Unveiling Insights: Navigating VCIS Analytics Roadmap for Targeting, Classification and Forecasting

In today's data-driven world, organizations heavily rely on analytics to gain valuable insights and make informed decisions. This necessitates the use of sophisticated tools and methodologies to effectively navigate through vast amounts of data. One such roadmap that aids businesses in leveraging analytics for targeting and classification is the VCIS Analytics Roadmap.

At VALOORES, we believe that Analytics is not simply a service to sell, but rather a culture to build. That's why actionable data production lies at the heart of our multi-service homegrown products, which offer comprehensive coverage of the pain-gain business life cycle.

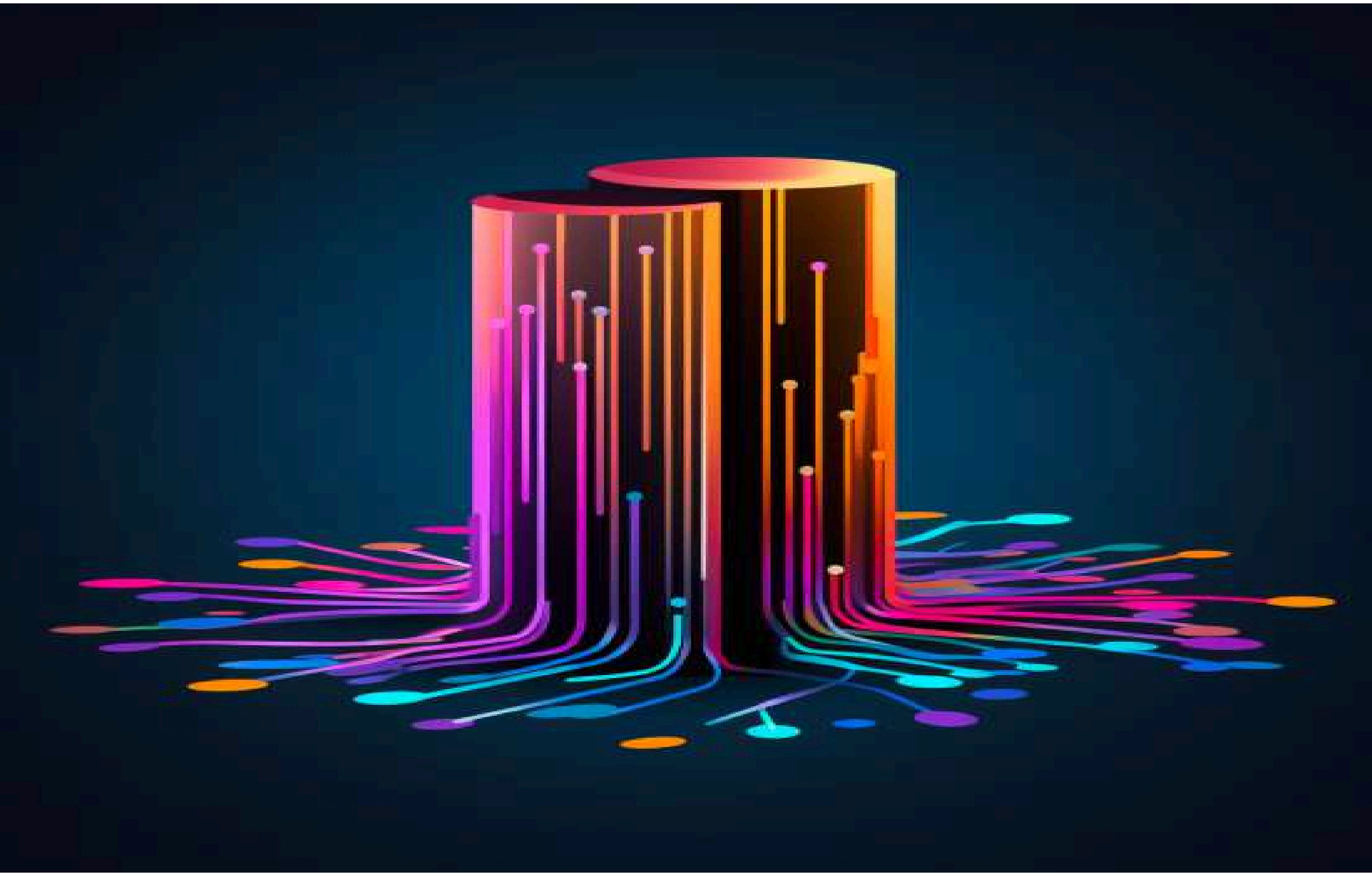

Table of Contents

Introduction

Analytics involves using data to make informed decisions and improve outcomes. By analyzing, we can gain valuable insights that can help us make better decisions, improve the processes, and drive better results. VCIS offers an advanced analytics system that provides investigators with an edge in crime investigation. Our AI and machine learning algorithms are being used to enhance the capabilities of VCIS by enabling it to analyze vast amounts of data quickly and accurately. By leveraging these technologies, geospatial data intelligence systems can identify correlations and anomalies that may not be immediately apparent to human analysts. Additionally, machine learning algorithms can be trained to recognize specific patterns and trends, allowing the system to identify potential risks or opportunities more efficiently.

Furthermore, AI and machine learning can help automate certain aspects of the analysis process, allowing analysts to focus on higher-level decision-making tasks. As a result, AI and machine learning are playing an increasingly important role in geospatial data investigation, providing analysts with the tools they need to make more informed decisions and uncover insights that would be impossible to identify through manual analysis alone.

Chapter 1: Our Analytics Roadmap

In what follows, we give a brief description about the process under which VCIS undergoes at the level of analytics. The system is designed to meet their expectations and is based on three key categories:

A. Descriptive Analytics

The important ingredient in descriptive analytics is grouping and classification. VALOORES has implemented diverse grouping functionalities in its in’DataCrowd application, enabling the creation of groups from multiple sources. Grouping entails the act of organizing people or objects into distinct groups. The criteria for grouping are derived from various factors such as business intelligence, simulation, rules, activities, and characteristics. While algorithms and engines initiate the process of grouping people or objects, there will always be a requirement for human intervention and manual updates to refine and reorganize the groups as new insights and information emerge. Groups are organized based on date and characteristics, and devices can be added to or removed from multiple groups as new data becomes available daily. Groups can range in size from massive, encompassing millions of devices, to being restricted to just two devices, depending on the availability of data. Each group is managed by a specific algorithm that determines which devices belong to it, enabling a more targeted approach to data analysis and management.

