Identification of Flight Maneuvers and Aircraft Types Utilizing Unsupervised Learning with Big Data

C1C Zachary Blanks, C1C Alyssa Sedgwick, C1C Brenden Bone, and C1C Andrew Mayerchak
US Air Force Academy, C17Zachary.Blanks, C17Alyssa.Sedgwick, C17Brenden.Bone, C17Andrew.Mayerchak@usafa.edu

Abstract - **Security professionals are interested in knowing with a high level of confidence the details of each flight beyond what the self-broadcasted data provides. A program run by the Air Force is attempting to better characterize flights. From various data sources, this program attempts to determine aircraft identities. The purpose of this study is to test unsupervised machine learning methods to improve the Air Force's aircraft identification process. These alternative methods cluster unlabeled aircraft flight data from which we identify different aircraft types and maneuvers. The alternate methods include k-means and k-medoids. To improve the clustering, we first perform feature engineering. Some of the features built include acceleration rates, G forces, specific excess power, flight path angle, flight efficiency, and winding number. We then cluster the data using the above features and determine the ideal number of clusters via the average silhouette width. The initial clusters identify discrete flight maneuvers, such as takeoff and landing, and from these initial clusters we utilize sub-clustering to identify aircraft. To validate our aircraft identification process, we implement web scraping to develop a labeled dataset to compare the distribution of aircraft within the sub-cluster versus the initial cluster. The final model uses k-means, and 12 of the 14 sub-clusters generated by it have statistically different distributions of aircraft at $\alpha = 0.01$ from the initial clusters. This indicates that we can better identify aircraft type given in which sub-cluster a data point resides. With some modifications, the method could be used by the Air Force to augment the current identification process.**

Index Terms - Clustering, Aircraft Classification, Unsupervised Machine Learning

INTRODUCTION

Everyday there are approximately 100,000 flights that take place all over the world. Data from most of these flights is recorded by organizations that collect automatic dependent surveillance-broadcast (ADS-B) data. Aircraft broadcast ADS-B data to air traffic control and other planes. This allows for greater situational awareness among pilots and ground controllers. This information, however, may not be accurate because the aircraft itself is transmitting it and unscrupulous actors may have an incentive to mislead others about their identity.

Those who are involved with national security are interested in knowing to a higher level of confidence the aircraft type, rather than what the self-broadcasted data provides. A program run by the Air Force is attempting to better identify unknown aircraft. From classified and unclassified data sources, they determine aircraft identities which then are pushed to the warfighters at their respective classification levels. The accuracy of this method is classified.

I. Problem Statement

The purpose of this study is to test new methods to improve the Air Force's process of identifying aircraft. These new methods cluster unsupervised aircraft movement tracks and phases of flight, which represent different types of airplanes and maneuvers. Unsupervised data refers to data that does not have a label of what group it belongs to; in this case, the data we are using does not tell us what type of aircraft it is. Supervised data, on the other hand, is labeled. After clustering the unsupervised data, we use a web scraped labeled version of the data to verify cluster accuracy. This study provides the Air Force with information of new research directions, which could augment their current system.

II. Related Work

Previous studies have been done on identifying aircraft through sound [1], radar [2, 3], and speed and acceleration data [4]. They establish a baseline minimum of 96.2% identification accuracy for supervised sound, 96% for supervised radar, and 63.55% for unsupervised radar; group identification based on speed and acceleration data was shown to work in a Monte Carlo simulation. The problem we consider involves data from a larger number of flights and types of aircraft than previous studies. Additionally, the problem we consider involves data that can be passively collected from a far distance in almost any location, which varies from sound and radar data which are limited in location (both) and are active (radar).

The first, and simplest, unsupervised learning method we use is k-means [5]. It details the pros and cons of the method, its computational complexity, and examples of domains where it has been applied. The authors concluded

by listing the challenges and limitations surrounding unsupervised learning, which include its inherent subjectivity and lack of certainty. In our study, we implement k-means clustering as a way to help group similar flight paths and maneuvers.

