ANOMALY DETECTION OF FLIGHT ROUTES THROUGH OPTIMAL WAYPOINT

M Y Pusadan[1,2], J L Buliali[1] and R V H Ginardi[1]

[1]Informatics Department, Institute Technology of Sepuluh Nopember
[2]Informatics Department, University of Tadulako

Abstract. Deciding factor of flight, one of them is the flight route. Flight route determined by coordinate (latitude and longitude). flight routed is determined by its coordinates (latitude and longitude) as defined is waypoint. anomaly occurs, if the aircraft is flying outside the specified waypoint area. In the case of flight data, anomalies occur by identifying problems of the flight route based on data ADS-B. This study has an aim of to determine the optimal waypoints of the flight route. The proposed methods: i) Agglomerative Hierarchical Clustering (AHC) in several segments based on range area coordinates (latitude and longitude) in every waypoint; ii) The coefficient cophenetics correlation (c) to determine the correlation between the members in each cluster; iii) cubic spline interpolation as a graphic representation of the has connected between the coordinates on every waypoint; and iv). Euclidean distance to measure distances between waypoints with 2 centroid result of clustering AHC. The experiment results are value of coefficient cophenetics correlation (c): $0,691 \leq c \leq 0974$, five segments the generated of the range area waypoint coordinates, and the shortest and longest distance between the centroid with waypoint are 0.46 and 2.18. Thus, concluded that the shortest distance is used as the reference coordinates of optimal waypoint, and farthest distance can be indicated potentially detected anomaly.

Keyword: Waypoint; Agglomerative Hierarchical Clustering(AHC); ADS-B; Anomaly Detection

1. Introduction

Flight is an integrated system consisting of the use of airspace, aircraft, airports, air transport, air navigation, safety and security, environment, and support facilities and other public facilities. The flight route is part of the data sources monitoring and surveillance of the flights. The flight route contains at least five parameters, including: a) the name of the flight path; b) the reference point and the coordinates (longitude and latitude); c) direction (track) leading to or from a point of reference; d) the distance between the reference point; and e) limits the lowest safe altitude (Minimum Safe Altitude / MSA) [1]. Waypoint on the flight is the reference point / set of coordinates (latitude and longitude) used for the purposed of air navigation. According to the Regulation of the Minister of Transportation No. 14 of 2009 on the Rules of Civil Aviation Safety (Civil Aviation Safety Regulations) sub part 170 051, that: observation data (surveillance) is data derived from radar equipment primary and secondary, or other systems (ADS-B, ADS-C) is used as air traffic services, should be recorded (record) automatically for use in accident and incident investigations, search and rescue, as well as the evaluation of air traffic control and system monitoring and surveillance [2].
Anomaly detection is part of data mining to determine a data (objects) that have different characteristics with the other data sets, called is exception mining. In the case of flight, one of the anomaly detection approach based on the anomalies that occur on the flight route. Anomaly detection of flight route can

be obtained by determining the optimal waypoint, therefore the flight route that is far from optimal waypoint can be indicated potential anomalies.

2. Literatures Review

Anomaly detection/outlier detection can be seen several approaches[3]. The detection by using of cluster analysis algorithm "agglomerative hierarchical clustering". These methods: 1). Begins by sampling data from the data source; 2) on the data samples, obtained by clustering the data based algorithms "agglomerative hierarchical clustering"; 3). The results of the clustering process as an accurate sample[4],

This process if found gaffe in the build of noise that value is less than δ, then released; 4) the mathematical approach μ and standard deviation σ obtained until accurate. For the determination of outliers:

$$\mu = E(X)\ dan\ \sigma = \sqrt{D(X)} \qquad (1)$$

The Outlier detection method with the approach distance[5] which is an outlier detection method of calculating the distance between data points. In the DB (Distance Based)-Outlier detection, which is the centre of Nearest Neighbourhood. At the time it was founded more than data points that are in the M-Neighbourhood D object, then the object is not an outlier. The dataset used in high-dimensional data characteristics.

