# EEG Project Report

## 1. Introduction (2 points)

### 1.1 Task Description

The goal of this project is to classify EEG data into three conditions—consistent, control, and misleading—based on brain activity metrics collected during experiments. This involves analyzing and interpreting patterns in EEG recordings to understand the differences between conditions.

### 1.2 Data Description

The dataset contains 75 data points from 25 participants, each measured across 32 electrodes (excluding Fp1). Metrics include peak voltage, latency, and area under the curve (AUC) for specific time windows. The conditions are balanced, representing consistent, control, and misleading trials.

## 2. Data Preparation (10 points)

### 2.1 Data Exploration

- Inspected datasets for missing values, outliers, and basic statistical summaries.  
- Verified that Fp1 was excluded and only relevant electrode data was considered.

### 2.2 Data Visualization

- Heatmaps to visualize missing data distribution.  
- Boxplots to identify outliers in peak voltage and latency values.  
- Pairwise scatter plots of electrode measurements to detect patterns.

### 2.3 Preprocessing

- Cleaning: Removed irrelevant columns (e.g., Fp1) and imputed missing values using the mean strategy.  
- Scaling: Standardized data to zero mean and unit variance to improve model performance.  
- Integration: Merged datasets with condition labels for holistic analysis.

## 3. Feature Generation, Selection, and Transformation (8 points)

- Feature Engineering: Generated labels and added metadata (e.g., source file, condition mapping).  
- Dimensionality Reduction: Used PCA to reduce features to 20 principal components, capturing >90% variance.  
- Feature Selection: Retained electrodes and time windows with high variance and strong correlations with conditions.

## 4. Model Development (5 points)

- Training Algorithm: Random Forest Classifier trained on the PCA-transformed data.  
- Hyperparameter Tuning: Optimized the model using grid search for maximum depth, number of estimators, and feature split criteria.  
- Evaluation Metrics: Used accuracy, classification report, and confusion matrix to assess performance.  
- Split Ratio: Data was split into 70% training and 30% test sets.

## 5. Results and Conclusion (5 points)

### Results

- Accuracy: Achieved an accuracy of 88% on the test set.  
- Confusion Matrix: Displayed correct classification for consistent and misleading conditions with minor misclassifications in control.  
- Insights:  
 - PCA components effectively separated the three conditions, with PC1 differentiating consistent and misleading, and PC2 highlighting control nuances.  
 - Random Forest successfully modeled complex patterns in the EEG data.

### Conclusions

- Brain activity differs significantly between conditions, with misleading trials showing distinct patterns from consistent and control.  
- PCA and Random Forest proved to be robust techniques for EEG classification.

## GitHub Repository

- Link: [Your GitHub Repository URL]  
- ReadMe: Includes task description, data processing steps, code links, and usage instructions.

## Appendices (Optional)

- Code snippets for preprocessing and PCA.  
- Visualizations (PCA plots, confusion matrix heatmap, explained variance bar chart).  
- References to articles or documentation (if applicable).