Background:

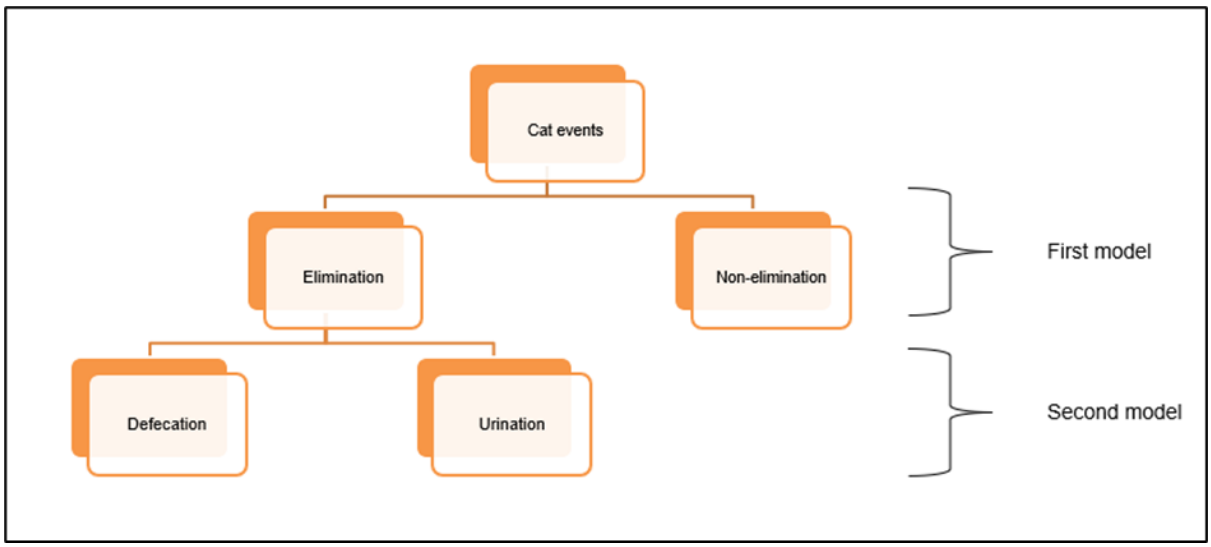
The client is a global company in the Pet Care and pet food space. The company owns a range of smart pet products aiming to improve the health of pets by monitoring the pet's activities and generating a portfolio for the pet owners to consume.

The Smart Litter System is a smart cat litter system that is equipped with four load sensors capturing any load disturbance happening on the litter box. The device functions at 40 Hz frequency rate which means it records 40 samples per second. Multiple devices are used in the client facility to capture all activities in the device in the form of load signals on a day-to-day basis. The client facility is also equipped with cameras which capture photo feeds and video feeds that allows manual tagging of the various cat and human interaction events with the litter system.

The cat interaction events are as shown below:

a. Elimination
    i. Defecation
    ii. Urination
b. Non-Elimination

The objective of this project is to develop a solution approach to correctly classify the events as elimination or non-elimination and further classify the elimination events into urination and defecation.



Note: For the first model, relabel defecation and urination events as "Elimination".

Data:

Client will provide the following data for the project:

- Event Load Sensor Data: Event named load sensor data containing timesteps and Load Sensor values. The schema of these files looks as below. *Note: File name is event name here*

CSV File

t	: Time elapsed after start of event (in days:hours:minutes:seconds)
lc0	: Load data from sensor 1
lc1	: Load data from sensor 2
lc2	: Load data from sensor 3
lc3	: Load data from sensor 4
timestamp	: Time steps
sum	: Sum of the indivial load cell values (lc0 + lc1 + lc2 + lc3)

- Tags Data: Event named label/tag data containing the information about the event such as cat name, event tags, free text etc.

