PTRAIL Dashboard
User Manual
Date: March 31st, 2023
Yaksh J Haranwala

Dashboard Version: 0.1
https://github.com/YakshHaranwala/PTRAIL yjharanwala@mun.ca

Table Of Contents

Dashboard Version: 0.1
https://github.com/YakshHaranwala/PTRAIL
yjharanwala@mun.ca

1. Introduction

The analysis of movement data, which consists of sequences of spatial positions recorded over time, can yield valuable insights, but it presents several challenges. The data is often stored in a complex format and produced in large volumes, and the technologies used to collect it, such as GPS, GSM, Wifi, or RFID, can introduce errors due to device failure or connectivity issues. As a result, performing mobility data mining requires a sequence of steps that can be handled by different computational libraries, making the data representation and feature extraction a non-homogeneous process with varying inputs and outputs. This can make the whole data analysis process cumbersome and time-consuming.

Fortunately, the PTRAIL mobility data processing dashboard is a cutting-edge visual tool that simplifies various trajectory data preprocessing tasks, including filtering, interpolation, trajectory visualization, and data comparison. It provides researchers in the mobility data field, who may be unfamiliar with technical tools like Python or Pandas, with new avenues for exploring their data. The PTRAIL Dashboard significantly reduces the time researchers spend preprocessing mobility data by offering a visual tool that streamlines the data preparation process, enabling them to perform machine learning tasks such as classification and regression with the click of a button, rather than through Python environments or Jupyter notebooks.

Disclaimer: The PTRAIL mobility data processing dashboard is like a superhero that swoops in to save researchers from the tedious task of preprocessing data. But, like all superheroes, it's not invincible. At least, not yet. You see, PTRAIL is currently in beta mode, which means it's still flexing its muscles and working out the kinks. So, while you're using it, you might encounter some bugs or lags, kind of like when Spiderman was still learning how to use his web-slingers. But don't worry, PTRAIL is still a hero in the making, and with your help, it'll become an even more powerful tool for analyzing mobility data.

2. Dashboard Overview

The PTRAIL Dashboard is developed using python's open source PyQt5 framework which provides multi-platform support in order to streamline the process of development. The core backend of the Dashboard uses PTRAIL (Haidri et al., 2022), a highly parallelized Python package for preprocessing trajectory data. In essence, the dashboard is the visual version of the PTRAIL library. For further information on PTRAIL and its architecture, refer to PTRAIL's github. In this section, we will briefly describe the dashboard and its main components. As we can see in the figure below, the dashboard has 5 main components mainly:

1. Command Palette
2. Map Viewer
3. Dataframe Viewer
4. Comparison Graph
5. Feature Importance Viewer

We will further describe each component, its functionality, usage and how to interpret the results in more detail in the upcoming sections.

