



**Analisa Data Rekam Medis Menggunakan Teknik Data Mining Association
Rules Dengan Algoritma Clustering**

TUGAS AKHIR

Muhamad Iqbal
41516010171

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2020**



**Analisa Data Rekam Medis Menggunakan Teknik Data Mining Association
Rules Dengan Algoritma Clustering**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:
Muhamad Iqbal
41516010171

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA

2020
UNIVERSITAS
MERCU BUANA

LEMBAR PERNYATAAN ORISINALITAS

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : Muhamad Iqbal

Nama : 41516010171

Judul Tugas Akhir : Analisa Data Rekam Medis Menggunakan Teknik Data Mining Association Rules Dengan Algoritma Clustering

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 02 September 2020



UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Muhamad Iqbal
NIM : 41516010171
Judul Tugas Akhir : Analisa Data Rekam Medis Menggunakan Teknik
Data Mining Association Rules Dengan Algoritma
Clustering

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).


Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 02 September 2020

UNIVERSITAS
MERCU BUANA



Meterai 1000000
METERAI
TEMPEL
11BEDAUX9438:4286
Muhamad Iqbal

LEMBAR PERSETUJUAN PENGUJI

NIM : 41516010171
Nama : Muhamad Iqbal
Judul Tugas Akhir : Analisa Data Rekam Medis Menggunakan Teknik
Data Mining Association Rules Dengan Algoritma
Clustering

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 Agustus 2020



LEMBAR PERSETUJUAN PENGUJI

NIM : 41516010171
Nama : Muhamad Iqbal
Judul Tugas Akhir : Analisa Data Rekam Medis Menggunakan Teknik
Data Mining Association Rules Dengan Algoritma
Clustering

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 – Agustus - 2020



LEMBAR PERSETUJUAN PENGUJI

NIM : 41516010171
Nama : Muhamad Iqbal
Judul Tugas Akhir : Analisa Data Rekam Medis Menggunakan Teknik
Data Mining Association Rules Dengan Algoritma
Clustering

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 – Agustus - 2020



(Herry Darajad Wijaya, S.Kom., MM)

UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

NIM : 41516010171
Nama : Muhamad Iqbal
Judul Tugas Akhir : Analisa Data Rekam Medis Menggunakan Teknik Data Mining Association Rules Dengan Algoritma Clustering

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 – Agustus – 2020

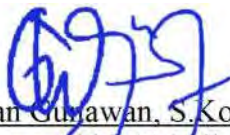
Menyetujui,



(DR. Mujiono Sadikin, MT, CISA, CGEIT)
Dosen Pembimbing

UNIVERSITAS
MERCU BUANA

Mengetahui,



(Wawan Gumawan, S.Kom., MT)
Koord. Tugas Akhir Teknik Informatika



(Ir. Emil R. Kaburuan, Ph D., IPM)
Ka. Prodi Teknik Informatika

ABSTRAK

Nama : Muhamad Iqbal
NIM : 41516010171
Pembimbing TA : DR.Mujiono Sadikin, MT, CISA, CGEIT
Judul : Analisa Data Rekam Medis Menggunakan Teknik
Data Mining Association Rules Dengan Algoritma
Clustering

Data informasi pasien yang terletak di pusat kesehatan dan rumah sakit ada di bentuk kode ICD. Data Rekam Medis tidak hanya berupa tumpukan berkas, tetapi dapat berupa digunakan untuk mengekstrak informasi yang terkandung di dalamnya. Beberapa informasi yang dapat dimanfaatkan adalah informasi kelompok penyakit yang sering diderita oleh pasien di rumah sakit. Kelompok informasi penyakit ini dapat dimanfaatkan oleh manajer dalam mengembangkan pelayanan kepada pasien. Dari sisi pasien dapat memanfaatkan tujuan pengobatan yang lebih tepat sasaran dalam sesuai dengan penyakit yang dideritanya. Untuk menentukan analisis ini tumpukan rekam medis menggunakan pengelompokan, termasuk pengelompokan wilayah, pengelompokan gender, pengelompokan usia, dan pengelompokan kode ICD. Dalam studi ini Kode ICD yang digunakan adalah ICD 10. Preprosesing data rekam medis dibagi menjadi dua : pertama data pelatihan dan kedua adalah data uji.

Kata kunci:

Algoritma Clustering, Data Mining, Medical Record, Association Rules, ICD.



ABSTRACT

Name : Muhamad Iqbal
Student Number : 41516010171
Counsellor : DR.Mujiono Sadikin, MT, CISA, CGEIT
Title : Analisa Data Rekam Medis Menggunakan Teknik
Data Mining Association Rules Dengan Algoritma
Clustering

Patient information data located in medical centers and hospitals are in the form of ICD codes. Medical Record data is not just a pile of files, but can be used to extract the information contained therein. Some information that can be utilized is the information group of diseases that often suffered by patients in the hospital. Information group of this disease can be utilized by the manager in developing services to patients. From the side of the patient can take advantage of treatment goals are more appropriate target in accordance with the disease he suffered. To determine the analysis of this stack of medical records using groupings, including grouping of regions, gender groupings, age groupings, and ICD code groupings. In this study ICD code used is ICD 10. Preposing medical record data is divided into two, first data training and second is test data.

Key words:

Clustering Algorithm, Data Mining, Medical Record, Association Rules, ICD.