In the case of aviation, conducted by the clustering[6][7]. The foremost step is outlier/anomaly analysis, by dividing of flight position (specific event) with two phases, namely the takeoff phase with 68 parameters and the approach phase with 69 parameters, then did the dimension reduction (using PCA), and to examine the cluster be used "DBSCAN (Density-based Spatial Clustering Algorithm with Noise) as the cluster method with density approach.

Some other approach is statistical[8]. The method used is ***two-side median filtering*** by calculating the median m_t, of the data point's local (signal altitude) y_t, then compared the specific threshold value.

$$if \left|y_t - m_t^k\right| < \tau, \text{keeps } y_t; \qquad (2)$$

$$\text{else } \left|y_t - m_t^k\right| \geq \tau,\ y_t \text{ as } \textbf{outlier} \text{ and replaced } y_t \text{ with } m_t^k$$

Along the flight routes, must be determined based on traffic of the aircraft. It intended that one aircraft to another aircraft that are in the same traffic determined by separation indicator value. So that anomalies (conflict/collision)[9] data on the detection based on the separation ($R_{CAZ} = 528ft$ and $H_{CAZ} = 200ft$, R = Radius, H = Horizontal, CAZ = Aircraft Collision Zone). Therefore, the flight position or flight level consists of takeoff, cruise, approach, and landing, which allows for the accurate detection of anomalies of the aircraft cruising level or in cruise phase[10].

The most important thing is also considered in the scope of anomaly detection of flight is the determination of flight waypoints[11] which is the specific location of the flights that the accurately to flying past the points of reference before reaching its destination.

The following point of finding the optimal waypoint based on the reference dot on the map of flight routes which occurs in one flight routes (call sign of flight). Established along the point of reference, the cubic spline interpolation process is carried out[12], so get waypoints.

From the waypoints, segmentation and clustering processed[13]. so that, the cluster can be analysed with to calculating the cophenetic correlation coefficient (C)[14][15] for each segment, so that obtained *centroids* that become the new points of reference in waypoints.

3. Proposed Method

The research framework shown in Figure 1, as the sequence of the process of determining the optimal waypoint:

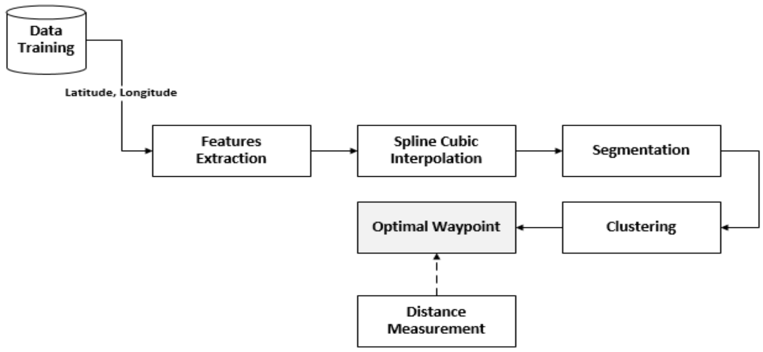

Figure 1. Block Diagram Process

3.1. The Algorithm Determination of the Optimal Waypoint

The algorithm that determination of the optimal waypoint is shown as follows:

```
pointsRf={(LatRf1,LongRf1),(LatRf2,LongRf2),…,(LatRfn,LongRfn)}
waypointRf←pointsRf
get dataTrain

datastart←waypointRf1
dataend←waypointRfk
i←1
while (dataTrain ≥ datastart and dataTrain ≤ dataend) do
  segment←dtTrain
  i←i+1
endwhile

[c,centroid]←clusterAHC(segment,k←2) //prosess clustering AHC
Result: centroid1 and centroid
//mengukur distance menggunakan euclIean distance
dist1←dist(waypointRf, centroid1)
dist2←dist(waypointRf, centroid2)
if dist1<dist2 then
     waypointoptimal ← dist1
  else
     waypointoptimal ← dist2
end
```

3.2. Data Source (ADS-B Data)

ADS-B data used in the research originated from flightradar24[16] with the data format is Comma Separated Values (CSV), with attributes as follows: timestamp, UTC, call sign, position, altitude, speed, and direction. The following on table 3.1 shows the attributes of data ADS-B:

Table 1. Data Source (ADS-B Data)

TIME-STAMP	UTC	CALL-SIGN	POSITION	ALTI-TUDE	SPEED	DIREC-TION
1472721080	2016-09-01T09:11:20Z	LNI860	-0.88474,119.856743	3650	173	161
1472721070	2016-09-01T09:11:10Z	LNI860	-0.87789,119.854507	3850	173	161
…	...	...	...	...	...	...