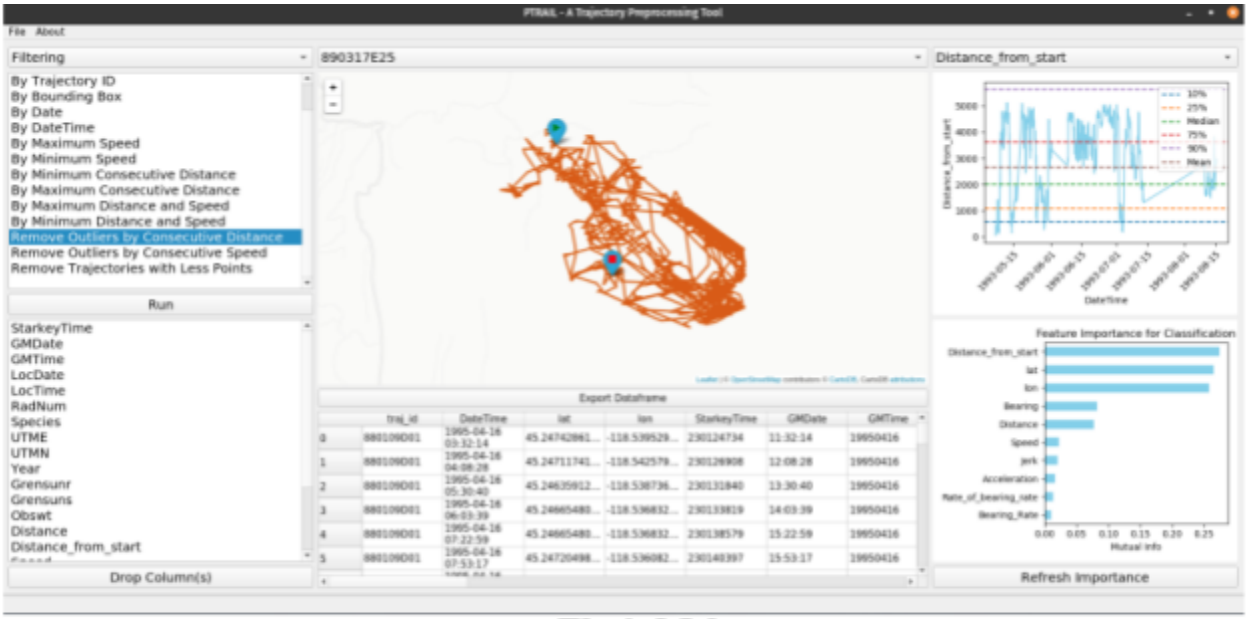

Fig 1: Main Dashboard View

2.1 Launching the Dashboard

As we mentioned earlier, the PTRAIL Dashboard is still in its beta stage. As a result, we have not yet released an executable file for the software in order to avoid system dependency issues at the moment. Therefore, the PTRAIL Dashboard is currently available to use by cloning PTRAIL's github repository to your local machine. The steps to launch the dashboard are as follows.

1. Download and Install Python 3.7+ on your machine from Python's Official Website.
2. Download and extract the PTRAIL repository from Github into your local machine.
3. Open terminal on your computer.
4. Navigate to the PTRAIL directory.
5. Install all the requirements using:
    a. Pip install -r requirements.txt
6. Run the GUI.py file with the following command:
   python3 dashboard.py

That's it, you will see the the PTRAIL Dashboard launch with the screen as seen below:

Fig. 2: PTRAIL Dashboard Launch View

In this help text, we will use an example of working on the Starkey dataset using PTRAIL Dashboard. The Starkey Forest Data is a wildlife dataset collected from the Starkey Experimental Forest and Range in northeastern Oregon, USA. It includes GPS locations and other information for several species, such as elk, deer, and wolves. The dataset was used in a study that focused on the use of spatial databases for managing wildlife data (Anthony et al., 2003).

As you can see above, there is nothing the moment the dashboard launches. To unleash the superhero that the PTRAIL Dashboard is, we first need to load a dataset into the dashboard. However, it is to be noted that PTRAIL currently supports only csv (comma separated values) files only as input. PTRAIL currently provides the Starkey dataset as one of its base datasets, and hence it can be found in the examples/data directory in the PTRAIL directory that you earlier extracted on your local machine.Once the file is loaded, we can start to see the full potential of the PTRAIL Dashboard. Once the file is loaded, it appears as we can see in the screenshot below where one trajectory is displayed at a time, along with the data frame that is supplying the data for all the operations right below it. Look at the

