Index

MINI PROJECT PROPOSAL

_____

Title of the Project

Emotion Drift Detection in Individuals Over Time using Machine Learning

_____

1. Introduction

Human emotions are dynamic and evolve continuously over time due to changes in personal, academic, social, and environmental factors. Most existing emotion analysis systems focus on identifying emotions at a single point in time, such as classifying a text as happy, sad, or angry. While such systems are useful, they fail to capture **long-term emotional behavior and gradual emotional changes**, which are often more meaningful than isolated emotional states.

This project proposes an **Emotion Drift Detection System** that analyzes emotional expressions over extended periods and identifies significant changes in emotional patterns using machine learning techniques. The system focuses on **temporal emotional behavior**, enabling deeper behavioral insights rather than momentary emotion recognition.

_____

2. Problem Statement

Conventional emotion detection systems suffer from the following limitations:

- They analyze emotions in isolation without considering historical context
- They fail to identify gradual emotional changes occurring over time
- They cannot detect behavioral shifts unless extreme emotional expressions occur
- They lack mechanisms to analyze emotional trends and consistency

As a result, meaningful emotional pattern changes often remain undetected, limiting the usefulness of traditional emotion analysis systems in long-term behavioral studies.

_____

3. Need

Emotional behavior is better understood as a **time-dependent process** rather than a sequence of independent events. Detecting long-term emotional changes can provide valuable insights in domains such as digital well-being research, behavioral analytics, and user experience evaluation.

The motivation behind this project is to design an intelligent AIML-based system that can **track, analyze, and detect emotional drift** over time without performing diagnosis or making psychological claims. The system aims to support early identification of emotional pattern shifts in a purely analytical and ethical manner.

---

4. Objectives

The main objectives of this project are:

- To classify emotions from textual data using machine learning
- To construct emotion timelines for individuals
- To analyze emotional patterns across time intervals
- To detect statistically significant emotional drift
- To visualize emotional trends and drift events
- To ensure ethical handling of emotional data

---

5. Scope of the Project

The scope of this project is limited to **emotion pattern analysis and drift detection** using textual data. The system does **not** attempt to diagnose mental health conditions or provide psychological evaluations. All outputs are analytical indicators intended for research and academic purposes only.

---

6. Proposed Solution

The proposed system processes time-stamped textual inputs (such as journal entries, reflections, or dialogue data) from individuals. Each entry is analyzed to detect emotional content using machine learning-based emotion classification. These emotions are then aggregated over time to form an emotion sequence.

Using time-series and statistical analysis techniques, the system identifies **changes in emotional distribution, frequency, and transitions**, which are interpreted as emotion drift. Detected drift events are visualized and reported for further analysis.

---

7. AIML Component

7.1 Emotion Classification

- Natural Language Processing techniques are used to extract emotional features from text
- Emotions such as Happy, Sad, Angry, Fear, and Neutral are classified
- Algorithms may include Logistic Regression, Support Vector Machines, or Naive Bayes with TF-IDF features

7.2 Emotion Time-Series Modeling

- Emotion labels are converted into numerical or categorical sequences
- Emotions are aggregated using daily or weekly windows
- Frequency distributions and transition probabilities are computed

7.3 Drift Detection Mechanism

- Sliding window comparison is used to detect distribution changes
- Statistical divergence measures identify deviation from baseline behavior
- Threshold-based deviation logic flags significant emotional drift

This AIML framework ensures that the project goes beyond simple emotion detection.

_____

8. Data Description

The system uses one or more of the following data sources:

- Publicly available emotion-labeled text datasets
- Anonymized personal journaling datasets
- Simulated diary or reflection data created for experimentation
- Movie or dialogue scripts annotated with emotional labels

All data used is anonymized and handled strictly for academic purposes.

_____

9. System Architecture and Methodology

Methodology Steps

i. Data collection and preprocessing
ii. Text cleaning and tokenization
iii. Emotion classification using ML models
iv. Emotion sequence generation per individual
v. Drift detection using temporal analysis

    vi. Visualization of emotion trends and drift points

Architecture Layers

- Input Data Layer
- NLP & Feature Extraction Layer
- Emotion Classification Layer
- Drift Detection Layer
- Visualization and Reporting Layer

---

10. Innovation and Uniqueness

The uniqueness of this project lies in its **temporal approach to emotion analysis**. Unlike traditional systems that classify emotions independently, this project focuses on:

- Long-term emotional behavior analysis
- Detection of gradual emotional changes
- Emotion consistency and transition patterns
- Application of drift detection concepts to emotional data

This methodological shift ensures originality and distinguishes the project from common sentiment or emotion classification systems.

