A Case Study of Epilepsy in Canines & Humans

By Amit Kumar, Immanuvel Vasanth, Gopi Chand Parupalli, Laxmi Nageswari

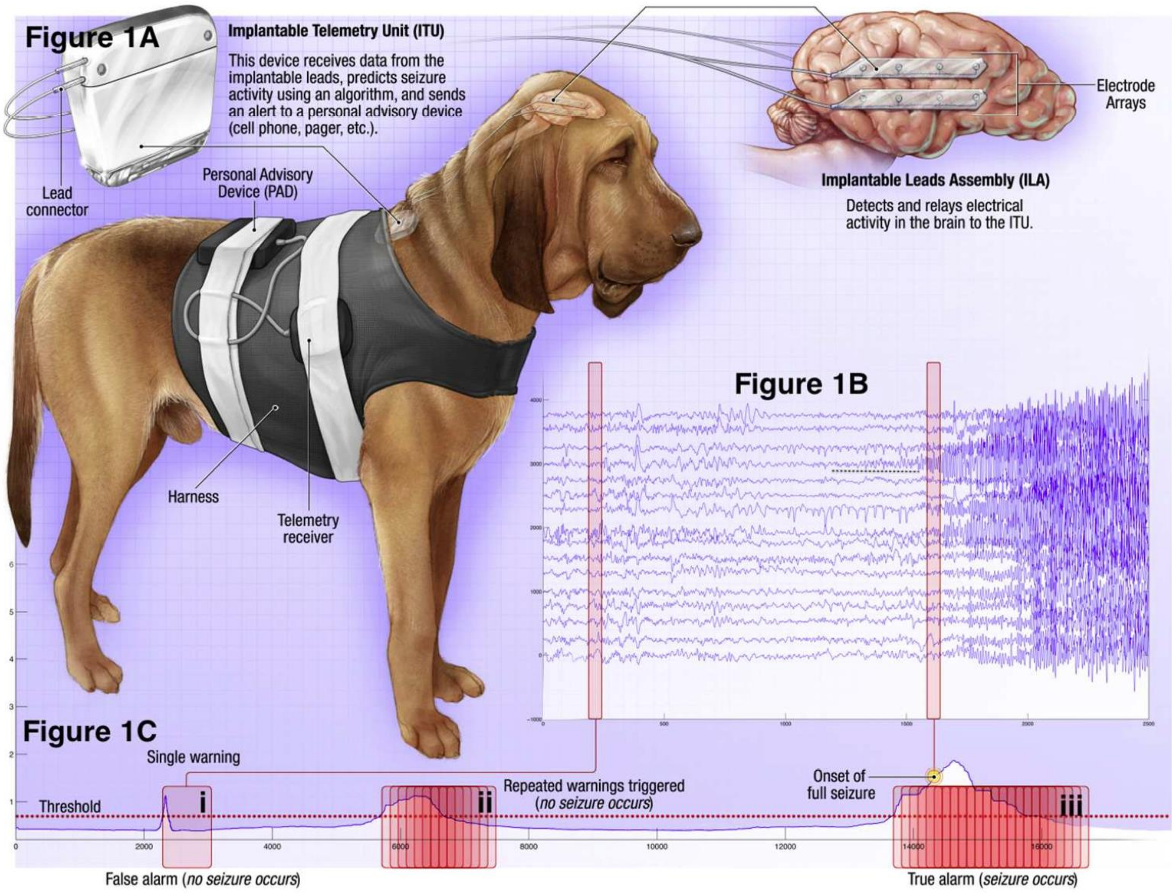
We would like to start with thanking Indian School Of Business and Cyient Insights for providing us an opportunity to work on this beautiful project. The project was interesting from viewpoint that it touched various aspects of data science like data visualization, feature engineering, predictive model building and fine tuning, data volume. It challenged us at every step moving forward making the whole journey exciting and overcoming each obstacle a memorable experience.

Project Description

"Seizure forecasting systems hold promise for improving the quality of life for patients with epilepsy"

Epilepsy afflicts nearly 1% of the world's population, and is characterized by the occurrence of spontaneous seizures. For many patients, anticonvulsant medications can be given at sufficiently high doses to prevent seizures, but patients frequently suffer side effects. For 20-40% of patients with epilepsy, medications are not effective -- and even after surgical removal of epilepsy-causing brain tissue, many patients continue to experience spontaneous seizures. Despite the fact that seizures occur infrequently, patients with epilepsy experience persistent anxiety due to the possibility of a seizure occurring. Seizure forecasting systems have the potential to help patients with epilepsy lead more normal lives. In order for EEG-based seizure forecasting systems to work effectively, computational algorithms must reliably identify periods of increased probability of seizure occurrence. If these seizure-permissive brain states can be identified, devices designed to warn patients of impeding seizures would be possible. Patients could avoid potentially dangerous activities like driving or swimming, and medications could be administered only when needed to prevent impending seizures, reducing overall side effects.