An alternative to k-means is the k-medoids clustering algorithm [6].They state that k-medoids is an attractive substitute because it is more robust to outliers and noise. Three implementations of k-medoids are Partitioning around Medoids (PAM), Clustering Large Applications (CLARA), and Clustering Large Applications based upon Randomized Search (CLARANS). According to the authors, due to PAM's computational complexity, it has not been shown to scale to larger datasets. However, both CLARA and CLARANS provide similar results at a lower computational cost compared to PAM. In our research, we use the CLARA algorithm to help cluster our data into discrete flight paths and maneuvers.

For all of the algorithms above the user has to specify the number of clusters. In the past a way of determining the ideal number of clusters was to rely on heuristics -- an example being the "elbow rule." This technique involves looking for an "elbow" in the clustering error graph. In response to this problem, Rousseeuw [7], developed the average silhouette width statistic as a way of determining the best number of clusters for a particular dataset. The silhouette shows which objects lie well within their cluster and which points are in between. Furthermore, the average silhouette width allows one to select the appropriate number of clusters for the dataset. We use the average silhouette width as a way of determining the ideal number of clusters in our research.

Since we are clustering phases of flight paths using the above methods, we need to understand the definitions of these different flight segments. The International Civil Aviation Organization (ICAO) [8] put together a list and explanation of each phase of flight for any aircraft that may fly in a given airspace. These phases of flight include takeoff, climb, en route, maneuvering, approach, landing, and some other uncommon phases where we occasionally observe aircraft. The ICAO provides sub-phases for each general phase of flight.

III. Organization

The remainder of this article is organized as follows. In the methodology section, we describe our data, data processing, feature engineering, and model creation. In results and analysis, we show our model verification, model validation, and results of the clustering. In conclusions and future research, we conclude with recommendations and suggestions for future work.

METHODOLOGY

I. Data Processing

The data we use is open source information gathered from ADS-B. This represents a small fraction of the data available to the Air Force, but allows us to publish and work in an unclassified environment. We have a month's worth of data provided by the Air Force where every day has an average of about 95,584 flights. This provides us with sufficient data for testing. Flight data is comprised of regular intervals (every 30 to 60 seconds, on average) that make up the record for each track. Each record is identified by a hexadecimal identification code. However, some of the tracks have too few entries and others have far too many to be a realistic flight. For the tracks with too many instances, it is possible that a collision of the hexadecimal flight identification code occurred. We identify these anomalies by computing the length of each flight and remove records that are shorter than ten minutes or longer than 18 hours [9, 10]. We record all of the track ID values for the aircraft records we remove.

In addition to collisions, we also have missing data. Approximately 2.00% of the data is missing, but around 65.4% of all flight tracks are missing at least one data point. To correct this issue, assuming there was enough data to impute on, we use k-Nearest Neighbor imputation with five neighbors for the same track. If the track ID is missing more than 15% of all its data, then we remove that entire flight. This is done as an error-handling exception so the imputation function does not break. We then add these removed track IDs to a list for record. We impute off of each track ID separately because this will ensure the data is pulled from the nearest points in the same flight and not the closest points in the entire dataset. This is because, for example, near an airport planes could all have similar latitudes, longitudes, and altitudes but completely different speeds based off of the aircraft type. Hence, we would rather have the function impute using the same flight to ensure more accurate data.

II. Feature Engineering

Since we have few initial features in our dataset, feature engineering is a key part of our analysis. The first set of variables we create includes turn rate, climb rate, and acceleration. To compute these features, we separate each track and put them in chronological order. We then find the first difference for time, heading, altitude, and ground speed between points t and t-1. Meaning, the first point in each flight receives a NULL for these features because the point t-1 does not exist. For heading, we assume that if the heading change is greater than 180° that it flies the more economical way, e.g. 10° to 350° would be a change of 20° and not 340°. This is because we are dealing with mostly civilian aircraft and we assume this smaller change to be more reasonable for them. To calculate the turn, climb, and acceleration rates we divide the differences in heading, altitude, and ground speed by the time difference, respectively.

