

KLASIFIKASI FASE PENERBANGAN PADA DATA ALIR AUTOMATIC DEPENDENT SURVEILLANCE-BROADCAST MULTI- RECEIVER DENGAN VARIASI FITUR

Adrianus Yoza Aprilio¹, R. V. Hari Ginardi²

^{1,2}Institut Teknologi Sepuluh Nopember

Email: yoza13@mhs.if.its.ac.id¹, hari@its.ac.id²

Abstrak. ADS-B merupakan teknologi pengawasan pesawat.. Data ini dapat digunakan untuk menentukan fase penerbangan. Fase penerbangan adalah kategorisasi pergerakan pesawat. Selama ini kategorisasi dilakukan secara manual, berdasarkan komunikasi radio analog antara pilot dan ATC. Dengan adanya data ADS-B, analisa dan klasifikasi fase penerbangan dapat dilakukan secara real-time. Tujuan penelitian ini adalah klasifikasi fase penerbangan berbasis data ADS-B. Metode yang digunakan adalah K-Nearest Neighbor dan Decision Tree. Pengujian metode akan ditunjang dengan pengujian variasi fitur dan transformasinya, termasuk penggunaan data input mentah dan data yang telah diolah, serta pengujian metode pembelajaran alternatif, sehingga hasil yang dicapai efektif untuk digunakan dalam kondisi real-time.

Kata Kunci: ADS-B, data alir, fase penerbangan, klasifikasi, kualitas data

Saat ini, transportasi merupakan salah satu kebutuhan manusia, termasuk transportasi udara. Peningkatan jumlah pengguna transportasi udara ini mulai memperlihatkan dampaknya dengan kepadatan lalu lintas di udara [1]. Kepadatan lalu lintas udara akan berefek terhadap besarnya resiko kecelakaan. Selama ini pergerakan pesawat hanya dapat dipantau oleh radar yang dimiliki oleh regulator perhubungan udara, dalam hal ini *Air Traffic Control* (ATC). Pada kenyataannya, banyak data yang disiarkan oleh pesawat dapat diakses oleh publik, salah satunya adalah *Automatic Dependent Surveillance-Broadcast* (ADS-B).

Automatic Dependent Surveillance (ADS) adalah sebuah teknologi yang diharapkan dapat menggantikan radar sebagai cara untuk melakukan *tracking* pada pesawat udara. ADS adalah mekanisme komunikasi baik antar pesawat maupun antara pesawat dengan stasiun di darat (*ground*) (FAA - Federal Aviation Administration, 2010) Data yang dikomunikasikan meliputi lalu lintas dan pergerakan pesawat, cuaca, informasi geografis termasuk *terrain*, dan informasi penerbangan seperti *Notice-to-Airmen* (NOTAM) dan *Temporary Flights Restriction* (TFR). Salah satu aspek dalam ADS adalah ADS-B yang dapat diakses secara bebas pada frekuensi 1.090MHz. Data ADS-B akan disiarkan oleh pesawat yang menggunakannya pada frekuensi tersebut [2].

Salah satu data terpenting dari ADS-B adalah data pergerakan masing-masing pesawat itu sendiri. Proses penerimaan data ADS-B secara umum akan dilakukan oleh >1 stasiun, dengan harapan, setiap stasiun akan memperbesar cakupan penerimaan data ADS-B. Penggunaan data ADS-B yang dimulai dari komunitas hobby telah banyak membantu proses pengambilan keputusan pada kejadian-kejadian tertentu pada dekade terakhir untuk mendukung data radar primer atau data yang terekam pada *Flight Data Recorder* (FDR), yang selama ini dimanfaatkan oleh otoritas. Misalnya, kejadian *incident hard landing*, *lost contact*, *crash*, maupun *delay* berkepanjangan beberapa maskapai di seluruh dunia. Data ADS-B akan digunakan untuk menentukan fase penerbangan.

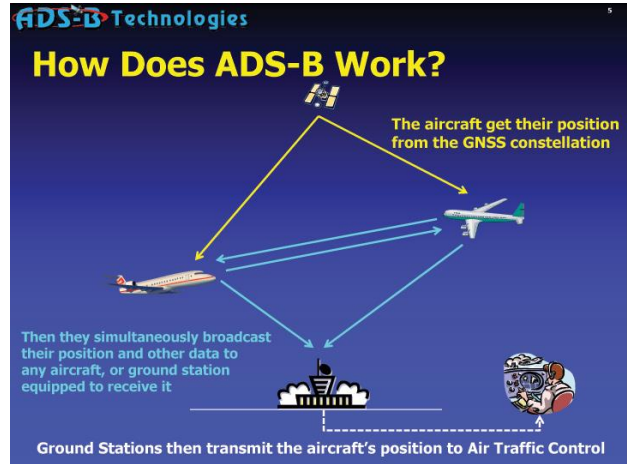
Fase penerbangan, adalah salah satu kategorisasi pergerakan pesawat. Mulai dari apron hingga nantinya kembali ke apron [3]. Selama ini kategorisasi dilakukan secara manual, berdasarkan komunikasi radio analog antara pilot dan ATC. Dengan adanya data ADS-B, analisa fase penerbangan dapat dilakukan secara real-time dengan melibatkan data ADS-B yang juga didapatkan secara real-time. Proses klasifikasi ini membutuhkan fitur-fitur dan metode yang tepat karena data berjumlah cukup banyak dalam rentang waktu yang cukup singkat dan keadaan kelas yang tidak setimbang

Pemilihan fitur pada klasifikasi data merupakan salah satu bidang penelitian yang cukup penting. Pemilihan fitur yang tepat akan menghasilkan klasifikasi data yang lebih baik dibandingkan penggunaan sebarang fitur. Penggunaan data ADS-B yang memiliki variabel dalam jumlah besar memiliki resiko dalam pengolahannya. Salah satu resiko penggunaan data dengan jumlah variabel dan jumlah data yang besar dalam klasifikasi adalah terjadinya *overfit* [4].

