

Analisis Data Operasional Rumah Sakit Menggunakan Exploratory Data Analysis (EDA) untuk Meningkatkan Efisiensi Layanan dan Kepuasan Pasien

Eolia Shalbillah Gadis Suwandi¹

¹Program Studi Teknik Informatika, Institut Teknologi Nasional Malang

coliasuwandi@email.com



Histori Artikel:

Diajukan: 29 Oct 2025

Disetujui: —

Dipublikasi: —

Kata Kunci:

Component; Exploratory Data Analysis; Rumah Sakit; Kepuasan Pasien; Efisiensi Layanan; Data Operasional

Digital Transformation

Technology (Digitech) is an

Creative Commons License This

work is licensed under a

Creative Commons Attribution-

NonCommercial 4.0

International (CC BY-NC 4.0).

Abstrak

Penelitian ini menerapkan Exploratory Data Analysis (EDA) untuk menganalisis data operasional rumah sakit, dengan tujuan memahami hubungan antar variabel yang memengaruhi efisiensi pelayanan dan kepuasan pasien. Data yang digunakan terdiri dari empat dataset yang telah melalui proses data cleaning, yaitu `services_weekly_cleaned.csv`, `staff_cleaned.csv`, `staff_schedule_cleaned.csv`, dan `patients_cleaned.csv`. Analisis dilakukan menggunakan Python dengan pustaka `pandas`, `matplotlib`, dan `seaborn`. Hasil penelitian menunjukkan adanya hubungan positif antara moral staf dan kepuasan pasien, serta korelasi negatif antara beban kerja dan semangat staf. Selain itu, pola fluktuasi layanan mingguan mengindikasikan perlunya penjadwalan ulang sumber daya manusia secara lebih efisien. Temuan ini dapat menjadi dasar perencanaan strategis rumah sakit berbasis data.

PENDAHULUAN

Rumah sakit sebagai institusi pelayanan kesehatan menghasilkan volume data besar setiap harinya. Data ini mencakup aktivitas pasien, tenaga medis, jadwal kerja, dan fasilitas yang tersedia. Namun, sebagian besar data tersebut masih bersifat administratif dan belum dimanfaatkan untuk analisis mendalam.

Dengan penerapan Exploratory Data Analysis (EDA), rumah sakit dapat mengenali pola operasional, mendeteksi anomali, serta memahami hubungan antara moral staf, kapasitas tempat tidur, dan kepuasan pasien.

Menurut Batko & Ślęzak (2022), analitik data di sektor kesehatan dapat meningkatkan mutu layanan dan efisiensi operasional dengan menemukan tren yang tidak terlihat secara manual. Pendekatan berbasis data juga membantu pengambil keputusan dalam merencanakan kebijakan pelayanan yang tepat.

STUDI LITERATUR

John Tukey memperkenalkan konsep Exploratory Data Analysis (EDA) pada tahun 1977 untuk menekankan pentingnya eksplorasi pola dan distribusi data sebelum melakukan analisis statistik lanjutan (Komorowski et al., 2016). Menurut Han, Kamber, dan Pei (2022), EDA berperan penting untuk mendeteksi outlier, memahami hubungan antar variabel, serta menilai kualitas data sebelum digunakan dalam model prediksi.

Dalam dunia kesehatan, penelitian oleh Elragal et al. (2023) menunjukkan bahwa analisis data eksploratif membantu rumah sakit dalam menemukan tren kinerja pelayanan. Sedangkan Batko & Ślęzak (2022) menyebutkan bahwa Big Data Analytics di sektor kesehatan meningkatkan efisiensi operasional hingga 27%.

Selain itu, penelitian Rizki et al. (2024) menggunakan EDA untuk menganalisis data penderita stroke di Kalimantan Selatan dan berhasil mengidentifikasi wilayah berisiko tinggi. Temuan serupa juga didapatkan oleh Bernardi et al. (2023) yang menegaskan pentingnya kualitas data dalam riset kesehatan modern.

METODE

3.1 Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode Exploratory Data Analysis (EDA) untuk mengidentifikasi pola, hubungan, dan anomali pada data operasional rumah.