---

11. Expected Outcomes

- Emotion timelines for individuals
- Visualization of emotional trends over time
- Identification of significant emotional drift events
- Analytical reports highlighting behavioral changes
- A reusable framework for temporal emotion analysis

---

12. Tools and Technologies

- **Programming Language:** Python
- **Libraries:** NumPy, Pandas, Scikit-learn
- **NLP Tools:** NLTK / spaCy
- **Visualization:** Matplotlib / Seaborn
- **Development Environment:** Jupyter Notebook

---

13. Applications

- Behavioral pattern analysis
- Digital well-being research
- User experience analytics
- Human-centered AI research
- Long-term emotion trend studies

---

14. Ethical Considerations

- No psychological diagnosis or health claims are made
- Data is anonymized and used only for academic purposes
- Outputs are analytical indicators, not judgments
- User privacy and data integrity are maintained

---

15. Future Scope

- Integration with multimodal data (voice, interaction logs)
- Advanced sequence models for emotion transitions
- Personalized emotion baseline modeling
- Cross-individual emotion pattern comparison
- Deployment as a research analytics tool

---

16. Conclusion

This project presents a novel AIML-based approach to emotion analysis by shifting focus from single-instance emotion detection to long-term emotional pattern monitoring. By incorporating time-series analysis and drift detection techniques, the system offers meaningful insights into emotional behavior evolution while maintaining ethical and academic integrity.

Diagrams :

BBC Guardian		Twitter Streaming API
↓		↓
News crawler		Tweets retrieval

News - tweet matching

News clustering

Sentiment analysis

```
┌─────────────────────┐                          ┌─────────────────────┐
│  Student behavior   │                          │        Data         │
│   data collection   │ ───────────────────────▶ │   preprocessing     │
└─────────────────────┘                          └─────────────────────┘
                                                             │
   ┌──────────────────┐                                      │
   │                  │                                      │
  ┌──────────────────┐│                                      │
  │   Lide data      ││                                      │
  │ consumption data ││                                      │
  │  academic data   ││                                      ▼
  └──────────────────┘                          ┌─────────────────────┐
                                                │                     │
┌─────────────────────┐    ┌──────────┐         │   Data feature      │
│  Model training     │ ◀─ │ Training │ ──────  │    extraction       │
│  improved K-means   │    │ data set │         │                     │
└─────────────────────┘    └──────────┘         └─────────────────────┘
          │                                               │
          │                                               │
          ▼                                               │
┌─────────────────────┐    ┌──────────┐                   │
│                     │    │ Training │                   │
│  Analysis model     │ ◀─ │ data set │ ◀─────────────────┘
│                     │    └──────────┘
└─────────────────────┘
          │
          │
          ▼
┌─────────────────────┐                          ┌─────────────────────┐
│                     │                          │       Results       │
│  Analysis result    │ ───────────────────────▶ │    management       │
│                     │                          │                     │
└─────────────────────┘                          └─────────────────────┘
```

2. Signal processing analysis

3. Emotion recognition

4. Time series representation

Audio file

Feature extraction

Low-level features array

Valence prediction → array

Arousal prediction → array

Selected part

Metadata

Time interval labeling

1. Song structure analysis

Data extraction

Plain text, JSON, XML, HTML, .xls, .csv

Data processing
- Preprocessing
- Features extraction
- Features selection

Classification
- Supervised methods
- Unsupervised methods
- Hybrid methods

Preprocessing

Common techniques of NLP such as:
- Stemming
- Lemmization
- Tokenization
- Online text cleaning
- Spell correction

Features extraction
- Text representation
- Bag of words,
- Vector Space Model
- Word embedding
- N-grams
- POS tagging
- Negations

Features selection
- Frequency-based selection
- Point-wise mutual Information
- Information Gain
- Gain Ratio