

# Teknik Data Science di Bidang Kesehatan : Menganalisis Faktor penting Ketidakhadiran Janji Medis Pasien

Amir Mahmud Husein<sup>1</sup> , Muhammad Khoiruddin Harahap<sup>2</sup>

<sup>1</sup>Universitas Prima Indonesia, Fakultas Teknologi dan Ilmu Komputer, Teknik Informatika

<sup>2</sup>Politeknik Ganesha Medan, Program Studi Teknik Informatika

<sup>1\*</sup>amirmahmud@unprimdn.ac.id

Received: 5 Januari 2022

Accepted: 9 Januari 2022

Published: 12 Januari 2022



\* Amir Mahmud Husein

**Keywords:** Ketidakhadiran janji medis, *Machine Learning*, Ilmu Data, Perawatan Kesehatan, Prediksi.

**DSI: Jurnal Data Science Indonesia** is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

**Abstrak :** Ketidakhadiran pasien dan pembatalan janji temu yang terlambat akan berdampak buruk pada kinerja keuangan dan kualitas layanan organisasi perawatan kesehatan, kesinambungan perawatan terganggu, kemanjuran obat tidak dapat dipantau secara teratur, layanan pencegahan dan pemeriksaan tidak dapat diberikan tepat waktu. Makalah ini menyajikan hasil analisis dalam kumpulan data menggunakan pendekatan ilmu data. Kerangka kerja menerapkan konsep CRIS-DM dengan tahapan pemahaman bisnis, pra-pemrosesan data, pembersihan data, rekayasa fitur, pemodelan dan evaluasi. Analisis data statistik deskriptif dan visual digunakan untuk memahami kumpulan data untuk menjawab delapan pertanyaan penting yang berhubungan dengan faktor ketidakhadiran janji temu pasien yang terjadwal, kemudian memprediksi apakah pasien akan datang pada hari janji temu yang dijadwalkan. Berdasarkan hasil evaluasi nilai matriks Accuracy, F1-Score dan ROC terhadap Lima model *machine learning* yang digunakan menghasilkan *Light GBM Classifier*, *Gradient Boosting Classifier* dan *XGBoost Classifier* adalah tiga model terbaik dengan akurasi diatas 78% sekitar 0.79968 (79.9%), 0.79883 (79.8%) dan 0.79806 (79.8%) pada masing-masing model.

## PENDAHULUAN

Pendekatan *Data Science* (ilmu data) dan Analisis Data menjadi semakin penting dari hari ke hari. Perusahaan dan eksekutif membutuhkan wawasan untuk membuat keputusan khususnya dalam layanan kesehatan (Leung et al. 2020). Kesehatan masyarakat dan fasilitas terkait sangat penting untuk kota dan masyarakat yang berkembang. Pemanfaatan sumber daya kesehatan yang optimal menghemat uang dan waktu, tetapi di atas semua itu, menyelamatkan nyawa lebih berharga (Schuler et al. 2020). Banyak pasien frustrasi dengan biaya perawatan kesehatan yang terus meningkat di berbagai negara berkembang dan kurangnya transparansi dalam sistem pembayaran medis: harga prosedur dan tes tidak diketahui sebelumnya, berbeda dari satu lokasi ke lokasi lain, dan sering kali tampak tidak proporsional dengan yang sebenarnya. tingkat perawatan yang diterima (Jödicke et al. 2019; Nomura et al. 2021). Alat untuk membantu menegosiasikan sistem ini dapat memungkinkan pasien menemukan solusi paling ekonomis untuk kebutuhan perawatan kesehatan (Huang et al. 2021; Morid et al. 2019). Selain itu, dampak pandemi COVID-19 telah membebani sumber daya medis yang ada. Khusus untuk penjadwalan janji pasien, sikap biasa dari melewatkan janji medis dapat menyebabkan kerusakan parah pada kesehatan pasien (Daghistani et al. 2020).

Tingginya tingkat pasien yang tidak datang untuk pemeriksaan dan janji medis adalah masalah yang berulang dalam perawatan kesehatan (Marbough et al. 2020). Ketidakhadiran pasien dan pembatalan janji temu yang terlambat akan berdampak buruk pada kinerja keuangan dan kualitas layanan organisasi

perawatan kesehatan (Dashtban and Li 2021), kesinambungan perawatan terganggu (Davies et al. 2016). Kemanjuran obat tidak dapat dipantau secara teratur (Ahmadi et al. 2019). Layanan pencegahan dan pemeriksaan tidak dapat diberikan tepat waktu. Penyakit akut lebih mungkin untuk tidak diobati dan menjadi kondisi kronis dengan komplikasi. Singkatnya, melewatkan janji temu bisa sangat merugikan kesehatan seseorang (L. H. A. Salazar et al. 2021). Di sisi lain, pasien yang menginginkan janji temu sesegera mungkin tidak bisa mendapatkannya. Jadi, ada dua kerugian: hilangnya waktu untuk dokter dan hilangnya janji untuk orang yang membutuhkan.

Dalam penelitian ini, fokus utama melakukan statistik deskriptif dan visual untuk menjawab pertanyaan dan menganalisis data janji temu yang dijadwalkan untuk menemukan faktor-faktor apa yang penting untuk masalah ini, kemudian memprediksi apakah pasien akan datang untuk janji temu yang dijadwalkan. Menemukan faktor-faktor alasan mengapa beberapa pasien tidak datang ke janji temu yang dijadwalkan dan kemungkinan melewatkan janji temu merupakan salah satu masalah penting yang harus dipertimbangkan oleh pemangku kepentingan di masa depan. Analisis difokuskan pada penemuan yang mempengaruhi pasien untuk datang atau tidak datang ke janji.

## TINJAUAN LITERATUR

Melewatkan janji temu pasien memberi beban berat pada sistem medis dan berdampak negatif pada perawatan pasien. Ketidakhadiran pasien dapat menyebabkan kesulitan penjadwalan dan operasional klinik maupun rumah sakit, dapat membatasi akses ke perawatan dan tentunya akan mengurangi produktivitas layanan kesehatan (Davies et al. 2016). Banyak penelitian telah dilakukan dengan mengusulkan model untuk memahami faktor risiko yang menentukan prediksi ketidakhadiran dalam banyak situasi yang berbeda. Sebagian besar penelitian di bidang ini melihat faktor-faktor yang berkontribusi terhadap ketidakhadiran dalam disiplin dan semua janji di rumah sakit atau klinik maupun praktek umum (Dashtban and Li 2021).

