R. Kurniati, R. Passarella, I. G. Afriansyah, O. Arsalan, A. Aditya, M. R. Fathan, R. S. Yousnaidi, H. Veny*

AN UNSUPERVISED LEARNING-BASED ANALYSIS OF THE TAKE-OFF BEHAVIOR OF THE A320 AND B738 AT SULTAN HASANUDDIN INTERNATIONAL AIRPORT

UDK 629.7
RECEIVED: 2023-06-26
ACCEPTED: 2023-12-10

SUMMARY: The purpose of this research was to look at the behavior of two well-known commercial aircraft types in Indonesia (the A320 and the B738) during the take-off phase. This was done to provide new information in the field of aviation, particularly flight safety. Observations were made at Sultan Hasanuddin International Airport by observing aircraft ADS-B data, which defines the behavior of the flight pattern. This ADS-B data is the subject of data analysis, which will subsequently be taught to the machine (computer) so that it can recognize the pattern and construct clusters. The purpose of this study is to utilize unsupervised learning, specifically K-Means clustering, to categorize and identify patterns in unlabeled ADS-B data obtained from AERO-TRACK. To prepare the raw data and create a dataset, data analysis techniques were employed. The machine learning model generates three distinct clusters: cluster 1 represents aircraft take-off on two-thirds of the runway, cluster 2 represents aircraft take-off on the entire runway, and cluster 3 represents aircraft take-off on one-third of the runway. The elbow method is utilized to analyze and interpret the three clusters produced by the model. An interesting observation is that the B738 aircraft dominate in all three clusters, while the A320 aircraft dominate in clusters 1 and 3. Notably, in cluster 2, there is a significant number of commercial planes taking off, accounting for 145 out of 628 flights. Based on the observed data spanning 91 days (September 26 to December 26, 2022), there is a 23% probability of runway excursion (overshooting the runway) in this cluster. Additionally, the research reveals that A320 aircraft demonstrate a safe zone take-off rate of 87%, whereas the B738 aircraft demonstrate a rate of 70.5%. These findings, derived from the analysis of ADS-B data such as GPS-Altitude and Coordinate, are intended to serve as valuable knowledge for aviation authorities, aviation users, and other stakeholders in the aviation industry.

Key words: *airport, data ADS-B, cluster, runway, K-Means*

INTRODUCTION

Airplanes serve as a means of transportation employed by people. In addition to its time-saving benefits, air travel is widely regarded as the most secure form of transportation *(Hernik et al., 2018, Šebjan et al., 2017)*. However, despite its esteemed safety record, incidents involving air transportation, particularly in the context of commercial flights, occasionally occur, leading to passenger fatalities or substantial aircraft damages *(Passarella et al., 2023a)*.

One of the flight phases of concern that can cause accidents is the take-off phase. This phase is important because many things affect the movement of the aircraft, such as the total weight of the aircraft, aircraft thrust, wind speed and direction, and runway length *(Airbus Accident Statistics, 2022, Huang, 2020)*.

*Rizki Kurniati, MA of Engineering, (rizkikurniati@ilkom.unsri.ac.id), Rossi Passarella, MA of Engineering, (passarella.rossi@unsri.ac.id), (corresponding author), Indra Gifari Afriansyah, BA of Computer Science, Osvari Arsalan, MA of Engineering, Aditya Aditya, BA of Computer Science, Muhammad Rizki Fathan, BA of Computer Science, Rani Silvani Yousnaidi, BA of Computer Science, Faculty of Computer Science, Universitas Sriwijaya, Indonesia, Harumi Veny, Ph.D. of Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia.

The process of taking off and landing airplanes can vary depending on different factors. In general, airplanes gain speed along the ground until there is enough lift to initiate take-off. The reverse process is followed for landing. However, there are variations in take-off procedures, including the ability of certain airplanes to take off at lower speeds, which is known as a short take-off. The International Civil Aviation Organization (ICAO) recognizes the importance of standardized take-off rules in aviation. These rules are in place to ensure that pilots adhere to proper procedures and leave the runway in a safe and appropriate manner. By following these regulations, pilots can prevent unexpected aircraft behavior that may lead to inappropriate reactions or unsafe conditions.

As per data provided by KNKT (Komite Nasional Keselamatan Transportasi - National Transportation Safety Committee - Indonesia), there were a total of 280 aircraft accidents categorized as runway excursions between 2007 and 2016. These incidents were further divided into two groups: accidents and incidents, with 105 cases accounting for 37.5 percent of the total *(Saputra, 2017)*. Another study conducted by Sandhyavitri et al. *(2014)* revealed that Sultan Hasanuddin Airport, located in Makassar City, ranked as the second most accident-prone airport after Wamena Airport. The airport had a total movement of 212,656 and a deviation value of 3.540, which is calculated by assessing the difference between expected and recorded occurrences. Another research mentions that runway excursions are common during the take-off and landing phases *(Jenkins et al., 2012, Chang et al., 2016, Distefano et al., 2017, Passarella and Nurmaini, 2022)*. In addition, the risk of aircraft excursions during the take-off and landing phases of flight at Indonesian airports shows very high results due to many aircraft landing outside the Touch Down Zone (TDZ); *(Passarella et al., 2023b)*.

The research was developed using new technology data, known as ADS-B (Automatic Dependent Surveillance-Broadcast) data, to comprehend and help the analysis from a standpoint other than the KNKT. When referring to the FAA (Federal Aviation Administration) decision requiring all airplanes to use or activate ADS-B by 2020 *(Federal Aviation Administration, 2019)*. This is the impetus for investigating whether Sultan Hasanuddin International Airport has optimized the ADS-B ground station and the behavior of commercial aircraft movements during the takeoff phase to estimate the risk of runway ejection while data collection is taking place.