KATA PENGANTAR

Puji syukur kita panjatkan kepada Tuhan Yang Maha Esa, karena dengan rahmat dan karunianya, penulis dapat menyelesaikan skripsi yang dibuat dan diajukan untuk memenuhi syarat guna untuk memperoleh gelar sarjana (S1) pada Fakultas Ilmu Komputer di Universitas Mercu Buana. Selama penulisan skripsi ini, penulis banyak menerima bantuan dan dukungan sehingga dapat menyelesaikan skripsi ini. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak DR.Mujiono Sadikin, MT, CISA, CGEIT. selaku Dosen Pembimbing skripsi yang membimbing dan memberikan saran untuk penulis sehingga dapat selesai tepat waktu. Dan selaku Dosen Pembimbing akademik.
2. Bapak Diky Firdaus, S.Kom, MM. selaku Koordinator Tugas Akhir Teknik Informatika.
3. Kedua Orang Tua yang selalu memberikan support terbaik berupa doa, motivasi dan materi sehingga penelitian ini dapat berjalan lancar sebagaimana mestinya.
4. Selanjutnya, penulis ingin mengucapkan terima kasih kepada teman-teman Teknik Informatika angkatan 2016 membantu dalam proses penulisan laporan skripsi ini yang tidak mungkin disebutkan satu per satu.

Akhir kata, dengan segala kerendahan hati penulis memohon maaf yang sebesar-besarnya atas kekurangan dan keterbatasan dalam penyusunan skripsi ini. Oleh karena itu saya mengharapkan saran dan kritik yang dapat menyempurnakan penulisan ini sehingga dapat bermanfaat dan berguna untuk pengembangan ilmu pengetahuan, Amiin.

Jakarta, 15 Agustus 2020

Muhamad Iqbal

I. TOPIK / BIDANG ILMU

Berisi topik / bidang ilmu yang akan dibahas, disesuaikan dengan topik / bidang ilmu yang ada pada Tugas Akhir, yaitu
Data Mining, Machine Learning

II. DAFTAR JURNAL (MINIMAL 20)

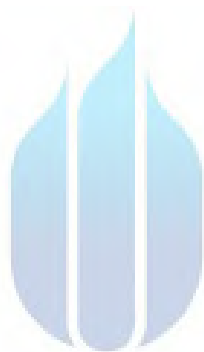
Bagian ini berisi **daftar judul artikel** yang akan di review. Dimana terdiri dari*

1. Minimal 5 Jurnal dari jurnal terakreditasi Nasional Sinta pada web <http://sinta.ristekbrin.go.id/>, bisa dicari di <http://garuda.ristekbrin.go.id/>
2. Minimal 5 Jurnal yang terindeks Scopus atau science direct dan jurnal bereputasi lainnya
3. Maximal 5 jurnal yang tidak termasuk dalam point 1,2

No	Judul Jurnal	Kategori (diisi dengan 1/2/3)*
1.	Personal and social patterns predict influenza vaccination decision.	Jurnal Internasional
2.	A data-driven approach to predicting diabetes and cardiovascular disease with machine learning.	Jurnal Internasional
3.	A Machine-Learning-Based Prediction Method for Hypertension Outcomes Based on Medical Data.	Jurnal Internasional
4.	Estimating summary statistics for electronic health record laboratory data for use in high-throughput phenotyping algorithms.	Jurnal Internasional
5.	Pathway analysis using XGBoost classification in Biomedical Data.	Jurnal Internasional
6.	How context affects implementation of the Primary Health Care approach: an analysis of	Jurnal Internasional

	what happened to primary health centres in India.	
7.	Predicting opioid dependence from electronic health records with machine learning	Jurnal Internasional
8.	Machine learning to predict rapid progression of carotid atherosclerosis in patients with impaired glucose tolerance.	Jurnal Internasional
9.	Using a machine learning approach to predict mortality in critically ill influenza patients: a cross-sectional retrospective multicentre study in Taiwan.	Jurnal Internasional
10.	Diagnosis of Cervical cancer using CLAHE and SGLDM on RGB Pap smear Images through ANN.	Jurnal Internasional
11.	Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning.	Jurnal Internasional
12.	An infrastructure for precision medicine through analysis of big data.	Jurnal Internasional
13.	Imputation and characterization of uncoded self-harm in major mental illness using machine learning.	Jurnal Internasional
14.	API design for machine learning software: experiences from the scikit-learn project.	Jurnal Internasional
15.	Aerial Mapping of Forests Affected by Pathogens Using UAVs, Hyperspectral Sensors, and Artificial Intelligence.	Jurnal Internasional
16.	A metabolite-based machine learning approach to diagnose Alzheimer-type dementia in blood: Results from the European Medical Information Framework for Alzheimer disease biomarker discovery cohort.	Jurnal Internasional
17.	Prediction of future gastric cancer risk using a machine learning algorithm and	Jurnal Internasional

	comprehensive medical check-up data: A casecontrol study.	
18.	Predicting urinary tract infections in the emergency department with machine learning	Jurnal Internasional
19.	Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: machine-learning algorithms and validation using national health data from Kuwait—a cohort study.	Jurnal Internasional
20.	Prediction of Acute Kidney Injury With a Machine Learning Algorithm Using Electronic Health Record Data.	Jurnal Internasional



III. TABEL REVIEW

No	1
Judul Artikel	Personal and social patterns predict influenza vaccination decision. [1]
Topik	Data Mining
Data	Anonymized Electronic Medical Records 2007 - 2017
Metode / Algoritma	Personal pattern analysis, Social pattern analysis, predictive modeling.
Abstrak	Secara umum, cakupan vaksinasi influenza musiman tetap suboptimal di sebagian besar negara maju rekomendasi lama dari organisasi kesehatan masyarakat. Keputusan individu mengenai vaksinasi adalah terletak di inti ketidakpatuhan. Menganalisis data skala besar untuk mengidentifikasi perilaku pribadi dan social pola untuk pengambilan vaksinasi influenza, dan mengembangkan model untuk memprediksi keputusan vaksinasi individu dalam suatu musim influenza mendatang.

