

OUTLIER DETECTION PADA SET DATA FLIGHT RECORDING (PRE-PROCESSING SUMBER DATA ADS-B)

Mohammad Yazdi Pusadan¹⁾

¹⁾ Teknik Informatika Institut Teknologi Sepuluh Nopember
Kampus ITS Sukolilo Surabaya, 60111
Email : yazdi.diyana@gmail.com¹⁾

Abstrak

Outlier detection (deteksi anomali) merupakan bagian dari data mining untuk menentukan suatu data/objek yang memiliki perbedaan karakteristik dengan set data lainnya (exception mining). Pada kasus data penerbangan (flight data), outlier diperlukan untuk mengidentifikasi permasalahan yang terjadi pada jalur penerbangan yang diperoleh dari sistem navigasi.

Data navigasi penerbangan yang digunakan pada penelitian ini bersumber dari ADS-B (Automatic Dependent Surveillance Broadcasting) merupakan sistem navigasi dalam dunia penerbangan yang dengan frekuensinya dapat di deteksi dengan menampilkannya dalam bentuk text, visual 2D dan 3D. Studi analisis terhadap data navigasi penerbangan tersebut digunakan untuk mendeteksi permasalahan penerbangan. Dengan melakukan pengelompokkan (clustering) data serta metode deteksi berbasis jarak (distance) dan kepadatan (density) dapat dihasilkan identifikasi anomali yang terjadi. Hasil yang diperoleh dari penelitian ini adalah frame deteksi outlier pada data penerbangan yang selanjutnya dapat dikembangkan.

Kata kunci: Deteksi Outlier, Data Mining, ADS-B..

1. Pendahuluan

Dalam kurun waktu 2007-2012 menurut data dari Komisi Nasional Keselamatan Transportasi (KNKT) Indonesia bahwa tingkat kecelakaan transportasi udara mengalami perhatian, walaupun sudah mengalami penurunan namun ditahun tertentu terjadi peningkatan. Berikut pada Gambar 1 data kecelakaan pesawat tahun 2007-2012 yang bersumber dari KNKT tahun 2011.



Gambar 1. Data Kecelakaan Pesawat (2007-2012)
Sumber: KNKT Indonesia 2011

Sampai sekarang, penyelesaian kasus penerbangan dilakukan dengan identifikasi berdasarkan analisis data

penerbangan yang bersumber dari *Flight Recorder* (kotak hitam). Hal tersebut belum optimal, disebabkan data kotak hitam hanya merekam pembicara pilot dengan menara bandara (*Air Traffic Controller/ATC*) bukan data jalur penerbangan dan identifikasi kotak hitam dilakukan setelah kecelakaan terjadi.

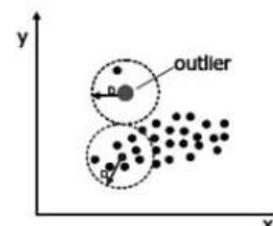
Berdasarkan hal tersebut, maka pada penelitian ini diharapkan dapat memberikan kontribusi untuk menyelesaikan permasalahan penerbangan melalui data penerbangan kompleks berbasis ADS-B *real-time*.

Masalah yang dijukan pada penelitian adalah mendeteksi permasalahan pada penerbangan dengan data set bersumber dari sistem ADS-B secara *real-time*. Sehingga diharapkan dapat menemukan penyelesaian baru terhadap kasus penerbangan dengan melakukan riset eksplorasi tentang data penerbangan secara komputasi (*soft computing*) berbasis *real-time* serta deteksi permasalahan dini (*early warning system*)

Penelitian ini bersumber dari beberapa referensi pendukung, diantaranya:

Deteksi Outlier dan Deteksi Outlier pada Penerbangan

Metode deteksi *outlier* dengan pendekatan jarak (*distance*) yang disebut dengan *Distance-Based Outlier Detection (DB-Outlier Detection)* merupakan suatu metode pendeteksian *outlier* dengan cara menghitung jarak antara titik data (*data points*). Pada gambar 2 menunjukkan metode deketsi *outlier*/anomali dengan pendekatan jarak (*distance*) data satu dengan lainnya.