3.3. Features Extraction

Data sources (ADS-B data), converted to the DBMS (MySQL) tables, to the computational. The following on table 3.2 below, shows the latitude and longitude attributes, parsing from attribute "position" that's shown on the coordinates x, y.

Table 2. Table (DBMS MySQL) with attributes: Latitude and Longitude are expressed individually

TIME-STAMP	DATE	CALL-SIGN	ALTI-TUDE	SPEED	LATI-TUDE	LONGI-TUDE
1472721080	01-09-16	LNI860	3650	173	-0.88467	119.8567
1472721070	01-09-16	LNI860	3850	173	-0.8779	119.8545
…	…	…	…	…	…	…

The following shows the flight routes to the destination of Surabaya to Palu (call sign: LNI860) and the distribution of data for 30 days (01^{st} Nov 2016 until 30^{th} Nov 2016).

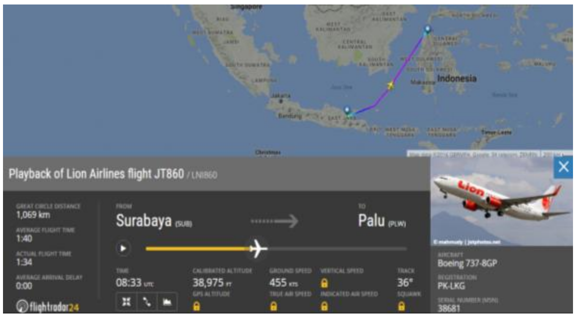

Figure 2. Flight route
(call sign: LNI860/SUB-PLW)

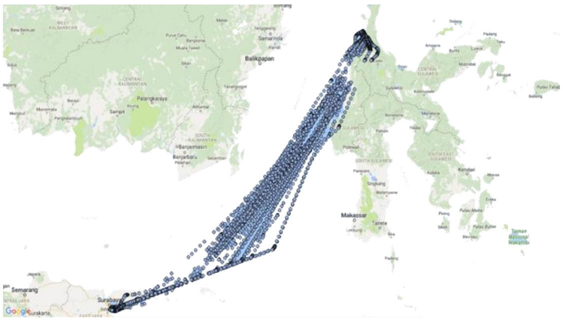

Figure 3. Mapping data distribution of flight
with based on call sign aircraft:
LNI860 (SUB-PLW)

3.4. Spline Cubic Interpolation

Assumed that the points of reference interpolation are: W_1 = (-2.4687,118.84); W_2 = (-4.172, 117.485001); W_3 = (-5.8717, 117.221703); W_4 = (-6.5922, 115.0785); and W_5 = (-7.0568, 113.8238). Established along the method of cubic spline interpolation equation:

$$S = \begin{Bmatrix} S_1(x), \ if \ x_1 \leq x \leq x_2 \\ S_2(x), \ if \ x_2 \leq x \leq x_3 \\ \vdots \\ S_{n-1}, \ if \ x_{n-1} \leq x \leq x_n \end{Bmatrix} \tag{3}$$

$$S_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i \tag{4}$$

$$S_i'(x) = 3a_i(x - x_i)^2 + 2b_i(x - x_i) + c_i \tag{5}$$

$$S_i''(x) = 6a_i(x - x_i)^2 + 2b_i \tag{6}$$

Obtainable, equation to interpolation of spline cubic:
$S_1(x) = -1.25x^3 + 2.3x^2 + 1.9x + 113.82$; $S_2(x) = -1.25x^3 - 0.56x^2 + 3.23x + 115.078$; $S_3(x) = 0.6x^3 - 2.15x^2 + 2.075x + 117.22$; $S_4(x) = -0.25x^3 + 0.1x^2 - 0.032x + 117.48$; and $S_5(x)$ = -0.25x3 - 0.36x2 + 0.1x + 118.84.

