

PREDICTION OF FUTURE FLIGHT DELAYS BASED ON CURRENT DATA ANALYSIS USING MACHINE LEARNING

¹Rena Nainggolan✉, ²Fenina A. T. Tobing, ¹Gortap Lumbantoruan,

¹Eva J. G. Harianja, ¹Resianta Perangin-angin

¹Universitas Methodist Indonesia, Medan, Indonesia

²Universitas Multimedia Nusantara, Tangerang, Indonesia

Email: renanainggolan@methodist.ac.id

DOI: <https://doi.org/10.46880/jmika.Vol8No1.pp110-116>

ABSTRACT

Airplane transportation has many advantages compared to other means of transportation, where this means of transportation is able to cover long distances in a short time. Besides having advantages, of course, air transportation often experiences flight delays which can be caused by several things, including delayed flights due to late arrival (Arrival Delay), delays of the airline itself (Carrier Delay), previously late arrivals (Late Aircraft Delay), air traffic congestion (Nas Delay), security issues (Security Delay) and weather conditions (Weather Delay). The tests carried out were 101,315 US flight delay data from 2017 to 2022. By using the Machine Learning method, the results obtained were that the largest flight delays were caused by late arrivals, namely 386,124,672, where the largest flight delays were by airlines. The airline, namely Southwest Airlines, is 61,474,379 of the total airlines, which is 20, and the biggest departure delay is at Chicago O'Hare International airport, which is in the city of Chicago, IL province, which is 20,912,928 of the total airports, which is 596. This research aims to predict future flights by analyzing current flight data, so that future flights can be better by overcoming or avoiding previous flight delay problems.

Keyword: Predict, Flight, Delay, Data Analysis.

ABSTRAK

Alat transportasi pesawat mempunyai banyak kelebihan dibandingkan dengan alat transportasi yang lain, dimana alat transportasi ini mampu menjangkau jarak tempuh dengan waktu yang singkat. Disamping mempunyai kelebihan tentunya alat transportasi udara sering mengalami penundaan penerbangan yang bisa disebabkan oleh beberapa hal diantaranya yaitu penerbangan yang tertunda disebabkan keterlambatan kedatangan (Arrival Delay), keterlambatan maskapainya sendiri (Carrier Delay), Kedatangan sebelumnya terlambat (Late Aircraft Delay), lalu lintas Udara yang padat (Nas Delay), masalah keamanan (Security Delay) dan kondisi cuaca (Weather Delay). Pegujian yang dilakukan adalah sebanyak 101.315 US data flight delay dari tahun 2017 s/d tahun 2022. Dengan menggunakan metode Mesin Learning, maka hasil yang diperoleh adalah bahwa penundaan penerbangan yang paling besar disebabkan oleh keterlambatan tiba yaitu sebesar 386,124,672 dimana penundaan penerbangan yang terbesar oleh maskapai penerbanga yaitu Southwest Airlines yaitu sebanyak 61.474,379 dari total maskapai yaitu sebesar 20 dan penundaan keberangkatan yang terbesar yaitu berada di airport Chicago O'Hare International yang berada di kota Chicago provinsi IL yaitu sebanyak 20.912.928 dari total airport yaitu berjumlah 596. Penelitian in bertujuan untuk memprediksi penerbangan di masa depan dengan menganalisis data penerbangan sekarang, sehingga penerbangan masa depan bisa lebih baik dengan mengatasi atau menghindari masalah-masalah penundaan penerbangan yang ada sebelumnya.

Kata Kunci: Prediksi, Penerbangan, Delay, Data Analisis.

PENDAHULUAN

Adanya perbedaan waktu antara kedatangan yang dijadwalkan dengan waktu keberangkatan dengan relasi antara kedatangan dan keberangkatan disebut dengan penundaan keterlambatan atau (Flight Delay). (Ardhana et al., 2022). Banyak kerugian yang terjadi jika terjadi penundaan penerbangan(Wahyuda,

2022). Pada penelitian ini akan dilakuan prediksi penerbagan yang akan terjadi pada masa depan dengan melakukan analisis pada big data penerbangan masa sekarang dengan menggunakan Mesin Learning. Dengan adanya penundaan penerbangan akan mengakibatkan ketidak nyamanan perjalanan karena merugikan dari sisi waktu dan biaya (Utami et al.,