Next, we determine the change in distance for each individual point in its respective flight track by using the latitude and longitude at times t and t-1, and the Law of

Haversines equation for great circle distance [11]. The formula is shown in (1).

$$d = 2r \arcsin\left(\sqrt{sin^2\left(\frac{\varphi_2-\varphi_1}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\, sin^2\left(\frac{\lambda_2-\lambda_1}{2}\right)}\right) \quad (1)$$

φ =latitude
λ = longitude

We find that due to irregularities in the recording of the latitude and longitude, some of our distances were unreasonable. If the distance calculated is more than 200km greater than the distance calculated by multiplying the initial ground speed by the time elapsed, we remove these data points. We use 200km as a buffer because if the flight was accelerating, then the distance would be greater than the basic calculation of how far it should have gone.

Calculating the distance allows us to determine flight path angle and flight efficiency. Let the flight path angle be γ. We use the inverse tangent to calculate this as seen in Figure 1.

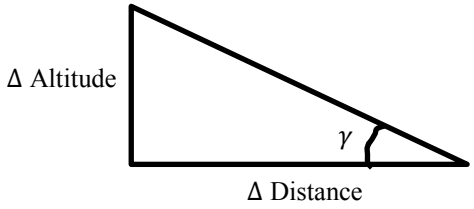

FIGURE I
FLIGHT PATH ANGLE DETERMINATION

$$\gamma = \arctan\left(\frac{\Delta Altitute}{\Delta Distance}\right) \quad (2)$$

If γ is greater in magnitude than 25°, then we remove those points because we determine this steep of a climb or descent is unreasonable for the civilian aircraft we are considering.

The flight path angle helps us determine the vertical G forces felt on the plane. We use the flight path angle to determine the magnitude of lift in terms of force. This can be seen in in Figure 2 and (3-5).

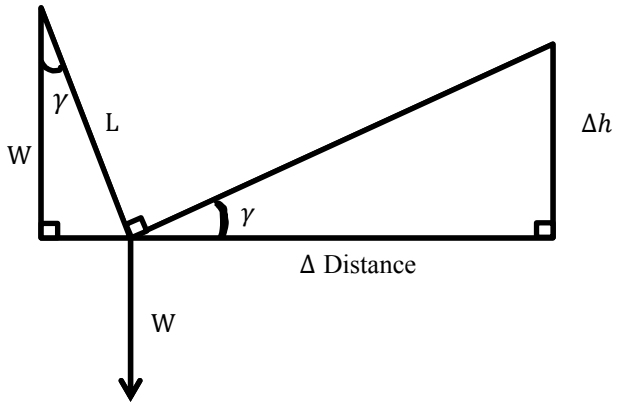

FIGURE II
VERTICAL G FORCE DETERMINATION

$$L = \frac{W}{\cos(\gamma)} \quad (3)$$

$$g_y = \frac{L}{W} \quad (4)$$

$$g_y = \frac{1}{\cos(\gamma)} \quad (5)$$

We also calculate the horizontal G forces to then combine with vertical G forces to determine overall G forces felt at every point in time. Horizontal G force comes from a level turn, which is calculated by (6).

$$g_x = \sqrt{\left(\frac{\omega V}{g}\right)^2 + 1} \quad (6)$$

We combine the vertical G forces with the horizontal G forces to determine overall G forces felt on the aircraft using (7).

$$g_t = \sqrt{g_x^2 + g_y^2} \quad (7)$$

As mentioned previously, we determine flight efficiency (ϵ) as an additional factor. Best flight path distance (β) is the direct distance between the first and last point of the flight using the Law of Haversines to calculate the shortest possible distance. Actual distance travelled (α) is the sum of the distances between every set of points in the flight record. This formulas for α and ϵ are shown in (8) and (9).

