

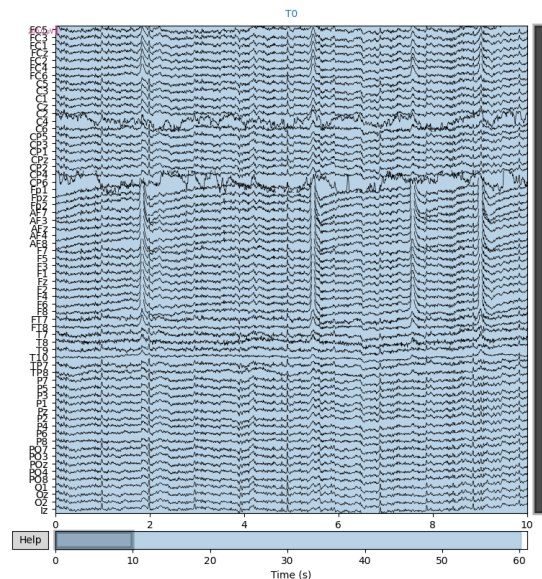
Data Engineering Report: Cleaning EEG Data

Steps of Cleaning EEG Data:

Raw Data	1
Applying Reference	1
Filtering	2
Downsampling	2
Handling Bad Channels	3
Independent Component Analysis	4
Comparison of the Raw and the Final Data	4

Raw Data

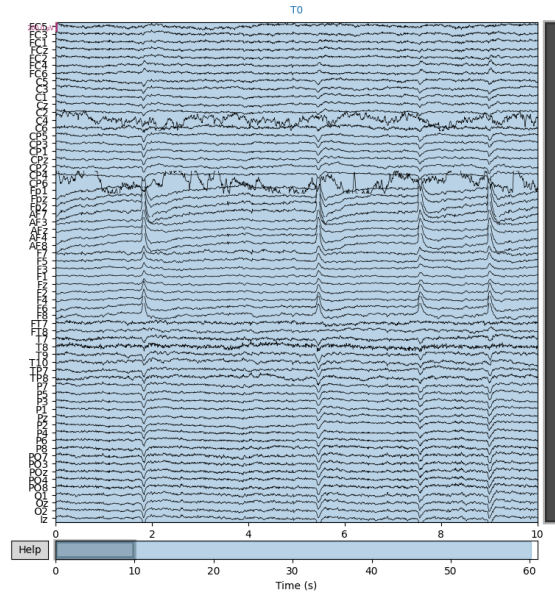
EEG atau *Electroencephalography* merupakan metode yang digunakan untuk merekam aktivitas sinyal listrik otak dengan menggunakan elektroda di area kepala. Sinyal listrik otak berukuran sangat kecil yaitu dalam satuan microvolt. Oleh karena itu, EEG data rentan terhadap *noise*. Dalam *project* ini, digunakan data EEG 64 channels dari *open dataset library* MNE, dapat diakses melalui link berikut [ini](#).



Gambar 1. Raw Data EEG

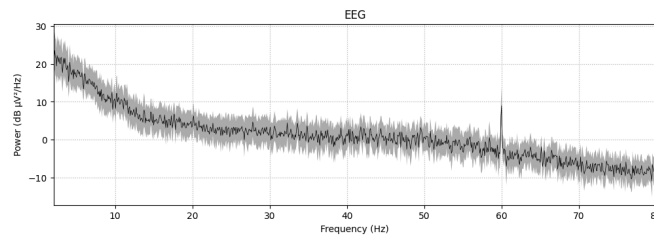
Applying Reference

Data EEG merupakan data dalam satuan volt, voltase sendiri berarti perbedaan tegangan antar suatu titik terhadap acuan. Titik acuan dalam EEG menjadi *baseline* sinyal data yang ada. Oleh karena itu, sinyal acuan yang ideal merupakan sinyal yang tidak terinterferensi dengan sinyal otak tapi dapat menangkap *noise* yang terjadi saat pengambilan data. Acuan lazimnya merupakan sinyal yang didapat dari telinga, hal ini dikarenakan telinga terletak dekat dengan elektroda untuk sinyal sehingga dapat menangkap *noise* di sekitar kepala.