Gambar 2. Deteksi Outlier pendekatan jarak (*distance*)

Pada deteksi DB-Outlier yang berada pusat *Nearest Neighborhood*. Pada saat itu ditemukan lebih dari titik data *M* yang berada pada objek *D-Neighborhood*, maka objek tersebut bukan *outlier*. Dataset yang digunakan dengan karakteristik data berdimensi tinggi [1].

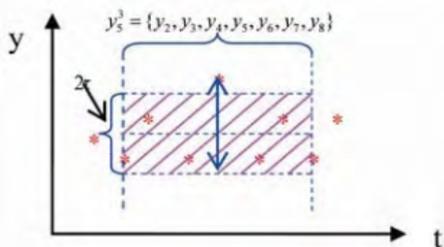
Beberapa metode yang digunakan pada deteksi *outlier* diantaranya: metode penentuan *outlier* berdasarkan

cluster dengan dua tahapan (*Two-stage Outlier Detection/TOD*), yaitu: mengelompokkan data set pada beberapa cluster yang memiliki kemiripan dan menentukan outlier pada cluster tersebut berdasarkan outlier factor/Cluster-Based Local Outlier Factor (*CBLOF*). CBLOF berhubungan dengan ukuran dataset, jumlah atribut, dan jumlah cluster[2].

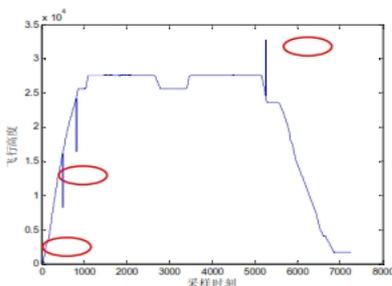
Terdapat beberapa tipe outlier pada data penerbangan, yaitu: i). kesalahan yang terjadi di luar sisten FDR(data penerbangan) diantaranya kerusakan mesin dan human error; ii). noise atau gangguan lain dari luar sistem; iii). perbedaan fase penerbangan yang berubah selama penerbangan, misal: dari posisi pesawat menanjak(*climb*) ke posisi meluncur(*cruise*). Dari semua tipe tersebut yang penting adalah menemukan jenis outlier yang ditentukan dari FDR dengan menganalisis berdasarkan ketinggian(*altitude*) yang ditentukan melalui lima tahapan, yaitu: *taking-off*, *climb*, *cruise*, *approach*, dan *landing*. Metode yang digunakan adalah **two-side median filtering** dengan menghitung median m_t^k dari local data point's(signal altitude) y_t , kemudian dibandingkan spesifik threshold τ [3].

$$\begin{aligned} &\text{if } |y_t - m_t^k| < \tau, \text{ keeps } y_t; \\ &\text{else } |y_t - m_t^k| \geq \tau, y_t \text{ as outlier and replaced } y_t \text{ with } m_t^k \end{aligned} \quad \dots\dots\dots(1)$$

Pada gambar 3 menunjukkan tahapan awal dalam penentuan outlier dengan melakukan metode *module two-side filtering* terlebih dahulu. Selanjutnya pada gambar 4 hasil penentuan outlier berdasarkan data signal dari posisi ketinggian(*altitude*) pesawat pada detik kecepatan 508km/jam, 826km/jam, dan 5254 km/jam



Gambar 3. Metode module of two-side filtering



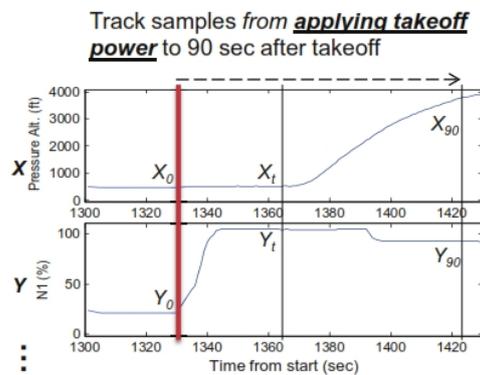
Gambar 4. Penentuan Outlier berdasarkan signal of altitude pada (508km/jam, 826km/jam, 5254 km/jam)