$$\alpha = \sum_{i=1}^{M} d_i \quad (8)$$

where $M$is the number of entries for each flight

$$\epsilon = \frac{\beta}{\alpha} \quad (9)$$

As an additional factor, we calculate specific excess power (SEP). This is a performance measure for aircraft.

$$SEP = \frac{dh}{dt} + \frac{V}{g} * \frac{dV}{dt} \quad (10)$$

Each flight also has a winding number (W) associated with it. The winding number is the number of times the aircraft turns a full 360°. To compute this value, we assume changes in the clockwise direction are positive and changes in the counter clockwise direction are negative. We still assume that the plane turns in the shorter direction – that is that from 10° to 350° is a change of -20° and not 340°.

$$W = \left|\left|\left(\frac{\left(\sum_{i=1}^{M} \Delta Heading_i\right)}{360}\right)\right|\right|, \quad (11)$$

where $M$is the number of entries for each flight

III. Summarizing Tracks

As an alternative to clustering every data point in a flight track, we also cluster a condensed dataset, where each row represents a summary of each complete flight. To create this dataset, we grab the flight number, track ID, flight efficiency, and winding number for every flight. Then we determine the average and max value for turn rate, climb rate, turn acceleration, climb acceleration, acceleration, vertical Gs, horizontal Gs, total Gs, specific excess power, and flight path angle. We also determine the minimum value of flight path angle and climb rate.

IV. Cluster Modeling

The models that we use are k-means and k-medoids. We utilize k-means and k-medoids because they are classical unsupervised machine learning methods with proven results across various domains. A requirement for both of these methods is to select the number of clusters. To do this, we use the average silhouette width. To be specific, we sample 5,000 data points 20 times, compute the average silhouette width of each sample, and determine the most commonly occurring result as the optimal number of clusters to use for the method. Before running k-means and k-medoids, the data is normalized such that the mean is 0 and all other data points are represented as the number of standard deviations either above or below the mean.

From the initial clusters we sub-cluster. We use the same process as before as many times as there are initial clusters. Every time we sub-cluster we only use the data from that cluster.

We cluster and sub-cluster on both all of the points individually and condensed tracks. To show the robustness of the method, we sample ten days of the data and follow the methodology described above to see how the number of optimal clusters changes.

VI. Web Scraping

Since the data utilized in this project was originally from Flight Aware, we knew that we could develop our own supervised dataset that would be mostly accurate by grabbing data about the flights from the internet. We say mostly accurate because airlines sometimes change which planes fly which routes. Using all of the flight numbers from the first day and the Flight Aware website for each flight we were interested in, we develop our own dataset that relates the flight number to the most likely aircraft type.

RESULTS AND ANALYSIS

I. Principal Component Analysis

We use principal component analysis to determine the most important features. We find that the first principal component explains 22.63% of the variance and the primary features are specific excess power, climb rate, flight path angle, turn rate, horizontal Gs, and total Gs. The second principal component captures 20.41% of the variance and the main features are the same. We feel that although these six features are the most important in clustering, since the first two principal components only explain 43.04% of the variance in the data and we are only considering 14 features, it would be best for us to cluster on all of the features. A graph showing the relative contributions to the first and second principal components can be seen in Figure 3. Anything above the red line is an important feature because it is doing more than the mean contribution.

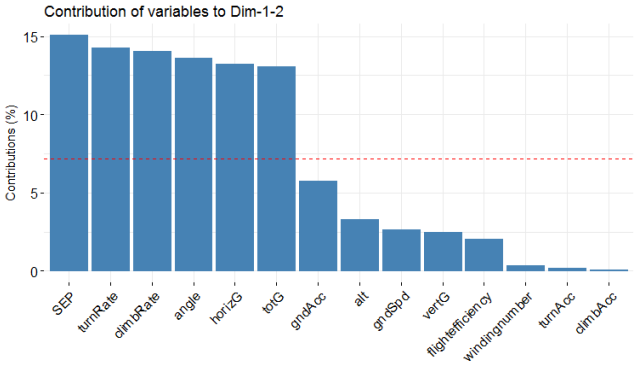

FIGURE III
FIRST PRINCIPAL COMPONENT FEATURE RELATIVE CONTRIBUTION

II. Whole Track k-Means

Using the average silhouette width for the individual points, k-means, and samples of data from ten days, we find that on average four clusters is best. The average silhouette widths for each number of clusters can be seen in Figure 4.

