Final Results: Nefarious Cell Phone Proximity Alerts

Agenda

- Problem Statement
- Data Set
- Scenario
- AI Model
- Results
  - Final Dataset
  - Model Accuracy
- Demo
- Conclusions
- Future Work

Problem Statement

A theoretical company, “Vormund”, has advertised an application with the capability to use device location data in order to flag devices as suspicious or nefarious. Our team was tasked with developing a method to test this theoretical application.

Application Claims

- Use only Geospatial Data
- Ability to identify a suspicious actor
- Notify User if a suspicious actor is within a given distance

Project Tasks

- Acquire or establish a geospatial dataset
- Define “nefarious actor”
- Scan through data points for actors in range
- Generate machine learning model(s) to classify an actor as suspicious

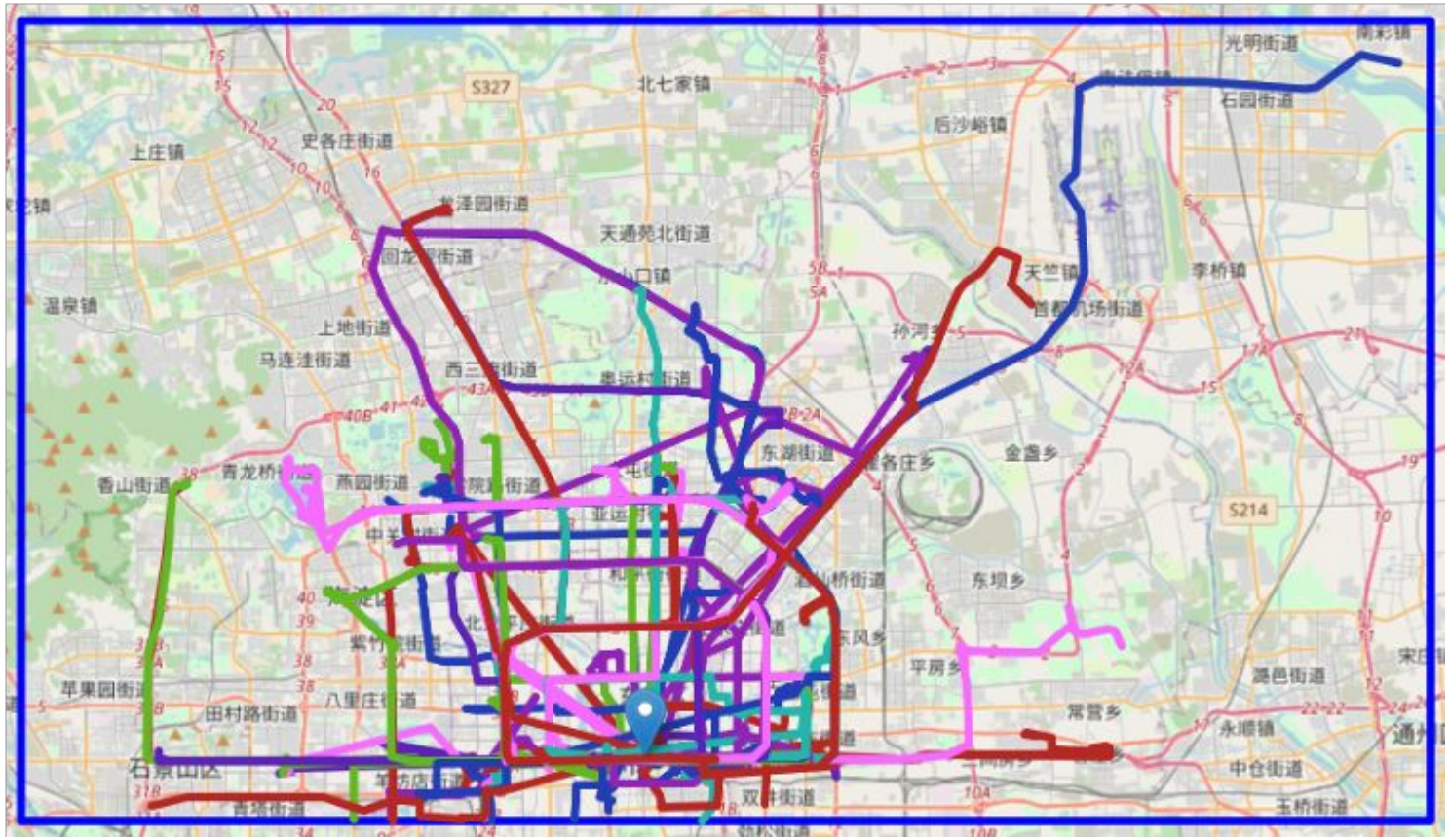
Data Set