The objective of this research is to monitor the departure of commercial airplanes from Sultan Hasanuddin Airport and analyze the collected ADS-B data at the take-off point. The analysis aims to provide useful information for aviation users and other stakeholders, allowing them to assess whether the aircraft conforms to the established standards of commercial flight procedures. Additionally, the research seeks to determine the precise GPS coordinates of the aircraft's natural position at the take-off point. This information can be inferred from the last altitude parameter of the flight's ADS-B time series data, which is typically zero. By applying the K-Means method to the flight dataset, the data points can be grouped into clusters. This helps in understanding the occurrence of various data clusters, the reasons for their clustering, and any similarities observed among different aircraft types. The conclusions drawn from the analysis can provide valuable insights for stakeholders involved in aviation operations.

In the field of runway excursion risks, previous research conducted by various researchers *(Komite Nasional Keselamatan Transportasi, 2021, Komite Nasional Keselamatan Transportasi, 2020, Komite Nasional Keselamatan Transportasi, 2017)* has primarily relied on primary data, specifically Flight Recorder (black box) data. In contrast, our research utilizes secondary data, namely ADS-B data. Several studies have indicated that ADS-B data *(Jun et al., 2011)* offers superior performance compared to radar. Additionally, research by Zhao et al. *(2020)* has demonstrated that ADS-B data exhibits high accuracy and meets the necessary requirements. Furthermore, the potential of using ADS-B data for pilot protection has been explored in the research conducted by Norman *(2021)*. The use of big data is an important issue in developing prediction models to reduce the incidence of workplace accidents *(Šrekl, 2022)*. ADS-B data is a form of big data.

This research is divided into the following sections: The background to the selection of the theme and title of the research is explained in the introduction section; the materials and methods section explains the data and methods used; and the results and discussion section explains each component of the approach used and the results obtained. All method results have been discussed for relevant values, and the findings have been summarized in the conclusion section.

MATERIAL AND METHODS

This section describes the data material used to assess the commercial aircraft take-off procedure at Sultan Hasanuddin International Airport using ADS-B data. This section also covers the processes implemented to convert data into information to provide a valued service to airlines and aviation transportation consumers. The research team collected the data for this study through their AERO-TRACK application (https://aerotrack.kioznets.id/). The team developed this application by accessing the ADS-B flight data API and placing a focus on airport boundaries *(Muhammad et al., 2023)*. The data on aircraft movement obtained through this application is classified as secondary data. It is worth noting that primary data on aircraft movement typically originates from the aircraft's BlackBox. Secondary data were obtained through the AERO-TRACK application in the form of flight ID, date, ICAO-24, latitude, longitude, heading, altitude, ground speed, squawk, radar, aircraft code, registration, time, departure, destination, number, airline International Air Transport Association (IATA), on the ground, vertical speed, callsign, and airline ICAO data. Data were collected from September 26 to December 26, 2022 (spanning 91 days). This dataset generally has metadata summarized in Table 1.

Table 1. Metadata Set

Tablica 1. Set metapodataka

Data Information	
Recording Period	September 26 – December 26, 2022, at 16:45:45
Download Date	December 27, 2022
Format	.csv
Data Size	30 MB (31.499.825 bytes)
Number of rows	192.622
Number of Columns	21

In this study, a data engineering technique was carried out by building a restricted data storage application for all aircraft performing take-off operations at Sultan Hasanuddin International Airport. This flight information is saved in the AERO-TRACK application database. After the data engineering process was completed, data analysis was performed to fill in the missing data and identify the variables utilized to derive insights or patterns from the data. The next step was to identify how to cluster this take-off data so that the machine could segregate data based on centroid closeness (K-Means) and develop an understanding and value for human learning. The next point of action was how to visualize the outcomes of commercial aircraft research during the take-off phase so that readers could comprehend the aim of this research. Figure 1 depicts our approach in greater detail.

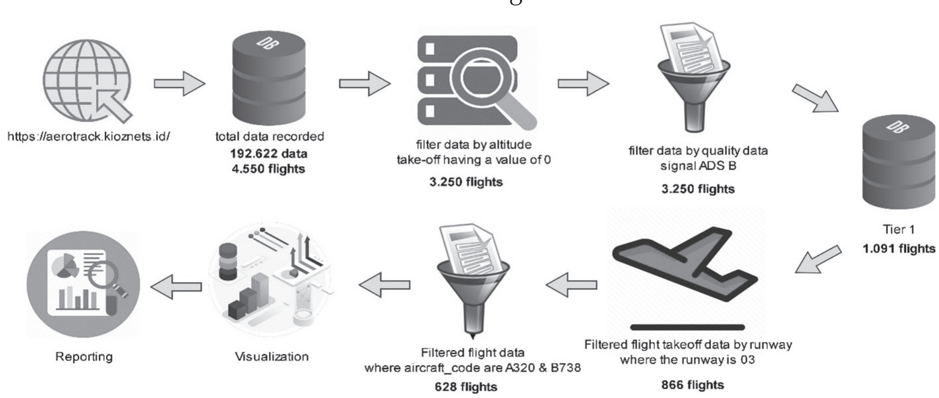


Figure 1. Proposed methods used for this study

Slika 1. Predložene metode korištene u ovom istraživanju

At first, the AERO-TRACK application records ADS-B data, and in this process, the API data is taken from flightradar24 by applying boundaries to the Sultan Hasanuddin International Airport area, implying that the AERO-TRACK application only records ADS-B data of commercial aircraft that enter the boundaries.

Secondly, AERO-TRACK'S data follows the *flightradar24* standard, with 192,622 data points and 4,550 flights stored over 91 days of observation. This information is kept in SQL format.