Teknologi ADS-B

Automatic Dependent Surveillance-Broadcast (ADS-B) adalah sebuah teknologi yang memungkinkan pesawat yang memilikinya, memancarkan data pada frekuensi tertentu. Pada saat ini, frekuensi yang umum digunakan adalah 1.090MHz. Data yang dipancarkan, antara lain, identitas pesawat, nomor penerbangan, lokasi pesawat, arah pergerakan, ketinggian, kecepatan, perhitungan arah angin, dan beberapa informasi lainnya. Pancaran data ini dapat diterima oleh stasiun darat dan pesawat lainnya. Informasi yang dikirimkan oleh sistem ini didapat dari informasi *Global Positioning System* (GPS) atau *Flight Management System* (FMS) yang ada di

pesawat [5]).



Gambar 1 Cara Kerja ADS-B

International Civil Aviation Organization (ICAO) dan International Air Transport Association (IATA) sebagai regulator internasional telah menerbitkan beberapa dokumen mengenai protokol dan perencanaan implementasi ADS-B di masa depan sebagai sarana navigasi sekunder. Cara kerja ADS-B secara umum ditunjukkan oleh Gambar 1 [6].

Tabel 1 Penjelasan Data ADS-B

Atribut	Kode	Penjelasan
8A043A	<i>Address Announced (AA)</i>	Identitas Mode-S pesawat, dapat diterjemahkan menjadi data berikut: 1. Negara asal: Indonesia (8A) 2. Registrasi: PK-GNE 3. Maskapai: Garuda Indonesia 4. Tipe Pesawat: Boeing 737-800 5. ICAO Type: B738 6. Nomor Seri: 39936
GIA321	<i>Callsign (CS)</i>	Nomor Penerbangan, dari informasi ini dapat diterjemahkan bahwa rute yang diterbangi adalah Surabaya (SUB) – Jakarta (CGK)
-7.02794	<i>Latitude (LAT)</i>	Koordinat Lintang
112.66039	<i>Longitude (LON)</i>	Koordinat Bujur
F223	<i>Flight Level (FL)</i>	Ketinggian pesawat (dalam 100 kaki)
22300	<i>Altitude Code (AC)</i>	Ketinggian pesawat (dalam satuan kaki)
2112	<i>Vertical Rate (VR)</i>	Kecepatan vertikal pesawat (dalam kaki/menit)
291	<i>True Track (TT)</i>	Arah pergerakan pesawat saat ini
290	<i>Heading (HDG)</i>	Arah pergerakan pesawat dengan patokan <i>magnetic-north</i>
279	<i>Indicated Airspeed (IAS)</i>	Kecepatan pesawat di udara yang ditampilkan dalam kokpit
280	<i>True Airspeed (TAS)</i>	Kecepatan pesawat di udara dengan referensi udara sekitarnya
405	<i>Ground Speed (GS)</i>	Kecepatan pesawat di darat
240	<i>Selected Altitude (MCP)</i>	<i>Flight Level</i> / ketinggian yang dituju (dalam 100 kaki)

8A043A:A-:GIA321:0:5:0:2:6317:-7.02794:112.66039:8:0:F223:22300:0:0:1375::0:2112:
 +21::291:290:0:279:280:405:632:0:::240::1013:0::0:::0:0:0:0:12:0:7:17:1:0:1:0:0:0:::13:13:13:5:
 415:9:3:11:::5:1726::1538:17:13:880:13:13:83:1719:1490:1453187889:60:WARR5:108542:60978:
 111175:19771

Gambar 2 Contoh Data ADS-B

ADS-B sendiri telah dimanfaatkan oleh banyak pihak, termasuk cara berkomunikasi menggunakan ADS-B dan penggabungannya dengan teknologi terdahulu telah dipatenkan dan efektif mulai tahun 2011 [6]. Di Amerika telah dipastikan FAA mewajibkan penggunaan ADS-B mulai 2020.

Beberapa negara lain termasuk Indonesia sudah berkomitmen untuk mengacu pada ADS-B untuk sektor penerbangan dan ketinggian tertentu di awal tahun 2018 [7]. Di Eropa, Eurocontrol sebagai pengendali aktivitas udara, juga telah mencanangkan modernisasi untuk teknologi *surveillance* menggunakan data ADS-B pada 750 *ground stations* di seluruh Eropa [9].

Identitas Mode-S

Mode-S adalah standar pertukaran data pada dunia aviasi. Standar ini berdasarkan identifikasi pesawat sesuai standar ICAO. Standar ini juga diratifikasi oleh FAA [10].

Setiap pesawat memiliki identitas unik sepanjang 24 bit (6 karakter heksadesimal) yang dapat membantu proses identifikasi masing-masing pesawat yang terlibat dalam komunikasi data. Contoh identifikasi data identitas pesawat dengan mode-S adalah sebagai berikut:

- 1) Identitas pesawat : 8A001A
- 2) Alokasi Negara : Indonesia
- 3) Data lain yang dapat dikaitkan:
 - a) Registrasi : PK-GSG
 - b) Maskapai : Garuda Indonesia
 - c) Tipe Pesawat : Boeing 747-400
 - d) ICAO Type : B744
 - e) Nomor Seri : 25704

Data ADS-B

Pesawat secara kontinu mengirimkan informasi berupa identitas, posisi, ketinggian, kecepatan, serta statusnya pada saat ini. Data ADS-B menyebarkan informasi identitas pesawat, posisi pesawat, kecepatan pesawat vertikal serta horizontal, ketinggian pesawat, ketinggian yang dipilih, serta kondisi pesawat seperti status penerbangan, arah gerak, status prioritas penerbangan, dan lain-lain [11]. Gambar 2 menunjukkan atribut dalam data ADS-B dan Tabel 1 menunjukkan penjelasan dari masing-masing atribut.

