tweets
- Enthusiasm: 759 tweets
- Boredom: 179 tweets
- Anger: 110 tweets

This dataset covers a wide range of emotions expressed on social media, providing a comprehensive foundation for training emotion classification models.

Table 6: Performance Metrics for Tweet Emotion Analysis Dataset

Model	Accuracy	Precision	Recall	F1 Score:
Logistic Regression	0.54	0.53	0.54	0.53
Random Forest:	0.79	0.80	0.80	0.80
Naive Bayes:	0.63	0.60	0.63	0.60
KNN	0.72	0.66	0.72	0.67

Based on the comparative analysis of various machine learning models, Random Forest has emerged as the most suitable algorithm for tweet emotion classification in our study. With an accuracy of 79%, along with balanced precision, recall, and F1 scores of 80%, it demonstrates superior performance compared to other models such as Logistic Regression, Naive Bayes, and K-Nearest Neighbors (KNN).

Random Forest's ability to handle complex data structures and its robustness against overfitting make it an ideal choice for this task. Therefore, we have selected Random Forest as the predictive model for the Tweet Emotion Analysis phase, ensuring reliable and accurate sentiment classification for diverse tweet data.

Flask based Backend Server

Based on the final performance metrics of different algorithms evaluated for each dataset, we will develop a Flask-based backend server to manage the prediction tasks. This server will leverage the best-performing model for each dataset to process incoming data and respond with a JSON file containing the predicted insights.

These prediction results will then be securely stored in a Google Firebase database, ensuring seamless data integration and management. The stored data will subsequently be utilized by the website to derive meaningful insights and present them to users through intuitive visualizations and analytics, enhancing the overall user experience.

Table 7: Overall Performance Metrics

DataSet	Algorithm	Accuracy	Precision	Recall	F1 Score
Emotion Analysis	Logistic Regression	90%	90%	90%	90%
Financial Sentiment Analysis	Logistic Regression	69%	59%	55%	55%
Mental Health Analysis	SVM	75%	79%	66%	71%
Toxic Language Analysis Phase	Logistic Regression	91%	92%	90%	91%
Tweet Emotion Analysis Phase	Random Forest	79%	80%	80%	80%

VII. FRONT-END DEVELOPMENT WITH ADVANCED VISUALIZATION

The EmoDiary front end leverages modern web technologies to deliver an intuitive, engaging, and user-friendly interface for emotional journaling. This paper delves into the technical and design aspects of the front end, focusing on the integration of Firebase for real-time data storage and retrieval, advanced charting techniques for visual analytics, and seamless user interaction features. Each feature is meticulously crafted to provide users with meaningful insights and a delightful experience.

Need for Interface?
The EmoDiary project emphasizes emotional self-awareness through sentiment analysis of daily diary entries. Its front end plays a pivotal role in bridging the gap between complex backend processes and a seamless user experience. By combining robust data handling capabilities with creative visualizations, the EmoDiary front end ensures that users can interactively explore their emotional trends and insights.

Features and Implementation
User Interface
The EmoDiary interface is built using HTML, CSS, JavaScript, and Bootstrap. It offers the following components:

Home Page: Features options like login and registration powered by Firebase Authentication.

Figure 7: Home Page

Profile Page: Displays four primary actions: Write, Latest Vibes, Performance Snapshot, and DeepDive Analytics.

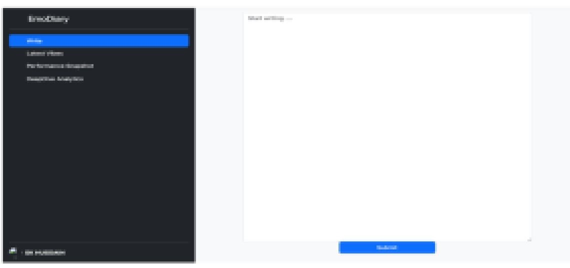
Figure 8: Profile Page

Analytics and Progress Pages: Provide in-depth visualizations of user data.

Real-Time Data Integration with Firebase
Firebase Firestore is utilized for seamless storage and retrieval of diary entries and predictions. The integration ensures that visualizations dynamically update as new data is added.