a) Group Creation

There are several ways in which groups can be added to our system:

1. The end-user manually creates a group by selecting a group name and group status and then adding devices. Each group has its own set of characteristics that are linked to the device types in it.
2. Create a group based on a rule or data simulation. For example, the system can retrieve hits and devices based on a set of criteria such as longitude and latitude, and automatically create a group.
3. Create groups through BI classification, which involves categorizing groups as either "Normal" or "Massive" based on their size and complexity.
4. Groups can also be created based on hits limited by a selected area on the map. This means that a group can be formed by selecting a specific geographic area and adding all devices that fall within that area.
5. Groups can be created based on segmentation, which refers to the previously configured settings that divide the devices into specific categories based on their characteristics.

b) Types of Groups

Different types of groups can be created; devices and users can be classified in different ways for different purposes.

1. Descriptive grouping: Users of devices can be grouped as employees, family members, suspicious, regular visitors of certain and specific places, co-travelers, etc...
2. Hotspots: One can create groups of hits from devices that occur in geographic "hotspots" where security incidents are most likely to occur. This could be based on historical data or on current trends in security threats.

3. Restricted zones: One can create groups of hits from devices that occur in restricted zones, such as areas where sensitive data is stored or where critical infrastructure is located.
4. Proximity-based groups: One can create groups of hits from devices that occur within a certain distance of specific locations, such as a company's headquarters, a government building, or a high-value target.
5. Traffic analysis: One can create groups of hits from devices based on traffic analysis, such as identifying devices that are generating a large amount of hits in a certain area or connecting with suspicious devices.
6. Mobile devices: One can create groups of hits from mobile devices that are frequently on the move, such as company vehicles or employees who work remotely. This can help you identify potential security risks associated with mobile devices.
7. Monitoring borders: One can create groups of hits from devices that occur near national borders, ports, or airports, where there is a higher risk of security threats such as smuggling or trafficking.
8. Emergency response: One can create groups of hits from devices that occur in areas where emergency response teams may need to be dispatched quickly, such as areas prone to natural disasters or where public safety incidents are more likely to occur.

c) Places of Interest (POI)

There are different kinds of places of interest. Places of interest are determined based on requests of users. It is also determined as a crucial part of the analytics by obtaining the areas of interest (AOI) and the restricted areas. Places of interest feed groups and classification and increase the input for enhancing classifications. There is a basic category for AOIs in our system and additional AOIs can be added based on the user's need.