I. Metodologi

Metodologi penelitian yang digunakan pada penelitian ini, yang terdiri dari (1) studi literatur, (2) desain dan implementasi, dan (3) analisis pengujian. Ilustrasi alur metodologi penelitian dapat dilihat pada Gambar 3.



Gambar 3 Metodologi Penelitian

Penelitian Sebelumnya

Pada penelitian ini, referensi yang digunakan diantaranya ICAO ANNEX 13 [3], penelitian sebelumnya [12], serta dokumen lain yang

menjelaskan peraturan tersebut [14] [15]. Dari studi literatur yang telah dilakukan maka diperoleh informasi yang berkaitan dengan penelitian yang dilakukan, seperti berikut:

4) Saat ini, data ADS-B hanya digunakan sebagai data sekunder, melengkapi radar primer, sehingga aspek pengolahannya belum menjadi perhatian.

5) Lingkup pemanfaatan data ADS-B kedepannya adalah sangat luas.

6) Data ADS-B adalah berupa data alir, yang akan berubah seiring berjalannya waktu.

7) Penerimaan sinyal ADS-B dengan banyak penerima akan memperluas area/wilayah cakupan.

8) Dalam melakukan kategorisasi data, definisi standar kategori yang digunakan adalah dokumen *Flight Phase Definition ICAO ANNEX 13* revisi April 2013 [16], meliputi:

- a) *STANDING (STD)*
- b) *PUSHBACK/TOWING (PBT)*
- c) *TAXI (TXI)*
- d) *TAKEOFF (TOF)*
- e) *INITIAL CLIMB (ICL)*
- f) *EN ROUTE (ENR)*
- g) *MANEUVERING (MNV)*
- h) *APPROACH (APR)*
- i) *LANDING (LDG)*
- j) *EMERGENCY DESCENT (EMG)*
- k) *UNCONTROLLED DESCENT (UND)*
- l) *POST-IMPACT (PIM)*
- m) *UNKNOWN (UNK)*

Catatan: fase yang ditebalkan adalah fase yang mungkin dipantau dari data ADS-B

9) Dari studi literatur, dapat disimpulkan juga mengenai kondisi saat ini bahwa:

- a) Pemantauan data ADS-B berbasis fase penerbangan sangat dibutuhkan dalam hal terjadinya *accident*, *incident*, maupun mendeteksi lalu lintas udara yang ada di sekitar bandara.
- b) Diperlukan sistem yang efektif dan efisien untuk mengolah data ADS-B

berbasis fase penerbangan tersebut, dikarenakan tipe data yang diolah adalah aliran (*stream*). Ketidakmampuan sistem mengolah dengan efisien akan menyebabkan penumpukan data dan *bottleneck*.

c) Data bersifat cuplikan. Tidak terdapat rekaman penerbangan penuh.

d) Pesawat yang terekam datanya adalah pesawat yang berasal dari bandara Juanda, menuju ke bandara Juanda, dan *crossing* (hanya lewat).

Modifikasi fitur data

Modifikasi masing-masing fitur akan ditunjukkan sebagai berikut:

1) Contoh data mentah, data ini berasal dari data gabungan penerima ADS-B. Contoh data mentah ditunjukkan oleh Tabel 2.

2) Contoh data olah (1), data ini merupakan contoh pengolahan dari data mentah, dengan menggunakan jarak relatif (sebelumnya lokasi *lat* dan *lon*), serta perubahan ketinggian (sebelumnya ketinggian), seperti ditunjukkan Tabel 3.

3) Contoh data olah (2), data ini merupakan contoh pengolahan tahap kedua dari data mentah, dengan mengolah *lat-lon* menjadi jarak dua titik, serta klasifikasi perubahan ketinggian (sebelumnya perubahan ketinggian), seperti ditunjukkan

4) Tabel 4.

5) Contoh data olah (3), data ini merupakan contoh pengolahan tahap ketiga dari data raw, dengan mengolah jarak yang ada sebelumnya menjadi perubahan jarak, serta klasifikasi perubahan altitude (sebelumnya perubahan altitude), seperti ditunjukkan Tabel 5.

Tabel 2. Contoh Data Mentah

<i>Reg</i>	<i>Type</i>	<i>AA</i>	<i>Callsign</i>	<i>Lat</i>	<i>Lon</i>	<i>Alt</i>	<i>Vert Speed</i>	<i>True Track</i>	<i>Hdg</i>	<i>IAS</i>	<i>TAS</i>	<i>GS</i>
PK-LJK	B739	8A036E	LNI315	-7,37252	112,74169	650	-768	99	97	149	152	151
PK-LJK	B739	8A036E	LNI315	-7,37278	112,74323	625	-832	99	97	149	152	151
PK-LJK	B739	8A036E	LNI315	-7,37292	112,74435	600	-832	99	97	149	152	151
PK-LJK	B739	8A036E	LNI315	-7,37308	112,74534	600	-704	99	97	149	152	152
PK-LJK	B739	8A036E	LNI315	-7,37308	112,74534	550	-704	99	97	149	152	152

Tabel 3 Contoh Data Olah (1)

Reg	Type	AA	Callsign	Dist (Lat)*	Dist (Lon)*	Alt Chg [^]	Vert Speed	True Track	Hdg	IAS	TAS	GS
PK-LJK	B739	8A036E	LNI315	0,006942	0,043714	-25	-832	99	97	149	152	151
PK-LJK	B739	8A036E	LNI315	0,006802	0,042594	-25	-832	99	97	149	152	151
PK-LJK	B739	8A036E	LNI315	0,006642	0,041604	0	-704	99	97	149	152	152
PK-LJK	B739	8A036E	LNI315	0,006642	0,041604	-50	-704	99	97	149	152	152