Pendeteksian outlier berbasis analisis cluster dengan menggunakan algoritma “*hierarchical agglomerative clustering*”. Pada metode ini: i). ditentukan terlebih

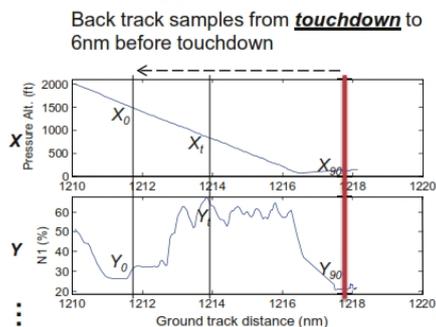
dahulu sampel data yang bersumber dari data asal; ii) pada sampe data tersebut, diperoleh clustering data berdasarkan algoritma “*hierarchical agglomerative clustering*”; iii). perhatikan hasil clustering yang ditentukan sebagai sampel akurat, proses ini jika ditemukan kejanggalan berupa noise yang nilainya lebih kecil dari δ , maka dikeluarkan; dan iv) dengan pendekatan matematika μ dan standard deviasi σ diperoleh sampai yang akurat[4]. Untuk penentuan outlier:

$$\mu = E(X) \text{ dan } \sigma = \sqrt{D(X)} \quad \dots\dots\dots(2)$$

Metode lain yang dilakukan untuk deteksi outlier data penerbangan adalah dengan pendekatan clustering. Terdapat beberapa tahapan, yaitu: i). *time series to vector* (vektor pada data penerbangan (V)= $[X_0, X_1, \dots, X_{90}, Y_0, Y_1, \dots, Y_{90}, \dots]$); ii). *dimension reduction* (dengan metode PCA/*Principle Component Analysis*; dan iii). *Cluster Analysis (Cluster-based Outlier Detection/Density-Based Spatial Clustering Algorithm with Noise)*. Proses analisis outlier diawali dengan membagi spesifikasi keadaan (*specific event*) pada 2(dua) fase: fase *takeoff (takeoff phase)* dan fase mendekati *landing (approach phase)* serta terdapat 68 parameter *takeoff* dan 69 parameter *approach*[5]. Berikut karakteristik tiap fase yang ditunjukkan pada gambar 5 merupakan fase *takeoff*, dan gambar 6 adalah fase *approach*.

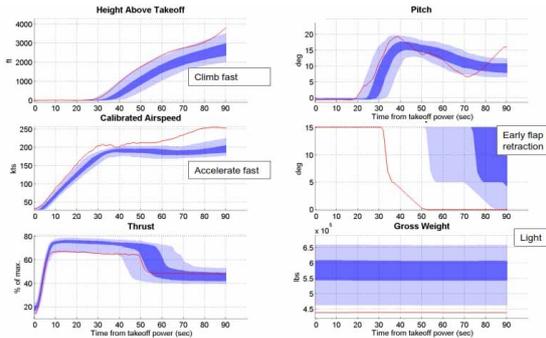


Gambar 5. Fase Takeoff

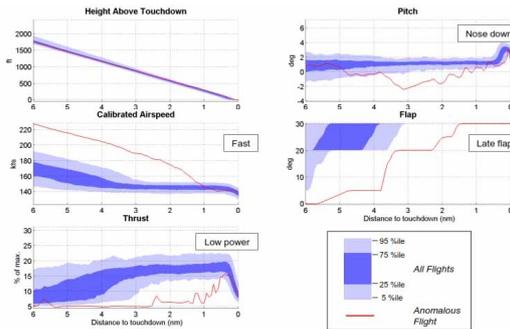


Gambar 6. Fase Approach

Dari hasil penelitian, menunjukkan anomali deteksi pada data penerbangan pada gambar 7 adalah anomali pada fase *takeoff* dan gambar 8 adalah anomali pada fase *approach*.



Gambar 7. Deteksi Anomali pada Fase Takeoff



Gambar 8. Deteksi Anomali pada Fase Approach

Deteksi outlier dengan pendekatan kepadatan (density) data, yaitu: penentuan deteksi outlier yang selama ini dilakukan oleh metode LOF (Local Outlier Factor) klasik, yaitu FastLOF[6].