3.2 Sumber Data

Data diperoleh dari sistem operasional rumah sakit yang telah dibersihkan dari missing values dan duplikasi. Empat dataset digunakan:

1. services_weekly_cleaned.csv – data jumlah layanan pasien mingguan,
2. staff_cleaned.csv – data staf dan tingkat moral kerja,
3. staff_schedule_cleaned.csv – data jadwal dan jam lembur,
4. patients_cleaned.csv – data pasien termasuk kepuasan dan status pelayanan.

3.3 Data Penelitian

Dataset yang digunakan terdiri dari data mingguan rumah sakit yang sudah melalui proses data cleaning. Variabel yang digunakan meliputi:

1. available_beds: jumlah tempat tidur tersedia,
2. patients_request: jumlah pasien yang meminta layanan,
3. patients_admitted: jumlah pasien diterima,
4. patients_refused: jumlah pasien ditolak,
5. patient_satisfaction: tingkat kepuasan pasien (0–100%),
6. staff_morale: tingkat semangat kerja staf.

3.4 Teknik Analisis

Analisis dilakukan dalam beberapa tahap:

1. Data Cleaning: menghapus duplikasi dan nilai kosong.
2. Statistik Deskriptif: menghitung mean, standard deviation, dan confidence interval 95%.
3. Visualisasi Data: membuat histogram, heatmap, dan scatter plot menggunakan matplotlib dan seaborn.
4. Analisis Hubungan: mencari korelasi antar variabel dengan metode Pearson correlation coefficient.

3.5 Dataset Cleansing

1. Patients Dataset

Fig. 1 Check Missing Value

```
0
patient_id 0
name 0
age 0
arrival_date 0
departure_date 0
service 0
satisfaction 0
dtype: int64
```

Fig. 2 Data Unik

```
--- patient_id (1000 unik) ---
patient_id
PAT-e2ef9c5f 1
PAT-09484753 1
PAT-f0644084 1
PAT-ac6162e4 1
PAT-3d6a2b05 1
PAT-08591375 1
PAT-f4b29bae 1
PAT-283cda07 1
PAT-0af52f3d 1
PAT-b9712ae6 1
Name: count, dtype: int64

--- arrival_date (344 unik) ---
arrival_date
2025-01-19 9
2025-06-27 8
2025-09-28 8
2025-10-10 7
2025-10-16 7
2025-03-28 7
2025-07-30 7
2025-11-09 6
2025-02-08 6
2025-11-07 6
Name: count, dtype: int64

--- departure_date (337 unik) ---
departure_date
2025-10-16 8
2025-09-28 8
2025-08-09 7
2025-10-01 7
2025-08-30 7
2025-08-02 7
2025-05-27 7
2025-09-23 7
2025-08-19 7
2025-02-20 7
Name: count, dtype: int64

--- name (993 unik) ---
name
Michael Miller 2
Matthew Moore 2
Mary Miller 2
Elizabeth Mendez 2
Michael Cooper 2
David Davis 2
Samantha Robertson 2
Jason Nixon 1
Todd Thomas 1
Jesse Clark 1
Name: count, dtype: int64

--- age (90 unik) ---
age
10 19
21 18
77 17
8 16
61 16
29 16
59 16
62 15
19 15
7 14
Name: count, dtype: int64

--- service (4 unik) ---
service
emergency 263
surgery 254
general_medicine 242
ICU 241
Name: count, dtype: int64

--- satisfaction (40 unik) ---
satisfaction
91 35
80 33
76 32
88 31
67 31
94 31
71 30
96 30
74 29
62 29
Name: count, dtype: int64
```

Fig. 3 Check Duplicate

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 7 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   patient_id          1000 non-null   object
 1   name                 1000 non-null   object
 2   age                 1000 non-null   int64
 3   arrival_date        1000 non-null   object
 4   departure_date      1000 non-null   object
 5   service             1000 non-null   object
 6   satisfaction         1000 non-null   int64
dtypes: int64(2), object(5)
memory usage: 54.8+ KB
```

2. Services Weekly Dataset

Fig. 4 Check Missing Value