Beberapa pekerjaan menggunakan pendekatan *machine learning* untuk memprediksi ketidakhadiran pasien dimana model *random forest* merupakan model yang menghasilkan tingkat akurasi yang terbaik (Dashtban and Li 2021; Praveena, Krupa, and Saipreethi 2019; L. H. Salazar et al. 2020). Dalam penelitian lainnya, (Davies et al. 2016) mengusulkan *Multifactor analysis of variance* (ANOVA) dengan tingkat ketidakhadiran sebagai variabel dependen, dan jenis kelamin, kelompok usia, usia janji temu, status pasien baru, dan jalur layanan sebagai faktor. Mereka melaporkan bahwa laki-laki memiliki tingkat ketidakhadiran yang lebih tinggi daripada perempuan hingga usia 65 tahun, di mana laki-laki dan perempuan menunjukkan tingkat yang sama. Rata-rata tingkat ketidakhadiran menurun dengan usia sampai 75-79, dimana tingkat meningkat. Seiring bertambahnya usia janji temu, pria dan pasien baru mengalami peningkatan tingkat ketidakhadiran. Pasien yang lebih muda sangat rentan untuk tidak hadir seiring bertambahnya usia janji temu.

Banyak peneliti telah mengeksplorasi berbagai teknik untuk meningkatkan akurasi model prediktif dalam konteks ketidakhadiran medis. Hasilnya menegaskan bahwa dampak dari memahami perilaku pasien terkait ketidakhadirannya dalam janji medis memiliki efek positif. Aspek lain yang perlu diperhatikan adalah bahwa tergantung pada spesialisasi medis, pasien memiliki perilaku yang berbeda dan harus diperhitungkan ketika mengembangkan model prediktif.

## BAHAN DAN METODE

### Himpunan Data

Dataset yang digunakan dalam penelitian ini terdiri dari data [Medical Appointment No Shows](#) yang diperoleh dan diambil dari kaggle (JoniHoppen 2017). Kumpulan data ini merupakan informasi dari 110.527 janji medis di negara Brasil dan difokuskan pada pertanyaan apakah pasien datang untuk janji temu atau tidak. Pertanyaan utama dalam penelitian ini adalah menemukan jawaban mengapa 30% pasien melewatkan janji temu yang dijadwalkan. Selanjutnya, menggunakan lima model machine learning untuk memprediksi faktor terpenting yang mempengaruhi kehadiran pasien. Data janji temu digabungkan dari lebih dari 45 klinik dan

rumah sakit di tingkat kotamadya di Brasil antara 29/4/2016 dan 6/6/2016. Dalam kumpulan data awal, setiap file terdiri dari 14 atribut, yaitu:

1. PatientId = Identifikasi pasien
2. AppointmentID = Identifikasi setiap janji
3. Gender = Pria atau Wanita
4. ScheduledDay = Hari janji temu yang sebenarnya
5. AppointmentDay = Hari seseorang menelepon atau mendaftarkan janji temu
6. Age = Umur pasien
7. Neighbourhood = Tempat janji temu.
8. Scholarship = Jika orang tersebut menerima bantuan kesejahteraan (Benar Salah)
9. Hipertensi = Benar atau Salah
10. Diabetes = Benar atau Salah
11. Alkoholisme = Benar atau Salah
12. Handicap = Benar atau Salah
13. SMS\_received = 1 atau lebih pesan yang dikirim ke pasien.
14. No-show = Benar atau Salah.

Kumpulan data ini difokuskan menyaring data untuk menemukan faktor-faktor apa yang penting kemudian memprediksi apakah pasien akan datang untuk janji mereka dan melakukan statistik deskriptif dan visual untuk menjawab pertanyaan dan melihat wawasan tersembunyi.

### **Metodologi**

Penelitian ini menggunakan pendekatan *data science* dengan menggunakan metode CRIS-DM yang meliputi beberapa tahapan, yaitu:

#### **1. Permasalahan Bisnis**

Pasien yang tidak datang tepat waktu menyebabkan terganggunya proses penjadwalan dan dapat menyebabkan dokter bekerja ekstra. Sebaliknya, apabila tidak datang sama sekali, akan membuang waktu dokter. Dalam kedua kasus, efisiensi sistem pelayanan kesehatan akan berkurang. Ketidakhadiran dapat menyebabkan penurunan kualitas perawatan, hilangnya produktivitas, kerugian finansial, dan gangguan hasil medis pasien. Dengan demikian, penelitian ini menjadi jawaban atas beberapa pertanyaan untuk membantu dalam menemukan faktor-faktor penting tentang ketidakhadiran pasien, yaitu:

- a) Apakah jenis kelamin pasien ada hubungannya dengan kehadiran?
- b) Berapa waktu tunggu rata-rata antara tanggal penjadwalan dan tanggal janji? Dan apakah ada hubungan antara waktu tunggu dan janji tidak hadir?
- c) Apakah usia tua lebih menjaga kesehatan daripada masa muda?
- d) Apakah jenis penyakit mempengaruhi kunjungan pasien?
- e) Berapa banyak pasien yang hadir dan tidak menghadiri janji temu melalui SMS yang diterima?
- f) Berapa proporsi janji temu versus tidak hadir?
- g) Apakah lingkungan berperan dalam membuat pasien tidak hadir?
- h) Faktor-faktor apa yang penting untuk memprediksi apakah pasien akan datang untuk janji temu yang dijadwalkan?

#### **2. Pra-pemrosesan Data**

Dataset berisi fitur target, yang diidentifikasi oleh variabel "No-Show" di mana: "true" mewakili pasien yang menghadiri janji medis, dan "false" mewakili pasien yang tidak muncul. Pada kolom ID\_PatientId tidak unik, yang menunjukkan bahwa pasien yang sama memiliki beberapa janji temu selama periode 6 minggu. Untuk menghindari kebocoran data (yang akan terjadi jika data pasien yang sama digunakan untuk validasi dan pengujian), hanya akan menyertakan janji temu terakhir pasien dalam model. Selanjutnya kolom PatientId ini akan diubah menjadi format bilangan bulat. Kolom "No-show" muncul dalam format string hal ini akan berdampak pada penggunaan memori. Jadi, kolom ini akan diganti nama menjadi "no\_show" dan mengubahnya menjadi tipe data bilangan bulat (1, 0) untuk digunakan dalam perhitungan statistik. Terakhir,

kebutuhan analisis selanjutnya perlu menambahkan kolom baru dari *ScheduledDay* dan *AppointmentDay* yaitu:

- ScheduledHour: jam dari Hari Terjadwal
- AppointmentTime: waktu Hari Janji
- ScheduleDate: tanggal Hari yang Dijadwalkan
- AppointmentDate: tanggal Hari Janji
- AppointmentWeekDay: hari dalam minggu dari Appointment Day
- AppointmentMonth: bulan dari Appointment Day
- AppointmentMonth: bulan dari Appointment Day
- WaitingDays: beberapa hari menunggu sebelum Hari Pengangkatan