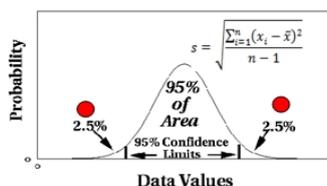
Adapun langkah-langkah yang dilakukan sbb: i). menentukan K-NN dengan formula Euclidean Distance untuk menentukan jarak tiap tetangga; ii). menggunakan formula LRD (Local Reachability Density) dengan menghitung semua data pada titik P di set data tetangga K dan $N_{min}(p)$:

$$LRD_{min}(p) = \frac{1}{\sum_{o \in N_{min}(p)} reach_dist_{min}(p,o) / |N_{min}(p)|} \dots\dots\dots(3)$$

dan iii) menentukan nilai LOF dengan menghitung LRD dari K-NN:

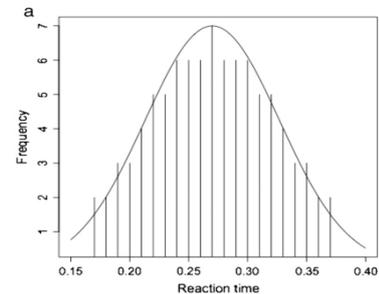
$$LOF_{min}(p) = \frac{\sum_{o \in N_{min}(p)} LRD_{min}(o)}{|N_{min}(p)|} \dots\dots\dots(4)$$

Selain itu terdapat pendekatan statistik lainnya, yaitu: dengan berbasis distribusi data (distribusi Gauss, Distribusi Poisson, dan Distribusi Gamma), parameter (mean, median, dan variance), hasilnya outlier ditemukan pada area diluar confidence limit. Berikut ini pada gambar 9 menggambarkan grafik distribusi data menggunakan standard deviasi, sehingga menunjukkan outlier.



Gambar 9. Grafik menentukan Outlier pada Distribusi Data Standard Deviasi

Selanjutnya pada gambar 10 menggambarkan penggunaan metode MAD (Media Absolute Deviation) untuk menentukan outlier. Dengan Normal distribution, $n=91$, $mean=0,27$, $median=0,27$, & $standard\ deviation=0,06$.



Gambar 10. Median Absolute Deviation (MAD)

$$MAD = b \cdot M_i(|x_i - M_j(x_j)|) \dots\dots\dots(5)$$

$b = 0,1482$ adalah konstanta pada data normal dan M_i adalah median series.

ADS-B. ADS-B adalah singkatan dari Automatic Dependent Surveillance Broadcasting yang merupakan sistem navigasi dalam dunia penerbangan yang dengan frekuensinya dapat di deteksi oleh antena receiver dengan berbagai data yang dapat ditampilkan dalam bentuk text, visual 2D dan 3D.

Data yang ditampilkan pada ADS-B adalah: Jadwal keberangkatan pesawat terbang, Flight Number, Route, Posisi pesawat lengkap dengan koordinatnya, Ketinggian pesawat / Altitude, Type pesawat dan nomor tanda registrasi pesawat, Destination, Airway route (3D), Posisi dan pergerakan pesawat real-time, Aircraft Taxing, take off, climbing, cruising, descent, landing, Weather info, Flight following (hanya satu pesawat), Flight info filter (hanya milik airline tertentu saja), dan Integrated system flight watch worldwide.