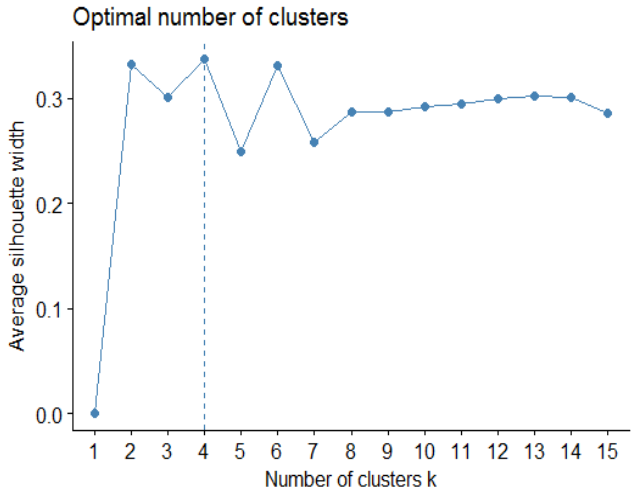

FIGURE IV
AVERAGE SILHOUETTE WIDTH

By looking at the centers of the clusters, we think that cluster one represents takeoff because of the high climb rate and flight path angle. The second cluster represents landing because of the negative climb rate and flight path angle. The third cluster represents cruising aircraft because of its low flight path angle, turn rate, and SEP. The final cluster represents maneuvering aircraft because of its high turn rate

and horizontal G forces. The various phases of flight can be seen in Figure 5.

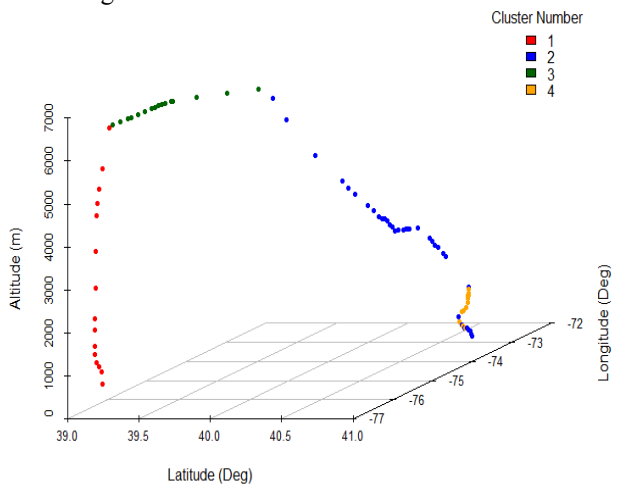

FIGURE V
FLIGHT PATH CLUSTER REPRESENTATION

After finding four centers to be the best number of clusters, on average, we cluster each of the ten days' worth of data into four groups. Once complete, we sub-cluster and find that the number of sub-clusters are robust across all ten days. The original takeoff cluster splits up into two sub-clusters. The first sub-clusters represents the initial part of takeoff with low altitude and ground speed, high turn rate climb rate, and the second represents the middle to final part of takeoff with low but slightly higher altitude and ground speed and is still climbing, but at a slower rate. The landing cluster generally breaks into two sub-clusters as well. In a similar manner to takeoff, they relate to the initial descent, where the plane is higher in altitude and faster with a steeper downward angle and more negative climb rate, and the actual landing portion, where the aircraft is lower and slower but still descending although at a slower rate than before and is turning more. For some of the days, landing had more than two clusters, though two was the mode. For these days, the additional sub-clusters appear to relate to smaller aircraft because of the higher winding number and lower flight efficiency we expect to see from small airplanes.