1. AOI: The basic AOIs defined in our system for each device are home location and work location. These two AOIs constitute vital pillars in any investigation procedure.
2. Restricted Areas: Users of the system can choose certain restricted areas where they can study the traffic of devices or mark devices as suspicious.

d) Classification of AOI

1. Classification by Time period to determine the pattern. Study of hits per device over days of week for every cluster to determine:
   - Nighttime activities: presence of hits over night time 12am - 7am according to the last hit at night and first hit in the morning in the clusters.
   - Work Activity: 8am - 5pm shift presence in the clusters
   - All day activity: Studying the movement pattern of the

device between the AOIs over all day 12am > 12pm

2. Category/Type Classification
    - Reverse geocoding of categories and types present in each AOI: categories' type will help us determine the AOI classification and approve the pattern of device inside it.
    - Comparison of new types with our already classified dataset to determine its classification (according to meaning and then by characters).

e) Classification of Device in AOI

1. Comparing opening-closing hours of each amenity/ other category to hours spent by the device in each classified AOI (check that the flow of hours of the device and AOI is consistent, example: not present after closing hour or on closing days.)
2. Comparing total number of hours spent by ID in each cluster weekly with the known number of working hours in each job according to the type of AOI.

B. Predictive Analysis

1. Time series analysis: We use this type to analyze geospatial hits over time to identify patterns and trends that can be used to make predictions about future activity in certain areas.
2. Regression analysis: We use regression analysis to identify correlations between geospatial hits and other variables, such as time of day, density-based allocations, etc.... One can use these correlations to make predictions about future activities in certain areas based on changes in these variables.
3. Machine learning: We use machine learning algorithms to identify patterns and trends in geospatial data and make predictions about future activity.
4. Spatial analysis: We use spatial analysis techniques, such as buffer analysis and proximity analysis, to identify spatial relationships between geospatial hits and other variables. We use these relationships to make predictions about future activity in certain areas.
5. Anomaly detection: We use anomaly detection techniques to identify unusual or unexpected geospatial hits. These anomalies can be used to make predictions about future activity, such as the likelihood of a crime occurring in a certain area based on unusual activity.

Chapter 2: Targeting Areas of Interest

In order to find the Areas of Interest for a device, we must identify where the device hits the cluster. We describe the different clustering methods we are using.

A. Clustering Methods

a) K-Means

K-means is a widely used unsupervised machine learning algorithm for cluster analysis. The basic idea of the K-means algorithm is to divide a set of n data points into k clusters, where k is a user-defined number, in such a way that the sum of the squared distances between the data points and their assigned cluster centroid (mean) is minimized. The algorithm works iteratively by first randomly initializing the centroids, then assigning each data point to the nearest centroid, followed by re-computing the centroids based on the newly assigned cluster points. This process continues until the cluster assignments no longer change.

b) DBSCAN

DBSCAN or Density-Based Spatial Clustering of Applications with Noise is a density-based clustering algorithm that groups together points that are closely packed together, while labeling points that are isolated or far away as noise. The algorithm starts by defining a neighborhood around each data point and grouping together points that are close to each other based on a defined distance metric. Points that are not part of any cluster are considered as noise. This algorithm is useful in finding clusters of arbitrary shapes in a data set and is particularly useful when clusters are dense and well separated from each other.This algorithm has two main parameters; Epsilon & Minimum Points:

- **Epsilon (ε)**
  This parameter defines the radius within which to search for neighboring points. Points that are within the ε-distance from each other are considered to be part of the same cluster. Because we are dealing with geolocation data, we use the ‘haversine distance’ to compute the distance between two coordinates.

- **Minimum Points (MinPts)**
  This parameter defines the minimum number of points required to form a dense region. If a cluster contains fewer points than the MinPts threshold, then it is considered as noise.

B. Targeting Approach

We use the DBSCAN Clustering, which was an ideal fit for our application. By specifying the two main parameters; epsilon and minimum points and in order to find the number of samples (MinPts), we first compute the frequency of hits per minute, then we get the number of hits per hour. For epsilon = 0.3 km (radius of 300 meters found reasonable for a cluster), and MinPts as initialized, the algorithm attempts to find the clusters. In case it doesn't, the MinPts is reduced by 30% until it reaches the number 2. If, at this stage, no clusters are found, the epsilon value is increased by an increment of 0.1 km until it finds a cluster, or until epsilon reaches the value 1 km.

a) From Clusters to Areas of Interest

If the algorithm finds clusters and centroids, the outliers/noise are removed from the data. Then, we check to see if the number of clusters is ideal. That is accomplished by a function called 'filter points' which checks whether the centroids are close in proximity or not. It may reduce or increase the number of centroids until an ideal is reached.