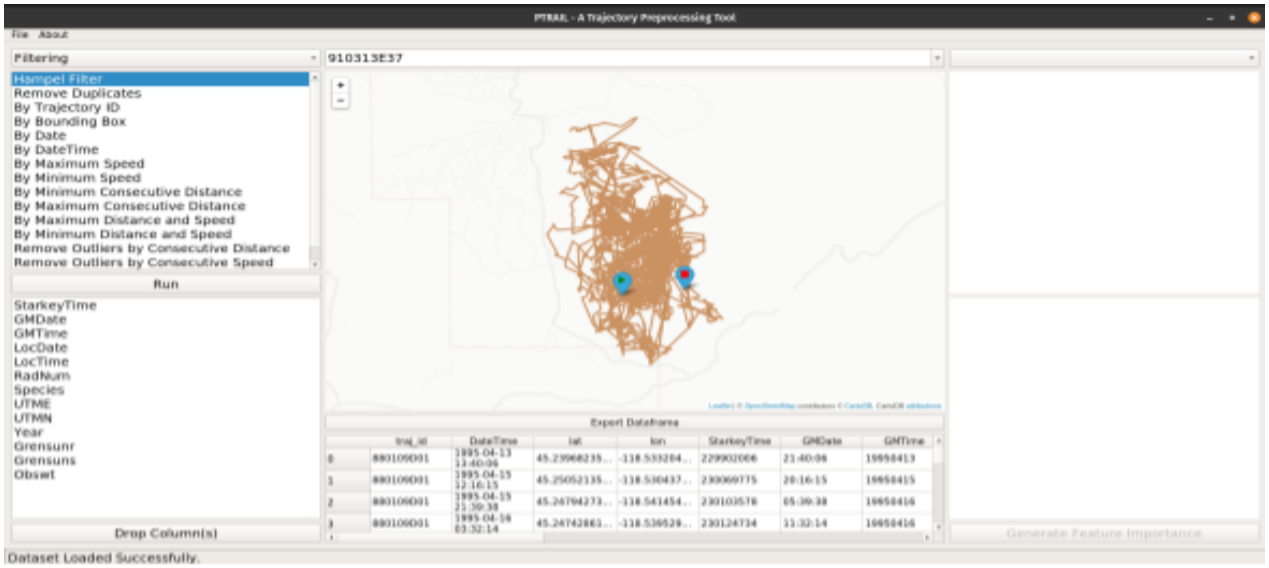

Fig. 3: Dashboard after Loading Starkey Data

screenshot below for reference.

3. Dashboard Components

As we earlier mentioned, the PTRAIL dashboard is mainly into 5 coherent, yet closely related components. Each component either serves as a command palette to modify or enhance the data or as a panel that gives the visualization of the current state of the loaded data. Without further ado, let's dive into each component in detail while continuing our tutorial on Starkey data.

3.1 Command Palette

The command palette is contained entirely within the left side of the window as we can see in Fig. 3. The top half of the palette is considerably the place where the most amount of interaction takes place. The top half of the palette provides the functionalities for filtering, feature extraction and data enhancement using interpolation. To switch between filtering, interpolation, and feature extraction, one can open the dropdown box and switch between the various modules. Upon selecting each module, the list of functions available in each module are displayed in the box below the drop down box. The magic of the Dashboard is that these functions are available at the click of a button, once the desired function is selected and the Run button is pressed, the PTRAIL Dashboard immediately processes the data and updates all the relevant components in the dashboard.

Let's look at the functions that each module provides. Further documentation and specifications of each function can be found in the PTRAIL documentation. Each relevant module is linked to its relevant documentation.

1. Filtering
    a. The filtering module is mainly used to remove the noise and irregularities in the data. As previously mentioned, mobility data is prone to irregularities due to a number of reasons. Hence, this module provides functionalities that can be used to reduce those irregularities and smoothen the data. The filtering module provides the following function.

i. Hampel Filter
ii. Duplicate Removal
iii. Filtering by metrics such as Bounding box, Date, Date and time.
iv. Filtering based on kinematic features such as speed, distance or a combination of both.
v. Removing outliers from the data based on distance or speed.

2. Interpolation
    a. Interpolation is the process of filling large gaps in data points by using various techniques such as averaging, curve fitting etc. In the context of trajectory data, it is especially important to use interpolation in order to fill gaps if the data has large jumps between two distinct observations. This helps prevent abrupt jumps in the path of the trajectory and provides a smoother flow of data. PTRAIL is the first package in the community that provides *four* different kinds of interpolation techniques to enhance trajectory data.
        i. Linear Interpolation
        ii. Cubic Interpolation
        iii. Kinematic Interpolation
        iv. Random-walk Interpolation

3. Kinematic Feature Generation
    a. Kinematic features, derived from positional observations in trajectory data, are valuable tools for enhancing the predictive power of machine learning algorithms. By extracting features such as speed, acceleration, direction, and distance, researchers can uncover hidden patterns and relationships that may not be immediately visible in raw data. Kinematic features can greatly increase the number of variables available for analysis, providing a more comprehensive view of movement patterns. These features unlock the full potential of trajectory data, enabling deeper insights and more accurate predictions. The PTRAIL Dashboard offers the following kinematic features:

i. Distance between consecutive points
ii. Distance from the start of the trajectory
iii. Distance from a specific coordinate point
iv. Speed
v. Acceleration
vi. Jerk
vii. Bearing
viii. Bearing Rate
ix. Rate of Bearing Rate
x. Generation All features above