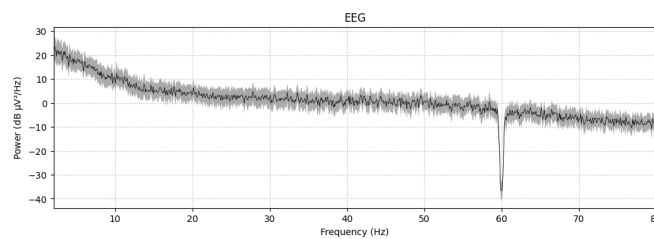
Gambar 2. Data EEG setelah Direferensikan

Filtering



Gambar 3. Visualisasi Data dalam Frequency Domain

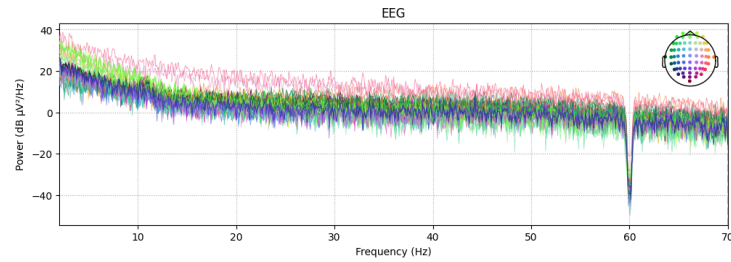
Berdasarkan gambar, terdapat anomali sekitar 60 Hz. Untuk mengatasi hal tersebut dilakukan *filtering* dari 59-61 Hz.



Gambar 4. Visualisasi Data dalam Frequency Domain setelah *Filtering*

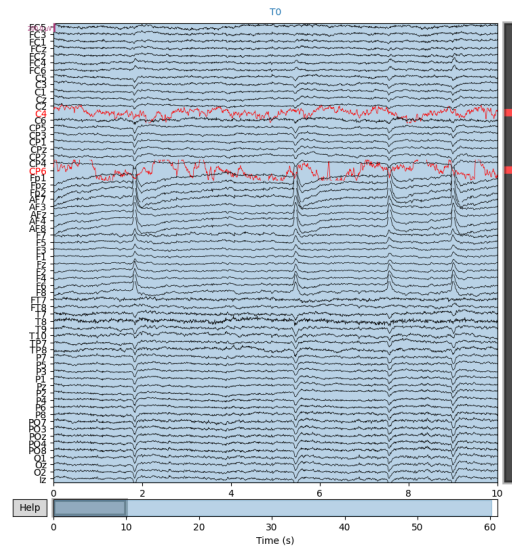
Downsampling

Downsampling adalah proses mengurangi jumlah data untuk analisis, penyimpanan, atau pemrosesan. Hal ini dilakukan dengan mengurangi rate cuplikan data.



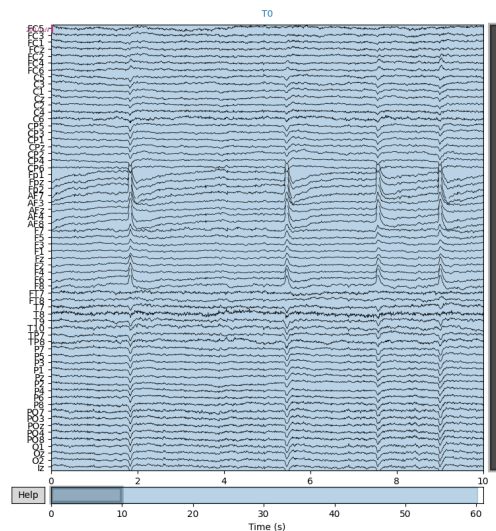
Gambar 5. Visualisasi Data dalam Frequency Domain setelah *Downsampling*