Thus, graph of equation to interpolation of spline cubic is generated:

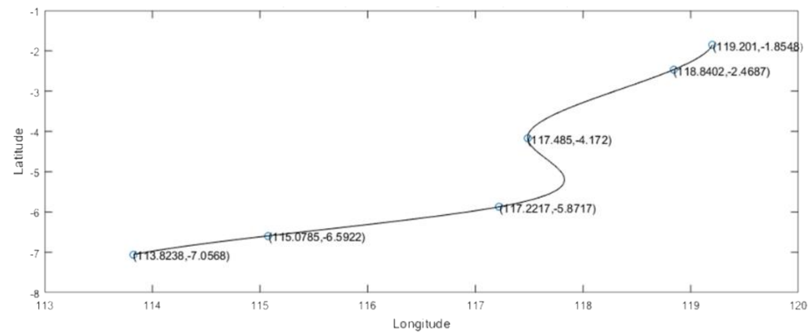

Figure 4. Cubic spline interpolation on the call sign: LNI860 (SUB-PLW)

3.5. Segmentation

Segmentation is the area located between the one waypoint to another waypoint in the same flight routes. Its formulation, as follows: $W_1 \leq SEGMENT_1 \leq W_2$, $W_2 \leq SEGMENT_2 \leq W_3$,…,$W_n \leq SEGMENT_n \leq W_{n+1}$ (n=1,2, … , k). Hence, range area was obtained:

Table 3. Segmentation based on area range waypoints

RANGE AREA (Latitude, Longitude	SEGMENTASI	WAYPOINTS
(-2.4687, 118.84) – (-1.8548, 119.20)	$SEGMENT_1$	$W_1 - W_2$
(-4.1720, 117,48) – (-2.4687, 118.84)	$SEGMENT_2$	$W_2 - W_3$
(-5.8717, 117.22) – (-4.1720, 117.48)	$SEGMENT_3$	$W_3 - W_4$
(-6.5922, 115.08) – (-5.8717, 117.22)	$SEGMENT_4$	$W_4 - W_5$
(-7.0560, 113,82) – (-6.5922, 115.07)	$SEGMENT_6$	$W_5 - W_6$

3.6. Clustering

Cluster method used is Agglomerative Hierarchical Clustering (AHC), because: the data to be grouped in a cluster is determined by the distance (Euclidean distance), so that the proximity between data become determined to forming a cluster. In this experiment, the "complete linkage" in the method of agglomerative hierarchical clustering (AHC) is used. The AHC formula with complete linkage clustering:

$$d(A,B) = \frac{max}{x \in A, y \in B} \{S_{xy}\} \qquad 7$$

S_{xy} : the distance between the two data (x and y) are each derived from cluster A and B.
then, analyze the cluster based on the value of cophenetic correlation coefficient (c). And then, that the value: $0 \leq c \leq 1$, formula c:

$$c = \frac{\sum_{i<j}(x(i,j)-x)(t(i,j)-t)}{\sqrt{[\sum_{i<j}(x(i,j)-x)^2][\sum_{i<j}(t(i,j)-t)^2]}} \qquad 8$$

x(i,j) : distance between object i and object j in x;
t(i,j): cophenetic distance between object i and object j of t(:,3);

3.7. Clustering Segmentation

Some of the steps are performed on clustering segmentation, such as: 1). Data training from the first day until thirtieth day, each the segment divided based on the range area of waypoints; 2). Clustering (AHC Clustering) in "data training" each the segment for one day of flight. Hence, obtained centroid of each cluster, as follows:

IOP Conf. Series: Journal of Physics: Conf. Series **801** (2017) 012041 doi:10.1088/1742-6596/801/1/012041