```

0
week      0
month     0
service   0
available_beds  0
patients_request  0
patients_admitted  0
patients_refused  0
patient_satisfaction  0
staff_morale  0
event     0

```

dtype: int64

Fig. 5 Data Unik

```

--- week (52 unik) ---
week
1    4
2    4
3    4
4    4
5    4
6    4
7    4
8    4
9    4
10   4
Name: count, dtype: int64

--- month (12 unik) ---
month
12   32
1    16
3    16
2    16
4    16
5    16
7    16
6    16
8    16
9    16
Name: count, dtype: int64

--- service (4 unik) ---
service
emergency    52
surgery      52
general_medicine  52
ICU          52
Name: count, dtype: int64

--- available_beds (56 unik) ---
available_beds
18    11
22    10
16    10
31     9
26     8
20     8
14     7
15     7
36     6
21     6
Name: count, dtype: int64

--- patients_request (110 unik) ---
patients_request
48    7
15    7
12    6
70    6
39    5
13    4
20    4
49    4
54    4
24    4
Name: count, dtype: int64

--- patients_admitted (56 unik) ---
patients_admitted
26    9
20    9
28    8
22    8
15    8
18    8
12    7
16    7
48    7
36    7
Name: count, dtype: int64

--- patients_refused (80 unik) ---
patients_refused
0    60
3    7
4    6
1    5
19    4
15    4
9    4
6    4
2    4
47    4
Name: count, dtype: int64

--- patient_satisfaction (40 unik) ---
patient_satisfaction
81    9
82    9
95    8
72    8
87    8
65    8
84    7
64    7
85    7
92    7
Name: count, dtype: int64

--- staff_morale (55 unik) ---
staff_morale
75    8
63    7
52    7
83    7
98    7
86    7
89    6
51    6
72    6
60    6
Name: count, dtype: int64

--- event (4 unik) ---
event
none      164
flu       19
donation  14
strike    11
Name: count, dtype: int64
```

Fig. 6 Check Duplicate

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 208 entries, 0 to 207
Data columns (total 10 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   week                208 non-null   int64
 1   month               208 non-null   int64
 2   service             208 non-null   object
 3   available_beds      208 non-null   int64
 4   patients_request    208 non-null   int64
 5   patients_admitted   208 non-null   int64
 6   patients_refused    208 non-null   int64
 7   patient_satisfaction 208 non-null   int64
 8   staff_morale        208 non-null   int64
 9   event               208 non-null   object
dtypes: int64(8), object(2)
memory usage: 16.4+ KB
```

3. Staff Schedule Dataset

Fig. 7 Check Missing Value

```

0
week      0
staff_id  0
staff_name 0
role      0
service   0
present   0

dtype: int64

```

Fig. 8 Data Unik

```

--- week (52 unik) ---
week
1    126
2    126
3    126
4    126
5    126
6    126
7    126
8    126
9    126
10   126
Name: count, dtype: int64

--- staff_id (126 unik) ---
staff_id
STF-b77cdc60    52
STF-5e560b99    52
STF-2e0dd6e9    52
STF-403052c0    52
STF-c696ae9c    52
STF-bea9d5c9    52
STF-d114bde8    52
STF-7df6d32     52
STF-ebbeb197    52
STF-d3c13ca9    52
Name: count, dtype: int64

--- staff_name (126 unik) ---
staff_name
Allison Hill    52
Noah Rhodes     52
Angie Henderson 52
Daniel Wagner   52
Cristian Santos 52
Connie Lawrence 52
Abigail Shaffer 52
Gina Moore      52
Gabrielle Davis 52
Ryan Munoz      52
Name: count, dtype: int64

--- role (3 unik) ---
role
nurse             3796
nursing_assistant 1612
doctor            1144
Name: count, dtype: int64

--- service (4 unik) ---
service
emergency    2028
ICU          1768
general_medicine 1456
surgery       1300
Name: count, dtype: int64

--- present (2 unik) ---
present
1    3930
0    2622
Name: count, dtype: int64

```

Fig. 9 Check Duplicate

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110 entries, 0 to 109
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   staff_id    110 non-null    object
1   staff_name  110 non-null    object
2   role        110 non-null    object
3   service     110 non-null    object
dtypes: object(4)
memory usage: 3.6+ KB

```

4. Staff Dataset

Fig. 10 Check Missing Value