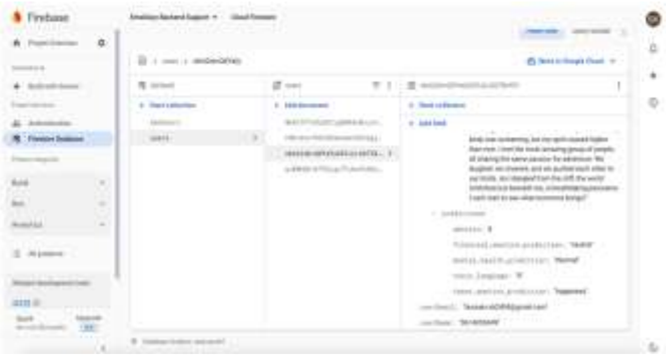

Figure 9: Firebase Interface for Database Support

Data Format
Date and Time

- The date field indicates when the diary entry was created. It's in ISO 8601 format and can be parsed to extract the day, month, or year for time-based visualizations.

Diary Content

- The diary field contains the full text of the entry. You can use this to generate word clouds, perform text analysis, or display diary excerpts.

Predictions

- **Emotion:** Numerical indicator of emotion (e.g., a scale from 0 to 6 or similar).
- **financial_emotion_prediction:** Categorizes financial sentiment (e.g., positive, negative, neutral).
- **mental_health_prediction:** Categorizes mental health status (e.g., Normal, Depression).
- **toxic_language:** Binary indicator or percentage of toxicity detected in language.
- **tweet_emotion_prediction:** Simplified emotion prediction for social media compatibility.

User Information

- userEmail and userName help associate entries with specific users.

Advanced Visualizations
Latest Vibes Section Key Features

International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 1, Jan-Feb-2025, ISSN (Online): 2395-566X

Figure 10: Today's Insights Page

Firebase Initialization
- Uses Firestore to fetch user-specific diary data.
- Authentication ensures the user is logged in before accessing their data.

Diary Entry Fetching
- Retrieves the latest diary entry from the diaries array.
- Error handling is implemented to manage cases where no diary entries exist.

Predictions Display
- Maps predictions (emotion, mental_health_prediction, toxic_language, tweet_emotion_prediction) to UI elements for user feedback.
- Categorizes emotions into labels like "Bright Day" or "Cloudy Day".

Word Cloud Rendering
- Uses D3.js to generate a word cloud from the diary text.
- Words are sized proportionally to their frequency, with random rotation and colors.

Mathematical Representation
For a word w, its size S(w) is proportional to its frequency f(w):

$$S(w) \propto f(w)$$

Performance Snapshot Section Key Features

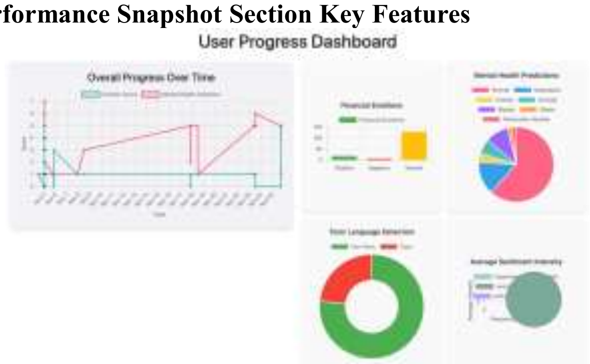

Figure 11: Performance Snapshot Page

Overall Progress Over Time Plot

Figure 12: Overall Progress Over Time Plot

Description
The Overall Progress Chart visualizes the user's emotional score and mental health predictions over time. It uses a line chart to plot the Emotion Score and Mental Health Prediction on the same timeline, helping users track changes and patterns in their emotional well-being.

Formula:

1. **Emotion Score** for an entry i:

$$E_i = \text{entry.predictions.emotion}$$

2. **Mental Health Mapping** M_i:

$$M_i = \text{mapping(entry.predictions.mental\_health\_prediction)}$$

Where M_i maps categories like *Normal (1)*, *Depression (2)*, etc., to numeric values for visualization.

Both are plotted as:

$$\text{Point} = (t_i, Y_i), \quad \text{where } t_i = \text{date of entry}, Y_i \in \{E_i, M_i\}.$$

Financial Emotions
The Financial Emotion Chart visualizes the distribution of financial-related emotions across diary entries. It uses a bar chart to display the count of three categories: Positive, Negative, and Neutral financial emotions, providing insights into the user's financial sentiment trends.

Figure 13: Financial Emotions Insights

Formula:

1. **Financial Emotion Count** for each category C:

$$\text{Count}_C = \sum_{i=1}^{N} \mathbb{I}(\text{entry}_i.\text{predictions.financial\_emotion\_prediction} = C)$$

Where $\mathbb{I}(x)$ is the indicator function:

$$\mathbb{I}(x) = \begin{cases} 1 & \text{if } x \text{ is true} \\ 0 & \text{if } x \text{ is false} \end{cases}$$

and $C \in \{\text{Positive, Negative, Neutral}\}$.