After the ideal number of centroids is reached, each data point is assigned to an area of interest in the non-noisy data. Then, for each AOI, some information is recorded such as the Latitude and Longitude, number of hits attached to the AOI, number of days spent by the device at the AOI, the AOI's address as well as its classification f. The classification of AOIs still remains a work in progress.

b) Clustering with respect to period of time

Choosing the timeframe for which the clustering occurs can offer us more insights in the results. We distinguish between clustering the hits over the entire period of time, and clustering data points day by day.

- **Overall Period of Time**

By clustering the hits from the entire period of time, more importance is attributed to the hits' density. The AOIs resulting from this technique are the areas which have the greatest amount of hits.

	index	NAME_ID	LAT	TYPE	LNG	CNT	percentage	Adress	NBR_DAYS	coords
0	0	84a31c9d-57ef-472e-bff3-8621ac6205c1	33.874960	percentage	35.517689	734	63.221361	McDonald's, Alam, Badaro, Park, Mazraa, Beirut...	24	[[35.51791, 33.87473], [35.517721, 33.874788],...
1	1	84a31c9d-57ef-472e-bff3-8621ac6205c1	33.878762	percentage	35.577023	76	6.546081	Street 46, Fanar, Matn District, Mount Lebanon...	9	[[35.57714, 33.87881], [35.57712, 33.87881], [...
2	2	84a31c9d-57ef-472e-bff3-8621ac6205c1	33.885002	percentage	35.522811	78	6.718346	American University of Science and Technology,...	9	[[35.5229, 33.885166], [35.522983, 33.885246],...
3	3	84a31c9d-57ef-472e-bff3-8621ac6205c1	33.903949	percentage	35.585979	54	4.651163	Street 62, Hai er Roum, Jal El Dib, Matn Distr...	4	[[35.58594, 33.90395], [35.58596, 33.90396], [...
4	4	84a31c9d-57ef-472e-bff3-8621ac6205c1	33.906692	percentage	35.601081	219	18.863049	Mezher, Antelias, Matn District, Mount Lebanon...	12	[[35.6011, 33.90673], [35.6011, 33.90674], [35...

Table 1: AOI Summary Table resulting from clustering hits over the entire period of time

Figure 1: Map showing AOIs resulting from clustering hits over the entire period of time

- **Day by day**

By clustering the hits on a day-by-day basis, the focus is given to the most important clusters per day, and by aggregating the centroids from each day over the entire period of time, we can observe which centroids are most frequently visited, filtering the clusters which have a lot of hits but are only visited once or twice.

	index	NAME_ID	LAT	TYPE	LNG	CNT	percentage	Adress	NBR_DAYS	coords
0	0	84a31c9d-57ef-472e-bff3-8621ac6205c1	33.874804	percentage	35.517682	18	54.545455	Alam, Badaro, Park, Mazraa, Beirut Governorate...	14	[[35.5177155, 33.874779000000004], [35.5177025...
1	1	84a31c9d-57ef-472e-bff3-8621ac6205c1	33.878775	percentage	35.577047	6	18.181818	Street 46, Fanar, Matn District, Mount Lebanon...	5	[[35.57713, 33.87881], [35.57702820000001, 33....
2	2	84a31c9d-57ef-472e-bff3-8621ac6205c1	33.906694	percentage	35.601109	9	27.272727	Mezher, Antelias, Matn District, Mount Lebanon...	9	[[35.601106800000004, 33.9067364], [35.6011623...

Table 2: AOI Summary Table Resulting from day-by-day clustering

Figure 2: Map showing the AOIs resulting from day-by-day clustering

Figure 3: AOI (left) resulting from clustering overall hits does not appear in the Day-by-day Clustering (right)

Chapter 3: Suspicious Areas

The solution allows the product user to enter suspicious areas as polygon objects and assigns a percentage of suspiciousness for each area. Then, we find how suspicious each device's behavior is, based on the percentage of hits and the percentage of clusters which are in suspicious areas. Flagged area contribution of each device is calculated in accordance to the percentage of area specified and total suspicious percentage allows us to classify the risk level of each device.