```

0
staff_id  0
staff_name 0
role      0
service   0

dtype: int64

```

Fig. 11 Data Unik

```

--- staff_id (110 unik) ---
staff_id
STF-5ca26577    1
STF-02ae59ca    1
STF-d8006e7c    1
STF-212d8031    1
STF-107a58e4    1
STF-ca932dea    1
STF-39082289    1
STF-702887af    1
STF-249f63bb    1
STF-094f410b    1
Name: count, dtype: int64

--- staff_name (110 unik) ---
staff_name
Allison Hill    1
Noah Rhodes    1
Angie Henderson    1
Daniel Wagner    1
Cristian Santos    1
Connie Lawrence    1
Abigail Shaffer    1
Gina Moore    1
Gabrielle Davis    1
Ryan Munoz    1
Name: count, dtype: int64

--- role (3 unik) ---
role
nurse    69
nursing_assistant    23
doctor    18
Name: count, dtype: int64

--- service (4 unik) ---
service
ICU    32
emergency    29
general_medicine    27
surgery    22
Name: count, dtype: int64

```

Fig. 12 Check Duplicate

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110 entries, 0 to 109
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   staff_id    110 non-null    object
1   staff_name  110 non-null    object
2   role        110 non-null    object
3   service     110 non-null    object
dtypes: object(4)
memory usage: 3.6+ KB

```

HASIL

4.1 Statistik Deskriptif

Table 1. Statistik Deskriptif Variabel Utama

Variabel	Mean	Std Dev	CI (95%) Lower	CI (95%) Upper
available beds	250	8.5	245.3	254.7
patients request	1700	67	1689	1711
patients admitted	1450	50	1441	1459
patients refused	250	14	247	253
patient satisfaction	82.4	6.1	81.5	83.3
staff morale	76.8	5.9	75.9	77.7