*Dist (Lat) dan dist (Lon) adalah jarak pesawat dari Bandara Juanda (-7,379722 112,786944)

[^]Alt Chg didapatkan dari perubahan altitude

Tabel 4 Contoh Data Olah (2)

Reg	Type	AA	Callsign	Dist**	Class Alt Chg ^{^^}	Vert Speed	True Track	Hdg	IAS	TAS	GS
PK-LJK	B739	8A036E	LNI315	4.8902510	-1	-832	99	97	149	152	151
PK-LJK	B739	8A036E	LNI315	4.7658463	-1	-832	99	97	149	152	151
PK-LJK	B739	8A036E	LNI315	4.6552334	0	-704	99	97	149	152	152
PK-LJK	B739	8A036E	LNI315	4.6552334	-1	-704	99	97	149	152	152

** Dist adalah jarak pesawat dari Bandara Juanda (-7,379722 112,786944) dengan rumus Haversine (dalam kilometer)

^{^^}Class Alt Chg pada pengolahan data kedua hanya terdapat 1,0,-1

Tabel 5 Contoh Data Olah (3)

Reg	Type	AA	Callsign	Chg Dist**	Class Alt Chg ^{^^}	Vert Speed	True Track	Hdg	IAS	TAS	GS
PK-LJK	B739	8A036E	LNI315	-0,1722629	-1	-1	99	97	149	152	151
PK-LJK	B739	8A036E	LNI315	-0,1244047	-1	-1	99	97	149	152	151
PK-LJK	B739	8A036E	LNI315	-0,1106129	0	-1	99	97	149	152	152
PK-LJK	B739	8A036E	LNI315	0,0000000	-1	-1	99	97	149	152	152

**Chg Dist adalah perubahan jarak pesawat dari Bandara Juanda (-7,379722 112,786944)

^{^^}Class Alt Chg dan Vert Speed pada pengolahan data ketiga hanya terdapat 1,0,-1

Akuisisi Data ADS-B

Sistem server akan menerima masukan dari >1 receiver, hasil dari masing-masing receiver. Sistem akan mengolah, membersihkan, dan membuang data-data yang tidak valid dari masukan sistem sebelum dilakukan klasifikasi. Detil data yang digunakan dari proses logging adalah sebagai berikut:

1. Pengambilan Data : 14 Juni 2015 17:00 – 25 Juni 2015 00:00
2. Record valid : 997.661
3. Penerbangan : 2.018
4. Maskapai : 27
5. Pesawat : 405
6. Jenis Pesawat : 15

Adapun sampel data yang diambil dari data tersebut dan diberi label untuk digunakan sebagai data *training* dan *testing* adalah sebagai berikut:

1. Record : 32.243
2. Penerbangan : 500
3. Maskapai : 17
4. Pesawat : 197
5. Jenis Pesawat: 11

Pemberian Label Kelas

Lima orang ahli yang berprofesi sebagai ATC, pekerja penerbangan, dan pengamat penerbangan memberikan label kelas. Proses ini

menghasilkan 6 label kelas pada data, yakni *Climbing* (kode kelas: L), *Cruise* (C), *Descend* (D), *Approach* (A), *Holding* (H) dan *Landing* (G). Label kelas fase *take-off* (T) yang telah disiapkan, tidak muncul pada data sampel. Data yang telah diberi label akan digunakan untuk pembelajaran dan pengujian. Kelima orang ahli memberikan pendapatnya. Pada perbedaan pendapat, para ahli akan mengacu pada referensi “Aircraft Operations : Procedures for Air Navigation Services” [16].

Data yang telah diberi label

Data logging berupa data CSV sebanyak 32.243 record. Varian yang akan diuji adalah 4 tipe *dataset* yang merupakan *subset* dari data sampel dengan tetap mempertahankan keenam kelas ada dalam *dataset*. *Dataset_1* memuat 1.500 *record*, *Dataset_2* memuat 10.000 *record*, *Dataset_3* memuat 20.000 *record*, serta *Dataset_4* memuat 32.243 (seluruh) *record*.

Adapun karena sifat alami dari data yang diarsip dan diberi label, data bersifat *unbalanced*. Label *Cruise* mendominasi sekitar 45-56% dari seluruh data, sehingga karakteristik data perlu diperiksa kembali untuk *specificity* dan *sensitivity*. Statistik mengenai data tersebut dapat dilihat pada Tabel 6.

Terdapat tambahan tiga varian fitur pada *dataset* untuk perbandingan variasi/modifikasi fitur, yakni *Dataset* 1A, 1B, dan 1C, fitur pada *dataset* 1A, 1B, dan 1C mengacu pada Tabel 3, Tabel 4, dan Tabel 5. Perubahan dilakukan pada *Altitude Code*, *Latitude-Longitude*, dan *Vertical Rate*. *Dataset* 1A, 1B, 1C hanya memuat 1.476 data karena terkait dengan sifatnya yang membutuhkan histori (selisih dengan data sebelumnya).

Metode Klasifikasi

Proses pembelajaran akan dilakukan dengan *dataset* yang telah dilakukan proses variasi fitur. Data akan dimodelkan sesuai dengan variasi fitur dan disesuaikan proses pembelajarannya dengan tipe data (kontinu atau diskrit). Setiap variasi fitur, variasi jumlah, dan kombinasinya akan dilakukan pembelajaran dan menghasilkan model. Metode pembelajaran yang digunakan adalah *K-Nearest Neighbor* dan *Decision Tree* sebagai metode pembandingan.