INDEX_NUMBER	NAME_ID	ID	TYPE	NAME	CNT	Potential_house_coords	%_Suspicious	%_area	Flagged area contribution	Total Suspicious
0	84a31c9d-57ef-472e-bff3-8621ac6205c1	2214	Polygon	Lebanon	1161	['(33.885002243589746,35.52281130769231)', '(3...	100.000000	33.333333	33.333333	55.856446
1	84a31c9d-57ef-472e-bff3-8621ac6205c1	2916	Circle	Raqqa	219	['(33.90669238356164,35.60108108675799)']	19.545220	33.333333	6.515073	55.856446
2	84a31c9d-57ef-472e-bff3-8621ac6205c1	2956	Circle	Baghdad	349	['(33.90394933333334,35.58597890740741)', '(33...	48.024117	33.333333	16.008039	55.856446

Table 3: Suspicious Areas

Chapter 4: Classification of Areas of Interest

In this section, we describe the different classifications methods for the areas of interest.

A. Classification of AOIs based on Time & Movement Pattern

By studying the hits of a device over the days of a week in the discovered AOIs, we aim to determine whether a location is a person's home, work location or a frequently visited site. Here are the different approaches we have tried to tackle this problem:

a) Predicting the probability that an AOI is a Home, Work or Other based on the Day of Week and Hour of Day

We sampled the behavior and movement patterns of people from varying backgrounds, with the goal of predicting the probability that a person is in their home, work, or another location, based on general data. This was achieved by asking them about their whereabouts; whether their home, work or another location, with respect to the day of week and hour of day. With the collected data, we trained multiple machine learning models to classify a location based on day of week and hour of day. This approach however, was discontinued, given that the results were biased, as the sampling was not representative of the entire population.

b) Analyzing & Classifying the AOIs of a Device based on its Historical Data

By manipulating the historical geolocation data of a device, we managed to extract, for each day, the different stays and changes in location made for that device, as well as the time period spent in each location.

From this data, we learn the number of hits per each AOI with respect to the day of week. We also learn the number of days spent in each AOI per day of week. By cross-referencing the acquired information with the timeframe in which we have hits for the device, we get an idea on the most important clusters for a device.

- **Home AOIs**

The classification of home AOIs involves assessing the likelihood that a particular AOI represents a person's residential location. This assessment is based on various metrics and weightings, which are used to compute the likelihood that an AOI is a person's home.

cluster_label Day of week	0	1	2	3	4
Monday	148.000000	0.000000	75.000000	19.000000	20.000000
Tuesday	70.000000	0.000000	27.000000	0.000000	28.000000
Wednesday	127.000000	29.000000	4.000000	16.000000	21.000000
Thursday	60.000000	0.000000	42.000000	4.000000	9.000000
Friday	122.000000	16.000000	48.000000	27.000000	0.000000
Saturday	57.000000	9.000000	0.000000	2.000000	0.000000
Sunday	150.000000	0.000000	23.000000	8.000000	0.000000

Table 4: Number of Hits in each AOI with respect to Day of Week over entire period

Cluster Day of week	0	1	2	3	4
Monday	100.000000	0.000000	40.000000	40.000000	60.000000
Tuesday	66.666667	0.000000	66.666667	0.000000	66.666667
Wednesday	100.000000	25.000000	50.000000	25.000000	50.000000
Thursday	75.000000	0.000000	50.000000	25.000000	50.000000
Friday	100.000000	50.000000	50.000000	75.000000	0.000000
Saturday	100.000000	33.333333	0.000000	33.333333	0.000000
Sunday	60.000000	0.000000	40.000000	20.000000	0.000000

Table 5: Percentage of Days spent at each AOI From Total Days; with respect to Day of Week

Percentage of stays in an AOI with respect to Day of Week

This helps us track the percentage of visits to each AOI with respect to each day of the week. A person is more likely to have regular patterns, such as visits to work from Monday through Friday, and is more likely to frequent an AOI all days of the week, which most probably is the person's Home.

Likelihood that an AOI is Home based on Day-of-Week pattern in AOI

Each day of week is given a weight, with days from Monday to Thursday having less importance than Friday through Sunday. This is to reflect the increased likelihood of people being at home during the weekend than during the weekdays.

Percentage of Overnight Stays

This metric calculates the percentage of overnight stays in each AOI from the total number of overnights recorded for the device.

Percentage of Days spent in each AOI from Total Days

This metric determines the percentage of days spent in each AOI from the total number of days recorded for the device.