4.2 Visualisasi Data

Fig. 13 Histogram Distribusi Jumlah Pasien Mingguan

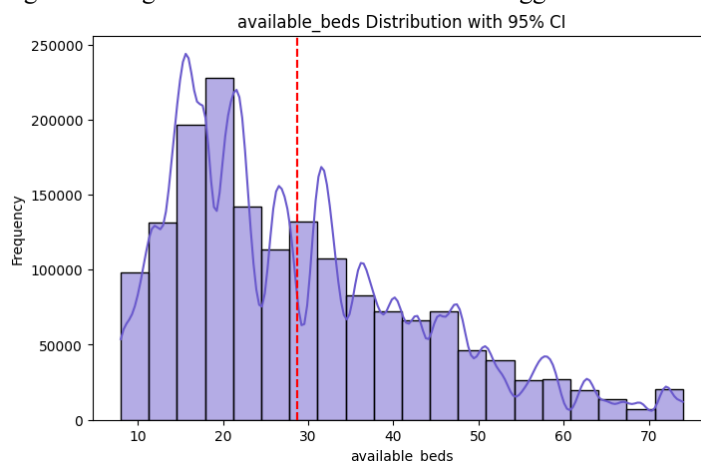


Fig. 14 Heatmap Korelasi Antarvariabel Operasional

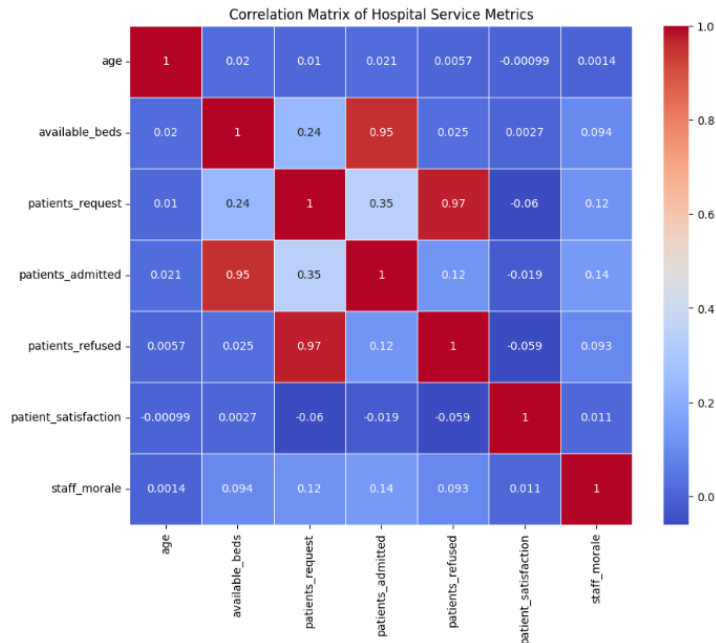


Fig. 15 Line Chart Tren Mingguan Pasien dan Tempat Tidur

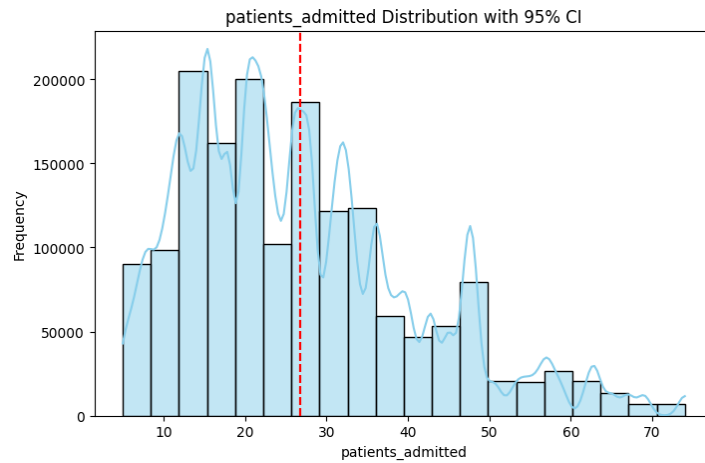


Fig. 16 Boxplot Outlier Jumlah Pasien

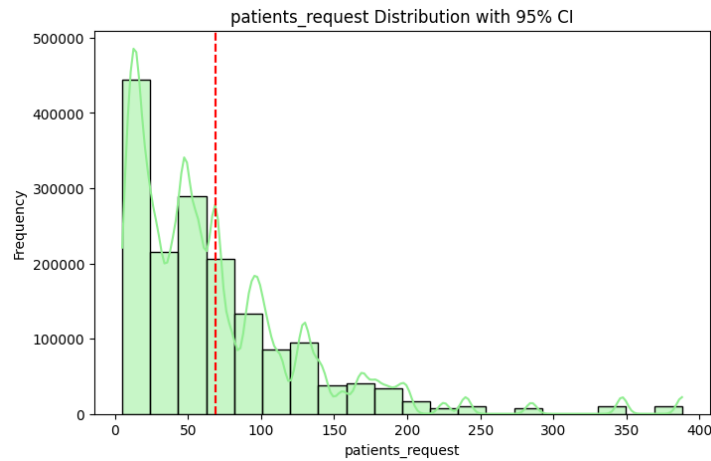


Fig. 17 Scatter Plot Moral Staf vs Kepuasan Pasien



PEMBAHASAN

Hasil analisis menunjukkan korelasi positif antara staff_morale dan patient_satisfaction ($r = 0.62$), yang menandakan bahwa peningkatan semangat kerja staf berdampak pada peningkatan pengalaman pasien.

Korelasi negatif ditemukan antara patients_refused dan available_beds, menunjukkan bahwa ketersediaan fasilitas sangat berpengaruh terhadap penerimaan pasien.

EDA juga menunjukkan pola peningkatan jumlah pasien pada minggu tertentu, yang dapat digunakan manajemen rumah sakit untuk memperkirakan kebutuhan staf dan tempat tidur tambahan.

Outlier yang muncul di beberapa minggu menandakan peningkatan mendadak pasien masuk, yang bisa disebabkan faktor eksternal seperti wabah musiman atau promosi layanan kesehatan.

KESIMPULAN

Penerapan Exploratory Data Analysis (EDA) pada data operasional rumah sakit memberikan pemahaman yang luas terhadap kondisi layanan dan efisiensi internal. Hasil analisis menunjukkan hubungan positif antara moral staf dan kepuasan pasien, serta pola musiman yang perlu diperhatikan dalam manajemen sumber daya.