Percobaan diimplementasikan dengan menggunakan bahasa pemrograman Python versi 2.7 dan beberapa pustaka eksternal yakni *scipy*, *numpy*, *matplotlib* dan *scikit-learn* [17]. Pengujian klasifikasi dilakukan dengan menggunakan *K-Nearest Neighbor* serta *Decision Tree* (menggunakan fungsi *DecisionTreeClassifier* dari pustaka *scikit-learn*

/ *sklearn.tree*) yang digunakan sebagai perbandingan.

80% *dataset* akan digunakan sebagai bahan pembelajaran dengan kondisi proporsi data pembelajaran akan mendekati proporsi label yang diberikan oleh ahli. 80% dipilih setelah pengujian karena menghasilkan data yang cukup stabil untuk kedua metode *KNN* dan *Decision Tree*.

Metode Pengujian

Pengujian dilakukan dengan membandingkan data keluaran dari model yang dihasilkan. Seluruh data pengujian dihitung nilai Akurasi, Presisi, *Recall*, *Specificity*, dan *F-Measure*, untuk menentukan variasi fitur yang paling sesuai untuk klasifikasi fase penerbangan pada Data ADS-B.

Pada saat pengujian juga akan dihitung waktu yang dibutuhkan untuk melakukan proses pemberian / prediksi label pada data uji, sehingga faktor waktu akan dimasukkan dalam kesimpulan pengujian. Pada saat analisis pengujian akan diputuskan bahwa penggunaan fitur dan metode yang dapat menghasilkan proses klasifikasi yang baik (dan cepat, apabila memungkinkan) untuk dapat digunakan dalam proses penentuan fase penerbangan.

Tabel 6 Statistik *Dataset*

Label Kelas	Fase	<i>Dataset_1</i>		<i>Dataset_2</i>		<i>Dataset_3</i>		<i>Dataset_4</i>	
		N	%	N	%	N	%	N	%
C	Cruise	840	56,00%	5.518	55,18%	10.115	50,58%	14.742	45,72%
D	Descent	184	12,27%	585	5,85%	1.776	8,88%	2.906	9,01%
L	cLimb	274	18,27%	2.179	21,79%	3.907	19,54%	6.746	20,92%
A	Approach	157	10,47%	1.024	10,24%	2.826	14,13%	5.676	17,60%
G	landinG	24	1,60%	219	2,19%	654	3,27%	1.116	3,46%
H	Holding	21	1,40%	475	4,75%	722	3,61%	1.057	3,28%
Banyak Data		1.500		10.000		20.000		32.243	

II. Hasil dan Pembahasan

Sampel Data ADS-B

Jumlah sampel yang cukup banyak, ditambah dengan kondisi data yang *unbalanced* (kelas C mendominasi mendekati 50%) memunculkan kecenderungan data yang harus dipastikan sebelum melakukan pengujian secara intensif. Kecenderungan terkait dengan data tersebut antara lain:

1. Data yang *unbalanced* akan menyebabkan model condong ke salah satu kelas dan jumlah sampel yang mendominasi akan menghilangkan signifikansi kelas yang lain
2. Model *overfit*

Skenario Ujicoba

Keseluruhan percobaan menggunakan data *training* dan *testing* yang sama pada setiap *dataset*, sebesar 80% data *training* dan 20% data *testing* kecuali pada ujicoba 1 dan 2 (pembagian bervariasi). Parameter yang diujicoba akan dijelaskan sebagai berikut:

- a) **Ujicoba 1:** Ujicoba KNN dengan *Dataset_1* untuk mengukur kestabilan data pada parameter jumlah data *training*:*data_testing*, parameter 25%:75%, 50%:50%, 60%:40%, 70%:30% dan 80%:20%
- b) **Ujicoba 2:** Ujicoba Decision Tree dengan *Dataset_1* untuk mengukur kestabilan data pada parameter jumlah data *training*:*data_testing*, parameter 25%:75%, 50%:50%, 60%:40%, 70%:30% dan 80%:20% serta parameter = gini index dan entropy
- c) **Ujicoba 3:** Ujicoba KNN dengan *Dataset_1*, membandingkan penggunaan Euclidean Distance, Manhattan Distance, dan Cosine Distance. Parameter K=1.
- d) **Ujicoba 4:** Ujicoba KNN dengan *Dataset_1* menggunakan Manhattan Distance dan Parameter K=1,3,7,11,13,17.
- e) **Ujicoba 5:** Ujicoba KNN dengan *Dataset_1* menggunakan Cosine Distance dan Parameter K=1,3,7,11,13,17.
- f) **Ujicoba 6:** Ujicoba KNN dengan *Dataset_1*, *Dataset_2*, *Dataset_3*, *Dataset_4* menggunakan Manhattan Distance dengan parameter K terbaik dari ujicoba (2).
- g) **Ujicoba 7:** Ujicoba KNN dengan *Dataset_1*, *Dataset_2*, *Dataset_3*, *Dataset_4* menggunakan Cosine Distance dengan parameter K terbaik dari ujicoba (3).
- h) **Ujicoba 8:** Ujicoba KNN dengan parameter terbaik (dari ujicoba 3-7) menggunakan varian *dataset* 1, 1A, 1B, 1C (*dataset* 1 dengan 1.500 data yang telah divariasikan fiturnya).
- i) **Ujicoba 9:** Ujicoba KNN dengan parameter tercepat (dari ujicoba 3-7) menggunakan varian *dataset* 1, 1A, 1B, 1C (*dataset* 1 dengan 1.500 data yang telah divariasikan fiturnya).
- j) **Ujicoba 10:** Ujicoba Decision Tree dengan *Dataset_1*, *Dataset_2*, *Dataset_3*, *Dataset_4*, parameter=Gini Index dan Entropy.
- k) **Ujicoba 11:** Ujicoba Decision Tree dengan varian *Dataset_1*, 1A, 1B, 1C, parameter=Gini Index dan Entropy.