2. **Data for Bar Chart:**

$$\text{Data} = [\text{Count}_{\text{Positive}}, \text{Count}_{\text{Negative}}, \text{Count}_{\text{Neutral}}]$$

The chart maps these counts to bars corresponding to each emotion category.

Mental Health Predictions

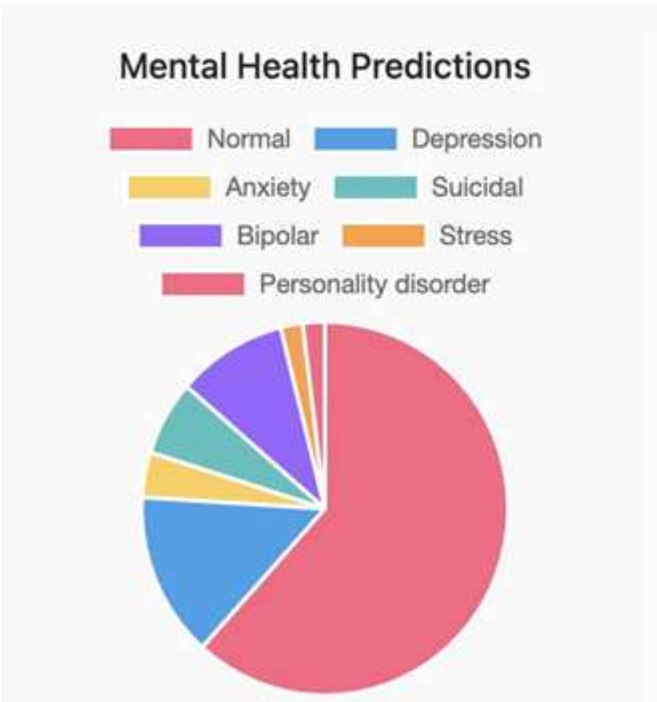

Figure 14: Mental Health Predictions Insights

The Mental Health Chart presents the distribution of various mental health predictions derived from the diary entries. It employs a pie chart to visually represent the proportion of different mental health states, such as Normal, Depression, Anxiety, etc., in the dataset. This visualization provides an intuitive overview of the user's mental health trends over time.

Formula:

1. **Mental Health Count** for each state S:

$$\text{Count}_S = \sum_{i=1}^{N} \mathbb{I}(\text{entry}_i.\text{predictions.mental\_health\_prediction} = S)$$

Where $\mathbb{I}(x)$ is the indicator function:

$$\mathbb{I}(x) = \begin{cases} 1 & \text{if } x \text{ is true} \\ 0 & \text{if } x \text{ is false} \end{cases}$$

and S represents a specific mental health state (e.g., Normal, Depression, Stress).

Figure 15: Formula

2. **Proportions for Pie Chart:**

Each mental health state S is visualized as a sector of the pie chart, with its size determined by:

$$\text{Proportion}_S = \frac{\text{Count}_S}{\sum_{S'} \text{Count}_{S'}}$$

Where S' iterates over all possible mental health states.

The chart maps these proportions to sectors, with distinct colors for easy differentiation.

Figure 16: Properties

Toxic Language Detection

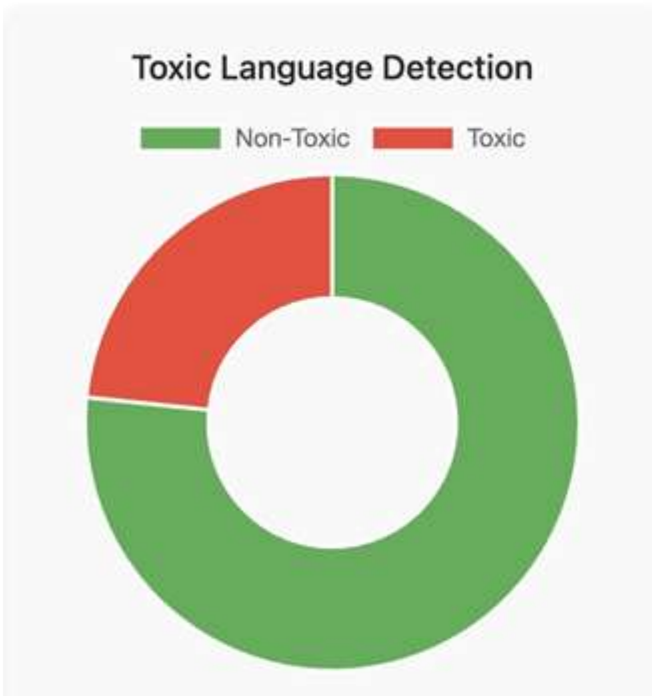

Figure 17: Toxic Language Analysis

The Toxicity Chart provides a visual representation of the presence of toxic language in the diary entries. Using a