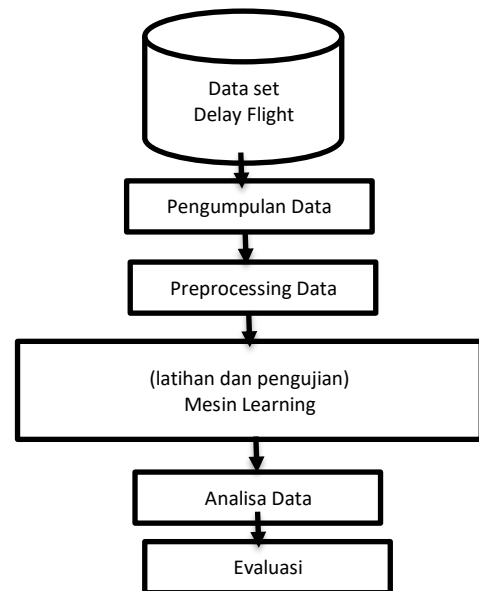
2020). Bagi pihak pengelola bandara penundaan penerbangan akan sangat mempengaruhi operasional Bandar udara. Bagi maskapai penerbangan, penundaan penerbangan membuat manajemen operasional tidak sesuai dengan jadwal penerbangan yang sudah ditentukan sebelumnya sehingga berpengaruh terhadap reputasi maskapai penerbangan yang mengakibatkan berdampak terhadap kerugian ekonomi atau pendapatan (Zámková et al., 2017).

Pengurangan penundaan keberangkatan adalah fokus utama dari pihak pengoperasian bandara (Mining et al., 2017), maskapai penerbangan maupun pihak penjadwalan kontrol lalu lintas udara (Setiawan et al., 2017). Namun saat ini urutan atau factor yang mempengaruhi penundaan penerbangan belum tergambar dengan jelas (Wang et al., 2020), sehingga kurangnya perbaikan atau peningkatan kualitas pelayanan dari industry penerbangan (Cheng et al., 2019), dengan semakin berkembangnya teknologi sehingga menuntut di semua sektor harus bergerak dengan cepat (Cao et al., 2019) untuk bertahan secara kompetitif di era globalisasi saat ini sehingga alat transportasi udara merupakan salah satu pilihan yang paling tepat untuk melakukan perjalanan karena jarak jauh bisa ditempuh dengan relatif sangat cepat (Zhou et al., 2022). Oleh karena itu menuju pesatnya perkembangan indsutri penerbangan, sangatlah penting untuk mengeksplorasi atau menganalisis apa yang mendasari penundaan penerbangan dan bagaimana mengurangi penundaan keberangkatan (Yi et al., 2021) sehingga memperoleh pelayanan penumpang lebih berkualitas (Mining et al., 2017). Fitur fitur yang berhubungan dengan penundaan penerbangan dapat digabungkan sebagai model prediksi penundaan penerbangan menggunakan Data Mining. Penelitian sebelumnya menunjukkan bahwa dapat memberikan hasil prediksi yang real time dan cepat melalui porses pembelajaran dan pelatihan yang efisien.(Zámková et al., 2017)

Teknik Kecerdasan buatan (AI) dan pembelajaran Mesin (ML) semakin banyak digunakan untuk mengatasi masalah penundaan penerbangan di industry penerbangan, karena kemampuannya menyelesaikan berbagai masalah di dunia nyata (Utami et al., 2020), pendekatan berbasis mesin learning melibatkan penggunaan data historis untuk membangun model yang prediktif yang dapat memprediksi potensi penundaan penerbangan (Roihan et al., 2020). Mesin learning dapat menghasilkan prediksi dimasa yang akan datang. Ada 2 tahap proses pembelajaran pada mesin learing yaitu latihan atau training dan pengujian atau testing.(Lv et al., 2022)

METODE PENELITIAN

Metodologi dan tahapan penelitian dan penyelesaian masalah dapat diuraikan dalam gambar berikut ini.