### 3. Pembersihan Data

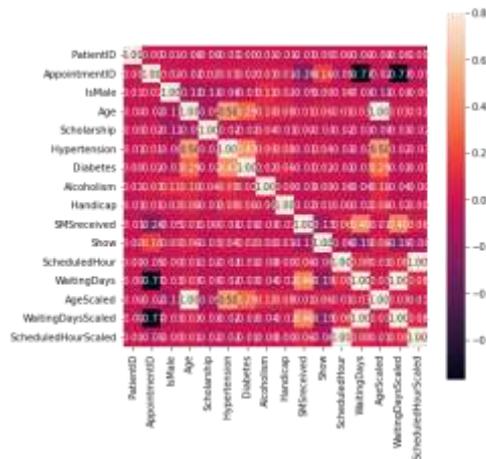
Kumpulan data kemudian diproses sebelumnya, menghilangkan catatan yang tidak lengkap dan salah, menangani nilai yang hilang dan memecahkan ketidakkonsistenan. Transformasi antara tipe data kategorikal atau numerik dilakukan dengan cara normalisasi atau penskalaan. Dalam normalisasi, penskalaan ulang nilai atribut dari rentang asli untuk menjaga rentang nilai antara [0, 1]. Dalam diskritisasi, atribut numerik usia diubah menjadi atribut kategoris dengan memilih lima sebagai kelompok. Selanjutnya, mengidentifikasi faktor-faktor yang paling penting dalam prediksi dan secara signifikan mempengaruhi kinerja model. Metode perolehan informasi digunakan untuk mengurutkan faktor berdasarkan dampaknya pada pasien yang datang dan tidak datang dan menghilangkan faktor yang tidak relevan.

### 4. Eksplorasi data analisis

Eksplorasi data adalah salah satu langkah dalam mengeksplorasi dan memvisualisasikan data sehingga memungkinkan untuk mengidentifikasi pola yang terkandung dalam kumpulan data. Dengan cara ini, akan mengaktifkan kesimpulan yang dapat berkontribusi pada pemahaman masalah yang dimaksud. Dalam tahapan ini digunakan statistik deskriptif. Penggunaan statistik deskriptif memungkinkan untuk meringkas karakteristik utama dari karakteristik numerik dataset (kontinu atau diskrit), seperti top, frekuensi, mean dan standar deviasi (std). Namun, dalam skenario dengan banyak variabel kategori, pendekatan lain mungkin lebih tepat. Ikhtisar atribut kategoris ada pada Gambar 1. Gambar 2 menyajikan peta panas matriks korelasi. Gambar 2 menggambarkan korelasi antara semua variabel. Secara khusus, ini menunjukkan korelasi positif atau langsung di mana variasi satu karakteristik secara langsung mempengaruhi yang lain dan korelasi negatif atau terbalik di mana fluktuasi satu atribut berbanding terbalik dengan atribut lainnya.

	PatientID	AppointmentID	Age	Scholarship	Hypertension	Diabetes	Alcoholism	Handicap	SMS_received
count	1.105260e+05	1.105260e+05	110526.000000	110526.000000	110526.000000	110526.000000	110526.000000	110526.000000	110526.000000
mean	1.474934e+14	5.675304e+06	37.089219	0.098266	0.197248	0.071965	0.030400	0.022248	0.321029
std	2.560943e+14	7.129544e+04	23.110026	0.297676	0.397923	0.258266	0.171686	0.161543	0.468874
min	3.921784e+04	5.030230e+06	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	4.172536e+12	5.840285e+06	18.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	3.173184e+13	5.690572e+06	37.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	9.438963e+13	5.725523e+06	55.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
max	9.999816e+14	5.790484e+06	115.000000	1.000000	1.000000	1.000000	1.000000	4.000000	1.000000

Gambar 1 Statistik deskriptif data



Gambar 2 Korelasi antar variabel

Pada gambar 2 merupakan *heat map* korelasi antar variabel dimana variabel Hipertensi-Usia, Hipertensi-Diabetes dan Diabetes-Usia merupakan variabel yang berkorelasi kuat, namun tidak ada korelasi kuat antara fitur apa pun dengan no-show.

### 5. Rekayasa Fitur

Rekayasa fitur digunakan untuk mengkode ulang fitur *datetime* dan menyimpan informasi yang diambil dari janji sebelumnya. Ketidakhadiran dan janji temu sebelumnya dihitung dengan mengurangkan 1 dari jumlah total ketidakhadiran dari janji temu masing-masing pasien. Tanggal yang dijadwalkan dan tanggal janji temu dibagi menjadi hari dalam seminggu dan hari dalam sebulan. Waktu yang dijadwalkan juga dibagi menjadi jam dalam sehari. Perbedaan hari adalah perbedaan dalam hal hari antara tanggal yang dijadwalkan dan tanggal janji. Kondisi total adalah jumlah dari hipertensi, diabetes, cacat dan alkoholisme. Dengan kata lain, ini menunjukkan jumlah kondisi yang diderita pasien.

### 6. Pemodelan

Pembuatan model dijalankan dalam bahasa pemrograman python 3 dengan menggunakan jupyter notebook yang melibatkan lima model machine learning yaitu *Gradient Boosting Classifier*, *Random Forest Classifier*, *Light GBM Classifier*, *XGBoost Classifier* dan *AdaBoost*.

### 7. Evaluasi

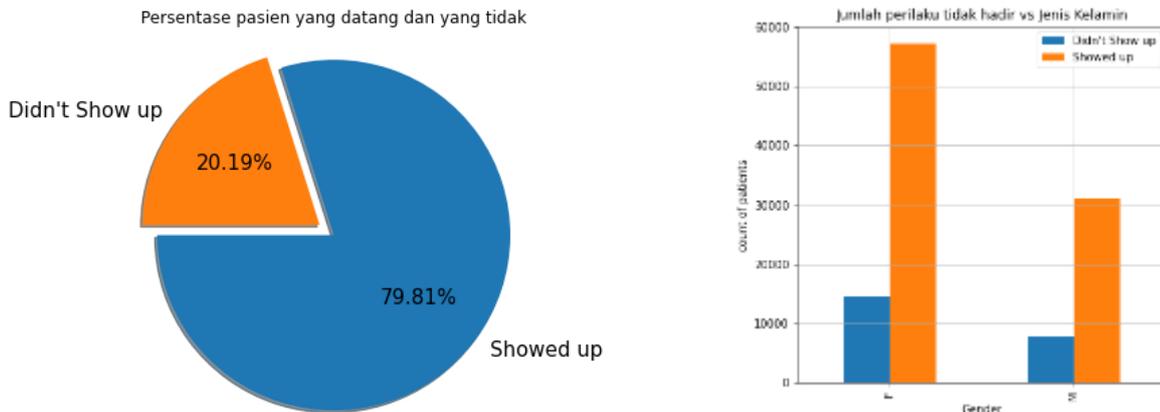
Keseluruhan model *machine learning* yang digunakan dalam penelitian ini akan dievaluasi dalam menentukan model terbaik berdasarkan nilai matrik *Accuracy*, *F1-score* dan *ROC (Receiver Operating Characteristics)*. *ROC* merupakan alat ukur *performance* untuk *classification problem* dalam menentukan *threshold* dari suatu model.