Cruising and turning both have more variation in them than takeoff and landing, but the mean number of clusters for both, when clusters holding anomalies are not considered, is five. These five sub-clusters for cruising generally relate to finishing the takeoff when the plane is high but still climbing, beginning to descend when the plane is high but has a negative climb rate, a generic cruise with a high winding number, a generic cruise with a low winding number, and those flights with a lower flight efficiency and often lower altitude. For turning, the five sub-clusters generally relate to turning up, turning down, lower and slower planes that are turning, fast level turns, and generic level turns. For clusters that have a high winding number and low flight efficiency, we generally think these refer to small aircraft that take off and land from the same airport. They could also refer to Remotely Piloted Aircraft or hurricane hunters that also fly around in a circle and take off and land at the same airport.

Using the supervised dataset, we are able to compare the proportion of each plane type (airliner or general aviation) found in each sub-cluster to the corresponding main cluster. To test statistical significance of our sub-clusters, we utilize a χ^2 test. In the hypotheses, X represents the distribution of planes in the particular sub-cluster or cluster. Furthermore, K is the set of all clusters for our model and I is the set of all sub-clusters corresponding to a particular cluster k. The hypothesis test for our sub-clusters is shown below.

Hypothesis Test

H_0: $X_{i,k} = X_k$

H_a: $X_{i,k} \neq X_k$

From the hypothesis test, we compute the p-values related to the distribution of planes compared to the cluster for all of our sub-clusters, $s_{i,k}$. The results are shown in Table 1. Cluster numbers do not correspond to the order explained previously.

TABLE I
CHI SQUARED RESULTS

Sub-cluster	P-value
Takeoff 1	4.36×10^{-3}
Takeoff 2	2.54×10^{-14}
Landing 1	$< 2.2 \times 10^{-16}$
Landing 2	$< 2.2 \times 10^{-16}$
Cruising 1	8.95×10^{-7}
Cruising 2	5.46×10^{-7}
Cruising 3	$< 2.2 \times 10^{-16}$
Cruising 4	0.42
Cruising 5	1.80×10^{-9}
Turning 1	$< 2.2 \times 10^{-16}$
Turning 2	$< 2.2 \times 10^{-16}$
Turning 3	$< 2.2 \times 10^{-16}$
Turning 4	0.60
Turning 5	$<2.2 \times 10^{-16}$

Looking at Table 1, we conclude that 12 of the 14 sub-cluster's distribution differs from the main cluster's distribution at $\alpha = 0.01$. This indicates that we are indeed better able to identify the general class of an aircraft based on in which sub-cluster the data point resides.

III. Other Methods

We also use k-means on the condensed track, k-medoids on all of the points, and k-medoids on the condensed track. The ideal number of clusters does not remain constant across the days randomly chosen to test – that is the results are not robust. From ten days' worth of sample data from, we use the average silhouette width to determine the ideal number of clusters for each day. The mean, minimum, and maximum number of clusters determined to be ideal for each sub-cluster across days can be seen in Table 2.

TABLE II
OTHER METHODS IDEAL NUMBER OF K RESULTS

Method	Min Cluster	Mean Cluster	Max Cluster
k-means: C	2	5	13
k-medoids: I	2	6	14
k-medoids: C	2	6	13

Because the results for the three methods are not robust across the sample of ten days, we do not sub-cluster. It is possible that erroneous data points still remain even though we use the same method of data removal due to the differences in keeping the entire track versus condensing it. If this hypothesis is true, this could be the cause of the large amount of variation in the ideal number of clusters. This represents an area of future work.

CONCLUSIONS AND FUTURE RESEARCH

I. Conclusions

This project makes incremental progress towards creating a solution to identifying aircraft using kinematic data. While three of the four methods we test in this project fail to yield robust results, using k-means on all of the individual points consistently cluster into four groups representing takeoff, maneuvering, cruising, and landing. These sub-cluster into consistent, smaller groups representing discrete flight maneuvers of which 12 of 14 have a statistically significant different distribution from the main cluster. This result indicates that we can better identify the general aircraft type (airliner or general aviation) depending on which sub-cluster the data point resides.

