Epileptic Seizure Prediction via EEG Analysis

- Aman Sahu

Introduction to Epilepsy & EEG

Epilepsy Overview: A serious, endemic neurological disorder caused by abnormal electrical discharging in the brain.

The Role of EEG: The electroencephalogram (EEG) is crucial for the evaluation, diagnosis, and accurate classification of neurophysiological disorders like epilepsy.

EEG Data Acquisition

Dataset: EEG recordings from 400 subjects (200 epileptic patients and 200 healthy individuals).

Methodology: Data collected at Dicle University Medical Faculty Hospital; used a PCI-MIO 16E DAQ card with LabVIEW; signals sampled at 173 Hz.

Channels/Bands: Analysis focused on four bipolar channels (F7–C3, F8–C4, T5–O1, T6–O2) across four frequency bands (delta, theta, alpha, beta).

Data Preprocessing & Segmentation

**Original Data:** 500 subjects, 23.6-second recordings, 4097 data points per recording.

Segmentation: Each 4097-point signal was shuffled and divided into 23 segments of 178 data points (1 second each).

Final Dataset Shape: 11,500 samples (trials), with 178 explanatory features (X_1-X_{178}).

Classification Task: Focused on binary classification: distinguishing seizure (Class 1) from non-seizure (Classes 2-5).

Time - domain visualizations

Goal: visually compare signal shapes across labels and compute an envelope (mean ± std) for each label so that differences are obvious.