<p>Hasil</p>	<p>Keputusan vaksinasi seseorang dapat dijelaskan dalam dua dimensi, Pribadi dan sosial. Itu dimensi pribadi sangat dibentuk oleh perilaku "default", seperti waktu vaksinasi di musim sebelumnya dan konsumsi kesehatan umum, tetapi juga dapat dipengaruhi oleh faktor-faktor temporal seperti penyakit pernapasan pada tahun sebelumnya. Dalam dimensi sosial, seorang pasien lebih mungkin untuk divaksinasi pada musim tertentu jika setidaknya satu anggotanya keluarga juga divaksinasi pada musim yang sama. Penyerapan vaksinasi sangat asertif dengan usia, sosial ekonomi skor, dan lokasi geografis. Model prediksi berbasis XGBoost mencapai skor ROC-AUC 0,91 dengan akurasi dan tingkat penarikan 90% pada set tes. Prediksi terutama bergantung pada vaksinasi individu dan rumah tangga pasien status di masa lalu, usia, jumlah pertemuan dengan sistem perawatan kesehatan, jumlah obat yang diresepkan, dan indikator penyakit kronis.</p>
<p>Kesimpulan</p>	<p>Kemampuan untuk membuat prediksi yang sangat baik tentang keputusan pasien membuat langkah besar menuju personalisasi kampanye vaksinasi influenza, dan akan membantu membentuk generasi berikutnya dari upaya vaksinasi yang ditargetkan.</p>
<p>Penulis</p>	<p>Adir Shaham, Gabriel Chodick, Varda Shalev Dan Yamin.</p>
<p>Nama jurnal, Volume, Nomor, Tahun</p>	<p>BMC Public Health, 20:222, 2016</p>
<p>Ulasan artikel</p>	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian tujuannya untuk: a) Mengidentifikasi perilaku pribadi dan sosial pola dan indikator serapan vaksinasi influenza, dan b) menggunakan indikator ini untuk mengembangkan pembelajaran mesin model yang akan memprediksi keputusan vaksinasi individu di musim influenza mendatang. Dengan menganalisis data skala besar untuk mengidentifikasi perilaku pribadi dan sosial pola penyerapan vaksinasi influenza, dan mengembangkan model untuk memprediksi keputusan vaksinasi individu di musim influenza yang akan datang. membuat prediksi yang sangat baik tentang keputusan pasien membuat langkah besar menuju personalisasi kampanye vaksinasi influenza, dan akan</p>

membantu membentuk generasi berikutnya dari upaya vaksinasi yang ditargetkan. Dalam artikel menggunakan metode Personal pattern analysis, Social pattern analysis dan predictive modeling sebagai pendukung pengambilan keputusan untuk sebagai pertimbangan keputusan pasien. Hasil yang di dapat mengidentifikasi beberapa pola perilaku terkait vaksinasi influenza dan mengembangkan model pembelajaran mesin yang memprediksi pengambilan keputusan individu di musim selanjutnya. Dan menunjukkan bahwa vaksinasi keputusan individu dapat dijelaskan dalam dua dimensi - pribadi dan sosial. Dimensi pribadi sangat dibentuk oleh perilaku "default" pendekatan, yang sering dimanifestasikan dalam keputusan vaksinasi berulang di musim berikutnya, lebih disukai waktu vaksinasi dalam satu musim. Pendekatan ini dikaitkan dengan pengukuran konsumsi perawatan kesehatan pasien. juga menemukan bukti yang dapat dipengaruhi oleh perilaku "default" ini efek sementara seperti penyakit pernapasan baru-baru ini diagnosis, menunjukkan bahwa mengalami penyakit pernapasan baru-baru ini mengubah persepsi pasien tentang risiko yang terkait dengan influenza. Dimensi sosial terbagi menjadi lingkungan sosial dan langsung keluarga. mengamati perbedaan yang signifikan dalam tingkat vaksinasi antara wilayah geografis dan pasien dengan skor sosial ekonomi yang berbeda. Hasil mengamati bahwa anggota keluarga cenderung memiliki kesamaan keputusan vaksinasi. Kemungkinan besar sementara sosial Lingkungan menetapkan pendekatan umum terhadap vaksinasi influenza sebagai konteks, kerabat dekat dapat mempengaruhi keputusan ad-hoc seseorang. Pola ini berfungsi sebagai dasar untuk model pembelajaran mesin prediktif , yang memanfaatkan data EMR dan data demografis untuk memberikan prediksi yang sangat baik tentang vaksinasi individu di masa mendatang keputusan. Keputusan vaksinasi sebelumnya, seiring dengan usia, jumlah pertemuan dengan sistem perawatan kesehatan, dan resep, berfungsi sebagai prediktor terkuat dari a keputusan vaksinasi di masa depan.