In the third step, the SQL data from the AERO-TRACK database is extracted and analyzed by focusing on the flight altitude point. Specifically, the data points where the flight altitude is 0 are identified, indicating that the aircraft is prepared for the takeoff phase until there is a change in the altitude value. After completing this step, the original dataset of 4,550 flights is narrowed down to 3,250 flights, resulting in a reduction in the amount of data. This decrease in data is due to the new aircraft's ADS-B data being turned on after take-off (starting the initial climb phase).

Fourth, the ADS-B data quality of the 3,250 flights was classified into tiers. The International Civil Aviation Organization (ICAO) has divided ADS-B data quality into three tiers: "tier 1" if the timestamp time difference is 0.5 to 9 seconds; "tier 2" if the difference is more than 19 seconds; and "tier 3" if the difference is greater than 60 seconds *(ICAO Asia and Pacific Office, 2014)*. The results of this fourth stage obtained tier 1 data for as many as 1,091 flights, tier 2 data for as many as 583 flights, and tier 3 data for as many as 1,576 flights. In this research, only tier 1 data is the focus of research; in other words, the dataset for this research is only 1,091 flight data.

Fifthly, the focus was placed on runway number 03 at Sultan Hasanuddin Airport, despite the presence of two runways with four arrival or departure angles each. The selection of runway number 03 was based on the fact that the takeoff aircraft data from this particular runway accounted for 79.4% of the total dataset.

Sixthly, considering that the research objectives were centered on A320 and B738 commercial aircraft, the filtering process was conducted once again to gather data exclusively for these two aircraft types. As a result, 628 flights were obtained as the final dataset. This data reduction indicates that this research only utilizes 14% of the 91-day observations of takeoff flights.

Seventhly, this step holds significance as it involves the application of a machine learning method. Unsupervised learning was chosen since there was no labeling procedure during phases 1 to 6. One of the unsupervised machine learning techniques utilized is K-means. The K-means preparation process incorporates the use of the Euclidean distance method *(Chouinard, 2023)*, which measures the distance between flight data, ultimately resulting in the grouping of data.

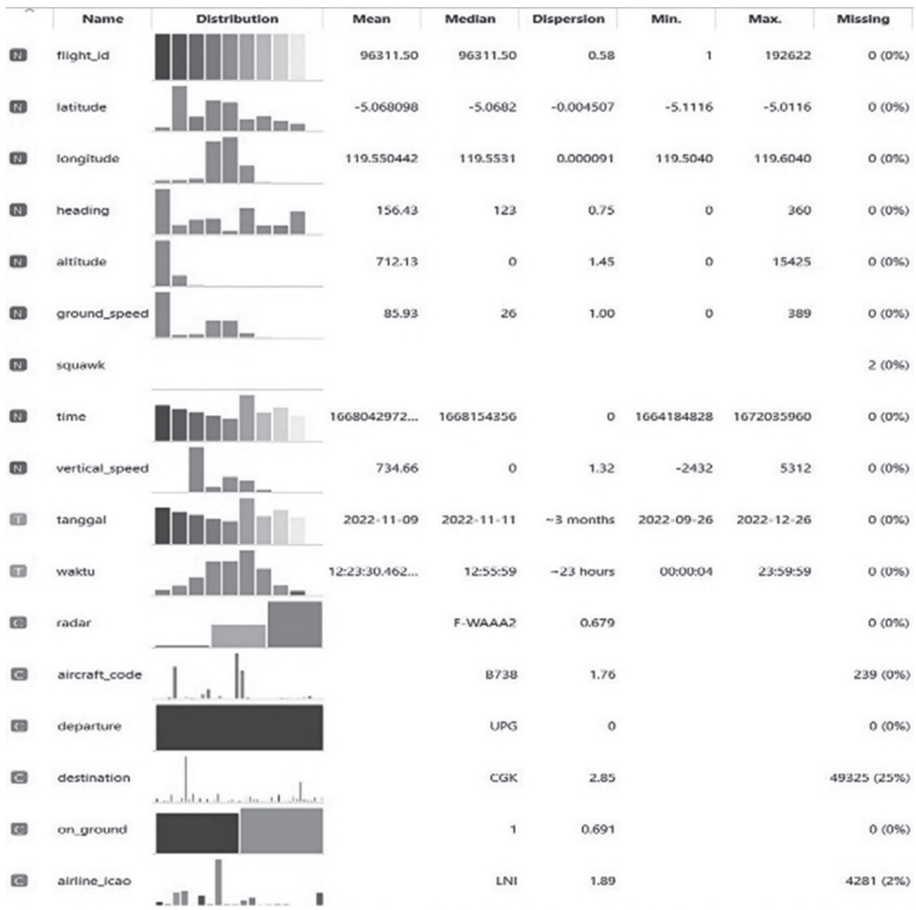
Eighthly, the number of groups is optimized using the elbow method, which helps determine the appropriate number of clusters. Furthermore, the cluster results are examined to uncover the reasons behind the organization of flight data into groups and to determine if these groupings can be utilized to draw conclusions or identify patterns.

RESULTS AND DISCUSSION

In this section, the process of pre-processing raw data to create an analysis dataset has been outlined. This process involved employing specific methods to extract two data samples representing the aircraft that have the highest frequency of takeoff flight activities at Sultan Hasanuddin International Airport. Subsequently, clustering techniques were applied to these two data samples in order to uncover patterns and gain insights from the dataset.

The process of converting raw data into a dataset, which can be used in research. At the outset of the research, the aircraft data recording process was conducted using the AERO-TRACK application *(Yousnaidi et al., 2023)*.