2. Pembahasan

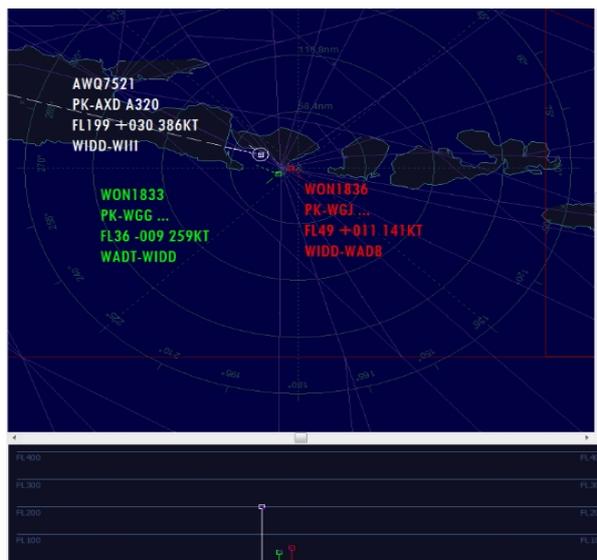
Pembahasan di paper ini ditekankan pada penjelasan tahapan sebelum proses deteksi outlier (pre-processing). Lingkup pembahasan adalah sumber data penerbangan, struktur data, karakteristik data ADSB (data script, data text, & data visual) dan mekanisme sistem data penerbangan.

Sumber data penerbangan adalah bersumber dari sistem ADS-B, telah dijelaskan sebelumnya. Struktur data dan Variabel ditunjukkan pada tabel 1 berikut ini:

Tabel 1. Struktur Data dan Variabel data ADS-B

ICAO code	Field name	Data type
CS	Callsign	String [0..8]
FS	Flight status	Integer (0..7)
CA	Capability	Integer (0..7)
LAT	Latitude	Float
LON	Longitude	Float

FL	Flight level	String[4]
AC	Altitude code	Integer (-1000..99999)
M	Metric altitude	Integer (0,1)
Q	Altitude resolution	Integer (0,1)
GNSS	GPS altitude difference	Integer (-99999..+99999)
HDG	Heading	Integer (0..359)
IAS	Indicated Air Speed (IAS)	Integer (0..999)
TAS	True Air Speed (TAS)	Integer (0..999)
GS	Ground Speed (GS)	Integer (0..999)
TIMEOUT		Integer
STATION ID		String



Gambar 12. Data Visualisasi 2D (garis putih, hijau, dan merah dibawah adalah menandakan Flight Level/FL)

Satuan pada FL adalah nm(nautical miles). 1nm sama dengan: 1.852km, 2027yard, 6076,10feet[8].

Karakteristik data ADS-B, yaitu: i) *data script*:

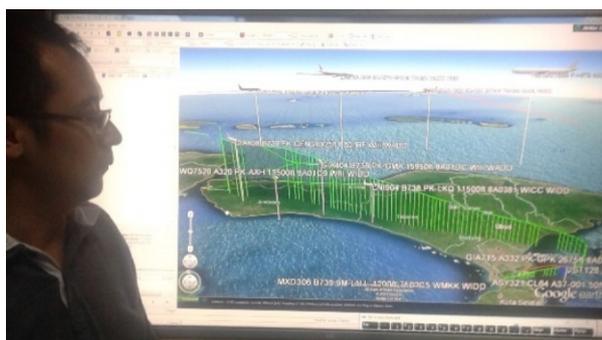
```
$PTA,20141122052905845,8A017C,,394.9,285.3,2048,
$PTA,20141122052905845,8A03CD,,4225,,,,,,,,
$PTA,20141122052905845,8A03BA,,225,,,,,,,,
$PTA,20141122052905845,8A039B,,,,,,,,-
8.8768,115.1734,,
$PTA,20141122052905845,8A039B,,4275,260.3,295.5,-
1088,,-8.8765,115.1728,,
$PTA,20141122052905845,8A039B,,4250,,,,,-
8.8762,115.1721,,
$PTA,20141122052907897,8A03BA,,275,,,,,,,,
$PTA,20141122052907897,8A03CD,,,,,,,,-
8.7535,115.2903,,
$PTA,20141122052907897,8A03CD,,4250,,,,,-
8.7537,115.2910,,
$PTA,20141122052907897,8A017C,,19000,394.7,285.1
,2176,,-8.5338,114.9197,4126,
```

....
 ii) data text yang ditampilkan pada gambar 11; iii) data visualisasi 2D yang ditunjukkan pada gambar 12; dan data visualisasi 3D yang ditunjukkan pada gambar 13. Pada paper ini data diperoleh dengan menggunakan *software AirNAV System*. Untuk data dengan visualisasi 3D adalah data *real-time* penerbangan, sedangkan untuk data 2D tidak *real-time*. Hanya pada data 2D dapat digunakan untuk *loading* data *record(data script)* penerbangan sebelumnya, sehingga sesuai untuk proses identifikasi penerbangan.