## HASIL PENELITIAN

Bagian ini menyajikan hasil yang diperoleh setelah melakukan pra-pemrosesan data, pembersihan data, eksplorasi data analisis, melakukan pemilihan fitur, dan menyetel hyperparameter tiga model terbaik dari hasil evaluasi model default. Hasil yang disajikan pada bagian ini terbagi menjadi dua bagian. Bagian pertama merupakan visualisasi hasil jawaban atas pertanyaan masalah bisnis.

### Hasil Analisis

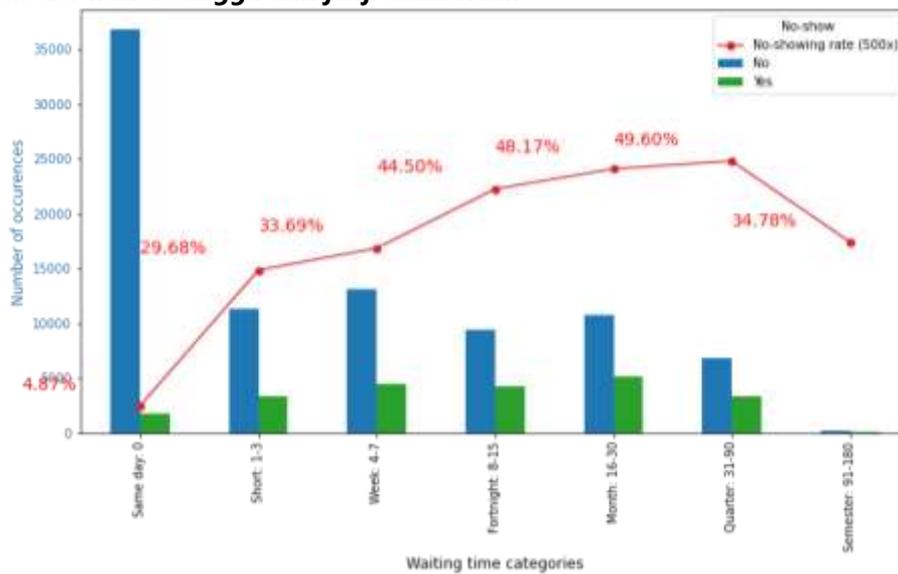
#### 1. Apakah jenis kelamin pasien ada hubungannya dengan kehadiran?



Gambar 3 Visualisasi hubungan jenis kelamin dengan kehadiran

Pada gambar 3 merupakan grafik yang menunjukkan bahwa kebanyakan wanita muncul pada janji mereka dan begitu juga pria. Kita perlu menghitung proporsi oleh masing-masing jenis kelamin untuk dapat membandingkan. Persentase Wanita yang datang pada janji adalah sekitar 79,69% dari semua pasien wanita dan persentase Laki-laki yang datang sekitar 80,03% dari semua pasien laki-laki. Kedua jenis kelamin memiliki komitmen yang sama terhadap jadwal medis. Persentase pasien yang datang pada janji mereka mewakili 79,8% dan 20,2% pasien tidak datang.

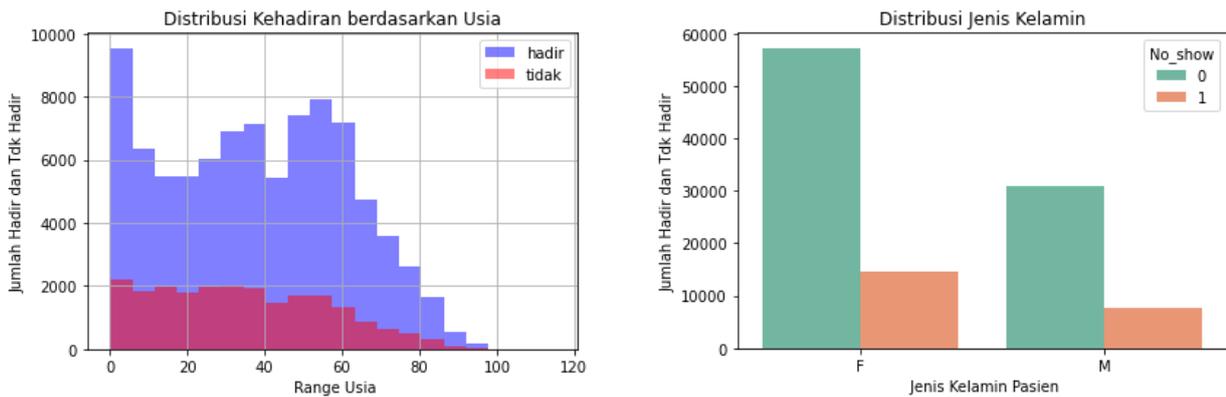
**2. Berapa waktu tunggu rata-rata antara tanggal penjadwalan dan tanggal janji? Dan apakah ada hubungan antara waktu tunggu dan janji tidak hadir?**



Gambar 4 Distribusi waktu tunggu

Melalui grafik pada gambar 4 di atas, menjadi jelas bahwa tingkat ketidakhadiran meningkat seiring dengan bertambahnya waktu menunggu. Ini mencapai tingkat yang lebih rendah ketika kehadiran terjadi pada hari yang sama dengan yang dijadwalkan. Kami kemudian dapat mengaitkannya dengan perawatan darurat atau bahkan ketika pasien pergi ke unit kesehatan tanpa penjadwalan sebelumnya. Menarik untuk diperhatikan bahwa setelah seperempat menunggu ada sedikit pengembalian ke ketekunan pasien, ketika tingkat ketidakhadiran turun dari 49,6% menjadi %34,78.