Percentage of Overnights spent in each AOI from Total Days

This value represents the percentage of overnight stays in each AOI from the total days recorded by this device.

Percentage of Overnights spent in AOI from Total Days spent in AOI

This metric calculates the number of overnight stays in each AOI from the unique days within each cluster. It helps identify the density of overnight stays in specific AOIs.

Determining the Home AOI

A new metric, Importance of Overnight Stays, is measured, to aggregate the results of the metrics related to overnight stays. Because a person almost always sleeps at their home, the location with the highest overnight importance earns a 70% weight when computing the likelihood than an AOI is the device's home AOI. The Likelihood that an AOI is home based on Day-of-Week pattern in AOI, as well as the percentage of days spent in each AOI from total number of days, are assigned 15% weight each.

The likelihood that an AOI is the person's Home AOI is measured for all AOIs, and the AOI with the highest likelihood is labeled as the home AOI. The result of this classification, however, heavily relies on the quality of the data at hand. A longer period of time and a higher number of geolocation hits increase the confidence in the home AOI classification. To reflect the confidence and reliability of this classification, we have measured a confidence Score, which is explained in further details in the "Importance of Classification" section.

- **Work AOIs**

To identify Work AOIs, we filter the geolocation data for hits between 8 AM and 4 PM, the period in which most people are at work. From this, we can learn which AOI is most probably a person's work, but working hours differ according to the job role and working institution. This solution is not ironclad and awaits further improvements.

B. Classification of AOI's based on the Location's Category and Type

Reverse geocoding an AOI's coordinates returns the location's address, as well as other attributes; most notably the location's category and type. One approach consisted of classifying an AOI from the Location's category and type.

a) Classification of AOIs based on the location's address

We attempted to identify the class of an AOI based on the full address of a location, by dissecting the components of the address, and classifying these entities using a Named-Entity Recognition NLP Model. This NER model is general and doesn't perform well on addresses from Arabic countries. Street names given after names of people are mistaken for people instead of streets. This hurdle could be overcome by training the model on address data, but collecting and labeling the training data takes a really long time, as well as training the model. Working on this approach has been discontinued.

b) Training ML Models to classify AOIs based on their location category and type

By extracting a list of all location categories and types, we attempted to manually label the classes of each category and type in order to train machine learning classifiers to predict the class of AOIs on new instances. But soon we discovered that this model would be rule-based and memorize the data. Not only would it perform poorly on new location categories and types, but the results are highly biased on the data labeling we did.

c) Finding the similarity of a location’s Category and Type with the words ‘Home’ and ‘Work’

We attempted to identify the classes of AOIs based on the location category/type’s similarity with the words ‘home’ and ‘work’, but this approach didn’t seem to be reliable and was discontinued.

d) Classifying an AOI by looking at its surrounding buildings and checking for similar opening hours

By scanning the surrounding buildings of an AOI, we aim to determine whether the location is a workplace. We start by selecting a radius around each AOI, equal to the maximum distance between the AOI centroid and the farthest hit belonging to it.
Using the google cloud API, we select the buildings and use a location sign to mark them. We also extract their opening hours if they have any. From the overpass turbo API, we extract the bounding boxes for these buildings. The markers and bounding boxes are then linked. After that, the bounding boxes are slightly expanded to cover a wider area around each building. Next, each box is assigned hits which are nearest to each bounding box. Then, we created a function that calculates the likelihood of an AOI being a workplace. This is determined by calculating the ratio of records within the AOI during the location's business hours (if applicable) to all records within that AOI.

We then assign a probability that an AOI is a workplace, considering the opening hours of the buildings surrounding it. If the hits of an AOI are only within a timeframe of the ‘opening hours’, then the AOI would be given a high probability that it is a workplace.