Gambar 1. Tahapan Penelitian

Data Set

Data diambil dari www.kaggle.com yaitu data delay penerbangan di Amerika Serikat dari tahun 2017-2022 dimana data tersebut berisi data delay penerbangan yang diantaranya yaitu penerbangan yang tertunda disebabkan keterlambatan kedatangan (Arrival Delay), keterlambatan maskapainya sendiri (Carrier Delay), Kedatangan sebelumnya terlambat (Late Aircraft Delay), lalu lintas Udara yang padat (Nas Delay), masalah keamanan (Security Delay) dan kondisi cuaca (Weather Delay)

Transform Data

Pada transform data kita mengubah tipe data csv menjadi data dalam berbentuk excel dimana data terdiri dari masing masing kolom atau field dan baris. Misalnya untuk tahun yaitu number, month yaitu number, carrier yaitu string atau teks,

a. Cleansing Data

Pada Proses cleansing data dilakukan untuk membersihkan data dari spasi atau karakter yang lain serta melakukan penggabungan dan pemisahan field sesuai kebutuhan.

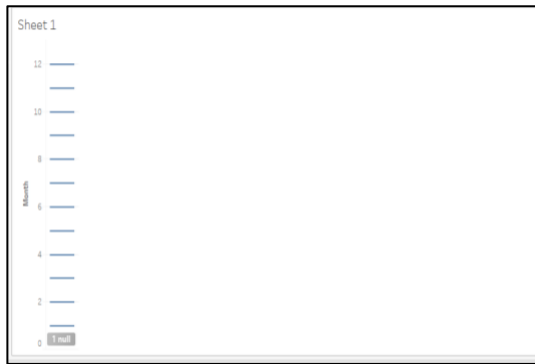
b. Penggabungan Field

Data disesuaikan dengan kebutuhan, seperti tahun dan bulan pada data set dibuat terpisah yakni di buat dalam 2 kolom, sebaiknya digabung dengan field name adalah date dan tipe data adalah date

- c. Pemisahan Field
 Pemisahan field yang sebelumnya nama airport name, Austin nama Kota, TX adalah nama provinsi. Data sebaiknya dipisah menjadi 2 kolom supaya bisa dianalisa secara tersendiri.
- d. Cleansing spasi berlebih
 Ada beberapa data dalam field yang mempunyai spasi diawal, maka dilakukan penghapusan spasi berlebih.

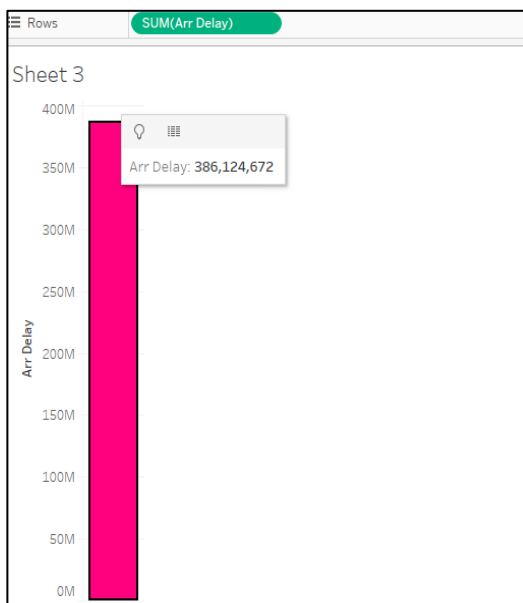
Visualisasi Data

- 1. Data Dimension
 Apabila terdapat data yaitu Data yang tidak untuk diagregasikan atau dijumlahkan atau dirata-ratakan
 Misalnya: Month



Gambar 2. Contoh Data Dimension

- 2. Data Measure
 Data yang bisa dilakukan untuk penjumlahan atau agregasi
 Misalnya: Sum Arrival Delay



Gambar 3. Contoh Data Measure

Nama Maskapai Penerbangan

Berdasarkan data set yang tersedia yaitu dari 101.315 data penundaan penerbangan ada 20 maskapai penerbangan yaitu seperti yang dijelaskan dalam table dibawah ini.