International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 1, Jan-Feb-2025, ISSN (Online): 2395-566X

doughnut chart, it shows the distribution of toxic versus non-toxic language based on the predictions from the model. The chart helps users quickly understand the proportion of entries with toxic content versus those without.

Formula:

1. **Toxic Language Count:**

$$\text{Count}_{\text{toxic}} = \sum_{i=1}^{N} \mathbb{I}(\text{entry}_i.\text{predictions.toxic\_language} = 1)$$

Where $\mathbb{I}(x)$ is the indicator function, and 1 represents toxic language.

Similarly, for non-toxic language:

$$\text{Count}_{\text{non-toxic}} = \sum_{i=1}^{N} \mathbb{I}(\text{entry}_i.\text{predictions.toxic\_language} = 0)$$

2. **Proportions for Doughnut Chart:**

The chart visualizes the following proportions:

$$\text{Proportion}_{\text{non-toxic}} = \frac{\text{Count}_{\text{non-toxic}}}{\text{Count}_{\text{non-toxic}} + \text{Count}_{\text{toxic}}}$$

$$\text{Proportion}_{\text{toxic}} = \frac{\text{Count}_{\text{toxic}}}{\text{Count}_{\text{non-toxic}} + \text{Count}_{\text{toxic}}}$$

These proportions are represented by the relative sizes of the sectors in the doughnut chart, with distinct colors for toxic and non-toxic labels.

Average Sentiment Intensity

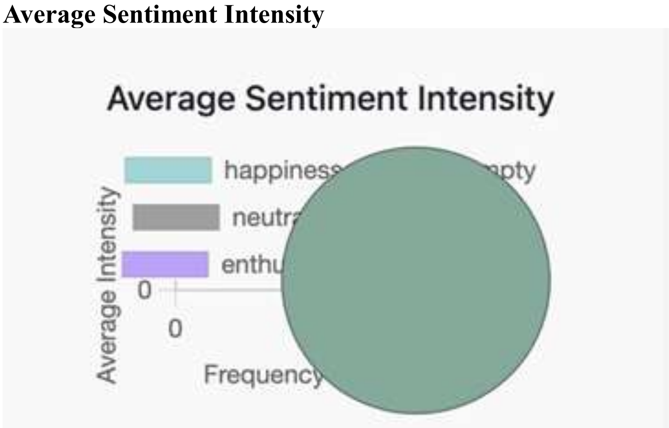

Figure 18: Average Sentiment Intensity

VIII. CONCLUSION: PERFORMANCE SNAPSHOT

The Performance Snapshot feature provides users with a clear visual representation of their emotional and mental health journey, based on their diary entries. It allows users to track emotional trends, identify patterns, and monitor fluctuations in mental health over time. By offering insights into key emotional metrics, the snapshot helps users reflect on their well-being and make informed decisions for emotional self-care. Overall, it serves as an effective tool for fostering emotional awareness and promoting better mental health management.

DeepDive Analytics Section

DeepDive Analytics offers a comprehensive exploration of your emotional journey, providing in-depth insights and detailed visualizations. This section goes beyond surface-level analysis to uncover hidden patterns and trends in your diary entries, allowing for a deeper understanding of your emotional and mental well-being. With advanced metrics and sophisticated charting, DeepDive Analytics helps you detect subtle shifts in mood, track recurring emotional states, and gain valuable perspectives on your emotional evolution over time.