Handling Bad Channels



Gambar 6. Data EEG setelah *Downsampling*

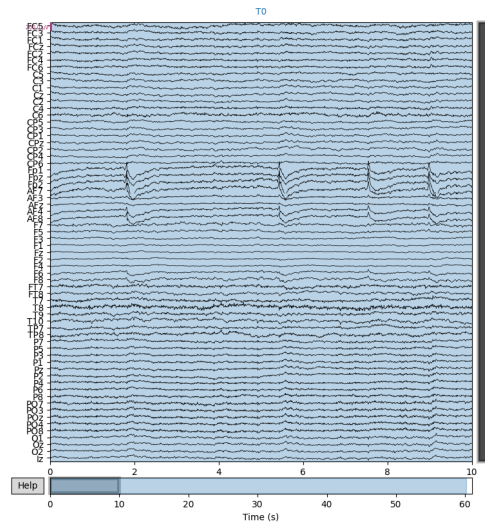
Dapat diperhatikan bahwa, terdapat dua channel yang mengalami anomali. Hal ini bisa disebabkan oleh lemahnya sinyal dalam channel tersebut atau faktor eksternal lainnya. Untuk mengatasi hal ini dilakukan *interpolation*. *Interpolation* dalam EEG data suatu channel diestimasi dari data EEG channel-channel lain di dekatnya.



Gambar 7. Data EEG setelah *Interpolation*

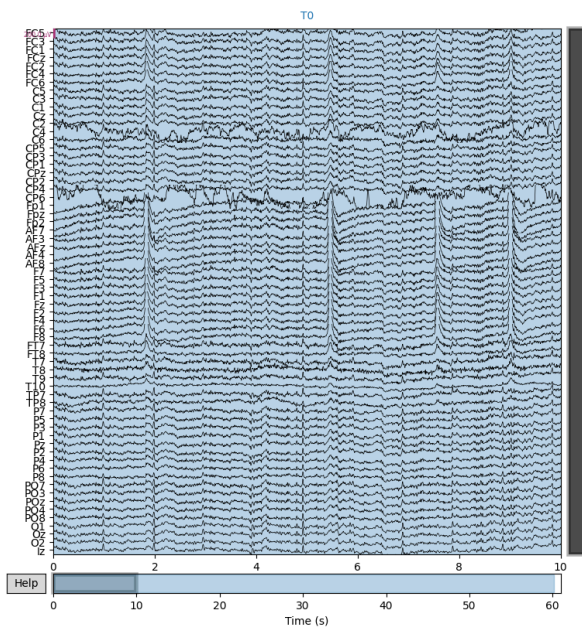
Independent Component Analysis

Pengambilan data EEG dilakukan menggunakan banyak sensor dengan area yang cukup berdekatan antara satu sensor dengan lainnya. Hal ini dapat menyebabkan *The Cocktail Party Problem*, dimana satu sensor dapat menangkap data sinyal EEG dari area lain. Permasalahan ini menyebabkan hasil pembacaan sinyal EEG dalam suatu sensor menjadi kusut karena sensor tersebut menangkap sinyal dari banyak tempat sekaligus. Untuk menyelesaikan permasalahan ini dilakukan *Independent Component Analysis* (ICA). Sederhananya, ICA mengurai data yang kusut dalam tiap sensor dan mendistribusikan sinyal-sinyal tersebut dalam sensor yang seharusnya.

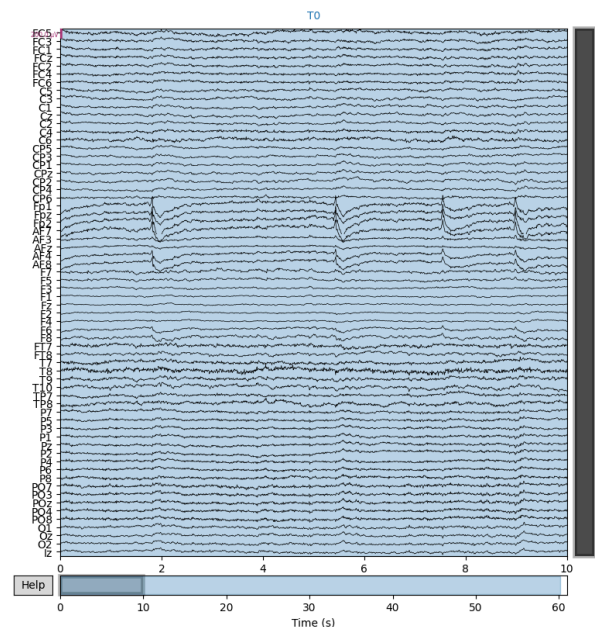


Gambar 8. Data EEG setelah dilakukan *Independent Component Analysis*

Comparison of the Raw and the Final Data



Gambar 9. Raw Data



Gambar 10. Final Data