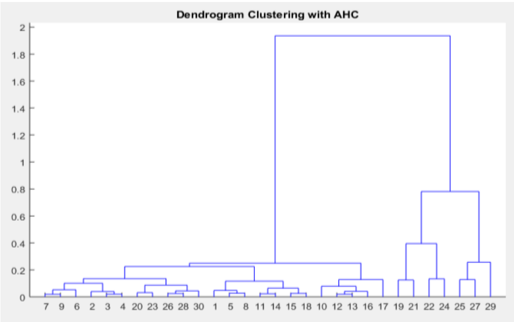

Figure 5. Dedrogram clustering AHC

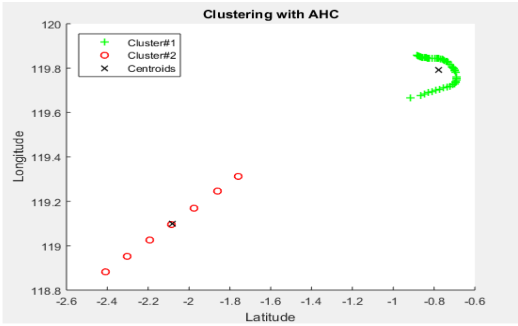

Figure 6. Cluster-1 and Cluster-2

(Data processing in segment-1)

4. Experiment Results

Based on the results of clustering AHC of data in segmentation (segment 1 to segment 5), it obtains the results shown in the following table:

Table 4. Centroid, Segmentation, and Cophenetic Correlation Coefficient

SEGMENTATION	Clustering AHC		
	Centroid-1	**Centroid-2**	**Cophenetic Corr. Coeff.**
$SEGMENT_1$	-0.784, 119.793	-2.083, 119.098	0.97424
$SEGMENT_2$	-2.764, 118.610	-3.626, 117.821	0.69170
$SEGMENT_3$	-5.544, 116.041	-4.730, 116.798	0.70207
$SEGMENT_4$	-5.999, 115.616	-6.450, 115.194	0.89212
$SEGMENT_5$	-7.3481, 113.012	6.8742, 114.340	0.88480

4.1. Distance Waypoints and Centroids Segment

To generate optimal waypoints, be determined by measuring the distance between the waypoint and the centroid. Smaller distance as points of reference for the new one. For the measure distance using the Euclidean Distance equation:

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \qquad 9$$

The following table shows the value of distance between waypoints and centroids.

Table 5. Distance waypoints and centroids

WAYPOINTS		CENTROIDS SEGMENTATION		DISTANCE
RANGE WAYPOINTS	**POINTS**	**SEGMEN-TATION**	**CENTROID**	
$W_1 - W_2$	(-1.8548, 119.201) –	$SEGMENT_1$	C_{11} = (-0.77843, 119.7926)	1.2282
	(-2.4868, 118.840)		C_{12} = (-2.08283, 119.0977)	**0.4638**
$W_2 - W_3$	(-2.4687, 118.840) –	$SEGMENT_2$	C_{21} = (-3.77079, 117.6875)	1.7378
	(-4.1720, 117.4850)		C_{22} = (-2.90686, 118.4808)	**1.6089**
$W_3 - W_4$	(-4.1720,117.4850) –	$SEGMENT_3$	C_{31} = (-5.40080, 116.1750)	1.7927
	(-5.8717, 117,2217)		C_{32} = (-4.58718, 116.9320)	**1.3166**
$W_4 - W_5$	(-5.8717,117,2217) –	$SEGMENT_4$	C_{41} = (-5.99902, 115.6165)	2.0971
	(-6.5922, 115.0785)		C_{42} = (-6.45084, 115.1944)	**0.7987**
$W_5 - W_6$	(-6.5922,115.0785) –	$SEGMENT_5$	C_{51} = (-7.34812, 113.0122)	2.1859
	(-7.056, 113.82380)		C_{52} = (-6.87422, 114.3403)	**0.5443**

4.2. Spline Cubic Interpolation pada Centroid

There are two groups of centroid is be determined based on the distance ($dist_{centroid}$), namely: $dist1_{centroid}$ and $dist2_{centroid}$, obtainable is: $dist1_{centroid} \geq dist2_{centroid}$.