Tabel 1. Maskapai Penerbangan

No	Nama Maskapai
1	Southwest Airlines
2	American Airlines
3	Skywest Airlines
4	Delta Airlines
5	United Airlines
6	JetBlue Airlines
7	Republic Airlines
8	Spirit Airlines
9	PSA Airlines
10	Mesa Airlines
11	Envoy Airlines
12	Frontier Airlines
13	Endeavor Airlines
14	Alaska Airlines
15	ExpressJet Airlines
16	Allegiant Airlines
17	ExpressJet Airlines LLC
18	Hawaiian Airlines
19	Virgin Airlines
20	Horizon Airlines

Penyebab Penundaan Penerbangan

Dalam penelitian ini dilakukan uji coba factor penyebab penundaan sebanyak 6 kriteria yaitu Arrival Delay (Keterlambatan disebabkan keterlambatan kedatangan), Carrier Delay (Keterlambatan disebabkan keterlambatan maskapainya sendiri), Late Aircraft Delay (Kedatangan sebelumnya terlambat), Nas Delay/system penerbangan (Keterlambatan disebabkan lalu lintas Udara yang padat), Security Delay (Keterlambatan disebabkan masalah keamanan). Weather Delay (Keterlambatan karena kondisi cuaca).

Tabel 2. Faktor Delay

No	Faktor Penyebab Penundaan
1	Arrival Delay
2	Carrier Delay
3	Late Aircraft Delay
4	Nas Delay
5	Security Delay
6	Weather Delay

Nama Airport dan Kota

Dari 101.315 US data flight delay dari tahun 2017 s/d tahun 2022 ada 596 airport, untuk beberapa

contoh nama airport dapat dilihat pada Tabel 3 dibawah ini:

Tabel 3. Daftar Nama Airport

No	Nama Airport	Nama Kota
1	Lehigh Valley International	Allentown
2	Southwest Georgia Regional	Albany
3	Nantucket Memorial	Nantucket
4	Alexandria International	Alexandria
5	Augusta Regional at Bush Field	Augusta
6	Albany International	Albany
7	Hartsfield-Jackson Atlanta International	Atlanta
8	Appleton International	Appleton
9	Austin - Bergstrom International	Austin
10	Asheville Regional	Asheville
11	Kalamazoo/Battle Creek International	Kalamazoo
12	Greater Binghamton/Edwin A. Link Field	Binghamton
13	Bangor International	Bangor
14	Birmingham-Shuttlesworth International	Birmingham
15	Central Regional Airport at Bloomington	Bloomington/Normal
16	Nashville International	Nashville
17	Logan International	Boston
18	Brunswick Golden Isles	Brunswick
..		
596	Tucson International	Tucson

Carrier Name (Maskapai Penerbangan)

Uji Coba dalam data set sebanyak 101.315 data, terdapat 20 maskapai penerbangan, untuk daftar nama maskapai penerbangan dapat dilihat seperti tabel dibawah ini:

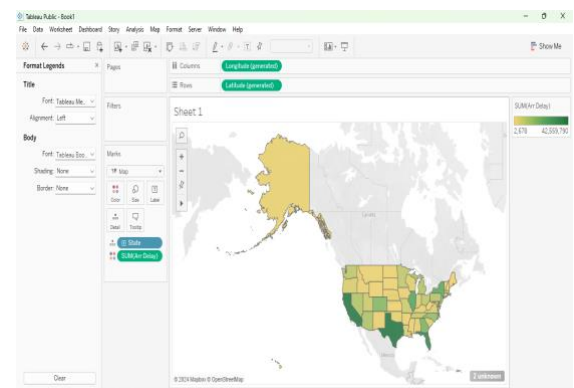
Tabel 4. Daftar Maskapai Penerbangan

No	Nama Maskapai	Total Delay
1	Southwest Airlines	61,474,379
2	American Airlines	55,238.907
3	Skywest Airlines	51,236,138
4	Delta Airlines	41,803,228
5	United Airlines	36,745,645
6	JetBlue Airlines	26,044,693
7	Republic Airlines	13,776,790
8	Spirit Airlines	12,770,142

9	PSA Airlines	11,220,158
10	Mesa Airlines	11,170,124
11	Envoy Airlines	11,080,592
12	Frontier Airlines	10,804,370
13	Endeavor Airlines	9,638,400
14	Alaska Airlines	9,386,123
15	ExpressJet Airlines	8,857,800
16	Allegiant Airlines	8,039,038
17	ExpressJet Airlines LLC	2,487,859
18	Hawaiian Airlines	1,830,964
19	Virgin Airlines	1,369,519
20	Horizon Airlines	1,149,803

HASIL DAN PEMBAHASAN

Untuk melihat jumlah data yang paling banyak delay dalam bentuk Maps yaitu seperti digambar dibawah ini, yaitu di map menunjukkan warna green dan gold, semakin gelap warnanya (green) semakin tinggi delaynya dan apabila semakin terang warnanya (gold) maka semakin rendah delaynya.