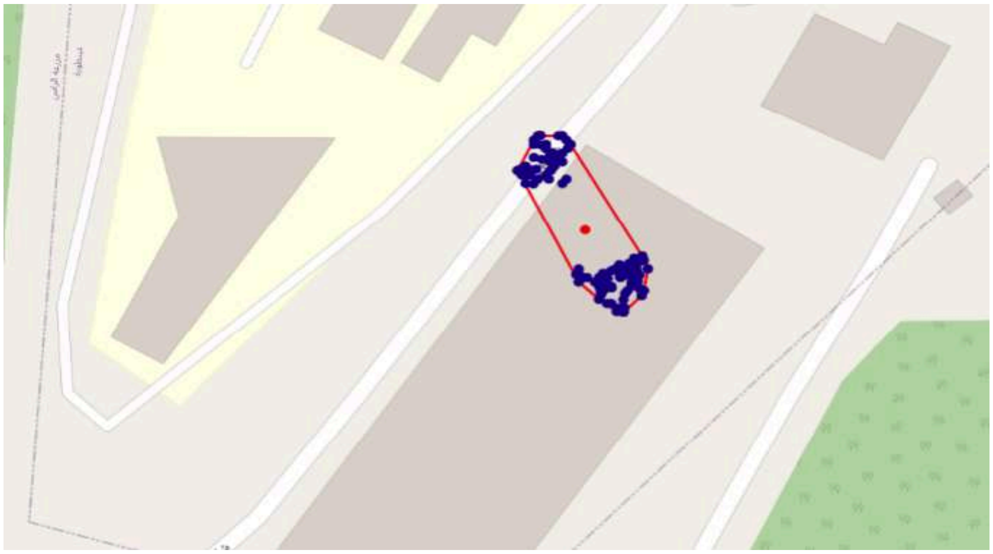

Figure 5: AOI with its centroid and hits

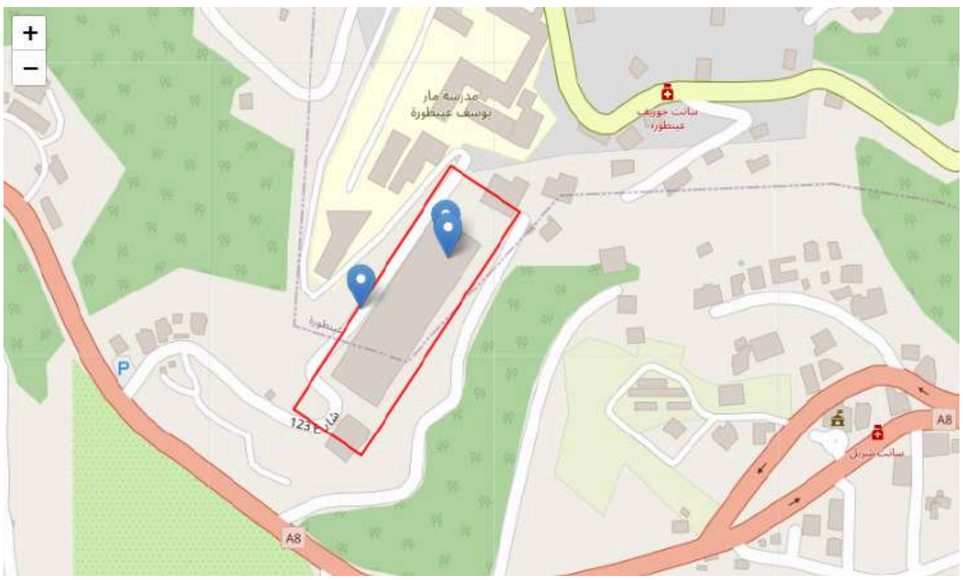

Figure 6: Blue Markers represent the locations which have opening hours, as returned by the Google Cloud API, and bounding box as returned by the Overpass Turbo API

Name	Location	Polygons	Opening_Hours	probability
Centre Sportif Saint Joseph Antoura	(33.9556062, 35.6336832)	[POLYGON ((33.95660825 35.633885750000005, 33....	{1: ('0800', '2200'), 2: ('0800', '2200'), 3: ...	1.000000
Catherina Moussa - Dietitian and Nutritionist	(33.955507, 35.6337086)	[POLYGON ((33.95660825 35.633885750000005, 33....	{1: ('0800', '0630'), 2: ('0800', '0630'), 3: ...	0.403509

Table 6: Determining whether the AOI is a workplace using the probability function

Chapter 5: Importance of Classification

In order to measure the confidence in the selected Home AOI, we examined the importance of the classification using metrics such as Time Period Consistency (TP), Hits Per Day Consistency (HD), Hour of Day Consistency (HOD), and Day of Week Consistency (DOW). These metrics collectively contribute to the calculation of a Confidence Score (CS) and an Adjusted Confidence Score (ACS), which help evaluate the reliability of the classification.

A. Time Period Consistency

This metric measures the extent to which recorded data aligns with the expected time period. It is calculated as the ratio of the number of recorded days to the number of expected days within a given time period. A TP value of 1 indicates perfect consistency, meaning that all the expected days are covered by the recorded data. On the other hand, a TP value closer to 0 suggests poor consistency, indicating the absence of data for the expected days.