	Dalam kasus dimana keluarga Keputusan vaksinasi di musim sebelumnya disediakan, itu juga berfungsi sebagai prediktor yang kuat. Modelnya sederhana menunjukkan bagaimana prediksi yang sangat akurat dicapai bahkan ketika hanya sebagian kecil dari pasien informasi pribadi dan sosial tersedia.
--	--

No	2
Judul Artikel	A data-driven approach to predicting diabetes and cardiovascular disease with machine learning.[2].
Topik	Data Mining
Data	National Health and Nutrition Examination Survey (NHANES) dataset.
Metode / Algoritma	Machine Learning Models
Abstrak	Diabetes dan penyakit kardiovaskular adalah dua penyebab utama kematian di Amerika Serikat. Mengidentifikasi dan memprediksi penyakit ini pada pasien adalah langkah pertama untuk menghentikan perkembangan mereka. Mengevaluasi kemampuan model pembelajaran mesin dalam mendeteksi pasien berisiko menggunakan data survei (dan hasil laboratorium), dan mengidentifikasi variabel kunci dalam data yang berkontribusi terhadap penyakit ini di antara pasien.
Hasil	Model ensemble yang dikembangkan untuk penyakit kardiovaskular (berdasarkan 131 variabel) mencapai Area Bawah - Skor Receiver Operating Characteristics (AU-ROC) sebesar 83,1% tanpa menggunakan hasil laboratorium, dan akurasi 83,9% dengan hasil laboratorium. Dalam klasifikasi diabetes (berdasarkan 123 variabel), model eXtreme Gradient Boost (XGBoost) mencapai skor AU-ROC sebesar 86,2% (tanpa data laboratorium) dan 95,7% (dengan data laboratorium). Untuk pra-diabetes pasien, model ensemble memiliki skor AU-ROC tertinggi 73,7% (tanpa data laboratorium), dan untuk berbasis laboratorium data XGBoost melakukan yang terbaik di 84,4%. Lima prediktor teratas pada pasien diabetes adalah 1) ukuran pinggang, 2) usia, 3) berat badan yang dilaporkan sendiri, 4) panjang kaki, dan 5) asupan

	natrium. Untuk penyakit kardiovaskular, model diidentifikasi 1) usia, 2) tekanan darah sistolik, 3) berat badan yang dilaporkan sendiri, 4) terjadinya nyeri dada, dan 5) tekanan darah diastolik sebagai kunci kontributor.
Kesimpulan	Makalah ini menyimpulkan model yang dipelajari mesin berdasarkan kuesioner survei dapat memberikan otomatis mekanisme identifikasi untuk pasien yang berisiko diabetes dan penyakit kardiovaskular. Juga mengidentifikasi kontributor utama untuk prediksi, yang dapat dieksplorasi lebih lanjut untuk implikasinya pada catatan kesehatan elektronik.
Penulis	An Dinh, Stacey Miertschin, Amber Young dan Somya D. Mohanty.
Nama jurnal, Volume, Nomor, Tahun	BMC Medical Informatics and Decision Making, 19:211, 2019.
Ulasan artikel	Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk Mengidentifikasi dan memprediksi penyakit Kardiovaskular dan Diabetes. Latar belakang yang terjadi adalah Diabetes dan penyakit kardiovaskular adalah dua penyebab utama kematian di Amerika Serikat. Dan penelitian ini bertujuan untuk Mengidentifikasi dan memprediksi penyakit ini pada pasien adalah langkah pertama untuk menghentikan perkembangannya. Dengan mengevaluasi kapabilitas model pembelajaran mesin dalam mendeteksi pasien berisiko menggunakan data survei (dan hasil laboratorium), dan mengidentifikasi variabel kunci dalam data yang berkontribusi terhadap penyakit ini di antara pasien. Makalah ini menyimpulkan model yang dipelajari mesin berdasarkan kuesioner survei dapat memberikan otomatis mekanisme identifikasi untuk pasien yang berisiko diabetes dan penyakit kardiovaskular. Juga mengidentifikasi kontributor utama untuk prediksi, yang dapat dieksplorasi lebih lanjut untuk implikasinya pada catatan kesehatan elektronik. Mengevaluasi kemampuan model pembelajaran mesin dalam mendeteksi pasien berisiko menggunakan data survei (dan hasil laboratorium), dan mengidentifikasi variabel kunci dalam data

	<p>yang berkontribusi terhadap penyakit ini di antara pasien. Lima prediktor teratas pada pasien diabetes adalah 1) ukuran pinggang, 2) usia, 3) berat badan yang dilaporkan sendiri, 4) panjang kaki, dan 5) asupan natrium. Untuk penyakit kardiovaskular, model diidentifikasi 1) usia, 2) tekanan darah sistolik, 3) berat badan yang dilaporkan sendiri, 4) terjadinya nyeri dada, dan 5) tekanan darah diastolik sebagai kunci kontributor. Penelitian ini mengeksplorasi pendekatan berbasis data yang memanfaatkan model pembelajaran mesin yang diawasi untuk mengidentifikasi pasien dengan penyakit tersebut. Menggunakan dataset National Health and Nutrition Examination Survey (NHANES), dan melakukan pencarian menyeluruh dari semua variabel fitur yang tersedia dalam data untuk mengembangkan model kardiovaskular, pradiabetes, dan deteksi diabetes. Menggunakan kerangka waktu yang berbeda dan set fitur untuk data (berbasis pada data laboratorium), beberapa model pembelajaran mesin (regresi logistik, mesin vektor dukungan, hutan acak, dan peningkatan gradien) dievaluasi berdasarkan kinerja klasifikasinya. Model tersebut kemudian digabungkan menjadi mengembangkan model ansambel berbobot, yang mampu meningkatkan kinerja model yang berbeda akurasi deteksi. Keuntungan informasi dari model berbasis pohon digunakan untuk mengidentifikasi variabel kunci dalam pasien data yang berkontribusi pada deteksi pasien berisiko di setiap kelas penyakit dengan model data-learn.</p>
--	--

No	3
Judul Artikel	A Machine-Learning-Based Prediction Method for Hypertension Outcomes Based on Medical Data.[3]
Topik	Data Mining
Data	-
Metode / Algoritma	Support Vector Machine (SVM), C4.5 decision tree, random forest, dan XGBoost.
Abstrak	Hasil dari hipertensi merujuk pada kematian atau komplikasi serius yang mungkin terjadi pada pasien dengan hipertensi.