- Used the Geolife Dataset
  - Collected by Microsoft Research
  - 182 devices' GPS data in Beijing
  - April 2007 to August 2012
- Data comprised of latitude, longitude, time, altitude, and device ID.
  - Each lat-long coordinate is recorded on intervals ~1-5 seconds
- The participants are presumably students
  - a lot of the actors are adjacent to universities.

Heatmap of the Geolife GPS Data

Data Processing

- Geolife data was filtered to only include devices within the perimeter of Beijing
  - Movement patterns are very different inside vs. outside the city
- This included the removal of devices that would travel to other cities
  - via plane, boat, or other means
- More consistent data, better modeling



Data Processing

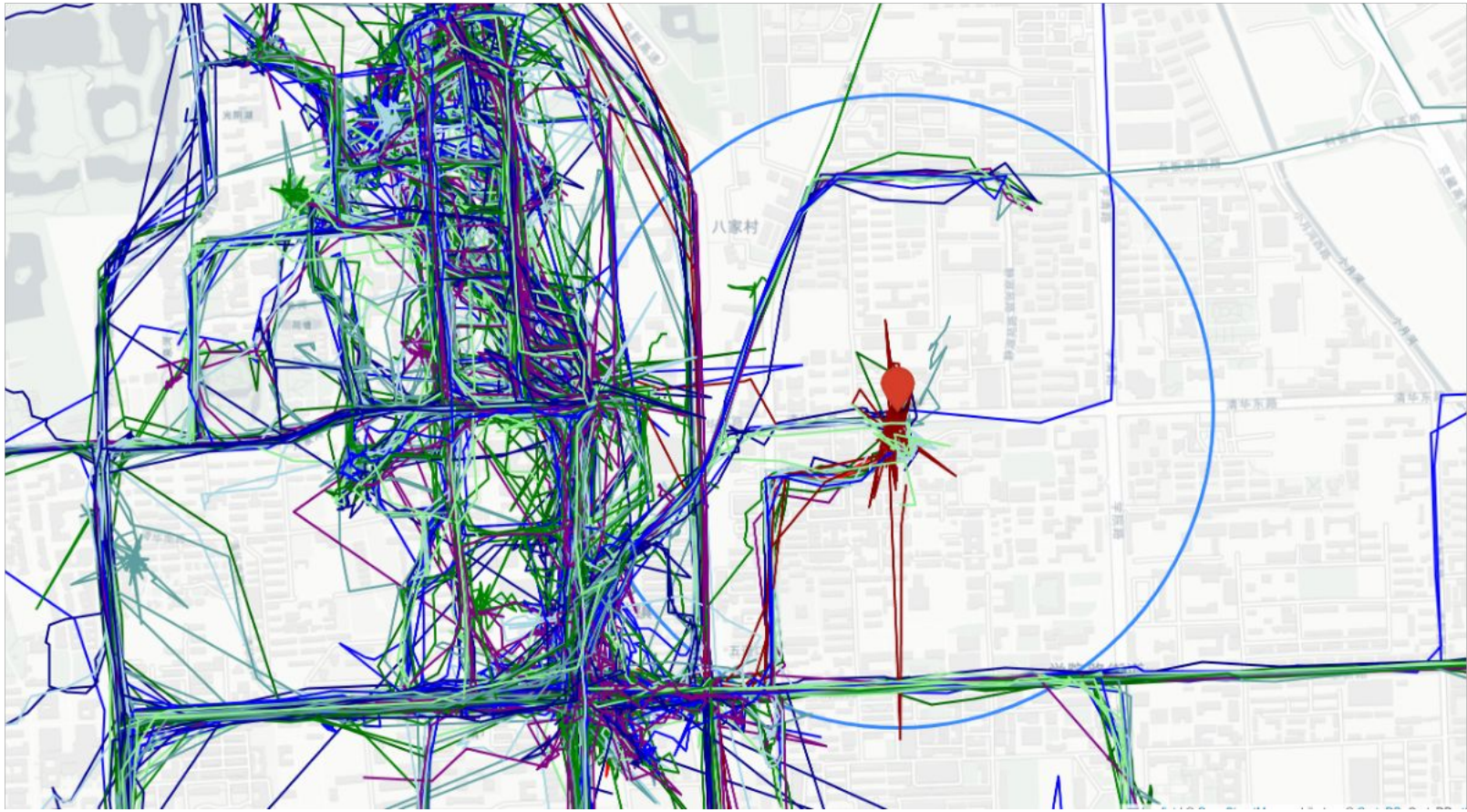
- The mode of transportation and duration of travel can vary for participants over their journeys.