Comparison in Time Domain

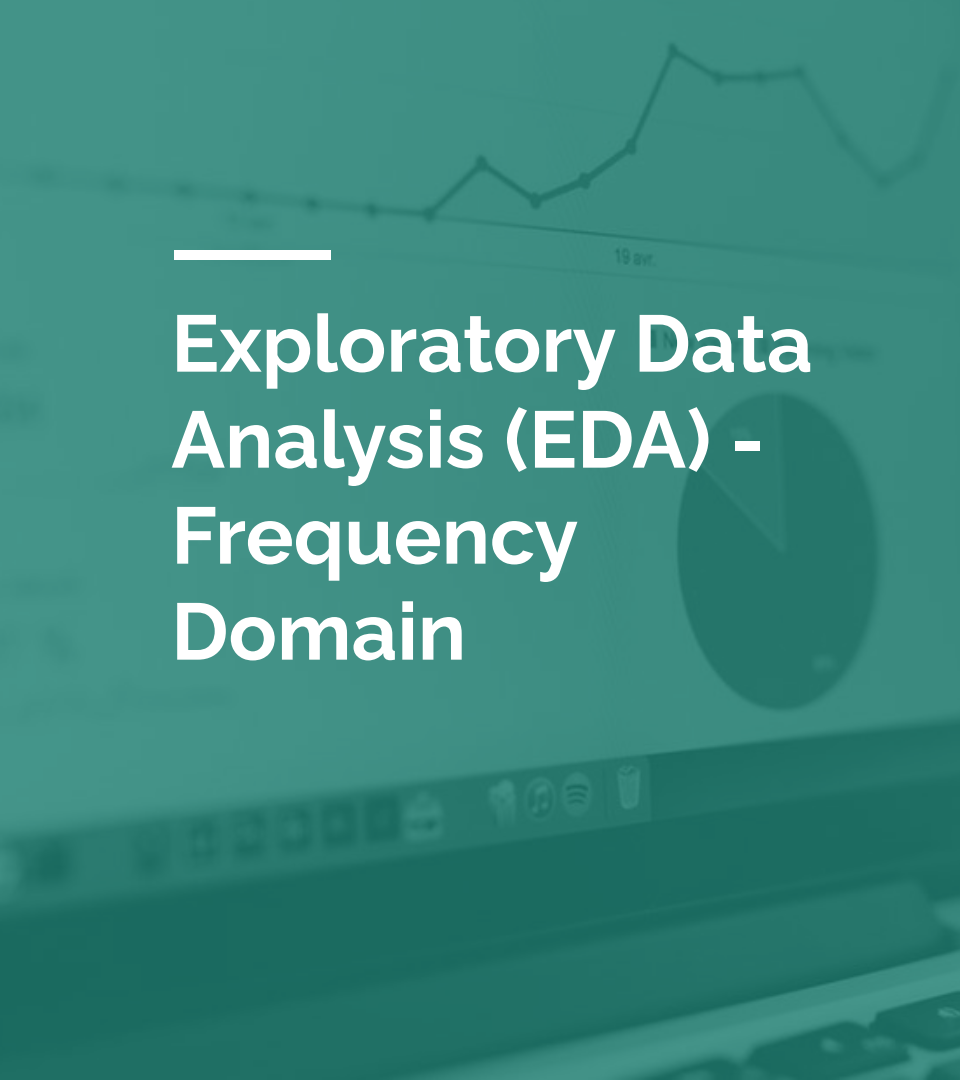

Mean Signal per Class - Time Domain
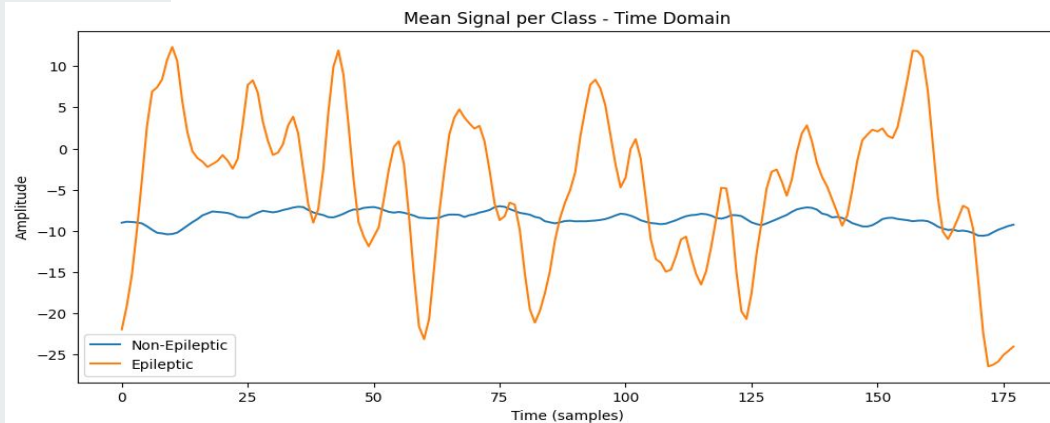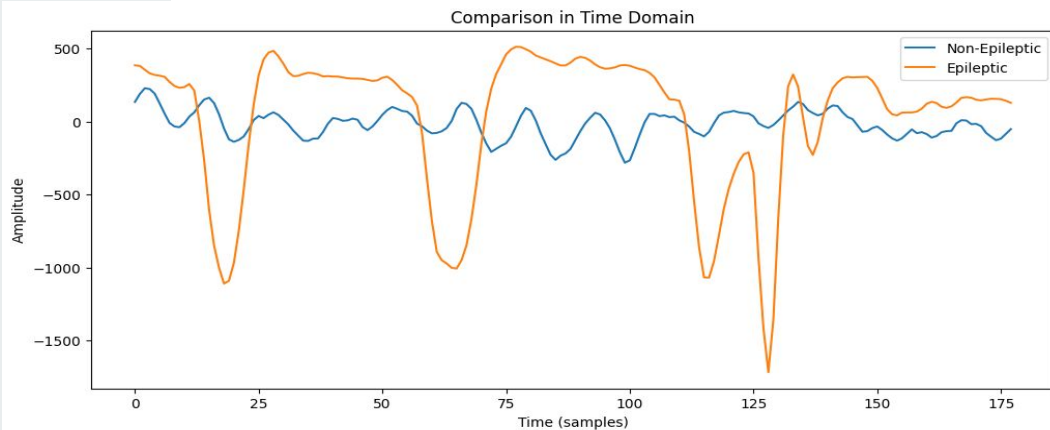
Exploratory Data Analysis (EDA) - Time Domain

- **Key Observation:** Epileptic signals appear to be significantly more deviating than non-epileptic signals.
- **Statistical Differences:** The Mean and Standard Deviation (Std) values showed a large difference between classes (e.g., Non-Epileptic Mean: ~1260 vs. Epileptic Mean: ~290).
- **Action:** The significant difference in values necessitates normalizing/scaling the data for some models.
- **Visual:** *Include a graph comparing the mean ± std envelopes of epileptic vs. non-epileptic signals.*

Frequency-domain visualizations

Goal: compute and plot average Power Spectral Density (PSD) per class and compare band-powers across classes.

Frequency Spectrum (FFT)