dist1$_{\mathbf{centroid}}$ = {C_{11} = (-0.77843, 119.7926), C_{21} = (-3.77079, 117.6875), C_{31} = (-5.40080, 116.1750), C_{41} = (-5.99902, 115.6165), C_{51} = (-7.34812, 113.0122)}

Equation Cubic Spline Interpolation in **dist1**$_{\mathbf{centroid}}$ **:**
$S_{11} = 0.27x^3 - 1.40x^2 + 3.33x + 113.01$; $S_{12} = 0.27x^3 - 0.31x^2 + 1.02x + 115.62$; $S_{13} = -0.02x^3 + 0.05x^2 + 0.90x + 116.04$; $S_{14} = -0.02x^3 - 0.08x^2 + 0.82x + 118.61$

dist2$_{\mathbf{centroid}}$ = {C_{12} = (-2.08283, 119.0977), C_{22} = (-2.90686, 118.4808), C_{32} = (-4.58718, 116.9320), C_{42} = (-6.45084, 115.1944), C_{52} = (-6.87422, 114.3403)}

Equation Cubic Spline Interpolation in **dist2**$_{\mathbf{centroid}}$**:**
$S_{21} = 0.19x^3 - 0.99x^2 + 2.40x + 114.34$; $S_{22} = 0.19x^3 - 0.75x^2 + 1.66x + 115.19$; $S_{32} = -0.07x^3 + 0.23x^2 + 0.76x + 116.80$; $S_{42} = -0.07x^3 - 0.01x^2 + x + 117.82$

The graph of Cubic Spline Interpolation:

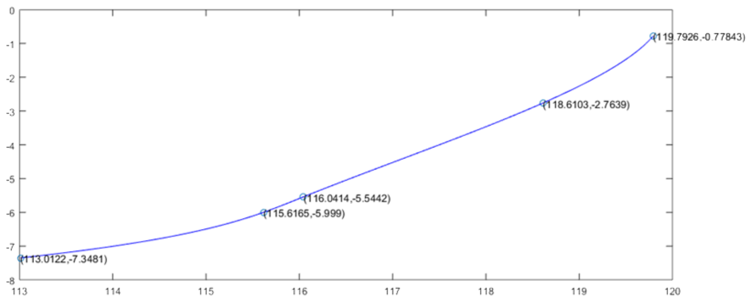

Figure 7. Graphic Spline Cubic Interpolation (**dist1**$_{\mathbf{centroid}}$)

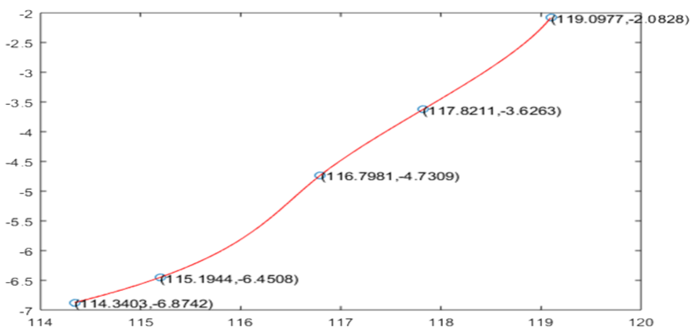

Figure 8. Graphic Spline Cubic Interpolation (**dist2**$_{\mathbf{centroid}}$)

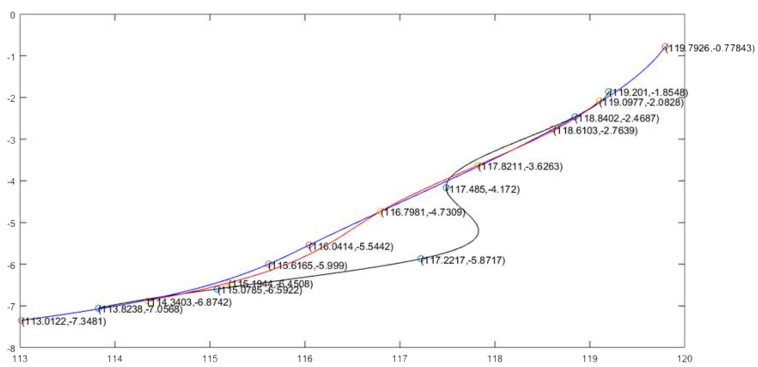

Figure 9. Graphic Spline Cubic Interpolation (waypoint, **dist1**$_{\mathbf{centroid}}$, **dist2**$_{\mathbf{centroid}}$)