- We broke the filtered dataset up by days, meaning that path data that extended over multiple days became separate paths in the final dataset.

- This method aimed to split up those paths into several shorter paths while preserving the overall pattern and distance covered by an actor

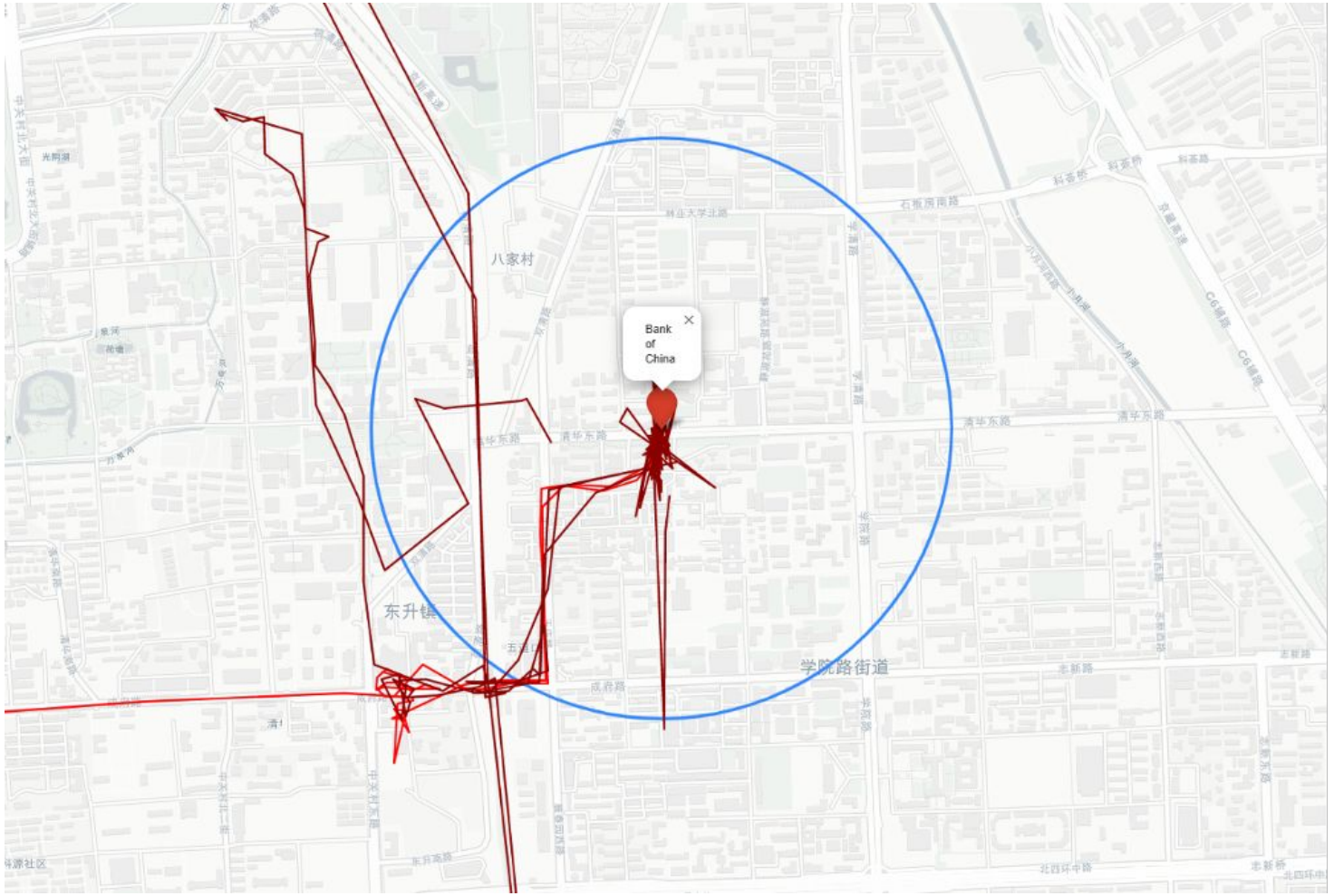
Scenario

- We set this problem up by defining a scenario and a point of interest.
  - Nefarious actions are heavily context dependent
- Additionally reduces scope of the problem