Distribution of Mental Health Predictions

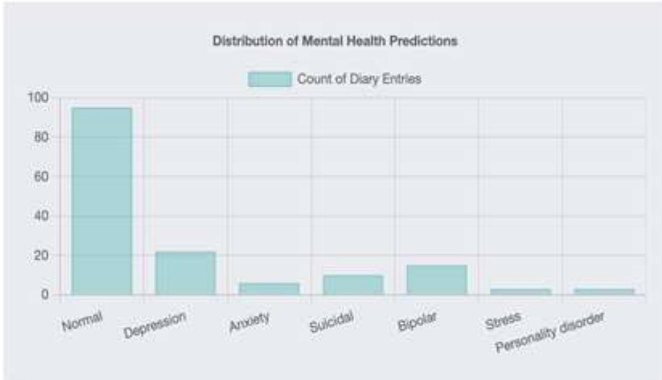

Figure 19: Distribution of Mental Health Predictions

This bar chart visualizes the distribution of mental health predictions based on diary entries. Each bar represents the count of diary entries classified into different mental health categories, such as 'Normal', 'Depression', 'Anxiety', etc. The chart provides insights into how frequently each mental health status appears across the dataset, helping users better understand the overall distribution of mental health conditions. The y-axis shows the number of entries, while the x-axis lists the mental health categories.

Feelings Palette

We use a radar chart to display emotions by plotting different emotion categories along the axes of the chart. Each axis represents one emotion. The data points on the chart show how strong or frequent each emotion is in the dataset. These points are connected to form a shape, giving us a visual representation of the overall emotional profile. The larger the area covered by the shape, the stronger or more frequent the emotions in that category. This makes it easy to see which emotions are most dominant and how they relate to each other.

International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 1, Jan-Feb-2025, ISSN (Online): 2395-566X

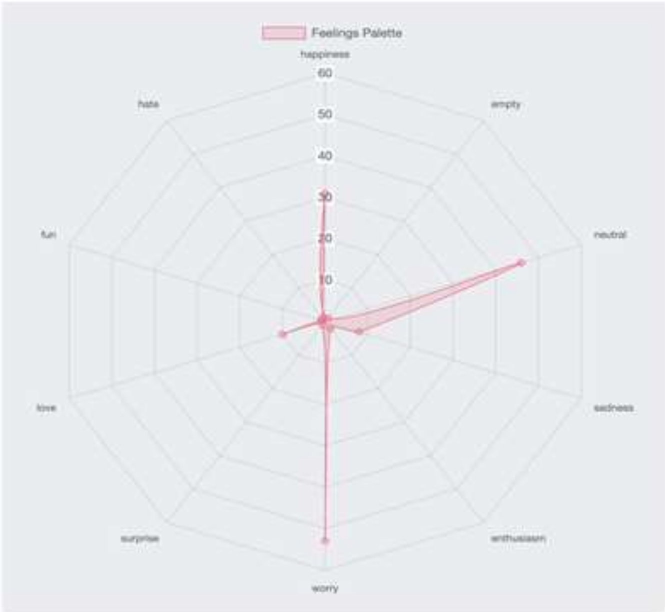

Figure 20: Feelings Palette

Radial Plot of Emotions

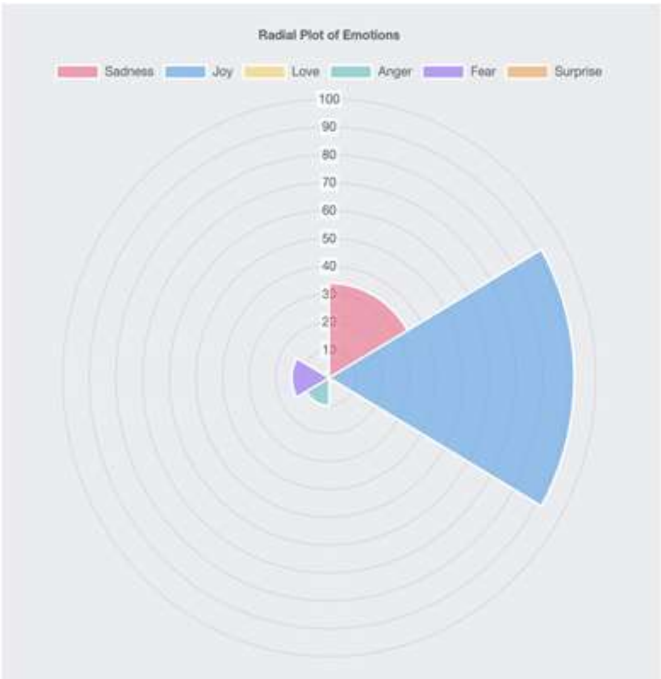

Figure 21: Radial Plot of Emotions

This radial (polar area) chart displays the distribution of different emotions (Sadness, Joy, Love, Anger, Fear, and Surprise) in a dataset. Each segment of the chart represents one emotion, with the size of the segment indicating the frequency or intensity of that emotion. The chart visually highlights the relative proportions of each emotion, making it easy to compare how they are distributed across the data.