Additionally, data were extracted from the AERO-TRACK database. Following the extraction, a series of pre-processing steps were conducted to examine the data distribution using Orange. This involved employing Python for data transformation, reduction, cleaning, and filling. The data were then visualized using Orange, highlighting statistical features and focusing on the identification and correction of missing values. The outcome of applying Orange to the raw data is presented in Figure 2.



	Name	Distribution	Mean	Median	Dispersion	Min.	Max.	Missing
N	flight_id		96311.50	96311.50	0.58	1	192622	0 (0%)
N	latitude		-5.068098	-5.0682	-0.004507	-5.1116	-5.0116	0 (0%)
N	longitude		119.550442	119.5531	0.000091	119.5040	119.6040	0 (0%)
N	heading		156.43	123	0.75	0	360	0 (0%)
N	altitude		712.13	0	1.45	0	15425	0 (0%)
N	ground_speed		85.93	26	1.00	0	389	0 (0%)
N	squawk							2 (0%)
N	time		1668042972...	1668154356	0	1664184828	1672035960	0 (0%)
N	vertical_speed		734.66	0	1.32	-2432	5312	0 (0%)
T	tanggal		2022-11-09	2022-11-11	~3 months	2022-09-26	2022-12-26	0 (0%)
T	waktu		12:23:30.462...	12:55:59	~23 hours	00:00:04	23:59:59	0 (0%)
C	radar			F-WAAA2	0.679			0 (0%)
C	aircraft_code			B738	1.76			239 (0%)
C	departure			UPG	0			0 (0%)
C	destination			CGK	2.85			49325 (25%)
C	on_ground			1	0.691			0 (0%)
C	airline_icao			LNI	1.89			4281 (2%)

Figure 2. Statistic feature from the orange data mining program

Slika 2. Statistička značajka iz narančastog programa za rudarenje podataka

Before reconstructing the missing value data, we sorted the date attribute, which had previously been split from DateTime into a standalone date format. Then, we proceeded to individually select the Day and icao24 attributes. We conducted sorting and selection to prevent recording overlapping flights. Subsequently, we filtered the data based on an altitude attribute value of zero and above, indicating that the data represented moments before and after each aircraft's takeoff. We repeated this process until we reached the end of the selected data row. Throughout this procedure, we identified incomplete flight data, including instances where altitude values were either missing or only zero.

The subsequent step involved reorganizing the flight database into a new data table, enabling the observation of the aircraft type and the frequency of takeoff flights from Sultan Hasanuddin International Airport. The outcomes of this data grouping process are presented in Table 2.

Table 2 explains that only 3250 of the 4550 flights have complete data. After obtaining data before and after take-off, data quality calculations were performed by calculating the timestamp difference in the time attribute by reducing the take-off data time on the second line of one flight with the previous line. The result of this calculation is the data quality value. This study relied on less than 10 seconds of data, or "tier 1."

Table 2. Data Calculation Results
Tablica 2. Rezultati izračuna podataka

No	Type of Aircraft	Manufacturer	Number of Flight	%
1.	A320	Airbus	797	24.52
2.	A20N	Airbus	29	0.89
3.	A333	Airbus	1	0.03
4.	A339	Airbus	31	0.95
5.	AT45	ATR	8	0.25
6.	AT75	ATR	94	2.89
7.	AT76	ATR	219	6.74
8.	B733	Boeing	76	2.34
9.	B735	Boeing	1	0.03
10.	B738	Boeing	1202	36.98
11.	B739	Boeing	767	23.60
12	BE20	Hawker Beechcraft	5	0.15
13	BE40	Hawker Beechcraft	1	0.03
14	C212	CASA	2	0.06
15	CRJ2	CANADAIR	2	0.06%
16	E35L	Embraer	1	0.03
17	GL7T	Bombardier Aviation	1	0.03
18	GLEX	Bombardier Aviation	2	0.06
19	L410	Let Kunovice	4	0.12
20	RJ85	British Aerospace	2	0.06
21	Unknown	Unknown	5	0.15
Total			3250	100%

Table 3. Data Quality Calculation using Complete Data
Tablica 3. Izračun kvalitete podataka korištenjem potpunih podataka

Type of Aircraft	Quality Tier 1	Quality Tier >1	Total Data
A320	278	519	797
A20N	4	25	29
A333	-	1	1
A339	4	27	31
AT45	-	8	8
AT75	20	74	94
AT76	31	188	219
B733	28	48	76
B735	1	-	1
B738	480	722	1.202
B739	240	527	767
BE20	1	4	5
BE40	-	1	1
C212	2	-	2
CRJ2	-	2	2
E35L	-	1	1
GL7T	1	-	1
GLEX	-	2	2
L410	-	4	4
RJ85	-	2	2
Unknown	1	4	5
Total Data	1.091	2.159	3.250
Percentage	33.57%	66.43%	100%

Table 3 presents the results of data grouping based on data quality, using the updated data from 3,250 flights. According to Table 3, out of the 3,250 flights with complete data, 1,091 flights, or 33.57 percent, have data categorized as tier 1 data quality. Among the Boeing (B738) aircraft type departures from Sultan Hasanuddin Airport, 480 flights (40 percent of the total B738 departures) have completed and up-to-date data falling under tier 1 quality. The Airbus (A320) series ranks second, with 278 out of 797 total flight data (35 percent) meeting the tier 1 quality criteria.

