# Emotion Analysis and Prediction

This program is designed to analyze and predict emotions based on various health-related metrics. It processes a dataset containing emotional and physiological data, performs data cleaning, and builds machine learning models to classify emotions using SVM and Random Forest classifiers. The program follows a structured approach to data preprocessing, feature engineering, and model training and evaluation.

## Data Cleaning or Wrangling

Data cleaning or wrangling is essential for transforming raw data into a usable format. It involves handling missing values, filling in default values, and preparing the dataset for further analysis. This ensures that the data is accurate and ready for analysis or model training.

Data:

* Raw Dataset: The initial dataset (daily\_fitbit\_sema\_df\_unprocessed.csv) containing various metrics related to health and emotions.
* Cleaned Dataset: The processed dataset (data\_updated.csv) with missing values handled and unnecessary columns removed.

Dependencies:

* Pandas (pandas): For data manipulation and analysis, including handling missing values and transforming the dataset.
* NumPy (numpy): Provides support for numerical operations and array handling.

## Data Analysis

Data analysis involves examining and modelling data to discover useful insights, which supports decision-making and understanding of patterns within the data. This includes summarizing data, visualizing distributions, and analyzing trends.

Descriptive Statistics:

* Provides summaries of numerical data, such as counts and medians, to understand dataset characteristics.

Data Visualization:

* Emotion Analysis Plot: A bar plot showing the total counts of different emotions, visualizing the distribution of emotional states.

A graph of emotions

Description automatically generated

Dependencies:

* Matplotlib (matplotlib): Essential for creating plots and visualizations to represent data trends.
* Pandas (pandas): Used for data analysis and preparation before visualization.

## Prediction Model

Prediction models use historical data to predict future outcomes or classify data into predefined categories. In this context, machine learning models are used to classify emotions based on various features.

Steps Involved:

1. Data Preparation: Involves filling missing values, encoding categorical features, and splitting data into training and testing sets.
2. Model Training: Fits machine learning models, including SVM and Random Forest classifiers, to the training data.
3. Prediction and Evaluation: Evaluates model performance on the test data using metrics such as accuracy and classification reports.

Results:

The machine learning models are trained to predict various emotions based on the following features. The independent variables (X) used in this process include nightly\_temperature, spo2, full\_sleep\_breathing\_rate, stress\_score, sleep\_points\_percentage, exertion\_points\_percentage, responsiveness\_points\_percentage, daily\_temperature\_variation, calories, filteredDemographicVO2Max, distance, bpm, lightly\_active\_minutes, moderately\_active\_minutes, very\_active\_minutes, sedentary\_minutes, sleep\_duration, minutesToFallAsleep, minutesAsleep, minutesAwake, minutesAfterWakeup, sleep\_efficiency, sleep\_deep\_ratio, sleep\_wake\_ratio, sleep\_light\_ratio, sleep\_rem\_ratio, steps, minutes\_in\_default\_zone\_1, minutes\_below\_default\_zone\_1, minutes\_in\_default\_zone\_2, minutes\_in\_default\_zone\_3, age, gender, ENTERTAINMENT, GYM, HOME, HOME\_OFFICE, OTHER, OUTDOORS, TRANSIT, and WORK. The dependent variables (y), which represent the emotions being predicted, are ALERT, HAPPY, NEUTRAL, RESTED, SAD, TENSE, and TIRED.

The models' performance is evaluated with accuracy scores ranging from 89% to 98%. For each emotion, both SVM and Random Forest models are trained and tested. The confusion matrices are plotted and saved for visual assessment, and accuracy scores along with detailed classification reports are provided.

Dependencies:

* Scikit-learn (sklearn): Provides tools for building and evaluating machine learning models, including classifiers and performance metrics.
* Matplotlib (matplotlib): For generating visualizations such as confusion matrices.

## Conclusion

This program effectively analyzes and predicts emotions from health-related metrics. By cleaning and preprocessing the data, it prepares it for accurate analysis and model training. Using machine learning models like SVM and Random Forest, the program achieves high accuracy scores ranging from 89% to 98% in classifying emotions. With dependencies on Pandas, NumPy, Scikit-learn, and Matplotlib, it ensures robust data manipulation, analysis, and visualization, delivering valuable insights into emotional patterns and trends.