Power Spectral Density (PSD)
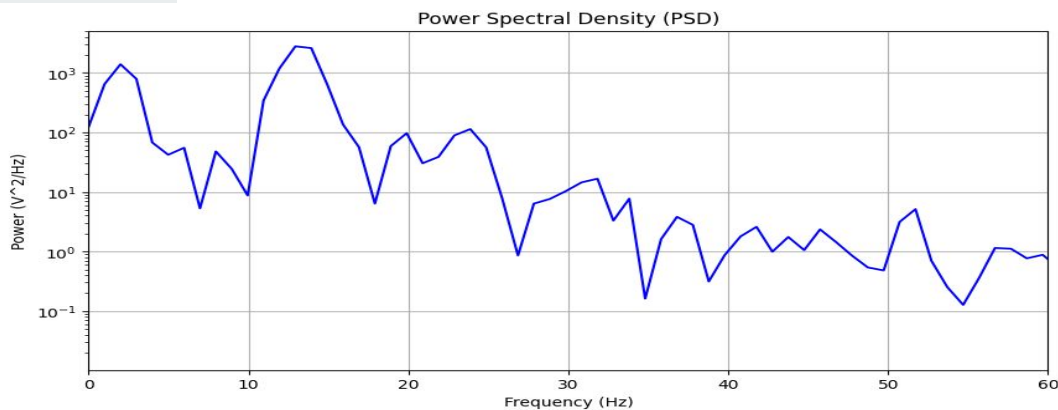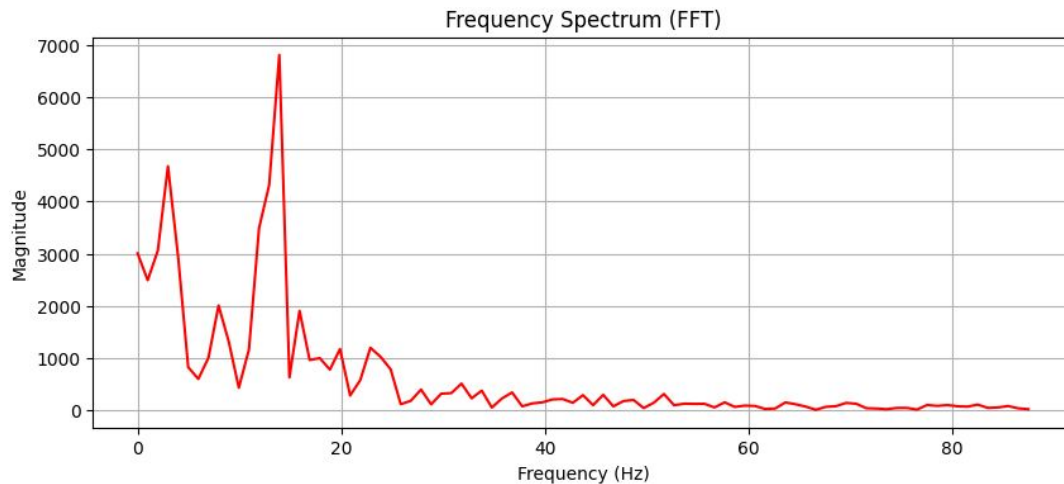
Exploratory Data Analysis (EDA) - Frequency Domain

Goal: Compute and compare average Power Spectral Density (PSD) and band-powers per class.

Key Finding: The PSD plot clearly shows dominant energy peaks around 14-15 Hz and a notable peak near 5 Hz.

Significance: Frequency analysis (like PSD) provides a clearer picture of where the signal's energy is concentrated.

**Visual:** *Include a plot of the average Power Spectral Density (PSD).*

Model Building & Evaluation

Models Used: Multiple machine learning models were trained and evaluated, including Logistic Regression, SVM, Random Forest, XGBoost, and MLP.

Preprocessing for Models: Features were standardized for scaling-sensitive models (LR, SVM, MLP).

Top Performer: XGBoost delivered the highest performance.

Confusion Matrix
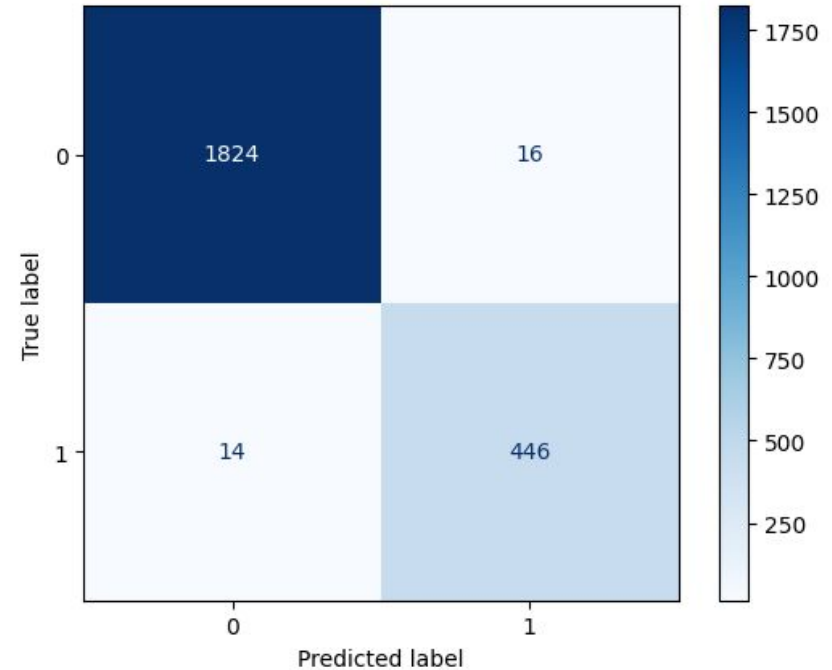
- The model **rarely misses seizures** (very low FN = 14), which is crucial for patient safety.
- **False alarms are minimal** (FP = 16), reducing unnecessary clinical interventions.
- Balanced high sensitivity and specificity indicate the model is **well-calibrated and reliable**.

Results and Conclusion

Model Performance: 98.7% accuracy.

Clinical Relevance: The model has very high recall (minimizing missed seizures, FN=14) and minimal false alarms (FP=16).

Conclusion: EEG-based binary classification using machine learning is highly effective, and ensemble models like XGBoost are particularly well-suited for this task.

Future Potential: Strong potential for clinical decision-support systems, pending further validation.