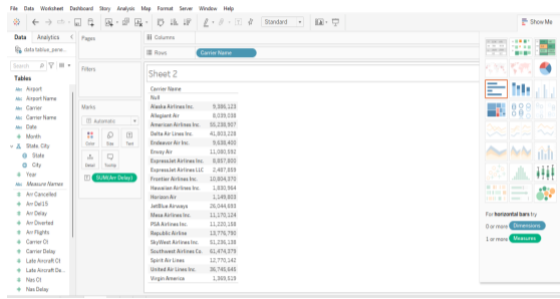


Gambar 4. Jumlah Sum Arrival Delay berdasarkan peta

Texas adalah kota yang paling tinggi delaynya yaitu sebesar 42.559.790. Kemudian kita akan menampilkan data point untuk kotanya

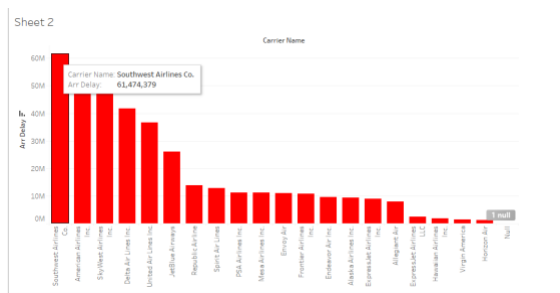
Carrier Name (Airlines)

Pada Gambar berikut ditampilkan grafik untuk nama maskapai penerbangan dengan jumlah total delay. Uji Coba dalam data set sebanyak 101.315 data delay. Bisa dilihat dari data yang paling banyak delay adalah Southwest Airlines yaitu sebanyak 61.474,379, kemudian American Airlines dan yang paling rendah delay adalah Horizon Airlines



Gambar 5. Jumlah Carrier Name Delay

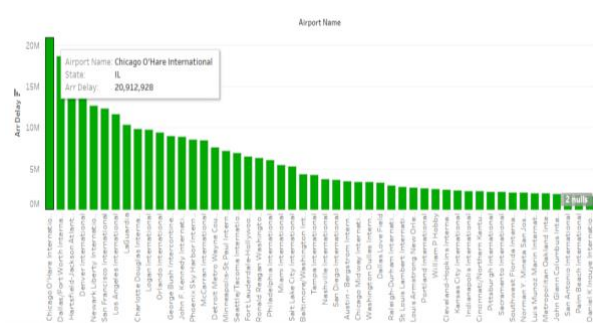
Jika dilihat dalam grafik hasilnya akan ditampilkan seperti gambar berikut ini.



Gambar 6. Grafik Carrier Name Delay

Airport Name

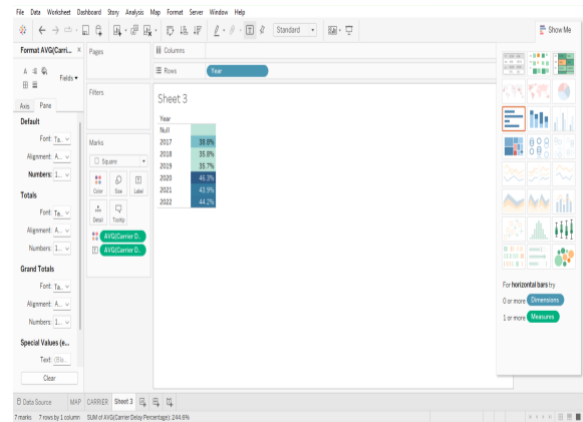
Uji Coba dalam data set sebanyak 101.315 data delay. Ada 596 airport, airport yang mengalami penundaan pesawat yang terbesar yaitu berada di airport Chicago O’Hare International yang berada di kota Chicago provinsi IL yaitu sebanyak 20.912.928



