Table of Contents

1. **Introduction:**

**+ Developed Countries:**

Psychology plays a vital role in understanding mental health. Developed countries like the United States, Europe, and Japan prioritize mental health development to improve their citizens' quality of life.

**+ Vietnam:**

In Vietnam, psychology has been a distant concern during the past 10-20 years of focusing on economic growth. However, at the present with the improved economy, there is now a growing interest in mental health among the population.

**+ General Motivation:**

This chatbot supports healthcare professionals in providing mental health therapy, improving quality of life, and promoting awareness of the value of mental well-being for a happier society.

1. **Methodology:**

**+Purpose:**

Create an app that can:

Image Sentiment Recognition:

Hugging Face Image Zero shot classification with label [‘Happy’,’Angry’, ‘Sad’]

Voice Sentiment Recognition:

Using kaggle dataset and train the model with modified label [‘’]

Psychology with LLM chatbot we should prepare:

Psychology knowledge:

Developmental Psychology

Social Psychology

Counseling Psychology

Industrial and Organizational Psychology

Educational Psychology

Model building:

Flowise interface with some key from OpenAI for Large Language Models (LLM), pinecone for storing data

1. **Idea:**

**+Emotion Recognition:**

Recognizing several emotions:

Angry, Disgust, Fear, Happy, Sad, Surprise,

Neutral

**+Survey Quests Suggestion:**

Recognizing several emotions:

Angry, Disgust, Fear, Happy, Sad, Surprise,

Neutral

**+Natural Language Preprocessing:**

Understanding human languages and Psychology concepts

**+Positive Solutions:**

Able to understand people emotion deeply and give some positive feedbacks

1. **Data Processing:**

+ **Collect Data:**

We collect dataset mostly from Kaggle.com with Facial Image Emotion, Voice Image Emotion,… (Adding more later)

+ **Understanding Data:**

**Data Exploration:**

**Examine the structure and format of the data**: Understand the data types (numerical, categorical, text), variable names, and any predefined relationships or hierarchies.

**Review summary statistics**: Calculate basic statistics such as mean, median, mode, standard deviation, and range to get an overview of the data distribution.

**Visualize the data**: Plotting charts, histograms, scatter plots, or box plots can provide insights into patterns, trends, and potential outliers.

**Data Profiling:**

**Assess data quality**: Identify missing values, inconsistencies, or anomalies that may affect the reliability or accuracy of the data.

**Check for data completeness**: Determine the percentage of missing or incomplete data in each column or variable.

**Investigate data distributions**: Understand the spread, skewness, or correlation of variables to gain insights into their relationships.

**Domain Knowledge:**

**Understand the context**: Familiarize yourself with the subject matter or domain to interpret the data more effectively.

**Consult with experts**: Seek input from domain experts who can provide insights into the data and its implications.

**Data Visualization:**

**Create meaningful visualizations:** Use charts, graphs, or interactive dashboards to present the data visually and facilitate understanding.

**Identify patterns and trends:** Look for recurring patterns, seasonality, or trends that can provide valuable insights.

**Data Relationships:**

**Analyze correlations**: Determine the relationships between variables using correlation coefficients or scatter plots.

**Conduct feature engineering:** Explore potential interactions or derived features that may enhance the predictive power of the data.

**Data Context:**

**Consider the data source:** Understand how and where the data was collected, including any biases or limitations associated with the source.

**Identify potential biases:** Assess whether the data may be skewed or incomplete, leading to biased conclusions or decisions.

+ **Preprocessing Data:**

**Data Cleaning:**

**Handle missing values:** Decide what to do with missing data, either by removing rows with missing values or filling in the missing values.

**Remove duplicates:** Identify and remove any duplicate records in the dataset.

**Handle outliers:** Identify and deal with extreme values that are significantly different from the rest of the data.

**Data Transformation:**

**Scale numerical features:** Adjust the numerical values to have similar ranges. This can involve techniques like normalization (scaling to a range of 0-1) or standardization (scaling to have zero mean and unit variance).

**Encode categorical variables:** Convert categorical variables into numerical representations that can be used in machine learning models. Common techniques include one-hot encoding, label encoding, or ordinal encoding.

**Extract meaningful features:** Create new features or derive useful information from existing features. For example, using techniques like Principal Component Analysis (PCA) or extracting features from text data using methods like bag-of-words or TF-IDF.

**Data Integration:**

**Combine multiple datasets:** Merge or join datasets if you have data from different sources that need to be combined.

**Standardize data:** Deal with inconsistencies in the data, such as differences in naming conventions or data formats, by making them consistent.

**Data Reduction:**

**Reduce feature dimensionality:** Decrease the number of features while retaining important information. Techniques like PCA or t-SNE can help with this.

**Sample data:** If the dataset is too large, you may need to work with a smaller subset. Techniques like random sampling or stratified sampling can be used.

**Handling Imbalanced Data:**

**Address imbalanced class distribution:** If one class is much more prevalent than others, techniques like oversampling (e.g., SMOTE) or undersampling can be used to balance the classes.

**Normalization:**

**Normalize the data:** Ensure that different features are on a similar scale. This is important for algorithms that are sensitive to differences in magnitudes, such as distance-based methods.

1. **Model:**
2. **Conclusion:**