**3. Apakah usia tua lebih menjaga kesehatan daripada masa muda?**

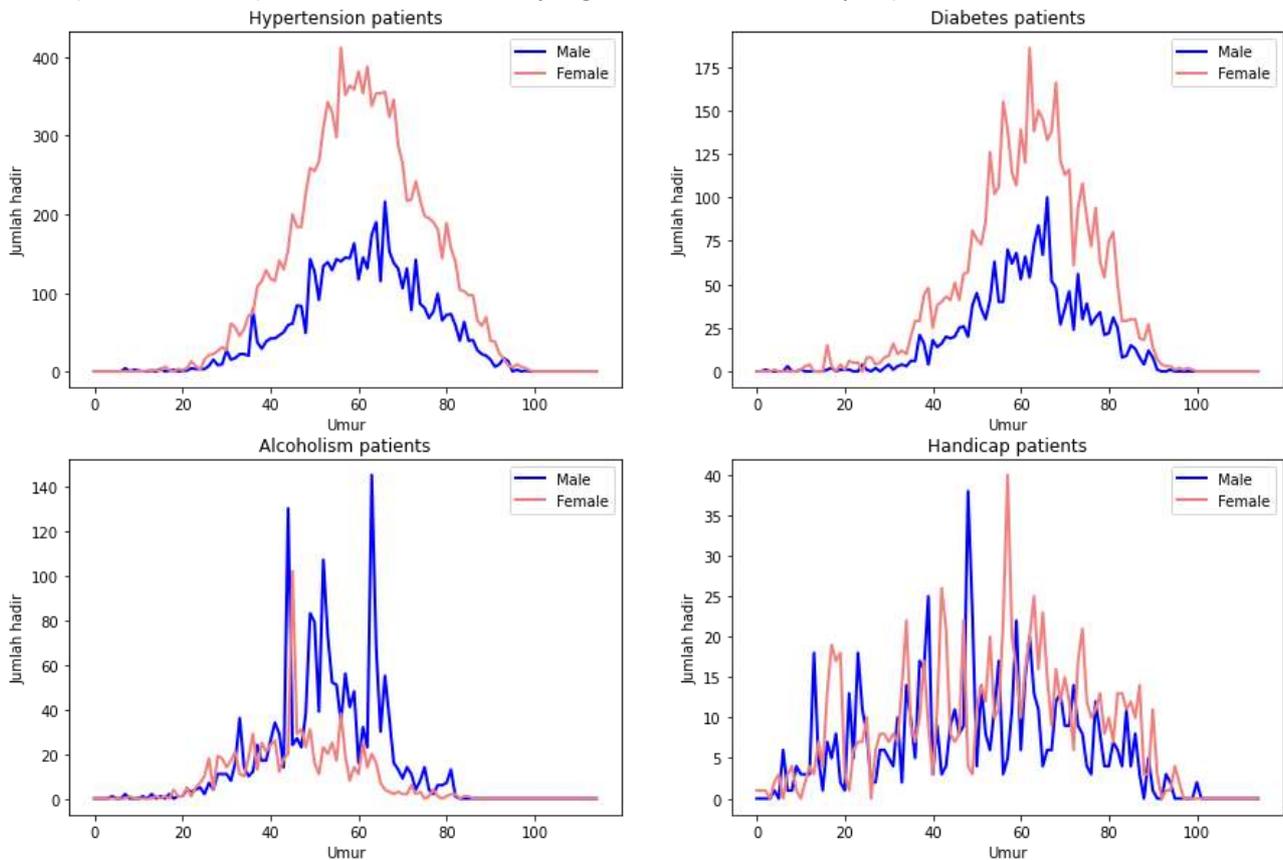


Gambar 5 Plot hubungan usia dengan kehadiran

Jadi, berdasarkan plot gambar 5, memang terlihat seperti anak-anak yang sangat kecil memiliki janji menghadiri pertemuan yang lebih tinggi daripada Usia dari 45-65 tahun. Pasien yang berusia rentang 23 hingga 40 dan 0-5 memiliki jumlah yang lebih tinggi tidak menghadiri janji. Jumlah perempuan yang menghadiri lebih banyak dari laki-laki dan perempuan yang tidak hadir lebih banyak dibandingkan laki-laki.

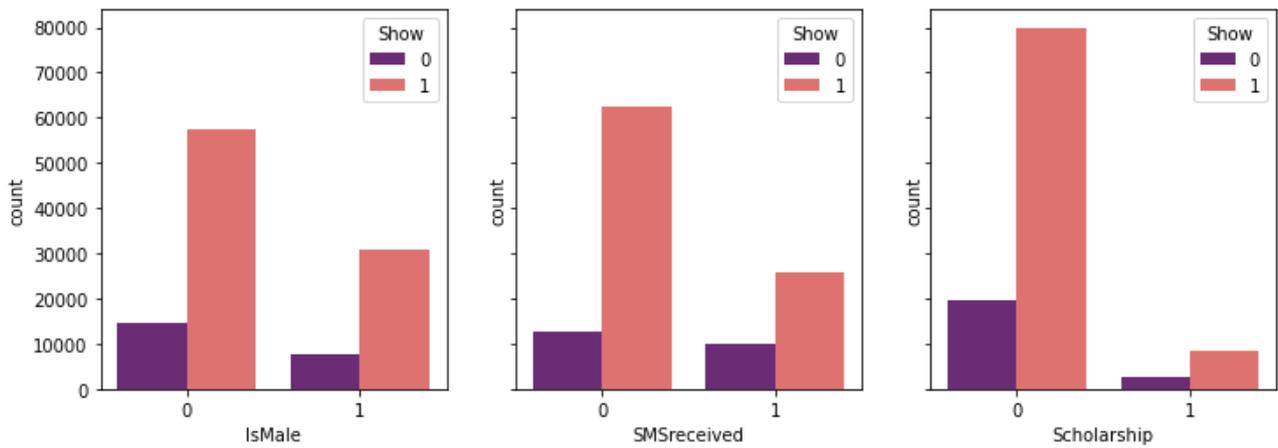
**4. Apakah jenis penyakit mempengaruhi kunjungan pasien?**

Berdasarkan gambar 6, menemukan bahwa pasien Hipertensi dan Diabetes lebih tinggi menghadiri janji temu bila dibandingkan dengan penyakit lain. Usia berkisar antara 40 hingga 80 dengan konversi wanita lebih baik daripada Pria. Pasien alkoholisme dari Laki-laki yang menghadiri janji temu lebih banyak dibandingkan dengan perempuan dan mereka berkisar antara 45 hingga 65. Kunjungan Pria dan Wanita untuk pasien Handicap berada dalam kisaran yang sama. Tidak ada banyak perbedaan.