	<p>Hasil dari hipertensi adalah sangat memprihatinkan bagi pasien dan dokter, dan idealnya dihindari. Langkah pertama adalah mengekstraksi fitur-fitur utama dari banyak indikator pemeriksaan fisik pasien. Langkah kedua adalah menggunakan fitur-fitur utama diekstraksi dari langkah pertama untuk memprediksi hasil pasien. Untuk tujuan ini, mengusulkan model menggabungkan eliminasi fitur rekursif dengan metode validasi silang dan algoritma klasifikasi. Metrik evaluasi kinerja prediksi model yang dipilih adalah akurasi, Ukuran F1, dan area di bawah kurva karakteristik operasi penerima. 10 kali lipat validasi silang menunjukkan bahwa C4.5, RF, dan XGBoost dapat mencapai hasil prediksi yang sangat baik dengan sejumlah kecil fitur, dan pengklasifikasi setelah penghapusan fitur rekursif dengan pemilihan fitur validasi silang memiliki kinerja prediksi yang lebih baik.</p>
<p>Hasil</p>	<p>Menurut metode pemilihan fitur yang diusulkan dalam makalah ini, subset fitur yang dipilih oleh RFECV dipengaruhi oleh dua aspek: satu adalah classifier yang dikombinasikan dengan RFECV, dan yang lainnya adalah kriteria evaluasi kinerja dari classifier. Saat memilih berbagai pengklasifikasi atau kriteria evaluasi kinerja, subset fitur seringkali tidak sama. Tabel 4 mencantumkan jumlah himpunan bagian fitur optimal untuk setiap classifier di bawah tiga kriteria evaluasi. Dari Tabel 5, dapat dilihat bahwa jumlah fitur redundan dapat sangat dikurangi dengan metode pemilihan fitur yang diusulkan dalam makalah ini, yang akan membantu menghemat efisiensi komputasi dan meningkatkan efek prediksi model. Performa prediksi masing-masing classifier menggunakan subset fitur optimal mereka. Gambar 3 menunjukkan tren akurasi untuk empat pengklasifikasi menggunakan berbagai fitur. Gambar 4 menunjukkan tren ukuran F1 untuk empat pengklasifikasi menggunakan jumlah fitur yang berbeda. Gambar 5 menunjukkan tren AUC untuk empat pengklasifikasi menggunakan jumlah fitur yang berbeda. Dari Tabel 6 dan Gambar 3, kita dapat menemukan akurasi tertinggi dari empat</p>

	<p>pengklasifikasi ini di bawah kriteria yang berbeda. Memotong atau menambahkan fitur lain akan mengurangi akurasi. Kemudian, dengan peningkatan jumlah fitur, kinerja memiliki tren penurunan dan volatilitas yang signifikan. Kinerja prediksi XGBoost kurang dipengaruhi oleh jumlah fitur.</p>
Kesimpulan	<p>Memprediksi hasil dari pasien hipertensi adalah pekerjaan penelitian yang sangat berarti. Untuk pekerjaan ini, makalah ini mengusulkan metode yang menggabungkan classifier dengan RFECV untuk secara akurat memprediksi hasil pasien secara otomatis. Metode RFE digunakan untuk menilai pentingnya indikator pemeriksaan fisik untuk hasil hipertensi. Eksperimen menunjukkan bahwa metode RFEVC yang dikombinasikan dengan C4.5, RF, dan XGBoost dapat mencapai kinerja prediksi yang lebih baik. Selain itu, melalui RFECV menemukan bahwa tekanan darah ekstremitas dan tekanan darah rawat jalan memiliki efek penting pada hasil hipertensi. Metode yang diusulkan dalam makalah ini dapat secara efektif membantu dokter untuk menentukan apakah akan ada hasil pada pasien dengan hipertensi. Dengan cara ini, dokter dapat memberikan intervensi yang ditargetkan untuk pasien dengan risiko hasil yang lebih tinggi dan mengurangi kemungkinan hasil. Pertama, kita akan mendapatkan kumpulan data yang lebih besar dari lebih banyak sumber data untuk menguji lebih lanjut kemampuan generalisasi dari metode yang usulkan.</p>
Penulis	<p>Wenbing Chang, Yinglai Liu , Yiyong Xiao, Xinglong Yuan, Xingxing Xu, Siyue Zhang dan Shenghan Zhou.</p>
Nama jurnal, Volume, Nomor, Tahun	<p>Diagnostics, 9, 178, 2019.</p>
Ulasan artikel	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk Memprediksi hasil dari pasien hipertensi adalah pekerjaan penelitian yang sangat berarti. Untuk pekerjaan ini, makalah ini mengusulkan metode yang menggabungkan classifier dengan RFECV untuk secara akurat memprediksi hasil pasien secara otomatis. Metode RFE digunakan untuk menilai pentingnya indikator pemeriksaan fisik untuk hasil hipertensi. Eksperimen menunjukkan bahwa metode RFEVC yang dikombinasikan</p>