Furthermore, data grouping based on the runway was conducted by rounding off the aircraft heading attribute to match the runway number. Sultan Hasanuddin International Airport has two runways, namely runways 03 and 13, as well as runways 21 and 31, each having two landing or take-off angles. The number of commercial aircraft departures for each runway angle was determined using engineering data. Table 4 displays the departure data collected during the observation period, categorized by runway numbers for the two observed aircraft types (B738 and A320). In Table 4, it can be observed that 628 flights departed from runway 03, while there were no flights from runway 13. Additionally, there were 96 flights from runway 21, 33 flights from runway 31, and 1 flight for which the runway could not be determined. Based on this data, runway 03 was selected for further observation, as it had a sufficient amount of data, accounting for 82.84 percent of the total data.

Table 4. Runway Data Separation by Aircraft Type
Tablica 4. Razdvajanje podataka o pisti prema vrsti zrakoplova

Aircraft	Runway					Total Number of Flight
	03	13	21	31	Error	
A320	248	-	22	7	1	278
B738	380	-	74	26	-	480
Grand Total	628	-	96	33	1	758

Furthermore, the data was added to the Haversine formula to accommodate the Euclidean Distance (ED) value, which calculates two coordinate points using the Haversine formula *(Robusto, 1957)*. The calculations have been performed to determine the aircraft's distance from the runway's start until the aircraft take-off.

To provide distance information between these two points, the calculation requires details regarding the aircraft's longitude and latitude as well as the runway's starting point. Based on longitude and latitude, the Haversine Formula calculates a wide circumference distance (radius) between two points on the surface of the sphere (Earth). The Haversine Formula is the correct formula for calculating the distance between two places, given their latitude and longitude *(Soe et.al., 2020)*. The Haversine formula is as follows:

$$\Delta lat(radian) = (lat_2 - lat_1) \cdot (\frac{\pi}{180}) \quad [1]$$

$$\Delta long(radian) = (long_2 - long_1) \cdot (\frac{\pi}{180}) \quad [2]$$

$$a = sin^2 \left(\frac{\Delta lat(radian)}{2}\right) + cos \left(lat_1 \cdot (\frac{\pi}{180})\right) \cdot cos \left(lat_2 \cdot (\frac{\pi}{180})\right) \cdot sin^2 \cdot \left(\frac{\Delta long(radian)}{2}\right) \quad [3]$$

$$c = 2.atan2 (\sqrt{a}, \sqrt{(1-a)}) \quad [4]$$

$$d = R \cdot c \quad [5]$$

Where: Δlat is latitude, $\Delta long$ is longitude, R is the radius of the earth, which is 6371e3(m), c = axis intersection point, d = distance (meters) and 1 = 0.0174532925 radian. At this stage, the K-Means algorithm was implemented by using the Orange application *(Demšar et al., 2013)* and determining the best K value using the elbow method *(Cui, 2020, Bholowalia et al., 2014)*. Before implementing K-Means, the best K value (centroid) was determined first *(Kodinariya et al., 2013, Pham et al., 2005)*, which clusters ten times with an increase in a class by one time in each experiment. The elbow experiment was then visualized in Figure 3. Based on the elbow technique visualization findings, the elbow angle was calculated at the point of the number of clusters (X-axis), namely 3 (as shown in Figure 3). This value indicated the best K value to be implemented in the K-Means algorithm on the dataset used.

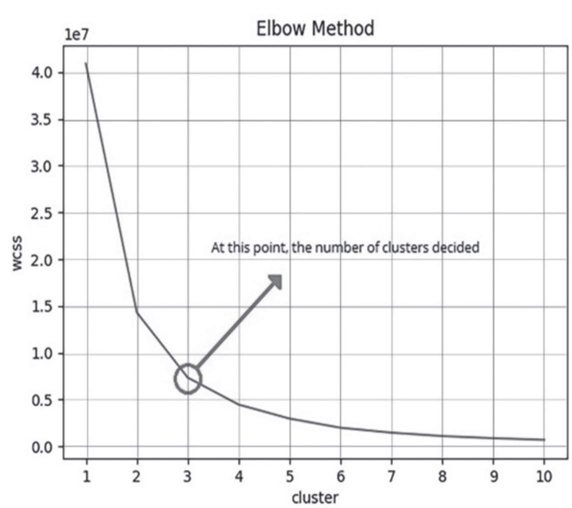


Figure 3. Line plotting elbow method
Slika 3. Metoda lakta za crtanje linija

The attributes used as features in the K-Means implementation are ground speed, aircraft code, and Haversine. Other attributes, on the other hand, were only used as metadata. The value of K was then calculated based on the results of determining the best K value, namely 3, and determining ten trials with 300 iterations.

After the K-Means algorithm was applied, the data was tabulated to see the data distribution so that the results of the formed clusters could be analyzed (Table 5).

Table 5. Number of Cluster Result Data

Tablica 5. Broj podataka rezultata klastera

Aircraft Type	C1	C2	C3
A320	128	33	87
B738	174	112	94
Total	302	145	181

Table 5 illustrates the distribution of data clusters for the two types of sample aircraft as follows: The A320 aircraft type data is distributed across 128 flights in Cluster 1 (C1), 33 flights in Cluster 2 (C2), and 87 flights in Cluster 3 (C3). In contrast, the B738 aircraft type is represented by 174 flights in Cluster 1 (C1), 112 flights in Cluster 2 (C2), and 94 flights in Cluster 3 (C3). The total number of take-offs (including both Boeing and Airbus) in each cluster is 302 in Cluster 1, 145 in Cluster 2, and 181 in Cluster 3. These figures provide insights into the distribution of flights within each cluster for the respective aircraft types. Additionally, Figure 4 visually presents the results of the cluster separation for better comprehension.