Example Json File

```
{
    Device_ID: "VA00000000000020",                    ID of the device on which the event was recorded
    Event_ID: "RXZlbnRUbmZvcm1hdGlvbjo=",             ID of the event
    Event_start_time: "2020-07-18T07:26:16-05:00",    Time at which the device started recording
    Tags: ["urination", "cat in box"],                Activities happening during the event and details of the event
    Free_text: "7:27:11\t7:28:17\tConfirmed activity, cat in the box, urination",   Activity start time ; Activity end time; details of the activity
    Name_of_cat: "Star Lord",                         Name of the Cat
    Weight_of_cat: "5193",                            Estimate of weight of Cat
    Tare_weight: [-138, 2121, -729, 1173]             Tare weight ( weight of the litter and box )
}
```

While cleaning the events ignore events with tags such as merged, split, combine, no camera, review and trimmed.

Demonstrate the portfolio of algorithms to correctly predict the ***Urination, Defecation and Non-elimination*** among the cat events***.***

Company will also provide a clear and detailed description of the dataset including all the fields present. Company will make available knowledgeable personnel to provide any necessary background on data and business context. These personnel will also help Tiger in working with the end consumers of our models and identifying which features could be actionable. They would also help in corroborating findings from the models and can help provide an insider's business perspective.

Deliverables:

Data Preparation

Demonstrates:

- Provide a report for data sufficiency
- Decisions on data selection
- Decisions on merging the data, joining key
- Decisions on the target variable

Data exploration and Hypothesis Validation

Demonstrates:

- Explore the support and time duration of activities

- How do the statistical measures (e.g., mean load value, variance etc.) of load sensor data differ for each activity?
- Describe and understand the activities' behaviour
    - Visualization of raw signals for different activities
    - For elimination (defecation and urination) usually the cat digs up the litter, eliminates and covers it up. Does digging up and covering activity differ between urination and defecation? If so, how?
    - Summarize the patterns you observe for each activity in a few lines
- Pre-process the load sensor data (cleaning, normalization, etc.)

Feature Engineering

Demonstrates:

- Perform appropriate data transformation/aggregation Hint: Look for time domain, frequency domain features
- Does each cat have a specific behaviour pattern? Are the patterns similar across different cat morphologies?
- Explain why and how the features would help classify the activities? (Hint: Explain the relationship of features with target)

Train model to predict the elimination and non-elimination activity:

- Identify the right modelling technique
- Use some appropriate feature selection technique(s) to select important features
- Justify the models based on
    - Model Performance using the appropriate model evaluation metrics (Hint: Confusion Matrix, Multiclass overall and class wise metrics)
    - Hold out validation (Hint: Create train, test and validation datasets)
- Residual analysis
    - Summarize the cause/patterns in misclassifications
    - Explore new features that would help in rectifying misclassifications

Train model to predict the urination and defecation:

- Identify the elimination events from the previous model to classify if it is urination or defecation
- Identify the right modelling technique
- Use some appropriate feature selection technique(s) to select important features
- Justify the models based on
    - Model Performance using the appropriate model evaluation metrics (Hint: Confusion Matrix, Multiclass overall and class wise metrics)
    - Hold out validation (Hint: Create train, test and validation datasets)
- Residual analysis
    - Summarize the cause/patterns in misclassifications
    - Explore new features that would help in rectifying misclassifications

Deliverable Best Practices:

- Structured code base with a few tests
- The submitted code should satisfy coding standards.
- The README should contain, or point to a doc that contains, information about the dataset and how to run the notebooks/code.
- Notebooks should only have relevant analysis and should run successfully in a sequential manner. Notebooks should have adequate documentation on the analysis using proper markdown.
- Please make use of the Templates. Use CT as a starting point and make customizations based on your solution (Optional)
- Share your final codes along with EDA and modelling reports generated with mentors ahead of the presentation date through GitHub repo
- Prepare the presentation in such a way to be able to cover it within 25-30 mins, excluding Q&A. Total number of slides should not exceed 25
- Presentation should cover the following sections: Objective, Data Preparation, Exploratory Analysis, Model results, Misclassification Analysis and Conclusion.
- Presentation should cover the answers to all the questions asked in the problem statement