4. Temporal Feature Generation
    a. Temporal features are essential for understanding the time-based behavior of individuals and groups in mobility data. These features, which are calculated from time-stamps present in the trajectory data, can include dwell time, start time, stop time, and more. By extracting these temporal features, researchers can generate important derived features that enhance the predictive power of machine learning algorithms. The ability to analyze temporal patterns and relationships is crucial for a wide range of applications, including transportation planning and resource allocation. The PTRAIL Dashboard provides the following temporal features:
        i. Date Extraction
        ii. Time Extraction
        iii. Day of the Week Extraction
        iv. Time of Day Extraction
        v. Weekend Indicator
        vi. All the features above!

5. Statistical Features
    a. The statistics module provided by PTRAIL Dashboard mainly deals with converting the trajectory data into segment-based form. In trajectory data analysis, it is often preferred to work with segment-based form of data to reduce the effect of noise that

point-based form representation of the data can contain. Therefore, the PTRAIL dashboard provides the following statistical features:

i. Generating statistic based on kinematic features
ii. Segmenting trajectory based on time interval
iii. Converting the data to segment based representation

6. Column Dropping Panel
7. The column dropping panel is located right below the run button on the command palette. Trajectory data can often contain features that are not very important and bear little weight in classification tasks. Therefore, it is often desired to remove those features. To aid this process, the PTRAIL Dashboard provides the column dropping panel which allows removing columns by selecting one at a time or selecting multiple columns at a time. It is to be noted that as a requirement in PTRAIL, the following columns "cannot" be deleted:

    i. Trajectory ID
    ii. Timestamp
    iii. Latitude
    iv. Longitude

8. Usage Example

With the PTRAIL dashboard, generating kinematic features and performing cubic interpolation on trajectory data has never been easier. In fact, this process can be completed with just a few clicks of a button.

First, to generate all available kinematic features, users simply need to click on the drop-down menu at the top of the command palette, switch to kinematic features, select the "All Kinematic Features" option, and click the run button. The dashboard then automatically generates all the kinematic features, which can greatly enhance the predictive power of machine learning algorithms for classification tasks.

Next, smoothening the data using cubic interpolation is just as simple. Users again click on the drop-down menu, toggle to the interpolation module, select cubic interpolation, and hit the run button to interpolate the trajectory data. This step can help to reduce noise and improve the accuracy of classification results.

Overall, PTRAIL's user-friendly interface makes generating kinematic features and performing cubic interpolation on trajectory data a straightforward process. Researchers no longer need to spend countless hours manually extracting features or writing complex code – with PTRAIL, these tasks can be accomplished with just a few clicks of a button, freeing up time for more important aspects of research.

3.2 Map Viewer

The Map Viewer panel is an essential tool for visualizing the trajectory of an object, from start to end, making it easier to detect sudden changes and jumps in its movement. With the PTRAIL dashboard, users can easily visualize the pattern of an object's movement and identify the need for further smoothening using one of the interpolation methods. To save time and machine load, the dashboard only plots one trajectory at a time on the viewer panel, and users can cycle through the different trajectories with ease by using the searchable dropdown menu at the top. This feature eliminates the need to scroll through the entire list to find a specific trajectory ID, thus making the data analysis process more efficient and user-friendly.

Continuing our tutorial on the Starkey data, viewing the trajectory in the screenshot below clearly highlights the jumps in the trajectory at various points. Without the map viewer panel, it is impossible to discern that pattern visually which demonstrates how useful the dashboard can be!