Based on the analysis in Figure 4, the haversine data shows the division of clusters into 3, with each cluster illustrated in Figure 5. To avoid runway overshoot at the runway's end, aircraft should take off between one-third and two-thirds of the way down the runway. According to this assumption, there are 628 A320 aircraft flights with 248 flights, or 40%, and 628 B738 aircraft flights with 60%. When looking at the A320 aircraft, it reveals that 87% of the aircraft take off in the safe zone (C1 and C3 zones), while the B738 aircraft indicates that 70.5% of the aircraft take off in the safe zone. Thus, 91 days of data show that A320 aircraft take off in a safe zone better than B738 aircraft.

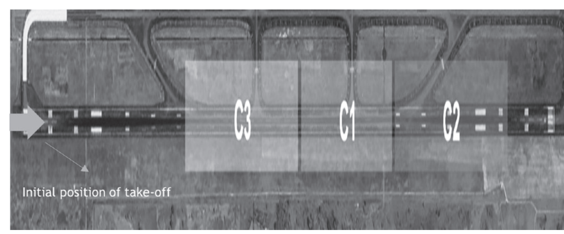


Figure 5. Runway Sultan Hasanuddin International Airport

Slika 5. Pista međunarodne zračne luke Sultan Hasanuddin

The results of data distribution show that the A320 and B738 aircraft types have a fairly contrasting amount of data. This is the guiding principle for limiting the analysis of aircraft types in the influence of data abnormalities on flights, especially in the take-off phase. It is also used to ascertain that data abnormality's influence is due to other factors.

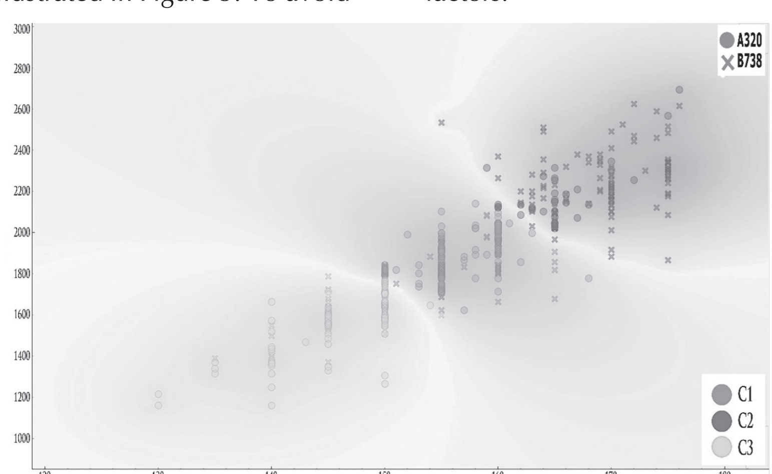


Figure 4. Scatter plot according to cluster data

Slika 4. Dijagram raspršenosti prema podacima klastera

To examine clustering outcomes beyond K means, modified K means, such as spherical K means and Gaussian mixture modeling (GMM), were employed. The resulting analysis shows that certain data points from the aircraft flight dataset underwent a centroid shift in each clustering algorithm. Figure 6 exhibits the findings of this clustering operation. Generally, for cluster 3, the three clustering algorithms illustrate that both aircraft types take off within a haversine distance of 1100 meters from the beginning of the runway. Furthermore, for cluster 2, both K-algorithms, means and spherical K-means, show the same thing as in cluster 1. Namely, both aircraft types display the same dominant takeoff. However, the GMM algorithm highlights the opposite; that is, the B738 aircraft is more dominant in taking off in this cluster compared to the A320.

Meanwhile, in cluster 2, which is located near the end of the runway, data from clustering results for both types of aircraft using three algorithms shows that the B738 aircraft has a very high percentage compared to the A320. In other words, the clustering results show that A320 aircraft were found to rarely take off at distances above 2000 m from the base of the runway.

The explanation of these three algorithms is: Spherical K-Means is a variation of K-Means that assumes uniform vector length in each cluster. Normalizing the vector length before executing the clustering process achieves this. The differences between spherical K-means and K-Means comprise better accuracy for spherical data by the former. Nevertheless, spherical K-means is more sensitive to outliers than K-Means. So, if dealing with spherical data with outliers, spherical K-means is the better choice. Meanwhile, GMM is a clustering algorithm that uses a probability model to divide data into clusters. GMM assumes that the data in each cluster follows a normal distribution.

The results obtained from this study show the same results as Narcizo et al *(2020)*, namely that Airbus aircraft take off on average at a distance of less than 1500 meters, while Boeing averages between 1500 and 2500 meters. This is influenced by many factors, including weight at takeoff, aircraft design, and the type of engine used. Thus, this study reveals that the behavior of commercial aircraft take-off patterns in Indonesia, namely Sultan Hasanuddin International Airport, has the same results as commercial aircraft in Brazil.

CONCLUSION

The data analysis reveals significant findings. Firstly, out of the 4,550 flights observed over 91 days, only 628 flights (14%) of A320 and B738 data were used, considering ADS-B data quality. This indicates that the analysis covers a mere 14% of the entire observation period. Secondly, the dataset of 628 flights is divided into three classes using the elbow method, a standard model evaluation technique. Class C1 has the highest data distribution, consisting of 302 flights (128 A320, 174 B738) representing aircraft taking off at around two-thirds of the runway length. Class C3, the second-largest dataset, includes aircraft taking off from the runway's base (1/3 of the runway), comprising 181 flights (87 A320, 94 B738). Class C2 represents aircraft taking off at three-thirds of the runway or at the runway's end, totaling 145 flights (33 A320, 112 B738). Lastly, the analysis indicates that 87% of A320 flights and 70.5% of B738 flights take off within the safe zone.