There is emerging evidence that the temporal dynamics of brain activity can be classified into 4 states: Interictal (between seizures, or baseline), Preictal (prior to seizure), Ictal (seizure), and Post-ictal (after seizures). Seizure forecasting requires the ability to reliably identify a preictal state that can be differentiated from the interictal, ictal, and postictal state. The primary challenge in seizure forecasting is differentiating between the preictal and interictal states. The goal of the competition is to demonstrate the existence and accurate classification of the preictal brain state in dogs and humans with naturally occurring epilepsy. Intracranial EEG was recorded from dogs with naturally occurring epilepsy using an ambulatory monitoring system. EEG was sampled from 16 electrodes at 400 Hz, and recorded voltages were referenced to the group average. These are long duration recordings, spanning multiple months up to a year and recording up to a hundred seizures in some dogs.

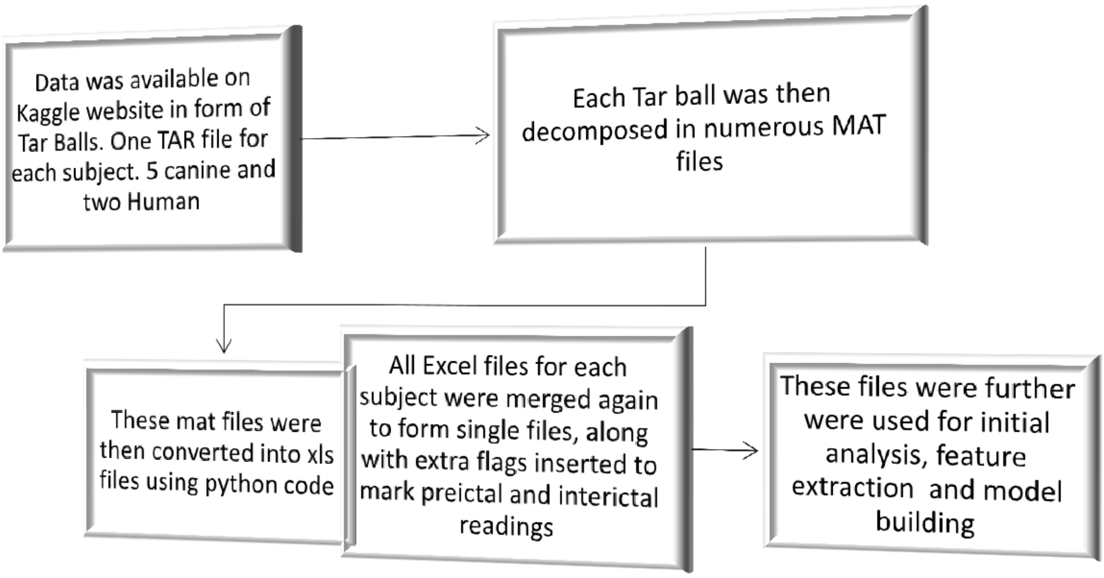


In addition, datasets from patients with epilepsy undergoing intracranial EEG monitoring to identify a region of brain that can be resected to prevent future seizures are included in the contest. These datasets have varying numbers of electrodes and are sampled at 5000 Hz, with recorded voltages referenced to an electrode outside the brain. The challenge is to distinguish between ten minute long data clips covering an hour prior to a seizure, and ten minute iEEG clips of interictal activity. Seizures are known to cluster, or occur in groups. Patients who typically have seizure clusters receive little benefit from forecasting follow-on seizures. For this contest only lead seizures, defined here as seizures occurring four hours or more after another seizure, are included in the training and testing data sets. In order to avoid any potential contamination between interictal, preictal, and post-ictal EEG signals interictal segments in the canine training and test data were restricted to be at least one week before or after any seizure. In the human data, where the entire monitoring session may last less than one week, interictal data segments were restricted to be at least four hours before or after any seizure. Interictal data segments were chosen at random within these restrictions for both canine and human subjects.