Gambar 6 Grafik hubungan jenis penyakit terhadap kehadiran

**5. Berapa banyak pasien yang hadir dan tidak menghadiri janji temu melalui SMS yang diterima?**

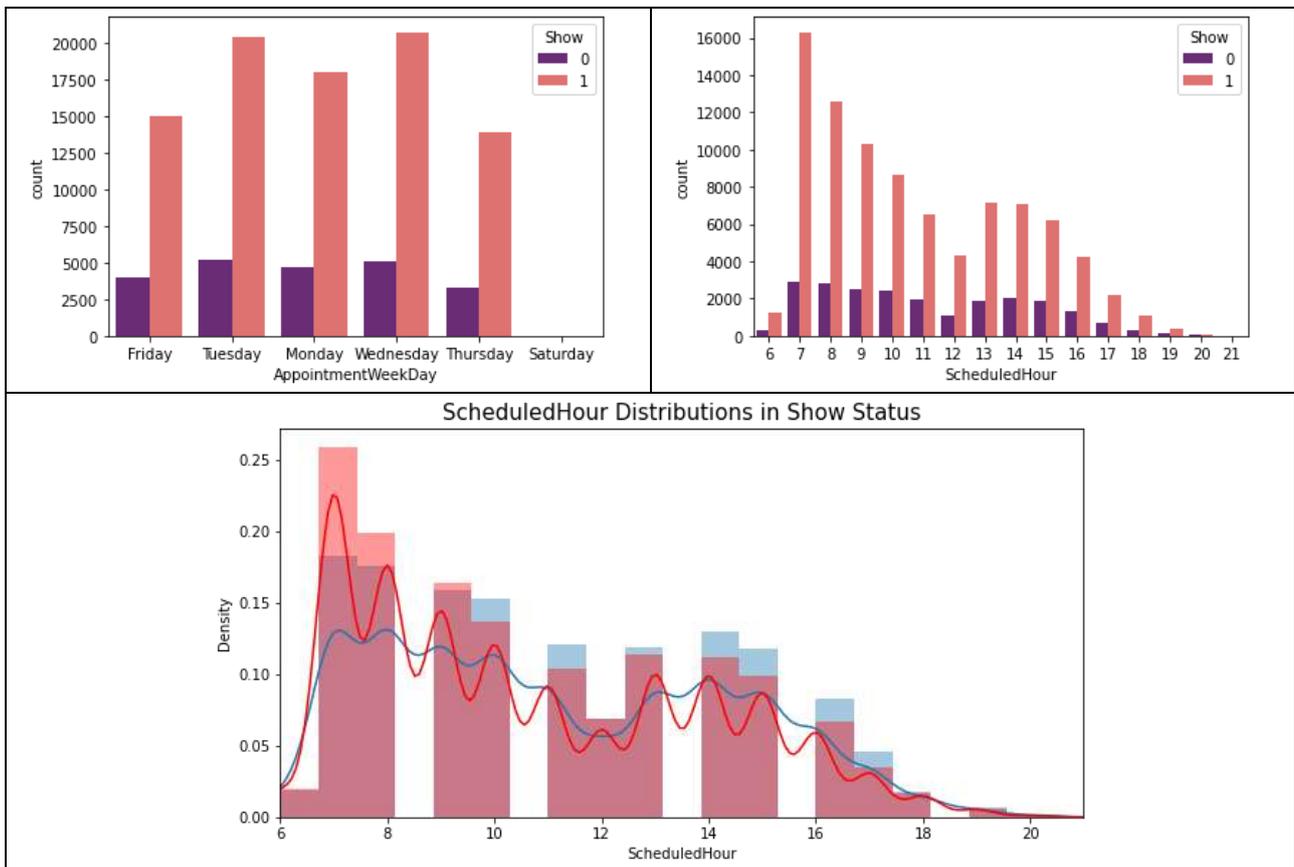


Gambar 7 Grafik hubungan SMS dengan kehadiran

Jika melihat gambar di atas, perempuan membuat janji lebih dari laki-laki. Anehnya, SMS pengingat pesan teks mungkin tidak efektif dalam mengurangi tingkat ketidakhadiran karena rasio ketidakhadiran lebih tinggi pada grup yang menerima SMS daripada grup yang tidak menerima.

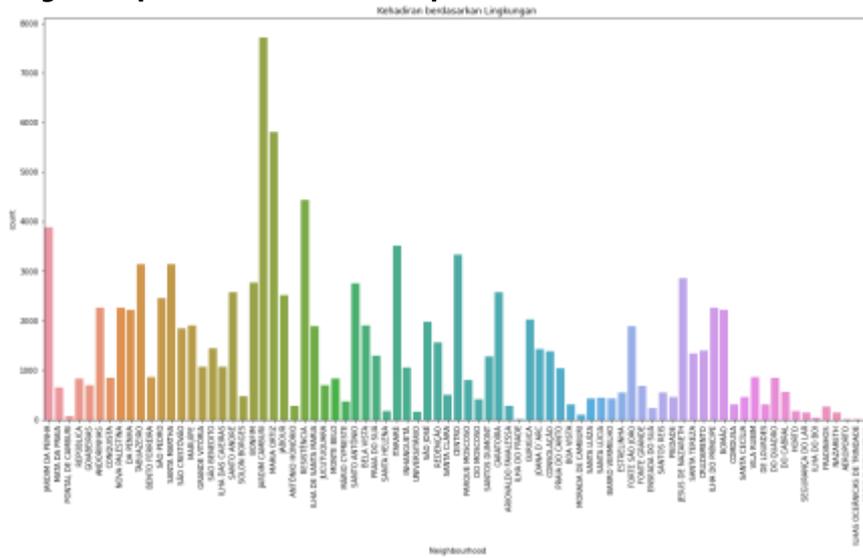
**6. Berapa proporsi janji temu versus tidak hadir?**

Distribusi janji antar hari dalam seminggu (Senin-Jumat) hampir sama dengan lebih sedikit kunjungan pada hari Kamis dan Jumat. Ada 24 kunjungan pada hari Sabtu dan tidak ada pada hari Minggu. Rata-rata 10 hari pasien menunggu janji. 50% pasien menunggu hingga 4 hari dan 75% hingga 15 hari untuk membuat janji.