  - Devices only need to be analyzed upon entering radius
- We selected the Bank of China (shown on the right) as our POI



Scenario

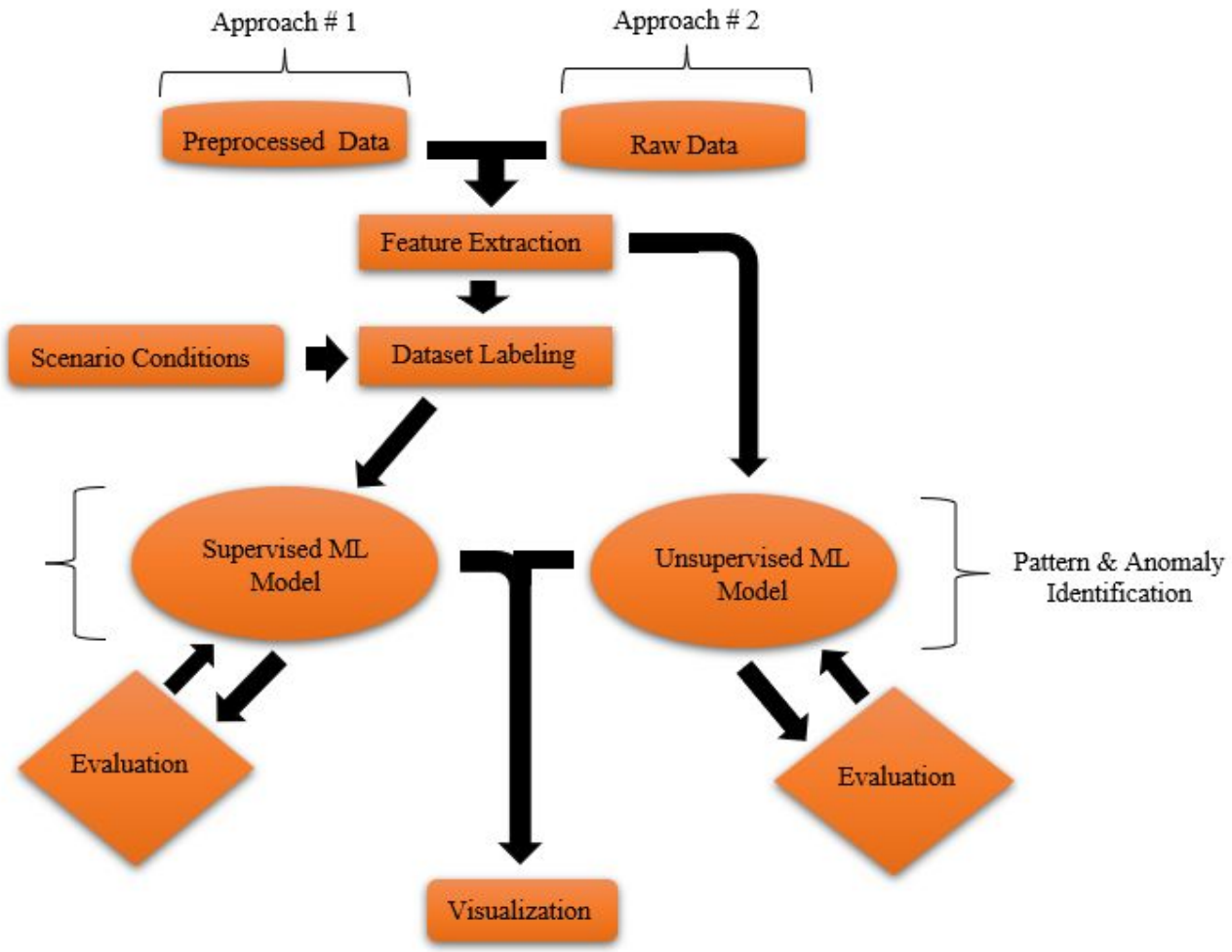
- Bank of China was selected as POI because
  - A lot of everyday visitors
  - Potential target for malicious activity
  - A lot of nearby device data in dataset
- Additionally, a public building such as this has defined no access areas
  - Can develop additional scenario specific features to help flag actors.