B. Hits Per Day Consistency

This metric assesses the fluctuations in the number of hits per day. It is determined by the coefficient of variation (CV), which is the ratio of the standard deviation (SD) to the mean (M) of hits per day. A higher HD value (closer to 1) implies minimal fluctuations, indicating a high level of consistency in the number of hits per day. Conversely, a lower HD value (closer to 0) indicates significant fluctuations, reflecting inconsistency in the daily hit count.

C. Hour of Day Consistency

This metric measures the consistency in the distribution of hits across different hours of the day. HOD is calculated using the coefficient of variation (CV), which is the ratio of the standard deviation (SD) over the average (M) number of hits per hour. A higher HOD value closer to 1 signifies minimal fluctuations, indicating a consistent distribution of hits across hours. Conversely, a lower HOD value closer to 0 suggests significant fluctuations, highlighting inconsistencies in the hourly hit count.

D. Day of Week Consistency

DOW evaluates the consistency in the distribution of hits across different days of the week. It is also calculated using the coefficient of variation (CV), considering the standard deviation (SD) and the average (M) number of hits per day of the week. A higher DOW value (closer to 1) indicates minimal fluctuations, reflecting consistent distribution of hits across days of the week. Conversely, a lower DOW value (closer to 0) suggests significant fluctuations, signifying inconsistencies in the daily hit count.

1. **Confidence Score**

To determine the importance of classification, these four metrics (TP, HD, HOD, and DOW) are assigned equal weights, but are tailored based on the application. The Confidence Score (CS) is computed by combining the weighted average of these metrics. A higher CS value signifies a higher level of overall consistency in the data.

2. **Adjust Confidence Score**

The confidence in the classification is further adjusted to reflect the data available for analysis. Two adjustment factors are considered: the number of hits and the number of days for a particular device. These adjustment factors help account for the quantity of data, considering that more data improves the classification confidence. The acceptable targets set for the number of hits is 20000, and the target set for acceptable number of days equals 60, providing benchmarks for what is considered acceptable for reliable classification.

The adjustment factors are computed as the logarithmic ratio between the actual number of hits and days and the acceptable targets. These adjustment factors range from 0 to 1, where 0 indicates a significant deviation from the targets and 1 indicates that the targets are met or exceeded. The overall Adjusted Confidence Score is obtained by multiplying the Confidence Score (CS) by the adjustment factors for number of hits and number of days. This adjusted score provides a more accurate reflection of the confidence in the classification, considering the quantity of data available.

Conclusion

The VCIS Analytics Roadmap presents a comprehensive approach to leverage analytics for targeting, classification, and forecasting. In today's data-driven world, organizations recognize the significance of analytics in gaining valuable insights and making informed decisions. The VCIS Analytics Roadmap offers a structured framework for businesses to effectively navigate vast amounts of data and extract actionable information.
At VALOORES, we believe that analytics goes beyond being a mere service to sell; it is a culture to foster. Our internally developed products are designed with a focus on generating

actionable data, providing comprehensive coverage throughout the entire business life cycle. By incorporating advanced tools and methodologies, we empower organizations to harness the power of analytics and drive improved outcomes.
The Analytics Roadmap encompasses multiple stages. It begins with descriptive analytics, involving data grouping and classification. Our in'DataCrowd application offers diverse grouping functionalities, enabling the creation of groups based on business intelligence, simulation, rules, activities, and characteristics. This targeted approach to data analysis facilitates more accurate insights and effective management.
Predictive analytics plays a vital role in the roadmap, utilizing techniques like time series analysis, regression analysis, machine learning, spatial analysis, and anomaly detection. These methods enable organizations to identify patterns, trends, and correlations in geospatial data, leading to more precise predictions and informed decision-making.
The roadmap also emphasizes the importance of targeting specific areas of interest. By employing clustering methods such as K-means and DBSCAN, organizations can group data points into clusters, identifying significant and concentrated areas. This helps understand user behavior and preferences. Analyzing data day by day or over an extended period provides deeper insights into the dynamics and frequency of hits.

Furthermore, the roadmap addresses the significance of identifying suspicious areas. Users can define polygon objects and assign a suspiciousness level to each area. By evaluating the percentage of hits and clusters in suspicious areas, the system can flag devices exhibiting suspicious behavior, assisting in crime investigation and security analysis.