II. Future Research

This project has been part of a much larger and ongoing attempt by the Air Force to better classify aircraft. Areas of future exploration could include parallelizing the code to allow it to run faster on a high-performance computer, engineering new features, and improving the feature selection. One possible feature selection technique is completing further principal component analysis. Another option for feature selection is sparse k-means clustering, which allows for every feature to have a different weight. A permutation approach similar to the gap statistic would be used to determine these weights [12].

Other areas for future research include looking into different modeling methods. One such option is utilizing methods that do not require the number of clusters to be chosen, for example, Gaussian Mixture Models. Another model that does not require the number of clusters to be chosen and is able to adapt to noise in the dataset is Density-based spatial clustering of applications with noise (DBSCAN).

Finally, as we note in Section III of our Results and Analysis, there may be issues with how we clean the data for the condensed track. One final area of future work can include identifying those erroneous data points and then removing them. This could yield greater stability for the methods involving the condensed data and lead to different insights that we are not getting from using the entire track.

ACKNOWLEDGMENT

We would like to extend thanks to the advisors involved, Lt Col Gerry Gonzalez, Lt Col Robert Harder, Lt Col Jesse Pietz, Dr. Brad Warner, and Capt Solomon Sonya, with this project as well as the client organization, the United States Air Force.

REFERENCES

[1] Barbarosou, M., Paraskevas, I., & Ahmed, A. (2016). Military aircrafts' classification based on their sound signature. Aircraft Engineering and Aerospace Technology: An International Journal, 88(1), 66-72.

[2] Pham, D. T. (1998, July). Applications of unsupervised clustering algorithms to aircraft identification using high range resolution radar. In Aerospace and Electronics Conference, 1998. NAECON 1998. Proceedings of the IEEE 1998 National (pp. 228-235).

[3] Zyweck, A., & Bogner, R. E. (1996). Radar target classification of commercial aircraft. IEEE Transactions on Aerospace and Electronic systems, 32(2), 598-606.

[4] Golmohammad, H., Bolandi, H., & Saberi, F. F. (2006, December). Air target classification in two dimensional feature space. In Industrial Technology, 2006. ICIT 2006. IEEE International Conference on (pp. 518-522).

[5] Friedman, J., Hastie, T., & Tibshirani, R. (2009). Unsupervised Learning. In The Elements of Statistical Learning (pp. 485-585). New York, NY: Springer

[6] Jin, X., & Han, J. (2011). K-medoids Clustering. Encyclopedia of Machine Learning, 564-565.

[7] Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53-65.

[8] Phase of Flight Definitions and Usage Notes. (2013). Retrieved November 18, 2016, from http://www.ntsb.gov/investigations/data/documents/datafiles/PhaseofFlightDefinitions.pdf

[9] The world's 10 longest flights. (2017, February 06). Retrieved February 27, 2017, from http://www.telegraph.co.uk/travel/maps-and-graphics/the-longest-flights-in-the-world/world-s-longest-flights-1/

[10] Drescher, C. (2016, August 19). The Longest (and Shortest) Flights in the World. Retrieved February 27, 2017, from http://www.cntraveler.com/stories/2015-10-28/the-worlds-longest-and-shortest-flights

[11] Haversine formula. (2017, February 22). Retrieved February 27, 2017, from https://en.wikipedia.org/wiki/Haversine_formula

[12] Witten, D. M., & Tibshirani, R. (2010). A framework for feature selection in clustering. Journal of the American Statistical Association, 105(490), 713–726.

AUTHOR INFORMATION

Zachary Blanks, Student, United States Air Force Academy.

Alyssa Sedgwick, Student, United States Air Force Academy.

Brenden Bone, Student, United States Air Force Academy.

Andrew Mayerchak, Student, United States Air Force Academy.