Gambar 7. Grafik Delay berdasarkan Airport Name

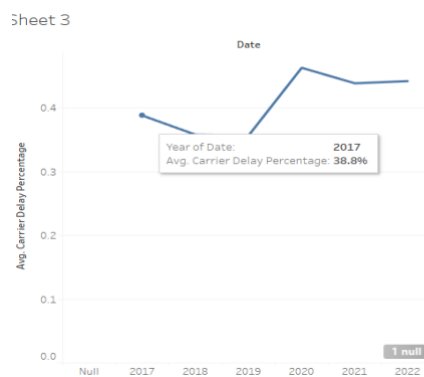
Visualisasi Date (Percentage)

Untuk persentase yang paling tinggi tingkat delay yang disebabkan maskapai penerbangan yaitu ada di tahun 2020 yaitu 46,3%, kemudian, pada tahun 2022 sebesar 44,2%, pada tahun 2021 sebesar 43,9 %, pada tahun 2017 sebesar 38,8 %, 2018 sebesar 35,8 % kemudian pada tahun 2018 sebesar 35, 8%. Dapat dilihat pada gambar berikut ini:



Gambar 8. Jumlah Delay Berdasarkan Tahun

Dapat digambarkan dalam bentuk grafik seperti dibawah ini



Gambar 9. Percentage Carrier Delay Berdasarkan Per Tahun

Untuk memunculkan setiap data tiap tiap pointnya, untuk persentase tiap bulan, rata-rata persentase perquarter pada setiap tahunnya dapat dilihat pada gambar dibawah ini:



Gambar 10. Percentage Carrier Delay Per Quarter

Total Delay

Untuk Total Delay secara keseluruhan dapat dilihat pada gambar dibawah ini yaitu:

Measure Names	Measure Values
Total Delay	
Air Delay	386,124,672
Carrier Delay	133,453,066
Late Aircraft Delay	145,618,944
Nas Delay	84,799,404
Security Delay	745,567
Weather Delay	21,506,968

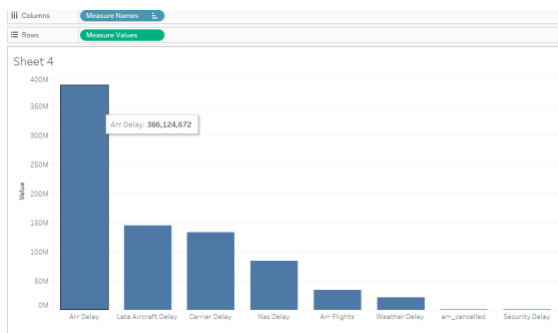
Gambar 11. Total Delay

Penundaan keberangkatan yang paling besar disebabkan oleh tiba keterlambatan yaitu sebesar 386,124,672, terbesar kedua yaitu penundaan keberangkatan disebabkan oleh keterlambatan pesawat terlambat yaitu 133,453,066 penyebab penundaan keberangkatan terbesar ketiga yaitu penundaan operator sebesar 145,618,944, kemudian penundaan terbesar keempat yaitu disebabkan oleh masalah lalu lintas udara yaitu 84,779,404, penundaan keberangkatan pada posisi kelima yaitu disebabkan oleh cuaca yaitu sebesar 21,506,968 kemudian yang terakhir yaitu penundaan keberangkatan yang disebabkan oleh masa masalah keamanan yaitu 745,567.

Tabel 5. Total Delay

No	Faktor Penyebab Penundaan	Jumlah
1	Arrived Delay	386,124,672
2	Late Aircraft Delay	145,618,944
3	Carrier Delay	133,453,066
4	Nas Delay	84,799,404
5	Weather Delay	21,506,968
6	Security Delay	745,567

Dalam bentuk grafik bisa dilihat seperti gambar dibawah ini:



Gambar 12. Grafik Total Delay

Untuk informasi secara keseluruhan bisa kita lihat di dashboard pada gambar dibawah ini.