Hasil Ujicoba

Tabel 7 Hasil Ujicoba Proporsi *Training-Testing*

Metode	Proporsi	Ujicoba	Tabel	Akurasi	Precision	Recall	Specificity	F1
K-Nearest Neighbor	25:75	1a	4-2	99,26%	98,00%	98,57%	99,43%	98,27%
	50:50	1b	4-3	99,51%	98,61%	99,18%	99,65%	98,88%
	60:40	1c	4-4	99,67%	98,93%	99,53%	99,78%	99,22%
	70:30	1d	4-5	99,63%	98,85%	99,43%	99,75%	99,13%
	80:20	1e	4-6	99,89%	99,55%	99,90%	99,94%	99,72%
Decision Tree Parameter: Gini Indeks	25:75	2a	4-7	99,20%	97,39%	98,62%	99,46%	97,98%
	50:50	2c	4-9	99,78%	99,16%	99,80%	99,87%	99,47%
	60:40	2e	4-11	99,83%	99,58%	99,24%	99,90%	99,40%
	70:30	2g	4-13	99,78%	99,79%	99,15%	99,77%	99,46%
	80:20	2i	4-15	100,00%	100,00%	100,00%	100,00%	100,00%
Decision Tree Parameter: Entropy	25:75	2b	4-8	99,35%	97,90%	99,23%	99,59%	98,51%
	50:50	2d	4-10	99,73%	99,42%	99,38%	99,77%	99,40%
	60:40	2f	4-12	99,72%	99,41%	99,26%	99,75%	99,33%
	70:30	2h	4-14	99,70%	99,42%	98,74%	99,78%	99,06%
	80:20	2j	4-16	100,00%	100,00%	100,00%	100,00%	100,00%

Tabel 8 Hasil Ujicoba Metode K-Nearest Neighbor

Ujicoba	Parameter Jarak	K	Dataset	Akurasi	Precision	Recall	Specificity	F1 Score	Runtime
3a	Euclidean			99,89%	99,55%	99,90%	99,94%	99,72%	22,97s
3b	Manhattan	1	1	99,89%	99,55%	99,90%	99,94%	99,72%	13,80s
3c	Cosine			100,00%	100,00%	100,00%	100,00%	100,00%	45,28s
4a		1		99,89%	99,55%	99,90%	99,94%	99,72%	16,47s
4b		3		99,56%	98,58%	99,23%	99,69%	98,89%	14,07s
4c		7		99,56%	98,58%	99,23%	99,69%	98,89%	15,95s
4d	Manhattan	11	1	99,44%	98,19%	99,13%	99,63%	98,63%	15,08s
4e		13		99,44%	98,19%	99,13%	99,63%	98,63%	15,67s
4f		17		99,44%	98,19%	99,13%	99,63%	98,63%	17,41s
5a		1		100,00%	100,00%	100,00%	100,00%	100,00%	40,05s
5b		3		99,67%	98,72%	99,69%	99,81%	99,18%	43,02s
5c		7		99,22%	97,46%	98,93%	99,50%	98,13%	41,22s
5d	Cosine	11	1	99,33%	97,81%	99,03%	99,56%	98,38%	58,13S
5e		13		99,22%	97,46%	98,93%	99,50%	98,13%	65,39s
5f		17		99,11%	97,12%	98,83%	99,43%	97,88%	60,38s
6a			1	99,89%	99,55%	99,90%	99,94%	99,72%	17,56s
6b			2	99,57%	97,37%	98,14%	99,67%	97,74%	704,37s
6c	Manhattan	1	3	99,53%	98,30%	97,48%	99,64%	97,88%	1.647,69s
6d			4	99,57%	97,57%	98,20%	99,71%	97,88%	3.216,87s
7a		1		100,00%	100,00%	100,00%	100,00%	100,00%	19,42s
7b			2	99,57%	97,39%	98,15%	99,67%	97,76%	787,96s
7c	Cosine	1	3	99,60%	98,46%	98,11%	99,71%	98,28%	3.269,94s
7d			4	99,58%	97,66%	97,92%	99,71%	97,78%	18.951,4s
8a			1	100,00%	100,00%	100,00%	100,00%	100,00%	18,62s
8b			1a	99,66%	98,98%	99,33%	99,74%	99,15%	17,41s
8c	Cosine	1	1b	99,89%	99,90%	99,52%	99,87%	99,71%	27,33s
8d			1c	99,89%	99,90%	99,52%	99,87%	99,71%	26,52s
9a			1	99,56%	98,58%	99,23%	99,69%	98,89%	8,51s
9b			1a	99,77%	99,43%	99,43%	99,80%	99,43%	4,04s
9c	Manhattan	3	1b	99,89%	99,90%	99,52%	99,87%	99,71%	10,07s
9d			1c	99,89%	99,90%	99,52%	99,87%	99,71%	5,86s

Tabel 9 Hasil Ujicoba Metode Decision Tree

Ujicoba	Parameter	Dataset	Akurasi	Precision	Recall	Specificity	F1 Score	Runtime	
10a		1	100,00%	100,00%	100,00%	100,00%	100,00%	0,06s	
10c		2	99,83%	98,99%	99,17%	99,88%	99,08%	0,06s	
10e	Gini Indeks	3	99,89%	99,24%	98,98%	99,93%	99,11%	0,07s	
10g		4	99,78%	98,72%	98,80%	99,86%	98,76%	0,11s	
10b		1	100,00%	100,00%	100,00%	100,00%	100,00%	0,06s	
10d		2	99,83%	98,99%	99,17%	99,88%	99,08%	0,05s	
10f	Entropy	3	99,91%	99,38%	99,05%	99,86%	99,21%	0,09s	
10h		4	99,79%	98,64%	98,96%	99,85%	98,79%	0,07s	
11a		1	100,00%	100,00%	100,00%	100,00%	100,00%	0,06s	
11c			1a	99,89%	99,90%	99,52%	99,87%	99,71%	0,06s
11e	Gini Indeks		1b	99,32%	87,44%	98,67%	99,56%	90,26%	0,04s
11g			1c	99,77%	99,81%	99,25%	99,74%	99,52%	0,03s
11b		1	100,00%	100,00%	100,00%	100,00%	100,00%	0,06s	
11d			1a	99,89%	99,90%	99,52%	99,87%	99,71%	0,06s
11f	Entropy		1b	99,77%	99,81%	99,25%	99,74%	99,52%	0,03s
11h			1c	99,77%	99,43%	99,43%	99,80%	99,43%	0,04s