Gambar 13. Data Visualisasi 3D

Berikut ini pada gambar 14, terlihat pengambilan data 3D secara *real-time*(ADS-B Data: 23/11/2014, Start Time: 22:15:10).

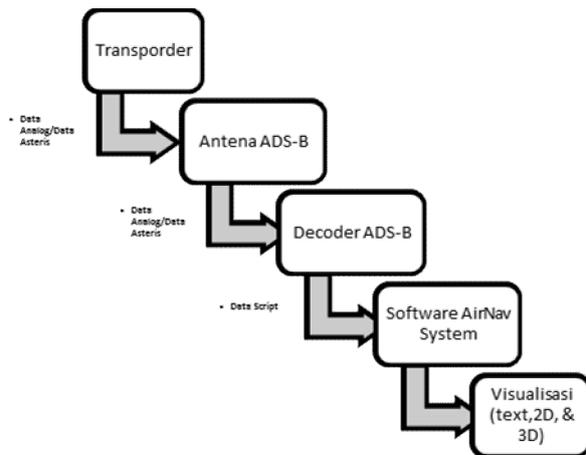


Gambar 14. Pengambilan Data *real-time* Visualisasi 3D

Selanjutnya, dijelaskan proses pengambilan data ADS-B: transponder(pilot) mengirim data berupa data analog ke ATC atau antenna ADS-B, data tersebut diteruskan ke dekoder ADS-B. Dari transmit data tersebut menghasilkan *data script*. Dari data tersebut diterjemah oleh *software AirNav System* menghasilkan data dengan visualisasi text, 2D, dan 3D. Berikut pada gambar 15 menunjukkan

Gambar 11. Data Text

tahapan pengambilan data dan visualisi berbasis sistem ADS-B



Gambar 15.Tahapan Pengambilan Data dan Visualisasi Data berbasis sistem ADS-B

Berikut terdapat perangkat yang digunakan pada proses pengambilan data:

Transponder. Transponder (*Transmitter Responder*) adalah alat yang dipasang di pesawat yang berfungsi memberikan data (posisi, ketinggian dan kecepatan pesawat) kepada radar sebagai alat bantu ATC (Air Traffic Controller) dalam memberikan pelayanan lalu lintas udara. Selain itu, transponder yang ditampilkan pada gambar 16, dapat berfungsi mengirimkan sinyal di darat saat bergerak di area bandar udara yang padat pergerakannya guna memantau posisi pesawat.



Gambar 16.Transponder

Antena ADS-B. Antena ADS-B adalah perangkat yang digunakan untuk menerima data dari *transponder*. Data yang diterima berupa data *analog/asterix*. Jangkauan data yang dapat diterima, yaitu: 150nm atau 277.800km. berikut pada gambar 17 memperlihatkan antena penerima data ADS-B dari transponder.



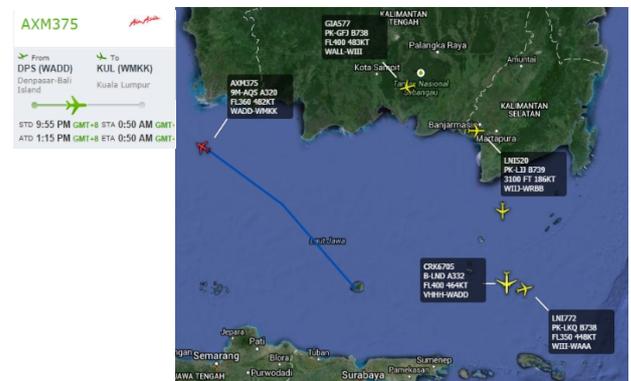
Gambar 17.Antena Penerima data ADS-B dari Transponder

Decoder ADSB. Decoder ADSB adalah perangkat yang digunakan sebagai penerima *data analog/asterix* dari antena, kemudian di konversi menjadi *data script*. Berikut pada gambar 18 memperlihatkan *Decoder ADS-B*.