	<p>dengan C4.5, RF, dan XGBoost dapat mencapai kinerja prediksi yang lebih baik. Selain itu, melalui RFECV menemukan bahwa tekanan darah ekstremitas dan tekanan darah rawat jalan memiliki efek penting pada hasil hipertensi. Metode yang diusulkan dalam makalah ini dapat secara efektif membantu dokter untuk menentukan apakah akan ada hasil pada pasien dengan hipertensi. Dengan cara ini, dokter dapat memberikan intervensi yang ditargetkan untuk pasien dengan risiko hasil yang lebih tinggi dan mengurangi kemungkinan hasil. Pertama, kita akan mendapatkan kumpulan data yang lebih besar dari lebih banyak sumber data untuk menguji lebih lanjut kemampuan generalisasi dari metode yang di usulkan. Hasil dari hipertensi merujuk pada kematian atau komplikasi serius yang mungkin terjadi pada pasien dengan hipertensi. Hasil dari hipertensi adalah sangat memprihatinkan bagi pasien dan dokter, dan idealnya dihindari. Langkah pertama adalah mengekstraksi fitur-fitur utama dari banyak indikator pemeriksaan fisik pasien. Langkah kedua adalah menggunakan fitur-fitur utama diekstraksi dari langkah pertama untuk memprediksi hasil pasien. Untuk tujuan ini, mengusulkan model menggabungkan eliminasi fitur rekursif dengan metode validasi silang dan algoritma klasifikasi. Metrik evaluasi kinerja prediksi model yang dipilih adalah akurasi, Ukuran F1, dan area di bawah kurva karakteristik operasi penerima. 10 kali lipat validasi silang menunjukkan bahwa C4.5, RF, dan XGBoost dapat mencapai hasil prediksi yang sangat baik dengan sejumlah kecil fitur, dan pengklasifikasi setelah penghapusan fitur rekursif dengan pemilihan fitur validasi silang memiliki kinerja prediksi yang lebih baik.</p>
--	--

No	4
Judul Artikel	Estimating summary statistics for electronic health record laboratory data for use in high-throughput phenotyping algorithms.[4]
Topik	Data Mining

Data	Electronic health record (EHR).
Metode / Algoritma	PopKLD
Abstrak	<p>Mempelajari pertanyaan tentang bagaimana untuk mewakili atau meringkas data laboratorium mentah yang diambil dari catatan kesehatan elektronik menggunakan pemilihan model parametrik untuk mengurangi atau mengatasi bias yang diinduksi melalui perawatan klinis. Sebelumnya telah ditunjukkan bahwa proses perawatan kesehatan, sebagaimana didefinisikan oleh konteks pengukuran dan pola pengukuran, dapat mempengaruhi bagaimana data EHR didistribusikan secara statistik. mengevaluasi metodologi dengan dua cara. Pertama, menerapkan metode ini ke data laboratorium yang dikumpulkan dalam dua konteks perawatan kesehatan yang berbeda, perawatan primer versus perawatan intensif. menunjukkan bahwa PopKLD mempertahankan fitur fisiologis yang diketahui dalam data yang hilang ketika meringkas data menggunakan ringkasan data laboratorium yang lebih umum seperti mean dan standar deviasi. menemukan bahwa ringkasan data laboratorium PopKLD secara substansial lebih baik dalam memprediksi keadaan penyakit. Algoritma PopKLD atau PopKLD-CAT tidak dimaksudkan untuk digunakan sebagai algoritma fenotip, tetapi menggunakan tugas fenotipe untuk menunjukkan informasi apa yang dapat diperoleh saat menggunakan ringkasan data laboratorium yang lebih informatif.</p>
Hasil	<p>Makalah ini semua nilai laboratorium dikumpulkan di klinik AIM dengan pengecualian satu, glukosa ICU. Glukosa terbatas ICU termasuk dalam upaya untuk mengisolasi data yang dihasilkan terutama karena fisiologi dan dengan bias proses perawatan kesehatan yang relatif minimal karena konteks pengumpulan. Pertama, tidak ada aturan umum yang jelas untuk memilih distribusi terbaik atau paling representatif untuk semua tipe data laboratorium. Semua model parametrik memiliki variabel laboratorium yang direpresentasikan secara buruk karena dicirikan oleh perbedaan-KL yang relatif besar</p>

sementara masih berada di antara yang terbaik untuk mewakili variabel laboratorium lainnya. Kedua, ada keragaman dalam berapa banyak model yang dapat memodelkan data laboratorium yang masuk akal. Beberapa jenis laboratorium memiliki pemenang yang jelas di antara model, e. Karena itu, menganggap normalitas dengan nilai-nilai laboratorium umumnya bukan ide yang baik. Selain itu, karena penyimpangan dari prediksi entropi maksimum, sebagian besar data variabel laboratorium memiliki informasi lebih banyak daripada yang terkandung dalam mean dan varians saja. Keempat, seringkali ketika pengukuran laboratorium dimodelkan dengan baik oleh distribusi normal, mereka juga dimodelkan dengan baik oleh beberapa model berparameter lainnya. Ketika data EHR tidak dipengaruhi oleh konteks pengumpulan atau proses perawatan kesehatan lainnya, mereka harus mewakili fisiologi pasien. Dalam pengaturan ini, PopKLD harus memilih distribusi yang mempertahankan fitur fisiologis. Untuk menguji ini mengevaluasi algoritma PopKLD dalam dua konteks pengumpulan data. Pertama menerapkan PopKLD untuk data glukosa yang dikumpulkan di ICU, sumber data konteks tunggal. berhipotesis bahwa data ICU mewakili sebagian besar fisiologi karena pengukuran seperti glukosa dalam ICU dikumpulkan sebagian besar independen dari keadaan pasien dibandingkan dengan konteks pengumpulan data EHR lainnya. Kedua, gunakan PopKLD untuk glukosa dari EHR terbatas pada pasien yang mengunjungi klinik Obat Penyakit Dalam Ambulatory, atau klinik AIM. Data ini mewakili sumber data konteks campuran karena data ini mencakup semua data untuk pasien AIM, termasuk data ICU, tetapi terutama berisi data rawat jalan. berhipotesis bahwa data AIM mewakili campuran fisiologi dan HCP. Dalam kedua konteks menunjukkan bahwa PopKLD menghasilkan ringkasan laboratorium yang menjaga fisiologi. Secara khusus, bahwa untuk glukosa, deviasi rata-rata dan standar terkait linier. Dalam hal data ICU, kita dapat mengevaluasi algoritma PopKLD lebih