Feelings Frequency

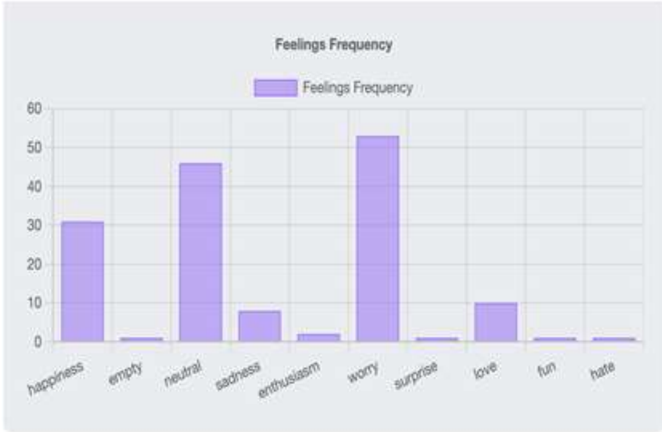

Figure 22: Feelings Frequency

The "Feeling Frequency" bar chart visualizes the frequency of various emotions or feelings in a dataset. Each bar represents a different feeling, and the height of the bar indicates how often that feeling occurs. This chart provides a clear comparison of how frequently each emotion appears, allowing for easy identification of dominant or less common emotions in the data.

About Google Firebase

Google Firebase is a comprehensive platform for building and managing web and mobile applications. It provides developers with a suite of tools and services designed to streamline app development, enhance functionality, and scale efficiently. Firebase is particularly popular for its ease of integration, robust real-time capabilities, and managed backend infrastructure.

Key Features of Firebase Used in EmoDiary

Authentication

Simplifies user authentication with pre-built UI libraries and integration options for email/password, phone number, Google, Facebook, Twitter, and more.

Ensures Secure and Efficient User Management.

Realtime Database

- A NoSQL cloud-hosted database that stores and syncs data in real time across clients.
- Ideal for collaborative apps and scenarios where low latency is critical.

Final Conclusion for EmoDiary Project

The EmoDiary Project successfully integrates the power of journaling with advanced machine learning to create a unique platform that promotes emotional awareness and mental well-being. By combining intuitive design with robust analytics, the platform allows users to capture their thoughts and feelings through text or voice diary entries while gaining meaningful insights into their emotional patterns and mental health.

With a strong technical foundation, the project employs cutting-edge technologies like Google Firebase for secure user authentication, real-time data storage, and seamless scalability. The backend, powered by a Flask-based API server, ensures efficient processing of diary entries to extract various emotional and mental health parameters. On the frontend, creative and interactive visualizations, built using jCharts, empower users to explore their emotional journeys and progress in depth.

Key features, such as emotion detection, toxic language analysis, and financial sentiment predictions, provide a comprehensive understanding of users' sentiments and mental states. The platform also includes detailed insights through DeepDive Analytics and a Progress Dashboard, making the application not only a diary but also a personal growth companion.

This project highlights the potential of technology in fostering self-awareness and emotional well-being. By bridging the gap between mental health and innovation, EmoDiary offers a personalized, user-friendly experience that supports users in tracking their mental health, managing stress, and improving their overall quality of life. The successful integration of various technologies and the emphasis on user engagement make EmoDiary a transformative tool for emotional well-being in the digital age.

REFERENCES

1. Kristina Machova, Martina Szaboova1, Jan Paralic, Jan Micko, 2023, “Detection of emotion by text analysis using machine learning” Front. Psychol., 20 September 2023, Sec. Emotion Science.
2. Shahd Alahdal, 2020, “Diary mining: predicting emotion from activities, people and places” Cardiff University Electronic Publication Form.
3. Prof. Dr. Friedhelm Schwenker, Prof. Dr. Mariofanna Milanova, 2022, “Data Analytics and Machine Learning in Artificial Emotional Intelligence”
4. Obinna Obeleagu, Yusuf Aleshinloye Abass, Steve A. Adeshina, 2019, “SENTIMENT ANALYSIS IN STUDENT LEARNING EXPERIENCE”, 2019 15th
5. International Conference on Electronics, Computer and Computation (ICECCO).
6. Kiran S Raj, Priyanka Kumar, 2021, Automated Human Emotion Recognition and Analysis using Machine Learning, 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE Xplore.