Gambar 18.Decoder ADS-B

Pada tampilan data ADS-B dapat ditampilkan secara visual dan text melalui website[9], hanya terdapat beberapa kelemahan, yaitu: i) Data yang ditampilkan *delayed ± 15* menit (tidak *real-time*), sehingga tidak dapat melakukan *Recording Flight Data*, dikarenakan data bersumber dari server yang dilakukan *autosave* dari proses pengambilan data menggunakan software AirNav, dan ii) sulit melakukan identifikasi dan interpretasi data apabila terjadi permasalahan pada penerbangan (anomali deteksi). Berikut pada gambar 19 menampilkan data penerbangan berbasis *website* bersumber dari radarbox24.



Gambar 19.Tampilan Data berbasis web

Selanjutnya, untuk kebutuhan deteksi *outlier* pada penerbangan terdapat pesan emergency yang menjadi standar internasional dalam penerbangan. Pada tabel 2 berikut memperlihatkan pesan darurat/*emergency* pada penerbangan.

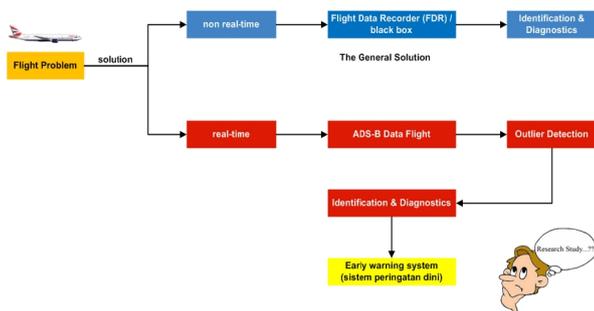
Tabel 2.Pesan Darurat/Emergency pada Penerbangan

Beacon Code	Allocated use
7500	Aircraft hijacking (ICAO, worldwide)
7600	Radio Failure (Lost Communications) (ICAO, worldwide)
7700	Emergency (ICAO, worldwide)

Sumber:

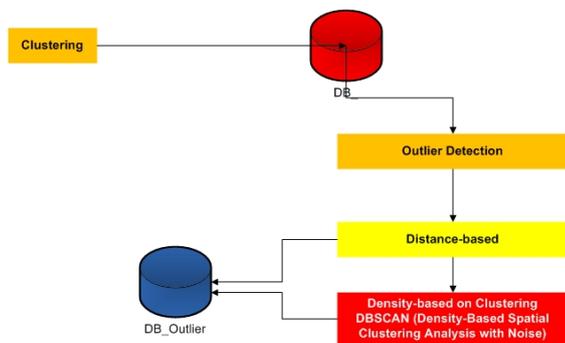
http://en.wikipedia.org/wiki/Transponder_%28aeronautics%29#cite_ref-JO711066D_15-28

Berikut ini terdapat diagram sebagai arah penelitian yang dicapai. Pada gambar 20 acuan dasar solusi masalah penerbangan sebagai *framework global*.

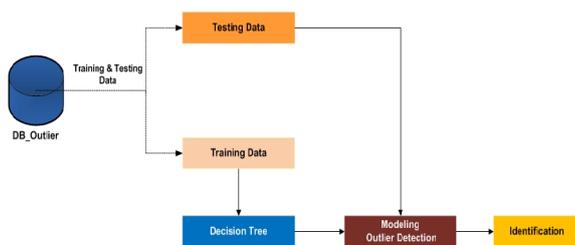


Gambar 20. Diagram Solusi Permasalahan Penerbangan

Selanjutnya pada gambar 21 merupakan diagram metode deteksi outlier sebagai acuan metode dan dilanjutkan pada model deteksi outlier sebagai identifikasi. Terdapat *training* dan *testing* data agar model yang dihasilkan akurat. Hal tersebut ditunjukkan pada gambar 22.