5. Conclusions

The results achieved in this study are: optimal waypoint that obtained from the centroid with the smallest distance from the waypoint (centroid$_2$) in every cluster of the segmentation. The resulting coordinates are C_{12} = (-2.08283, 119.0977), C_{22} = (-2.90686, 118.4808), C_{32} = (-4.58718, 116.9320), C_{42} = (-6.45084, 115.1944), C_{52} = (-6.87422, 114.3403). The coefficient cophenetics correlation (c): $0{,}691 \leq c \leq 0974$.

6. Reference

[1] K. Perhubungan, "Undang Undang No. 1 tahun 2009 tentang Perhubungan Udara." [Online]. Available: http://hubud.dephub.go.id/?en/uu/download/5.
[2] K. Perhubungan, "Peraturan Menteri Pehubungan No. 14 tahun 2009 tentang Peraturan Keselamatan Penerbangan Sipil." [Online]. Available: http://jdih.dephub.go.id/assets/uudocs/permen/2009/KM. 14 TAHUN 2009.pdf.
[3] J. Xi, "Outlier Detection Algorithms in Data Mining," *2008 Second Int. Symp. Intell. Inf. Technol. Appl.*, vol. 1, no. 4, pp. 94–97, 2008.
[4] Y. Zhang, J. Liu, and H. Li, "An Outlier Detection Algorithm based on Clustering Analysis," pp. 1127–1129, 2010.
[5] Y. Li and H. Kitagawa, "DB-Outlier Detection by Example in High Dimensional Datasets," pp. 73–78, 2007.
[6] L. Li, M. Gariel, R. J. Hansman, and R. Palacios, "Anomaly detection in onboard-recorded flight data using cluster analysis," *2011 IEEE/AIAA 30th Digit. Avion. Syst. Conf.*, p. 4A4-1-4A4-11, 2011.
[7] L. Li, S. Das, R. John Hansman, R. Palacios, and A. N. Srivastava, "Analysis of Flight Data Using Clustering Techniques for Detecting Abnormal Operations," *J. Aerosp. Inf. Syst.*, vol. 12, no. 9, pp. 1–12, 2015.
[8] X. Jiao, "One Effective Method of Outlier Detection in Flight Data If |," pp. 797–801, 2009.
[9] M. Gariel, F. Kunzi, and R. J. Hansman, "An algorithm for conflict detection in dense traffic using ADS-B," *AIAA/IEEE Digit. Avion. Syst. Conf. - Proc.*, pp. 1–12, 2011.
[10] E. Chu, D. Gorinevsky, and S. Boyd, "Detecting aircraft performance anomalies from cruise flight data," *AIAA Infotech Aerosp. Conf.*, no. April, p. 3307, 2010.
[11] Z. Liang, Q. Li, and Z. Ren, "Waypoint constrained guidance for entry vehicles," *Aerosp. Sci. Technol.*, vol. 52, pp. 52–61, 2016.
[12] Sky McKinley and Megan Levine, "Cubic Spline Interpolation." [Online]. Available: http://pages.intnet.mu/cueboy/education/notes/numerical/cubicsplineinterpol.pdf. [Accessed: 19-Nov-2016].
[13] J. F. Cui and H. S. Chae, "Applying agglomerative hierarchical clustering algorithms to component identification for legacy systems," *Inf. Softw. Technol.*, vol. 53, no. 6, pp. 601–614, 2011.
[14] O. Access, "Comparison of hierarchical cluster analysis methods by cophenetic correlation," pp. 1–8, 2013.
[15] MathWorks, "Cophenetic correlation coefficient." .
[16] Flight Radar 24, "ADS-B Data Source." .