AI-powered audio biomarker analysis for early detection and monitoring of depression and anxiety using the DAIC-WOZ dataset and LightGBM:

**Core Project Goals:**

**Develop an AI-Powered Audio-Based Depression/Anxiety Detection System:**

* 1. Your primary goal is to create a machine learning model that can accurately predict the presence or severity of depression and anxiety based on audio features extracted from speech.

**Demonstrate the Feasibility of Audio Biomarkers:**

* 1. You must show that audio features derived from conversational speech can serve as reliable biomarkers for mental health conditions.

**Implement a Robust Methodology:**

* 1. This involves:
     1. Properly preprocessing the DAIC-WOZ audio data.
     2. Extracting relevant audio features (MFCCs, OpenSmile, etc.).
     3. Training and optimizing a LightGBM model.
     4. Evaluating the model's performance using appropriate metrics.

**Specific Deliverables and Expectations:**

* **Data Preprocessing and Feature Extraction Pipeline:**
  + A well-documented and functional pipeline for loading, preprocessing, and extracting features from the DAIC-WOZ audio data.
  + This should demonstrate your understanding of audio processing techniques.
* **LightGBM Model Implementation:**
  + A trained and optimized LightGBM model for depression/anxiety detection.
  + This should include hyperparameter tuning and model selection.
* **Model Evaluation and Performance Analysis:**
  + A comprehensive evaluation of the model's performance using appropriate metrics (accuracy, precision, recall, F1-score, AUC).
  + An analysis of the model's strengths and weaknesses.
  + Compare your results to other research that has used the same dataset.
* **Feature Importance Analysis:**
  + An analysis of which audio features are most indicative of depression and anxiety.
  + This provides insights into the audio biomarkers.
* **Project Report/Presentation:**
  + A clear and concise report or presentation summarizing your project, including:
    - Introduction and background.
    - Methodology (data preprocessing, feature extraction, model implementation).
    - Results and discussion.
    - Limitations and future work.
    - Ethical considerations.
* **Code Repository (Optional but Recommended):**
  + A well-organized and documented code repository (e.g., on GitHub) that showcases your work.
* **Demonstration of Understanding:**
  + Show that you understand the ethical implications of AI in mental health.
  + Show that you understand the limitations of the data set that you are using.
  + Show that you understand the limitations of the model that you are using.

**In essence, your project should demonstrate:**

* Your ability to apply machine learning techniques to audio data.
* Your understanding of audio processing and feature extraction.
* Your ability to evaluate and interpret machine learning models.
* Your understanding of the practical implications of your work.

//

**Which features should I extract from DAIC-WOZ dataset**

When extracting features from the DAIC-WOZ dataset for depression detection, it's crucial to consider both the audio and textual modalities. Here's a comprehensive list of features to extract:

**Audio Features:**

* **Acoustic Features (using openSMILE or Librosa):**
  + MFCCs (Mel-frequency cepstral coefficients): Capture spectral characteristics of speech.
  + eGeMAPS (extended Geneva Minimalistic Acoustic Parameter Set): Designed for paralinguistic analysis, capturing emotional and psychological cues.
    - Includes features like:
      * Pitch (F0) and its variations.
      * Energy and loudness.
      * Speaking rate and articulation.
      * Spectral features (e.g., spectral slope, spectral flux).

* + **Prosodic Features:**
    - Pitch (mean, standard deviation, range).
    - Energy (mean, standard deviation, range).
    - Speaking rate (words per minute, syllables per second).
    - Pauses (duration, frequency).
    - Jitter and shimmer (measures of vocal instability).
  + **Voice Quality Features:**
    - Harmonic-to-noise ratio (HNR).
    - Vocal fold closure measures.

**Textual Features:**

* **Lexical Features:**
  + **Word Count and Diversity:** Number of words, unique words, and lexical richness.
  + **Function Words:** Frequency of pronouns, prepositions, and other function words.
  + **Sentiment Analysis:** Polarity and subjectivity scores using libraries like VADER or TextBlob.
  + LIWC (Linguistic Inquiry and Word Count): Measures psychological and emotional dimensions of language.
* **Semantic Features:**
  + **Word Embeddings:** Pre-trained embeddings like Word2Vec, GloVe, or FastText.
  + **Sentence Embeddings:** Embeddings from models like Sentence-BERT.
  + **Topic Modeling:** Latent Dirichlet Allocation (LDA) to identify underlying topics.
* **Discourse Features:**
  + **Turn-Taking Patterns:** Duration of turns, interruptions, and overlaps.
  + Dialogue Acts: Identifying the function of utterances (e.g., questions, statements, backchannels).
  + **Cohesion and Coherence:** Measures of how well the dialogue flows.
* **Transformer-Based Features:**
  + BERT/RoBERTa Embeddings: CLS token embeddings or averaged hidden state embeddings from pre-trained transformer models.
  + **Attention Weights:** Analyzing attention patterns to identify important words and phrases.

**Multimodal Feature Considerations:**

* **Alignment:** If combining audio and text features, consider aligning them temporally.
* Fusion: Experiment with early fusion (concatenating features) and late fusion (combining predictions).

**Tools and Libraries:**

* **openSMILE:** For extracting comprehensive acoustic features.
* Librosa: For audio analysis and feature extraction.
* **Transformers (Hugging Face):** For pre-trained transformer models and embeddings.
* **NLTK and spaCy:** For natural language processing tasks.
* **LIWC:** For linguistic analysis.

By extracting a comprehensive set of features from both audio and text, you can provide your model with rich information for accurate depression detection.