AI Model

- We have had two approaches
  the main difference is the dataset
  - Approach #1 uses a preprocessed dataset
  - Approach #2 uses a raw dataset

- Both approaches share the same algorithm
  - feature extraction
  - Scenario
  - data labeling
  - pattern & anomaly Identification



Results : Final Dataset

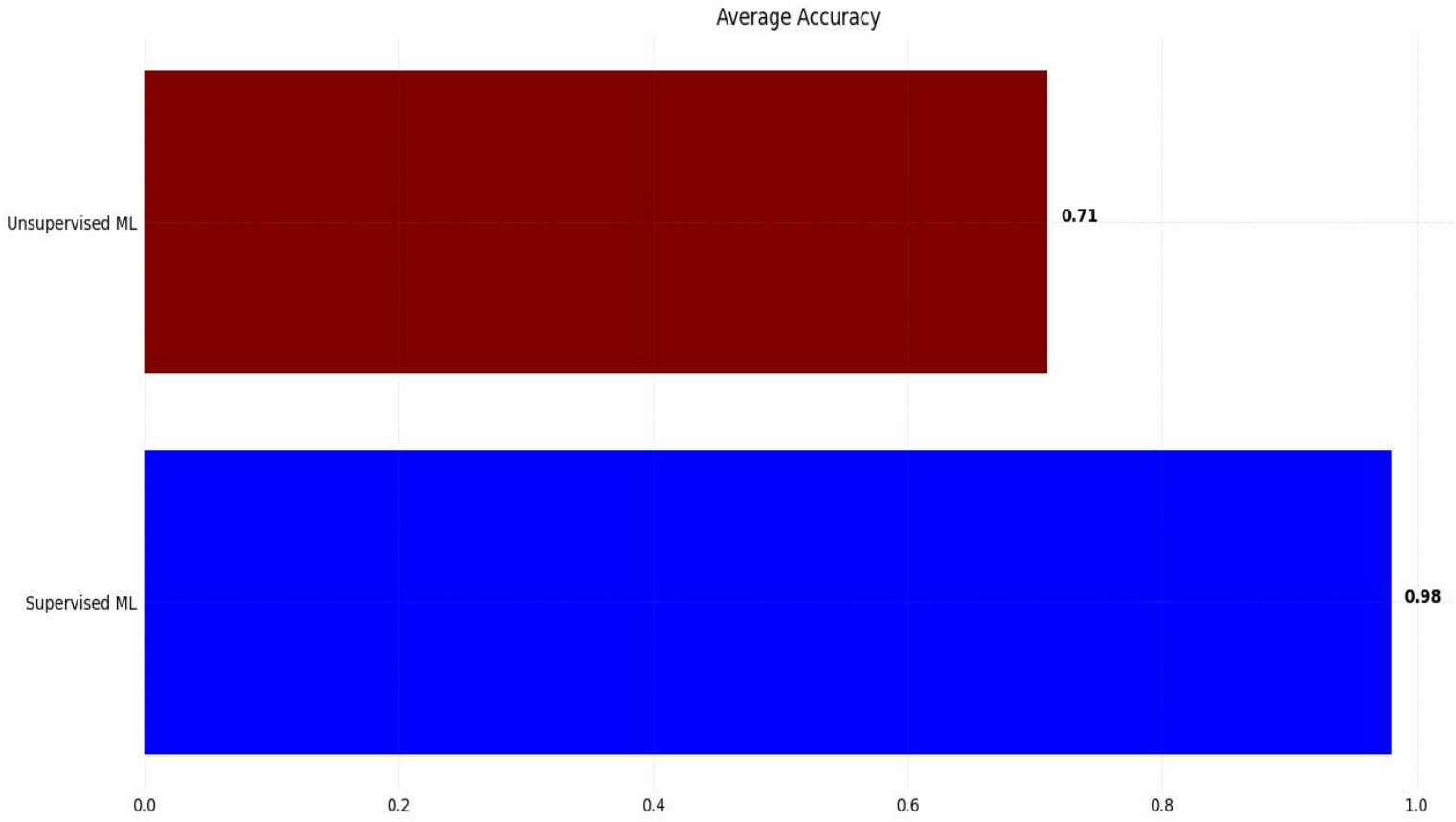
```
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   index               70 non-null     int64
 1   Latitude            70 non-null     float64
 2   Longitude           70 non-null     float64
 3   Altitude(Ft)        70 non-null     float64
 4   DaysSince12301899   70 non-null     float64
 5   DateString          70 non-null     object
 6   TimeString          70 non-null     object
 7   id                  70 non-null     int64
dtypes: float64(4), int64(2), object(2)
```