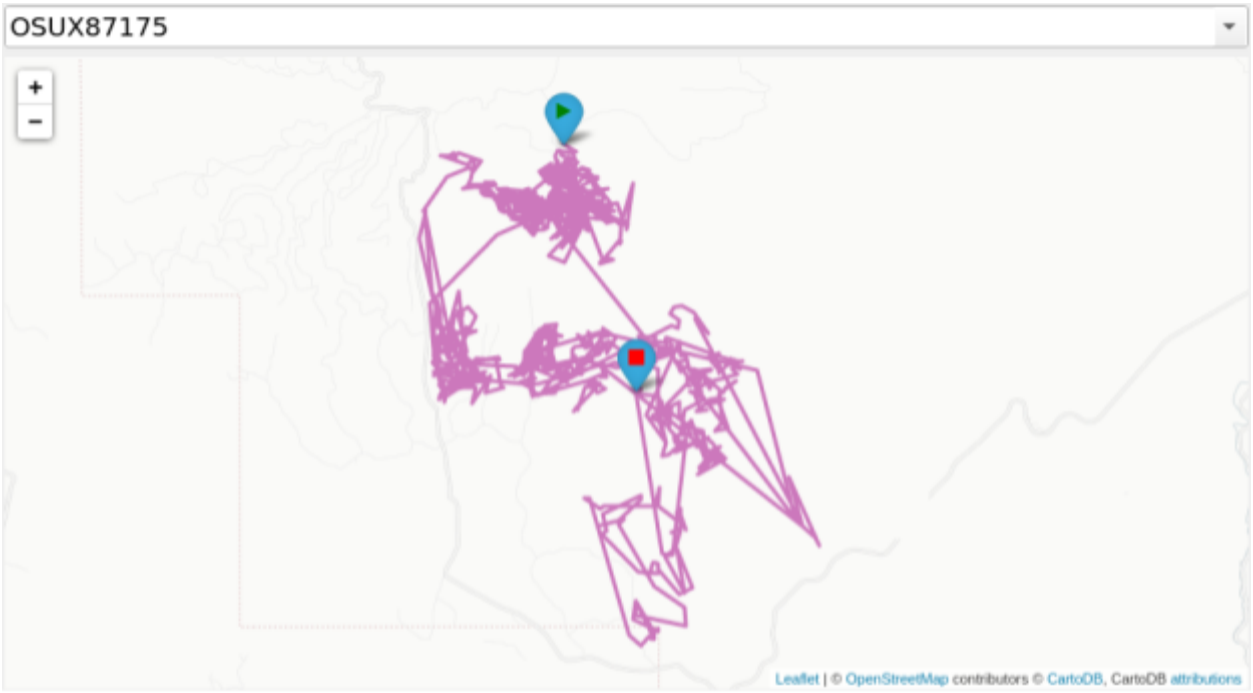

Fig. 4: Map Viewer Panel

3.3 Dataframe Viewer

The PTRAIL Dashboard provides users with powerful features to work with their trajectory data, but it's equally important to have a clear view of the raw data before performing any action on it. The Dataframe Viewer panel serves this purpose by displaying and updating the data in its original form, giving users a chance to inspect and verify the data quality. This is particularly useful for users who want to ensure that their data is ready for further analysis or machine learning tasks.In addition, the Dataframe Viewer panel also includes an "Export Dataframe" button, allowing users to quickly export their data to a CSV file with just one click. This feature saves users from the time-consuming task of manually copying and pasting data. To demonstrate this feature, let's take our Starkey example. After generating kinematic features and performing cubic interpolation, we can use the Dataframe Viewer to inspect the data and ensure it's ready for our analysis. Once we're satisfied, we can simply click on the "Export Dataframe" button and obtain a CSV file of our processed data without any hassle. Overall, the Dataframe Viewer

panel provides convenience and efficiency to the users, allowing them to focus more on their analysis and less on the manual work.

3.4 Comparison Graph Viewer

The PTRAIL Dashboard provides a visual comparison graph chart for each trajectory that is currently being viewed on the Map Viewer panel. The Map Viewer and the Comparison Graph Viewer are closely linked, as they allow users to analyze kinematic features of trajectories in a graphical format. Initially, the Comparison Graph Viewer is empty and does not display any data until a kinematic feature is generated on the trajectory data.

The Graph Viewer Panel serves the purpose of comparing kinematic features of the currently selected data with the entire dataframe. This helps users identify outliers by visually examining charts instead of manually interpreting numerical values. For example, the chart may display the distance traveled by the currently selected subject (active in the Map Viewer) compared to the rest of the data. The horizontal dashed lines represent various statistical measures for the selected metric, such as the mean, median, 10th percentile, 25th percentile, 75th percentile, and 90th percentile of the data.