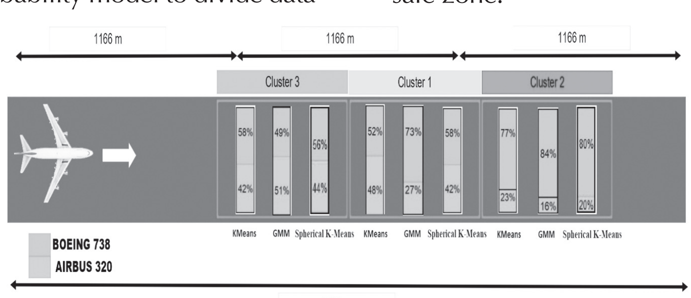


Figure 6. Comparison results of the K-means type algorithm for each cluster for B738 and A320 aircraft
Slika 6. Rezultati usporedbe algoritma tipa K-means za svaki klaster za zrakoplove B738 i A320

ACKNOWLEDGEMENT

Thank you to everyone who cooperated with this research so that we can examine the movement of aircraft take-off more clearly and learn how to avoid plane crashes in the future.

CREDIT AUTHOR STATEMENT

R. Kurniati: conceptualization, methodology, investigation.

R. Passarella: supervision, reviewing data curation, validation, Corresponding.

I. G. Afriansyah: writing-original draft preparation, investigation.

O. Arsalan: formal analysis.

A. Aditya: investigation.

M. R. Fathan: investigation, software.

R. S. Yousnaidi: investigation.

H. Veny: data curation, reviewing and editing, formal analysis.

LITERATURE

Airbus Accident Statistics: *Accident by Flight Phase-Most of the accidents over the last 20 years occurred during approach and landing phases.* Accessible at: https://accidentstats.airbus.com/statistics/accident-by-flight-phase. Accessed: 4.1.2023

Bholowalia, P., Kumar, A.: EBK-means: A clustering technique based on elbow method and k-means in WSN, *International Journal of Computer Applications*, 9, 2014, 105, 17-24.

Chang, Y.-H., Yang, H.-H., Hsiao, Y.-J.: Human risk factors associated with pilots in runway excursions, *Accident Analysis & Prevention*, 94, 2016, 227-237.

Chouinard, J. C.: *What is K Means Clustering Algorithm (with Example) – Python*. 2023. Accessible at: https://www.jcchouinard.com/kmeans/. Accessed : 12.6.2023

Cui, M.: Introduction to the k-means clustering algorithm based on the elbow method, *Accounting, Auditing and Finance*, 1, 2020.

Demšar, J., Curk, T., Erjavec, A., Gorup, C., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J., Žitnik, M., Zupan, B.: Orange: Data Mining Toolbox in Python, *Journal of Machine Learning Research*, 14, 2013, 71, 2349–2353.

Distefano, N., Leonardi, S.: *Investigation of the causes of runway excursions, Transport Infrastructure and Systems. 1. ed. Rome, Italy:* CRC Press, 2017, 227–237.

Federal Aviation Administration: Statement of Policy for Authorizations to Operators of Aircraft That are Not Equipped With Automatic Dependent Surveillance-Broadcast (ADS-B) Out Equipment. Accessible at: https://www.govinfo.gov/content/pkg/FR-2019-04-01/pdf/2019-06184.pdf. Accessed: 7.1.2023

Hernik, J., Mazur, R.: Innovations and safety in passenger railway transport - the travelers' perspective. Economic and Social Development. *33rd International Scientific Conference on Economic and Social Development – "Managerial Issues in Modern Business".* Warsaw, 26-27 September 2018, 71-81.

Huang, C.: Further Improving General Aviation Flight Safety: Analysis of Aircraft Accidents During Takeoff, *Collegiate Aviation Review International*, 38, 2020, 1, 88-105

ICAO Asia and Pacific Office: ADS-B Implementation and Operations Guidance Document. 14 ed. Montreal, CA: ICAO, 2014.

Jenkins, M., Aaron, R. F.: *Reducing Runway Landing Overruns.* Accessible at: https://www.boeing.com/commercial/aeromagazine/articles/2012_q3/3/. Accessed : 3.1.2023

Jun, Z., Wei, L., Yanbo, Z.: Study of ADS-B Data Evaluation, *Chinese Journal of Aeronautics,* 24, 2011, 4, 461-466.

Kodinariya, T. M., Makwana, P. R.: Review on determining number of Cluster in K-Means Clustering, *International Journal of advance Research in Computer Science and Management Studies*, 1, 2013, 6, 90-95.

Komite Nasional Keselamatan Transportasi: Aircraft Accident Investigation Report

KNKT.21.03.05.04, 2021. Accessible at: https://knkt.go.id/Repo/Files/Laporan/Penerbangan/2021/KNKT.21.03.05.04-Preliminary-Report.pdf. Accessed: 19.2.2023

Komite Nasional Keselamatan Transportasi: Aircraft Accident Investigation Report KNKT.20.09.19.04, 2020. Accessible at: https://knkt.go.id/Repo/Files/Laporan/Penerbangan/2020/PK-LEG%20Preliminary%20Report.pdf Accessed: 19.2.2023

Komite Nasional Keselamatan Transportasi: Aircraft Accident Investigation Report KNKT.17.05.19.04, 2017. Accessible at: https://knkt.go.id/Repo/Files/Laporan/Penerbangan/2017/9M-MXH%20Preliminary%20Report.pdf. Accessed: 19.2.2023

Muhammad, R. F., Aditya, A., Indra, G. A, Rani, S. Y., Rossi, P., Osvari, A., Rizki, K., Marsella, V.: Aero-Track: Perangkat Lunak Perekam Data Penerbangan Aeronautika, *Generic-Journal Ilmu Komputer dan Teknologi Informasi*, 15, 2023, 1, 15-19.