Gambar 21. Diagram Metode Deteksi Outlier



Gambar 22. Diagram Model Identifikasi Outlier

3. Kesimpulan

Paper ini menjelaskan tentang *pre-processing* deteksi outlier dengan pendekatan pemahaman tentang lingkup ADS-B sebagai sumber data pada identifikasi penerbangan baik secara *real-time* maupun *non real-time*. selanjutnya, dirancang *framework* untuk sebagai acuan untuk penelitian lanjutan mengenai deteksi outlier secara lengkap.

Untuk kelanjutan ada penelitian, yaitu: i) dilakukan pemahaman tentang struktur data dan variabel sumber data; ii) memahami *data script* ADS-B dan mampu melakukan konversi secara manual data script tersebut ke data text maupun visualisasi; iii) melakukan

clustering dari data penerbangan sehingga metode yang digunakan sesuai; dan iv) melakukan analisis outlier dengan identifikasi secara *real-time* maupun *non real-time*, sehingga dihasilkan model deteksi permasalahan penerbangan dini (*early warning system*).

UCAPAN TERIMA KASIH

Paper ini dibuat dengan melakukan pengambilan data di Asia Star Services, berlokasi di Jl. Kubu Anyar No. 22 Kuta Bali 80361, dengan memanfaatkan perangkat keras dan perangkat lunak untuk pengolahan data penerbangan berbasis *ADS-B System*.

Daftar Pustaka

- [1] Yuan Li¹ dan Hiroyuki Kitagawa^{1,2}. *DB-Outlier by Example in High Dimensional Dataset*. ¹Graduate School of System and Information Engineering, ²Center for Computational Science, University of Tsukuba, Tennoudai 1-1-1, Tsukuba, Ibaraki, 305-8573 Japan. IEEE 2007.
- [2] Sheng-yi Jiang^{1,2}, Qi-bo An¹. *Clustering-Based Outlier Detection Method*. School of Informatics, GuangDong University of Foreign Studies, 510006, Guangzhou, China; ²Guangdong Province Key Laboratory of Information Security, Sun Yat-sen University, Guangzhou, 510275, China. IEEE 2008.
- [3] Xiuzhen Jiao, Hui Lu, dan Rongling Lang. *One Effective Method of Outlier Detection in Flight Data*. Department of Electronic Information Engineering, Beijing University of Aeronautics and Astronautics, 100191, Beijing, China. 2009.
- [4] Yue Zhang*, Jie Liu*, Hang Li*. *An Outlier Detection Algorithm based on Clustering Analysis*. *Software College, Shenyang Normal University, Shenyang 110034, China. IEEE 2010.
- [5] Lishuai Li, Maxime Gabriel, R. John Hansman. *Anomaly Detection in Onboard-Recorded Flight Data using Cluster Analysis*. Massachusetts Institute of Technology (MIT), USA, dan Rafael Palacios dari Pontifical University, Spain. 2011.
- [6] Markus Goldstein. *FastLOF: An Expectation-Maximization based Local Outlier Detection Algorithm*. German Research Center for Artificial Intelligence (DFKI). IEEE 2012.
- [7] Christophe Leys^a, Christophe Ley^b, Olivier Klein^a, Philippe Bernard^a, Laurent Licata^a. *Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median*. ^aUniversité Libre de Bruxelles, Unité de Psychologie Sociale, Belgium, ^bUniversité Libre de Bruxelles, Département de Mathématique and ECARES, Belgium. Journal of Experimental Social Psychology. IEEE 2013.
- [8] <http://converticious.com/distance>.
- [9] <http://www.radardox24.com>

Biodata Penulis

Mohammad Yazdi Pusadan, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Matematika Program Studi Ilmu Komputer Fakultas MIPA Universitas Gadjah Mada Yogyakarta, lulus tahun 2004. Memperoleh gelar Master Engineering (M.Eng) Program Pasca Sarjana Magister Teknologi Informasi, Jurusan Teknik Elektro dan Teknologi Informasi, Fakultas Teknik Universitas Gadjah Mada Yogyakarta, lulus tahun 2011. Saat ini menjadi Dosen di Program Studi Teknik Informatika Universitas Tadulako Palu Sulawesi Tengah.