	<p>lanjut karena kita dapat membuat prediksi. Distribusi entropi maksimum untuk sistem apa pun dengan kendala yang berarti dan standar deviasi terkait secara linear adalah distribusi gamma. Oleh karena itu, dalam konteks ICU, jika data ICU terutama mewakili fisiologi, memperkirakan bahwa algoritma PopKLD akan memilih distribusi gamma untuk meringkas glukosa terbaik.</p>
Kesimpulan	<p>Makalah ini mengembangkan algoritma PopKLD dan PopKLD-CAT yang mengakui data laboratorium EHR mentah, kontinyu, berisik, outlier-ridden, bias dan muncul dengan ringkasan dimensi rendah yang kurang didominasi oleh bias proses perawatan kesehatan, pencilan, dan kompleksitas lainnya, siap digunakan oleh teknologi pembelajaran mesin saat ini.</p>
Penulis	<p>D.J. Albersa, N. Elhadada , J. Claassenb , R. Perottec , A. Goldsteina dan G. Hripcsaka.</p>
Nama jurnal, Volume, Nomor, Tahun	<p>Elsevier, 78, 87-101, 2018.</p>
Ulasan artikel	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk mengembangkan algoritma PopKLD dan PopKLD-CAT yang mengakui data laboratorium EHR mentah, kontinyu, berisik, outlier-ridden, bias dan muncul dengan ringkasan dimensi rendah yang kurang didominasi oleh bias proses perawatan kesehatan, pencilan, dan kompleksitas lainnya, siap digunakan oleh teknologi pembelajaran mesin saat ini. Mempelajari pertanyaan tentang bagaimana untuk mewakili atau meringkas data laboratorium mentah yang diambil dari catatan kesehatan elektronik menggunakan pemilihan model parametrik untuk mengurangi atau mengatasi bias yang diinduksi melalui perawatan klinis. Sebelumnya telah ditunjukkan bahwa proses perawatan kesehatan, sebagaimana didefinisikan oleh konteks pengukuran dan pola pengukuran, dapat mempengaruhi bagaimana data EHR didistribusikan secara statistik. Mengevaluasi metodologi dengan dua cara. Pertama, menerapkan metode ini ke data laboratorium yang dikumpulkan dalam dua konteks perawatan kesehatan yang berbeda,</p>

perawatan primer versus perawatan intensif. menunjukkan bahwa PopKLD mempertahankan fitur fisiologis yang diketahui dalam data yang hilang ketika meringkas data menggunakan ringkasan data laboratorium yang lebih umum seperti mean dan standar deviasi. menemukan bahwa ringkasan data laboratorium PopKLD secara substansial lebih baik dalam memprediksi keadaan penyakit. Algoritma PopKLD atau PopKLD-CAT tidak dimaksudkan untuk digunakan sebagai algoritma fenotip, tetapi menggunakan tugas fenotipe untuk menunjukkan informasi apa yang dapat diperoleh saat menggunakan ringkasan data laboratorium yang lebih informatif. Dalam Makalah ini semua nilai laboratorium dikumpulkan di klinik AIM dengan pengecualian satu, glukosa ICU. Glukosa terbatas ICU termasuk dalam upaya untuk mengisolasi data yang dihasilkan terutama karena fisiologi dan dengan bias proses perawatan kesehatan yang relatif minimal karena konteks pengumpulan. Pertama, tidak ada aturan umum yang jelas untuk memilih distribusi terbaik atau paling representatif untuk semua tipe data laboratorium. Semua model parametrik memiliki variabel laboratorium yang direpresentasikan secara buruk karena dicirikan oleh perbedaan-KL yang relatif besar sementara masih berada di antara yang terbaik untuk mewakili variabel laboratorium lainnya. Kedua, ada keragaman dalam berapa banyak model yang dapat memodelkan data laboratorium yang masuk akal. Beberapa jenis laboratorium memiliki pemenang yang jelas di antara model, e. Karena itu, menganggap normalitas dengan nilai-nilai laboratorium umumnya bukan ide yang baik. Selain itu, karena penyimpangan dari prediksi entropi maksimum, sebagian besar data variabel laboratorium memiliki informasi lebih banyak daripada yang terkandung dalam mean dan varians saja. Keempat, seringkali ketika pengukuran laboratorium dimodelkan dengan baik oleh distribusi normal, mereka juga dimodelkan dengan baik oleh beberapa model berparameter lainnya. Ketika data EHR tidak dipengaruhi oleh konteks