Pembahasan

Data dasar ADS-B yang digunakan memiliki 9 fitur, yakni ['lat','lon','AC','VR','TT','HDG','IAS','TAS','GS'], atau koordinat lintang-bujur, ketinggian, kecepatan vertikal, arah pesawat (real & magnetik), Kecepatan (terukur di udara,

real di udara, terukur di darat) dan 6 kelas [A,C,D,G,H,L]. Jumlah sampel terkecil dari *dataset* yang dimiliki, yang dapat mewakili keenam kelas tersebut dengan lengkap, adalah 1.500 *Dataset* (*Dataset_1*).

Proses pengujian (1) dan (2) menguji jumlah proporsi data sampel yang cocok untuk pembagian data *training*:*data_testing*. Hasil pengujian meningkat seiring dengan bertambahnya proporsi data *training* hingga mencapai maksimal pada pembagian 80% dan 20%, seperti terangkum dalam Tabel 7. Proporsi inilah yang dipilih menjadi parameter ujicoba pada pengujian (3) hingga (11).

Pada ujicoba (3) hingga (9) dilakukan 46 kali pengukuran dalam ujicoba menggunakan metode *K-Nearest Neighbor* dengan parameter yang berlainan sesuai dengan ujicoba yang telah didefinisikan pada bagian Skenario Ujicoba Pengujian ini menggunakan data *training* dan *testing* yang berada dalam satu subset. Mayoritas data *testing* memiliki kesinambungan terhadap data *testing*. Pada pengujian-pengujian ini, seperti tertera pada Tabel 8 mendapatkan hasil yang cukup baik, secara keseluruhan pengukuran menunjukkan nilai 80% atau lebih. Bahkan pada beberapa ujicoba mendapatkan nilai sempurna.

Namun pada metode KNN ini, terdapat kelemahan yang cukup signifikan untuk sampel data yang banyaknya signifikan, yakni waktu pengolahan data yang bersifat linear terhadap besarnya data bahkan kuadratik terhadap besarnya data. Meskipun hasilnya baik, metode KNN ini perlu di optimasi lebih lanjut karena pada setiap iterasi perlu membandingkan informasi / fitur antara data *training* dengan data *testing*. Apabila sampel cukup besar, data akurat, namun waktu pengolahan meningkat seiring dengan bertambahnya jumlah sampel.

Sebagai pembanding, digunakan metode Decision Tree yang membentuk model, kemudian model tersebut digunakan untuk memprediksi kelas data uji / data baru. Metode Decision Tree menghilangkan permasalahan pada jumlah sampel data. Pada pengujian (10) misalnya, seperti terlihat pada Tabel 9, waktu yang dibutuhkan untuk mengolah 1.500, 10.000, 20.000 dan 32.243 data tidak berbeda jauh dan waktunya tidak sampai 1s.

Sesuai pengujian, telah dibuktikan bahwa signifikansi kelas yang memiliki sampel lebih kecil tetap terjaga. Hal ini juga secara langsung membuktikan bahwa model tidak condong ke salah satu kelas. Penggunaan berbagai ujicoba

dengan jumlah data yang berbeda menguji *overfit* model, sehingga dapat dipastikan model tidak *overfit*.

Mengenai ekstraksi fitur, ternyata memberikan hasil yang bervariasi pada berbagai parameter yang diujicobakan. Secara umum, hasilnya tidak lebih baik dibandingkan data yang tidak dimodifikasi fiturnya, kecuali pada KNN dengan menggunakan jarak Manhattan. Namun pada beberapa pengujian, misalnya (8b), (9b), dan (9d) pada saat menggunakan metode KNN serta (11e) hingga (11h) pada saat menggunakan metode Decision Tree, proses eksekusi ujicoba berjalan dengan lebih cepat, meskipun perbedaan waktunya tidak terlalu signifikan.

III. Simpulan

Hasil penelitian menunjukkan bahwa skenario yang disiapkan dalam pengujian telah menguji setiap dugaan penulis baik dalam hal data maupun ekstraksi fitur. Jumlah sampel yang cukup banyak (32.243), ditambah dengan kondisi data yang *unbalanced* (kelas C mendominasi mendekati 50%). Sesuai pengujian, telah dibuktikan bahwa signifikansi kelas yang memiliki sampel lebih kecil tetap terjaga. Hal ini juga secara langsung membuktikan bahwa model tidak condong ke salah satu kelas. Penggunaan berbagai ujicoba dengan jumlah data yang berbeda menguji *overfit* model, sehingga dapat dipastikan model tidak *overfit*.

Usaha variasi ekstraksi fitur yang diujicobakan penulis juga mendapatkan hasil yang baik, namun tidak signifikan dan terkadang lebih buruk dari *dataset* pembanding. Perlu diujicoba lebih lanjut apabila fitur yang divariasikan lebih banyak atau alternatif fitur lain yang mungkin lebih efektif apabila divariasikan. Namun data ADS-B yang mentah tetapi lengkap (telah difilter) memberikan hasil yang cukup baik.