Additionally, users can compare all available kinematic metrics for a trajectory by toggling through the drop-down menu in the Comparison Graph Viewer. This feature allows for comprehensive analysis of different kinematic features for the trajectory being examined. The Comparison Graph Viewer provides a powerful tool for users to visually assess and compare kinematic features of trajectories, aiding in data interpretation and identification of patterns or outliers. Overall, the Comparison Graph Viewer in PTRAIL Dashboard enhances data analysis and decision-making by providing a visual and interactive way to compare kinematic features of trajectories. The combination of Map Viewer and Comparison Graph Viewer enables users to gain insights from trajectory data in a more intuitive and efficient manner.

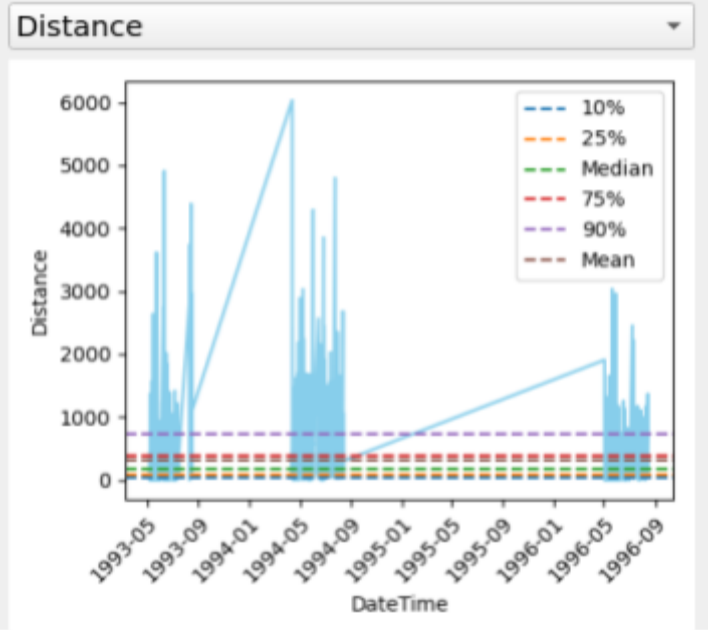

Fig. 5: Comparison Graph

When we load the data and generate kinematic features on the Starkey dataset in PTRAIL Dashboard, the Comparison Graph panel automatically populates with a graph, as shown in Fig. 5. This graph allows us to compare the distance traveled by the currently selected trajectory with the entire dataset.

Upon examining the graph, we can infer that there are two instances in the data where there is a noticeable gap in data collection. This suggests that there may be missing data points in those instances, and we may need to interpolate the data to smoothen out the trajectory. Furthermore, we can also observe that the distance traveled by the current subject is significantly higher than the average and midpoint of the dataset. This may indicate that this trajectory is an outlier compared to the rest of the data. As we conduct our experiments or analysis, we may need to consider removing this trajectory from our analysis to prevent it from skewing our results.

3.5 Feature Importance Viewer

The Feature Importance Viewer panel in PTRAIL Dashboard is similar to the Comparison Graph Viewer in that it is initially empty when data is loaded. However, once kinematic or temporal features are generated, the "Generate Feature Importance" button becomes active, allowing users to generate a bar chart that indicates the most important features in the dataset for classification tasks. This feature helps researchers identify which features carry more weight in the data and which features can potentially be trimmed.

Trimming features can significantly reduce the time required to train a machine learning model, as the model only needs to work on a reduced subset of the most important features. However, it's important to note that the feature importance bar chart is only applicable for datasets that are suitable for classification tasks, where there is a feature to be predicted by the machine learning model. This predicted feature serves as the Y variable in the machine learning model.

The feature importance calculation is done through scikit-learn integration in PTRAIL, and researchers can refer to the provided link for more information on feature importance. The Feature Importance Viewer panel provides valuable insights into feature importance, helping researchers optimize their machine learning models and improve classification accuracy. Overall, it serves as a useful tool for feature selection in classification tasks within PTRAIL Dashboard.

After generating kinematic and temporal features on the Starkey data, we can use machine learning algorithms like RandomForestClassifier to predict the species of the animal for each trajectory. However, due to the large number of features in the data, the training and testing of our model may become slow. To address this issue, we can utilize the feature importance chart in PTRAIL dashboard to identify the most important features and drop some of the less important ones using the column dropper panel.