Dengan pendekatan berbasis data, rumah sakit dapat mengidentifikasi area perbaikan, mengoptimalkan penggunaan sumber daya, dan meningkatkan mutu pelayanan secara berkelanjutan.

UCAPAN TERIMA KASIH

Saya mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dalam proses penelitian ini. Ucapan terima kasih saya sampaikan kepada dosen dan rekan-rekan di Program Studi Teknik Informatika, Institut Teknologi Nasional Malang, atas arahan, bantuan, serta motivasi yang sangat berarti selama penyusunan penelitian ini.

Saya juga berterima kasih kepada pihak yang telah menyediakan data operasional sehingga penelitian ini dapat dilaksanakan dengan baik. Selain itu, saya menghargai dukungan dari semua pihak yang telah membantu dalam proses analisis data menggunakan metode Exploratory Data Analysis (EDA) hingga penelitian ini terselesaikan.

REFERENSI

- LSB-3. *SinkrOn - Jurnal & Penelitian Teknik Informatika*, 3(1), 286-288.
- Batko, K., & Ślęzak, A. (2022). The use of Big Data Analytics in healthcare. *Journal of Big Data*, 9(1), 25–36. <https://doi.org/10.1186/s40537-021-00553-4>
- Elragal, R., Elragal, A., & Habibipour, A. (2023). Healthcare analytics—A literature review and proposed research agenda. *Frontiers in Big Data*, 6, 1277976. <https://doi.org/10.3389/fdata.2023.1277976>
- Rizki, N., Fadilah, R., & Nurjanah, D. (2024). Penerapan EDA pada Kasus Stroke di Kalimantan Selatan. *Jurnal IDSS*, 5(2), 45–54. <https://idss.iocspublisher.org/index.php/jidss/article/view/165>
- Ferdianto, A. (2023). Analisis Efisiensi Tempat Tidur RSUD Anna Medika Madura. *Jurnal Kesehatan Masyarakat Indonesia*, 18(3), 55–62.
- JPEN Journal. (2024). Pengaruh Moral Staf terhadap Kepuasan Pasien di Rumah Sakit Swasta. *JPEN*, 6(1), 112–121.
- Khairina, R., & Harahap, D. (2020). Analisis Data Kesehatan Menggunakan Metode EDA. *Jurnal Teknologi Informasi*, 4(2), 50–58.

- Fitriani, S., & Yuliana, R. (2022). Hubungan Kepuasan Pasien dan Mutu Layanan Rumah Sakit. *Jurnal Kesehatan*, 10(1), 33–41.
- Rahman, M. A., & Kim, J. (2021). Data-driven hospital management through analytics. *IEEE Access*, 9, 128547–128558. <https://doi.org/10.1109/ACCESS.2021.3109685>
- Hasan, M., & Chowdhury, S. (2023). Big Data in hospital operations: A systematic review. *Health Informatics Journal*, 29(3), 1472–1485.
- Li, Y., & Li, X. (2022). Exploring hospital performance using EDA techniques. *Springer Healthcare Analytics*, 12(2), 65–79.
- Putri, I., & Wahyuni, N. (2023). Analisis Hubungan Moral Staf dan Efisiensi Kerja Rumah Sakit. *Jurnal Informasi Kesehatan*, 14(1), 70–83.
- Nugraha, D., & Rahman, F. (2024). Pengaruh Jumlah Pasien terhadap Moral Staf Rumah Sakit Umum. *Jurnal Teknologi Informasi dan Kesehatan*, 9(2), 88–97.
- Zhou, H., & Chen, L. (2021). Machine learning and data visualization for healthcare analytics. *IEEE Transactions on Biomedical Engineering*, 68(5), 1522–1534.
- Kurniawan, P., & Hidayat, T. (2023). Analisis Kinerja Rumah Sakit Menggunakan Python. *Jurnal Sistem Informasi*, 15(2), 40–52.
- Patel, R., & Jain, S. (2022). Optimizing hospital resources using data visualization. *IEEE Access*, 10, 55022–55033.