Initial dataset

Final dataset

```
Data columns (total 21 columns):
 #   Column                                          Non-Null Count  Dtype
---  ------                                          --------------  -----
 0   Id                                              12000 non-null  int64
 1   Latitude(degree)                                12000 non-null  float64
 2   Longitude(degree)                               12000 non-null  float64
 3   Altitude(ft)                                    12000 non-null  float64
 4   Year                                            12000 non-null  int64
 5   Month                                           12000 non-null  int64
 6   Day                                             12000 non-null  int64
 7   Hour                                            12000 non-null  int64
 8   Second                                          12000 non-null  int64
 9   Distance_between_2_points_in_path(m)            12000 non-null  float64
 10  Durations_between_2_points_in_path(s)           12000 non-null  int64
 11  Speed_at_each_points_in_path(m/s)               12000 non-null  float64
 12  Directions                                      12000 non-null  int64
 13  IsCircularPath                                  12000 non-null  int64
 14  NumberOfLoops                                   12000 non-null  int64
 15  Distance_between_POI_and_P_in_path(m)           12000 non-null  float64
 16  Forward_azimuth_between_POI_and_P_in_path(degree)  12000 non-null  float64
 17  Back_azimuth_between_POI_and_P_in_path(degree)  12000 non-null  float64
 18  Forward_azimuth_between_2_points_in_path(degree)  12000 non-null  float64
 19  Fack_azimuth_between_2_points_in_path(degree)   12000 non-null  float64
 20  Threat_level                                    12000 non-null  int32
dtypes: float64(10), int32(1), int64(10)
```

Results: Model Accuracy



Unsupervised Learning

Supervised Learning

Conclusions

- Predicting nefarious actors using geospatial data would be highly beneficial as an application, as supported by our model and research.
- Our results are limited only by the quantity and quality of our data
  - Ground truth (nefarious or not) is unknown, so assumptions must be made
  - If the data were labeled, then we could better assess the validity of such application
  - Additional data could also help fine tune results

Future Work

- ➢ Extract more features
- ➢ Enhance dataset labeling algorithm
- ➢ Integrate other scenarios to the AI model
- ➢ Develop a Graphical User Interface (GUI)

Thank you

Demo

Switching to PyCharm IDE

Speakers

Bradley Tunks

MEng Computer Science
Virginia Tech
Blacksburg, VA
bradt@vt.edu

Mouad Ait Taleb Ali

MEng Computer Science
Virginia Tech
Blacksburg, VA
mouad@vt.edu