Gambar 13. Dashboard

KESIMPULAN

Dari hasil analisa data diatas Uji Coba dalam data set sebanyak 101.315 US data flight delay dapat disimpulkan bahwa penundaan penerbangan disebabkan oleh Penundaan keberangkatan yang paling besar disebabkan oleh keterlambatan tiba yaitu sebesar 386,124,672, terbesar kedua yaitu penundaan keberangkatan disebabkan oleh keterlambatan pesawat terlambat yaitu 133,453,066 penyebab penundaan keberangkatan terbesar ketiga yaitu penundaan operator sebesar 145,618,944, kemudian penundaan terbesar keempat yaitu disebabkan oleh masalah lalu lintas udara yaitu 84,779,404, penundaan keberangkatan pada posisi kelima yaitu disebabkan oleh cuaca yaitu sebesar 21,506,968 kemudian yang terakhir yaitu penundaan keberangkatan yang disebabkan oleh masa masalah keamanan yaitu 745,567. Dimana maskapai penerbangan yang paling banyak melakukan penundaan penerbangan yaitu Southwest Airlines yaitu sebanyak 61.474,379 dari total maskapai yaitu sebesar 20 maskapai penerbangan dimana maskapai penerbangan yang paling banyak mengalami penundaan pesawat yang terbesar yaitu berada di airport Chicago O’Hare International yang berada di kota Chicago provinsi IL yaitu sebanyak 20.912.928 dari total airport yaitu berjumlah 596.

DAFTAR PUSTAKA

Ardhana, V. Y. P., Syam, Muh. Y., Ramadani, E. F., Sampetoding, E. A. M., Syahril, M., Manapa, E. S., & Mardzuki, R. (2022). Prediksi Flight Delay Berbasis Algoritma Neural Network. *Journal of Informatics, Electrical and Electronics Engineering*, 2(1), 26–30. <https://doi.org/10.47065/jieee.v2i1.429>

Cao, Y., Zhu, C., Wang, Y., & Li, Q. (2019). A Method of Reducing Flight Delay by Exploring Internal Mechanism of Flight Delays. *Journal of Advanced Transportation*, 2019(5), 22–25. <https://doi.org/10.1155/2019/7069380>

Cheng, S., Zhang, Y., Hao, S., Liu, R., Luo, X., & Luo, Q. (2019). Study of Flight Departure Delay and Causal Factor Using Spatial Analysis.

- Journal of Advanced Transportation*, 2019.
<https://doi.org/10.1155/2019/3525912>
- Lv, Q., Zhang, S., & Wang, Y. (2022). Deep Learning Model of Image Classification Using Machine Learning. *Advances in Multimedia*, 2022.
<https://doi.org/10.1155/2022/3351256>
- Mining, D., Data, P., Access, Q., Keselamatan, P., & Sadono, M. (2017). *Jurnal Perhubungan Udara Penerbangan The Data Mining on Quick Access Recorder for Flight Safety Improvement*. 1–12.
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82.
<https://doi.org/10.31294/ijcit.v5i1.7951>
- Setiawan, I., Majid, S. A., & Yuliantini, Y. (2017). Airport Factor in Flight Delays in Indonesia. *Jurnal Manajemen Transportasi Dan Logistik*, 2(3), 365.
<https://doi.org/10.25292/j.mtl.v2i3.115>
- Utami, B. M., Latuconsina, R., & ... (2020). Prediksi Delay Pesawat Menggunakan Metode Exponential Smoothing. *EProceedings ...*, 7(2), 4891–4901.
- Wahyuda, D. V. (2022). Pengaruh Delay Karena Penanganan Flight Operation Dan Teknik Terhadap on Time Performance Di Maskapai Lion Air Di Bandar Udara Adi Soemarmo. *Flight Attendant Kedirgantaraan : Jurnal Public Relation, Pelayanan, Pariwisata*, 4(1), 116–122.
<https://doi.org/10.56521/attendant-dirgantara.v4i1.417>
- Wang, Y., Zheng, H., Wu, F., Chen, J., & Hansen, M. (2020). A Comparative Study on Flight Delay Networks of the USA and China. *Journal of Advanced Transportation*, 2020.
<https://doi.org/10.1155/2020/1369591>
- Yi, J., Zhang, H., Liu, H., Zhong, G., & Li, G. (2021). Flight Delay Classification Prediction Based on Stacking Algorithm. *Journal of Advanced Transportation*, 2021.
<https://doi.org/10.1155/2021/4292778>
- Zámková, M., Prokop, M., & Stolín, R. (2017). Factors influencing flight delays of a European airline. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 65(5), 1799–1807.
<https://doi.org/10.11118/actaun201765051799>
- Zhou, F., Jiang, G., Lu, Z., & Wang, Q. (2022). Evaluation and Analysis of the Impact of Airport Delays. *Scientific Programming*, 2022(1). <https://doi.org/10.1155/2022/7102267>