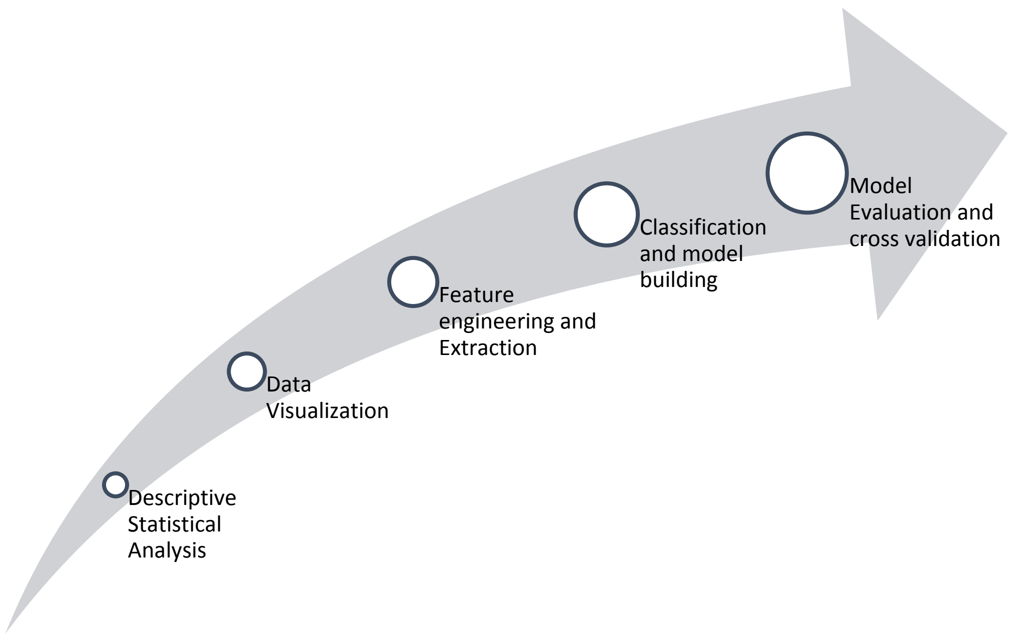
<u>Data Cleaning and Enrichment</u>

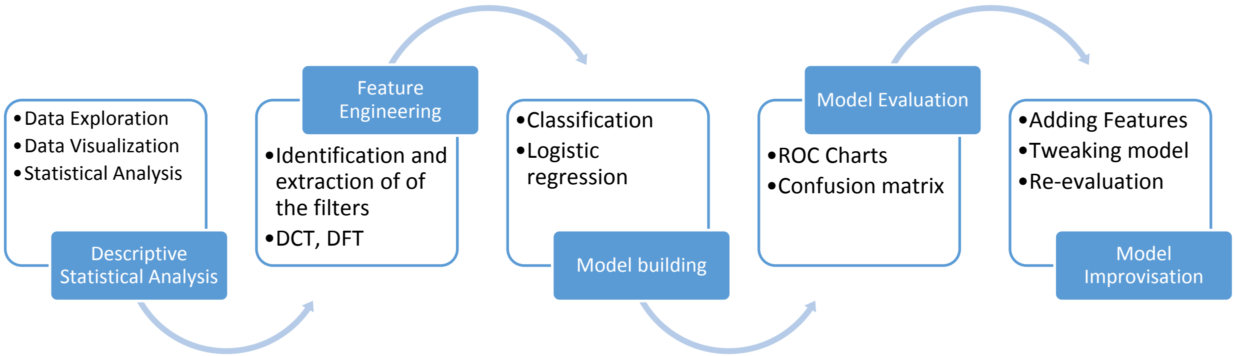
Data collection part was not involved in this project as we got the data directly from Kaggle site in form of Tar balls. But for data Exploration part we did some data manipulation. As data was in form of Tar balls so we needed to extract the same and extracted files were chunk of mat files corresponding to each tar ball. Once we extracted the files we further added some flags for aid of interpretation during data exploration.

We added interictal and preictal flags while converting the mat files in csv format and merging the separate files in one csv file per tar ball. The process can be summarized as follows.



<u>Overall Approach</u>





End Result

As an end result this project describes an automated classification of EEG signals for the detection of epileptic seizures using statistical pattern recognition. An overall classification of accuracy 65% was achieved. The results confirmed that the proposed algorithm has a potential in the classification of EEG signals and detection of epileptic seizures, and could thus further improve the diagnosis of epilepsy.

Learnings

Patience is the Key - Understanding the problem and preparing dataset for model building can take up to 70% of your total project time, and also it is most critical part of the whole project.

Communicate regularly– Meet regularly with project team members and stakeholders. Be disciplined about communication with your stake holders. Communicate to them every small progress that you make on project and it is fine if in some weeks you are not able to make progress. In those scenarios communicate what efforts you made and challenges you faced.

Work around your core competencies - Exploit the core competencies of your team members. Each project involves various aspects of data science let each member drive the area of their interest and expertise, so that you can work in parallel and utilize time.

Way Forward - Find "the thing" that motivates you to practice what you learned and to learn more, and then do that thing. That could be personal data science projects, Kaggle competitions, online courses, reading books, reading blogs, attending meetups or conferences.