Gambar 8 Grafik distribusi janji terjadwal

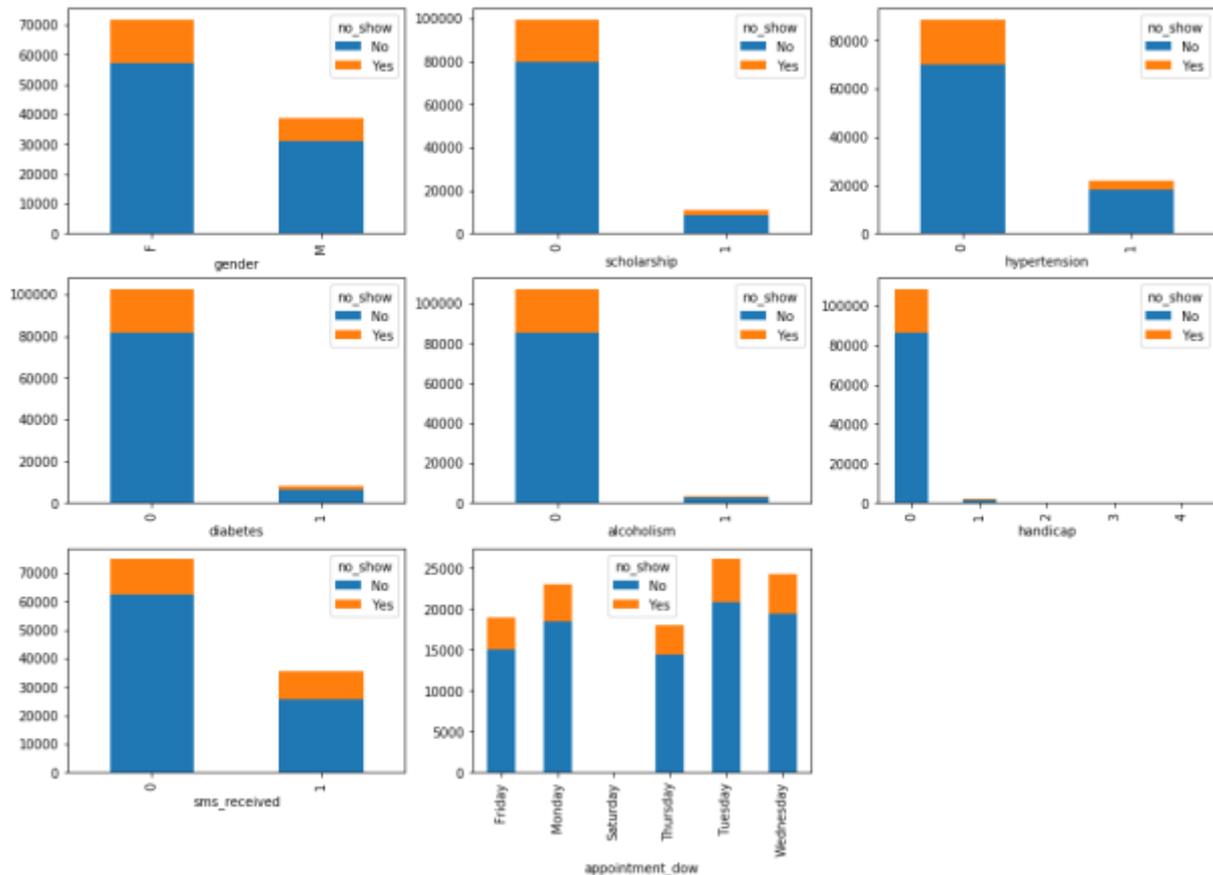
**7. Apakah lingkungan berperan dalam membuat pasien tidak hadir?**



Gambar 9 Grafik hubungan lingkungan dengan kehadiran

Pada gambar grafik tersebut menghasilkan bahwa beberapa lingkungan memiliki lebih banyak orang yang datang untuk janji mereka dan ini menunjukkan bahwa daerah ini memiliki peningkatan penyakit.

**8. Faktor-faktor apa yang penting untuk memprediksi apakah pasien akan datang untuk janji temu yang dijadwalkan?**



Gambar 10 Plot faktor penting kehadiran pasien

Dari hasil gambar di atas, dapat diamati bahwa hampir semua variabel kategori, distribusi show / no-show untuk kategori yang berbeda terlihat sangat mirip. Akhirnya, menganalisis dataset dan melakukan statistik deskriptif dan visual untuk menjawab pertanyaan faktor yang paling penting untuk memprediksi apakah pasien akan hadir pada janji temu adalah pasien yang tidak terdaftar dalam program kesejahteraan. Sebagian besar orang yang memiliki Beasiswa kemungkinan besar akan melewatkan janji temu dan pasien yang tidak memiliki beasiswa muncul untuk janji temu. Selanjutnya, jenis kelamin adalah perempuan, Usia dari (45–65) dan anak kecil karena lebih tua pengasuh bertanggung jawab untuk mereka dan jenis penyakit Pasien Hipertensi dan Diabetes menghadiri janji jika dibandingkan dengan penyakit lain. Namun, pengamatan ini mengkonfirmasi sekitar 20% tingkat ketidakhadiran untuk sebagian besar kategori. Hal ini harus diselidiki lebih lanjut dengan analisis statistik yang lebih maju untuk berpotensi mengungkapkan wawasan dan korelasi baru.

### Hasil Pemodelan

Bagian kedua merupakan hasil evaluasi yang menggambarkan kinerja saat menguji lima model berdasarkan nilai matrik *Accuracy*, *F1-score* dan *ROC*.

Tabel 1 Evaluasi Hasil Prediksi

Model	Accuracy	F1-Score	ROC
Gradient Boosting Classifier	0.79883	0.007843	0.501361
Random Forest Classifier	0.77029	0.274255	0.541362
Light GBM Classifier	0.79968	0.047706	0.509755
XGBoost Classifier	0.79806	0.115315	0.523906
AdaBoost	0.79702	0.838773	0.557904

Berdasarkan hasil evaluasi nilai matrik *Accuracy* pada tabel 1, model *Random Forest Classifier* menghasilkan

nilai terendah dibandingkan model lainnya. Namun model ini lebih tinggi nilai ROC dari yang lain. Model *Light GBM Classifier*, *Gradient Boosting Classifier* dan *XGBoost Classifier* adalah tiga model terbaik yang menghasilkan tingkat akurasi terbaik sekitar 0.79968 (79%), 0.79883 (78%) dan 0.79806 (78%) pada masing-masing model. Ketiga model masih perlu di tingkatkan dengan menggunakan penyetelan optimasi Hyperparameter seperti GridSearchCV, RandomizedSearchCV atau Bayesian.

## Diskusi

Berdasarkan evaluasi hasil analisis sebelumnya, terdapat beberapa temuan penting, yaitu penjadwalan kunjungan dimulai pada 10-11-2015 dan berakhir pada 08-06-2016. Distribusi janji antar hari dalam seminggu (Senin-Jumat) hampir sama dengan lebih sedikit kunjungan pada hari Kamis dan Jumat. Terdapat 24 kunjungan pada hari Sabtu dan tidak ada pada hari Minggu. Rata-rata 10 hari pasien menunggu janji. 50% pasien menunggu hingga 4 hari dan 75% hingga 15 hari untuk membuat janji. Waktu tunggu terlama adalah 179 hari. Hampir 40 ribu pasien menjadwalkan kunjungan di hari yang sama. Dari semua pasien yang menjadwalkan janji temu untuk hari yang sama (total 38561), 1792 pasien tidak datang (5%). Pasien rata-rata berusia 37 tahun, 25% pasien berusia di bawah 18 tahun dan sebagian besar berusia di bawah 55 tahun. Sebagian besar pasien bukan pecandu alkohol dan diantara pasien pecandu alkohol tidak memiliki riwayat penyakit diabetes. Rata-rata 20% dari janji tidak terjawab dimana 71.831 janji yang dibuat oleh perempuan 20% tidak terjawab atau sekitar 14.588 janji. Pasien laki-laki memiliki janji temu sebanyak 38.685 dan 7.723 tidak terjawab.