Narcizo, R., Alves, C.J., Caetano, M.: The Aircraft Choice Based on the Aircraft Take-Off Runway Length Requirement. *J Aerosp Technol Manag*, 12, 2020, e3120. https://doi.org/10.5028/jatm.v12.1167

Norman, J.: Flight Data Management and Pilot Protection in an ADS-B, *Aviation Faculty Publications*, 7, 2021, 33-35.

Passarella, R., Nurmaini, S.: Data Analysis Investigation: Papua is The Most Unsafe Province in Indonesia for Aviation: An Exploratory Data Analysis Study from KNKT-Database Accidents and Incidents (1988-2021), *Journal of Engineering Science and Technology Review*, 15, 2022, 3, 158-164.

Passarella, R., Veny, H., Fachrurrozi, M., Samsuryadi, S., Vindriani, M: Evaluating the Influence of the International Civil Aviation Organization on Aircraft Accident Rates and Fatalities: A Seven-Decade Historical Data Analysis. *Acadlore Trans. Appl Math. Stat.*, 1, 2023a, 1, 33-43.

Passarella, R., Nurmaini, S., Rachmatullah, M. N., Arsalan, O., Kurniati, R., Aditya, A., Afriansyah, I. G., Fathan, R., Yousnaidi, R. S., Veny, H.: Data analysis of commercial aircraft landing on the runway airports in Indonesia, *Scientific Journal of Silesian University of Technology. Series Transport*, 120, 2023b, 233-247.

Pham, D. T., Dimov, S. S., Nguyen, C. D.: Selection of K in K-means clustering, *Proc. IMechE Part C: J. Mechanical Engineering Science*, 219, 2005, 1, 103-119.

Robusto, C. C.: The Cosine-Haversine Formula, *American Mathematical Monthly*, 64, 1957, 38.

Sandhyavitri, A., Tjahjono, T., Khairumusa, A. R.: Identifikasi Tingkat Kerawanan Bandar Udara Di Indonesia, *Journal Transportasi*, 14, 2014, 1, 43-52.

Saputra, A. D.: Studi Analisis Penyebab Runway Excursion di Indonesia Berdasarkan Data Komite Nasional Keselamatan Transportasi (KNKT) Tahun 2007-2016, *Warta Ardhia*, 43, 2017, 2, 93-104.

Šebjan, U., Tominc, P., Širec, K: Multiple Conceptual Modelling of Perceived Quality of In-fight Airline Services, *Promet – Traffc & Transportation*, 29, 2017, 3, 311-319.

Soe, N. C., Thein, T. L. L.: Haversine Formula and RPA Algorithm for Navigation System, *International Journal of Data Science and Analysis*, 6, 2020, 1, 32-40.

Šrekl, J.: Occupational Safety Modelling, *Sigurnost*, 64, 2022, 2, 135 – 142.

Yousnaidi, R. S., Passarella, R., Kurniati, R., Asrsalan, O., Aditya, Afriansyah, I. G., Fathan, M. R., Vindriani, M.: Assessing Automatic Dependent Surveillance-Broadcast Signal Quality for Airplane Departure Using Random Forest Algorithm, *Mechatronics and Intelligent Transportation Systems*, 2, 2023, 2, 64-71.

Zhao, W., Liu, D.: ADS-B data source availability verification based on accuracy, *International Conference on Big Data and Informatization Education (ICBDIE)*, Zhangjiajie, China, 2020, 387-390.

NENADZIRANA PRAKTIČNA ANALIZA PONAŠANJA AVIONA A320 I B738 PROVEDENA U MEĐUNARODNOJ ZRAČNOJ LUCI SULTAN HASANUDDIN

SAŽETAK: Cilj istraživanja bio je motriti ponašanje dvaju dobro poznatih komercijalnih tipova aviona (A320 i B738) tijekom faze polijetanja. Ovo je način dobivanja novih saznanja na području avijacije, naročito u pitanjima sigurnosti leta. Motrenja su provedena u zračnoj luci Sultan Hasanuddin praćenjem ABDS-B podataka koji definiraju promjene u letu. ADS-B podaci podliježu analizi podataka koji će se onda ubaciti u računalo kako bi ono prepoznalo uzorke i sastavilo klastere. Svrha ispitivanja je koristiti nenadzirano učenje, posebno K-Means klastere, kategorizirati i utvrditi uzorke ponašanja u neoznačenim ADS-B podacima dobivenim iz AERO-TRACK-a. U pripremi sirovih podataka i skupa podataka korištene su tehnike za analizu podataka. Strojni model generira tri različita klastera: klaster 1 predstavlja polijetanje aviona na dvije trećine piste, klaster 2 na cijeloj pisti, a klaster 3 polijetanje na jednoj trećini piste. Elbow metoda koristi se za analizu i interpretaciju triju klastera što proizlaze iz modela. Zanimljivo je primijetiti da B738 dominira u sva tri klastera, dok A320 dominira u klasterima 1 i 3. Zanimljivo je da se znatan broj polijetanja nalazi u klasteru 2, tj. 145 od 628 letova. Na temelju podataka u 91 dan (26. rujna do 26. prosinca, 2022.) vjerojatnost izlijetanja s piste u ovom klasteru iznosi 23 %. Nadalje, istraživanje otkriva da A320 ima sigurnu zonu polijetanja od 87 %, dok B738 ima samo 70.5 %. Ovi nalazi, dobiveni analizom ADS-B podataka, npr. GPS-altituda i koordinata trebaju poslužiti kao važna saznanja upravama letenja, korisnicima letova i drugim sudionicima u avijaciji.

Ključne riječi: *zračna luka, ADS-B podaci, klaster, pista, K-Means*

Izvorni znanstveni rad
Primljeno: 26.6.2023.
Prihvaćeno: 10.12.2023.