To generate the feature importance chart, we need to click the "Generate Feature Importance" button in the dashboard. Once clicked, the dashboard will

prompt us to enter the name of the feature that we want to predict, which in this case is "Species", as it appears in the dataset. We can also specify how many features we want to compare, but it's important to note that the more features we compare, the longer it may take for the graph data to be computed and displayed.

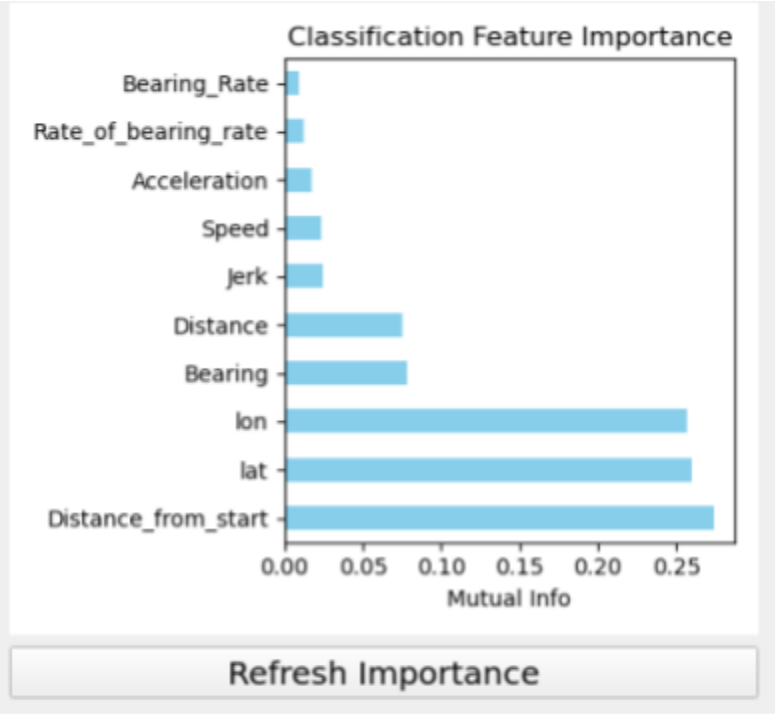

Fig. 6: Feature Importance

As we can clearly see in the figure above, Distance_from_start metric carries the most amount of weight in classification and Bearing Rate carries the least. From this chart, we can identify that we can possibly cut down some of the metrics such as Bearing Rate, Rate of Bearing Rate and Acceleration which can help us in optimizing our model.

4. Endnotes

As we come to the end of this user manual for PTRAIL dashboard, we hope you have found it helpful in understanding the features and functionalities of this powerful tool for analyzing trajectory data. From visualizing trajectories on the Map Viewer panel to comparing kinematic features on the Comparison Graph Viewer, and generating feature importance charts, PTRAIL dashboard provides researchers with valuable insights and streamlines their data preprocessing tasks.

However, as with any software in the universe, PTRAIL Dashboard may have a few quirks here and there. But hey, don't fret! We're here to help, even if our developers are fueled by caffeine and pizza. If you need assistance with setup or usage, drop us an email at yjharanwala@mun.ca. And if you spot a bug, head over to PTRAIL's GitHub repository and raise an issue. We'll be on it faster than you can say "trajectory data preprocessing is a mouthful!" So, let's tackle those data hurdles together, one quirky bug at a time! Happy PTRAILing! 🚀🐾

Bibliography

1. Haranwala, Y. J., Haidri, S., Tricco, T. S., da Fonseca, V. P., & Soares, A. (2022). *A dashboard tool for mobility data mining preprocessing tasks*. Retrieved April 1, 2023, from https://ieeexplore.ieee.org/document/9861122/
2. Haidri, S., Haranwala, Y. J., Bogorny, V., Renso, C., da Fonseca, V. P., & Soares, A. (2022, August 6). *PTRAIL - a python package for parallel trajectory data preprocessing*. SoftwareX. Retrieved March 31, 2023, from https://www.sciencedirect.com/science/article/pii/S2352711022001066

3. Starkey, E. E., Anthony, R. G., & McDonald, T. L. (2003). Spatial databases for managing wildlife data: a case study from Oregon. Journal of Wildlife Management, 67(2), 344-358. https://doi.org/10.2307/3802693