Mempertimbangkan bahwa rentang waktu kunjungan yang ditentukan mencakup lebih dari 3 bulan, janji temu kemungkinan besar setiap pemeriksaan atau setiap kunjungan dokter spesialis. Jadi, dalam satu kunjungan pasien di rumah sakit, mungkin ada beberapa janji temu yang dijadwalkan. Salah satu alasan ketidakhadiran adalah bahwa pasien mungkin terlalu lelah untuk mengikuti semua pemeriksaan selama kunjungan tertentu, atau jam buka tidak cukup untuk muncul di semua janji temu. Mungkin juga ada alasan lain. Tingginya jumlah janji temu dalam waktu yang begitu singkat harus dikonsultasikan dengan pihak manajemen untuk melakukan (atau tidak) sebagai analisis tambahan di bidang ini. Semakin pendek masa tunggu semakin banyak pasien yang datang. Pasien yang menjadwalkan janji temu untuk hari yang sama jauh lebih mungkin untuk datang sekitar 17% dari tidak datang.

## KESIMPULAN

Penelitian menyajikan analisis statistik deksriptif untuk menjawab beberapa pertanyaan penting berhubungan dengan ketidakhadiran pasien sesuai janji temu yang terjadwal. Hasilnya telah disajikan dalam bentuk visualisasi dimana faktor-faktor yang mempengaruhi ketidakhadiran pasien dengan lebih jelas. Jenis kelamin dan usia adalah faktor yang paling penting seperti yang terlihat sebelumnya bahwa perempuan dan pemuda lebih sering hadir daripada pria dan orang tua. Lingkungan dan hipertensi datang setelah jenis kelamin dan usia karena ada beberapa lingkungan yang penyakitnya menyebar dan pasien dengan hipertensi cenderung hadir sesuai dengan jadwal. Jadi, sangat perlu mencari lebih banyak faktor untuk membantu pasien mengingat janji mereka untuk hadir. Terakhir, penelitian ini menyajikan hasil analisis dalam kumpulan data dengan mengelola beberapa masalah seperti menyatukan nama, menghapus data yang salah, menambahkan fitur baru berdasarkan data yang ada dan menyelidiki sebagian besar variabel independen, membuat beberapa pengamatan yang membandingkannya satu sama lain serta dengan variabel dependen (no\_show). Karena ini hanya analisis eksplorasi, banyak korelasi potensial yang mungkin masih belum terungkap. Data harus diselidiki lebih lanjut dengan analisis statistik yang lebih maju untuk berpotensi mengungkapkan wawasan dan korelasi baru.

## REFERENCES

- Ahmadi, Ehsan et al. 2019. "A Metaheuristic-Based Stacking Model for Predicting the Risk of Patient No-Show and Late Cancellation for Neurology Appointments." *IISE Transactions on Healthcare Systems Engineering* 9(3): 272–91. <https://doi.org/10.1080/24725579.2019.1649764>.
- Daghistani, Tahani, Huda AlGhamdi, Riyad Alshammari, and Raed H. AlHazme. 2020. "Predictors of Outpatients' No-Show: Big Data Analytics Using Apache Spark." *Journal of Big Data* 7(1): 1–15. <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00384-9> (December 28, 2021).
- Dashtban, M., and Weizi Li. 2021. "Predicting Non-Attendance in Hospital Outpatient Appointments Using Deep Learning Approach." *Health Systems* 00(00): 1–22. <https://doi.org/10.1080/20476965.2021.1924085>.
- Davies, Michael L. et al. 2016. "Large-Scale No-Show Patterns and Distributions for Clinic Operational Research." *Healthcare (Switzerland)* 4(1). /pmc/articles/PMC4934549/ (December 28, 2021).
- Huang, He, Po-Chou Shih, Yuelan Zhu, and Wei Gao. 2021. "An Integrated Model for Medical Expense System Optimization during Diagnosis Process Based on Artificial Intelligence Algorithm." *Journal of Combinatorial Optimization*: 1. /pmc/articles/PMC8235905/ (November 26, 2021).
- Jödicke, Annika M. et al. 2019. "Prediction of Health Care Expenditure Increase: How Does Pharmacotherapy Contribute?" *BMC Health Services Research* 19(1): 1–11. <https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-019-4616-x> (November 29, 2021).
- JoniHoppen. 2017. "Medical Appointment No Shows | Kaggle." *Kaggle*. <https://www.kaggle.com/joniarroba/noshowappointments> (May 29, 2021).
- Leung, Carson K. et al. 2020. "Data Science for Healthcare Predictive Analytics." *PervasiveHealth: Pervasive Computing Technologies for Healthcare*.
- Marbough, Dounia et al. 2020. "Evaluating the Impact of Patient No-Shows on Service Quality." *Risk Management and Healthcare Policy* 13: 509–17. /pmc/articles/PMC7280239/ (December 28, 2021).
- Morid, Mohammad Amin et al. 2019. "Healthcare Cost Prediction: Leveraging Fine-Grain Temporal Patterns." *Journal of Biomedical Informatics* 91: 103113.
- Nomura, Yoshiaki et al. 2021. "Does Last Year's Cost Predict the Present Cost? An Application of Machine Learning for the Japanese Area-Basis Public Health Insurance Database." *International Journal of Environmental Research and Public Health* 18(2): 1–11. <https://www.mdpi.com/1660-4601/18/2/565/htm> (November 26, 2021).
- Praveena, M. D.Anto, J. Sai Krupa, and S. Saipreethi. 2019. "Statistical Analysis of Medical Appointments Using Decision Tree." *5th International Conference on Science Technology Engineering and Mathematics, ICONSTEM 2019* 1: 59–64.
- Salazar, L.H. et al. 2020. "Prediction of Attendance at Medical Appointments Based on Machine Learning | Predicção de Comparecimento Às Consultas Médicas Com Base Em Aprendizagem de Máquina." *Iberian Conference on Information Systems and Technologies, CISTI 2020-June(June)*: 24–27.
- Salazar, Luiz Henrique A. et al. 2021. "Application of Machine Learning Techniques to Predict a Patient's No-Show in the Healthcare Sector." *Future Internet* 14(1): 3. <https://www.mdpi.com/1999-5903/14/1/3/htm> (December 28, 2021).
- Schuler, Alejandro et al. 2020. "Assessment of Value of Neighborhood Socioeconomic Status in Models That Use Electronic Health Record Data to Predict Health Care Use Rates and Mortality." *JAMA Network Open* 3(10): e2017109–e2017109. <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2772019> (November 29, 2021).