Metode KNN dan *Decision Tree* yang digunakan pada penelitian ini juga menghasilkan pengukuran yang cukup baik, dengan rata-rata KNN mengungguli *Decision Tree*, namun dengan selisih yang tidak terlalu besar dalam hal pengukuran akurasi-presisi-*recall-specificity-F1*. Akan tetapi *Decision Tree* memiliki keunggulan dalam hal waktu pengolahan data, dimana jumlah sampel 1.500

hingga 32.243 dapat diolah dalam kurang dari satu detik, sementara untuk data yang sama, KNN membutuhkan waktu beberapa detik, menit, bahkan hingga 5 jam untuk ujicoba 7d, seperti tertera pada Tabel 8.

Dengan perbedaan waktu yang signifikan tersebut, dengan hasil yang serupa, metode *Decision Tree* lebih cocok dengan tipe *stream* data yang digunakan secara *real-time* pada ADS-B. Diperlukan pengembangan lebih lanjut untuk memberikan akses langsung data *real-time* kepada proses klasifikasi data, baik menggunakan KNN maupun *Decision Tree*. Hal ini karena sifat data *stream* ADS-B, yang meskipun telah dibersihkan, data yang secara signifikan dapat digunakan dan lengkap sembilan fiturnya, dari sekitar satu juta data, hanya 32.243 diantaranya, atau sekitar 3%. Proses filtrasi log data ini, dilakukan secara manual dalam penelitian ini.

Juga diperlukan mekanisme yang membuat data lebih lengkap, misalnya *fill-in* atau pendekatan serupa, mengingat data yang tidak memenuhi syarat, kebanyakan terjadi karena atribut data kosong. Pemilihan fitur yang divariasikan juga perlu menjadi perhatian dan dapat dijadikan sebagai pengembangan dari penelitian ini.

Penelitian ini juga hanya menggunakan data dari sekitar bandara Juanda, dimana proses *approach*, landing, dan sebagainya selama 9-11 bulan dalam setahun menggunakan salah satu ujung *runway* (R/W10) saja, sementara ujung lainnya tidak digunakan. Dibutuhkan pengujian lebih lanjut untuk bandara-bandara yang menggunakan lebih dari satu macam *approach*, prosedur, dan lain sebagainya.

IV. Daftar Pustaka

- [1] FAA - Federal Aviation Administration, "ADS-B Final Rule," 28 Mei 2010. [Online]. Available: <http://edocket.access.gpo.gov/2010/pdf/2010-12645.pdf>. [Diakses 1 Maret 2014].
- [2] A. B. Pradana, Automatic Dependent Surveillance (ADS) Controller-Pilot Data Link Communication (CPDLC), Curug-Tangerang: Civil Aviation Training Institute - Aviation Safety Training Department, 2012.
- [3] ICAO, "ICAO Annexes," 18 12 1975. [Online]. Available: <http://www.icao.int/safety/ism/ICAO%20Annexes/Annex%2013.pdf>.
- [4] I. Guyon dan A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research* 3, pp. 1157-1182, 2003.
- [5] FAA - Federal Aviation Administration, "Surveillance and Broadcast Services," 2010. [Online]. Available: <http://www.faa.gov/nextgen/implementation/programs/adsb/broadcastservices/>. [Diakses 10 Maret 2014].
- [6] ADS-B Technologies, "ADS-B Technologies Website," ADS-B Technologies, LLC, [Online]. Available: <http://www.ads-b.com/>. [Diakses 24 06 2014].
- [7] E. G. Rolfe, P. R. Drake dan P. L. Hoover, "Methods and Apparatus For Coordinating ADS-B with Mode-S SSR and/or Having Single Link Communication". United States of America Paten US008004452B2, 23 08 2011.
- [8] Dirjen Perhubungan Udara, Cetak Biru Transportasi Udara 2005-2024, Jakarta: Direktorat Jenderal Perhubungan Udara Kementerian Perhubungan Republik Indonesia, 2005.
- [9] Eurocontrol, "Surveillance Modernisation," 24 03 2015. [Online]. Available: <http://www.eurocontrol.int/articles/casade>.
- [10] FAA - Federal Aviation Administration, "The Mode S Team," FAA - Federal Aviation Administration, 06 Maret 2012. [Online]. Available: http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/safety_ops_support/nas_engineering/modes_digitizers_sbsm/modes/. [Diakses 10 Maret 2014].
- [11] jetvision.de, "RTL1090 Installation and Operating Manual," [Online]. Available: <http://rtl1090.web99.de/homepage/index.php?way=1&site=READOUT&DERNAME=Manual&dm=rtl1090&USER=rtl1090&goto=1&XURL=rtl1090&WB>

- =1&EXTRAX=X&PIDX=102385.
[Diakses 01 Maret 2014].
- [12] N. F. Azzahra, H. Ginardi dan A. Saikhu, "Praproses Data Alir ADS-B dari Multi-Receiver dengan Pengelompokan Agglomerasi Berbasis Konsistensi Jarak," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 4, no. 1, pp. 39-44, 2015.
- [13] NASA, "ARMD," 17 06 2009. [Online]. Available: http://www.aeronautics.nasa.gov/avsafe/iifd/pdf/iifd_glossary_of_terms.pdf.
- [14] Skybrary, "Flight Phase Taxonomy," 07 07 2010. [Online]. Available: http://www.skybrary.aero/index.php/Flight_Phase_Taxonomy. [Diakses 01 03 2015].
- [15] The CAST/ICAO Common Taxonomy Team, "Phases of Flight," 2014. [Online]. Available: <http://testapex.intlaviationstandards.org/Documents/PhaseofFlightDefinitions.pdf>.
- [16] Aircraft Operations : Procedures for Air Navigation Services, Fifth Edition, vol. 1, Montreal, Quebec: International Civil Aviation Organization, 2006 .
- [17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot dan E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.

Halaman ini sengaja dikosongkan.