	<p>pengumpulan atau proses perawatan kesehatan lainnya, mereka harus mewakili fisiologi pasien. Dalam pengaturan ini, PopKLD harus memilih distribusi yang mempertahankan fitur fisiologis. Untuk menguji ini mengevaluasi algoritma PopKLD dalam dua konteks pengumpulan data. Pertama menerapkan PopKLD untuk data glukosa yang dikumpulkan di ICU, sumber data konteks tunggal. berhipotesis bahwa data ICU mewakili sebagian besar fisiologi karena pengukuran seperti glukosa dalam ICU dikumpulkan sebagian besar independen dari keadaan pasien dibandingkan dengan konteks pengumpulan data EHR lainnya. Kedua, gunakan PopKLD untuk glukosa dari EHR terbatas pada pasien yang mengunjungi klinik Obat Penyakit Dalam Ambulatory, atau klinik AIM. Data ini mewakili sumber data konteks campuran karena data ini mencakup semua data untuk pasien AIM, termasuk data ICU, tetapi terutama berisi data rawat jalan. berhipotesis bahwa data AIM mewakili campuran fisiologi dan HCP. Dalam kedua konteks menunjukkan bahwa PopKLD menghasilkan ringkasan laboratorium yang menjaga fisiologi. Secara khusus, bahwa untuk glukosa, deviasi rata-rata dan standar terkait linier. Dalam hal data ICU, kita dapat mengevaluasi algoritma PopKLD lebih lanjut karena kita dapat membuat prediksi. Distribusi entropi maksimum untuk sistem apa pun dengan kendala yang berarti dan standar deviasi terkait secara linear adalah distribusi gamma. Oleh karena itu, dalam konteks ICU, jika data ICU terutama mewakili fisiologi, memperkirakan bahwa algoritma PopKLD akan memilih distribusi gamma untuk meringkas glukosa terbaik.</p>
--	---

No	5
Judul Artikel	Pathway analysis using XGBoost classification in Biomedical Data.[5]
Topik	Data Mining
Data	gene expression dataset.
Metode / Algoritma	XGBoost Classification

Abstrak	<p>Mengingat fakta bahwa keberadaan biologis kita berakar pada sistem yang kompleks dalam sel kita dengan ribuan interaksi antara gen dan metabolit, komunitas riset di bidang biologi dan medis telah mengalihkan minat mereka ke pendekatan berbasis jaringan. Berdasarkan ini, mengusulkan skema klasifikasi berdasarkan XGBoost, algoritma klasifikasi berbasis pohon baru-baru ini, untuk mendeteksi jalur yang paling diskriminatif terkait dengan suatu penyakit. Metode didemonstrasikan pada dataset ekspresi gen yang menua yang memberikan bukti bahwa XGBoost mengungguli metode klasifikasi terkenal lainnya dalam data biologis, sementara hasil yang disediakan oleh metode mencakup beberapa jalur lama yang terkait serta yang telah lama dilaporkan.</p>
Hasil	<p>Awalnya, jalur bawah tanah linier diekstraksi untuk setiap jalur yang diinginkan dan diukur dengan menjumlahkan semua skor VI dari setiap anggota gen dibagi dengan jumlah semua anggota. Sebagai hasilnya, daftar sub-jalur linear untuk setiap jalur diekspor dan dievaluasi melalui skor VI. Jalur bawah tanah dengan skor terbaik untuk setiap jalur minat dianalisis lebih lanjut berdasarkan penjelasan mereka dalam ontologi proses gen biologis dan istilah jalur. menggunakan Fungsional Anotasi Alat David, yang menyediakan anotasi bets khas dan analisis pengayaan istilah gen-GO untuk menyoroti istilah yang paling relevan, seperti GO dan jalur, yang terkait dengan daftar gen yang diberikan.</p>
Kesimpulan	<p>Makalah ini Singkatnya, metode klasifikasi berbasis jalur menggunakan algoritma XGBoost untuk menganalisis data ekspresi gen dijelaskan. Metode yang diusulkan mengidentifikasi jalur penting yang membedakan dua kasus yang diteliti, sementara itu mengisolasi jalur bawah tanah yang paling penting yang memainkan peran penting dalam seluruh topologi jalur, gangguan yang memunculkan penyakit yang diteliti.</p>
Penulis	<p>Georgios N. Dimitrakopoulos, Aristidis G. Vrahatis, Vassilis Plagianakos, Kyriakos Sgarbas.</p>
Nama jurnal, Volume, Nomor, Tahun	<p>SETN, 9-15, 2018</p>

Ulasan artikel

Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk analisis dan menjelaskan biologi yang mendasari untuk hasil yang tinggi pengukuran molekuler, karena lingkungan interaksi protein dalam jaringan seluler penting dalam mendefinisikan peran itu protein berperan dalam sistem secara keseluruhan [1]. Analisis Jalur milik keluarga metode Biologi Sistem, yang bertujuan untuk membuat model berbasis jaringan seluler yang mengekstraksi pengetahuan dari data diproduksi oleh teknologi pengurutan throughput tinggi. Lebar luas pendekatan berbasis jalur diterbitkan dalam dekade terakhir menciptakan berbagai pendekatan dan kategori analitik [2-4]. Dalam konteks yang lebih umum, pendekatan berbasis jaringan telah membanjiri komunitas penelitian di bidang Biologi Sistem dan Sistem Kedokteran, karena mereka mencakup banyak hal yang kompleks sistem di mana banyak saling ketergantungan di antara ribuan gen dan metabolit terjadi pada berbagai tingkat dan skala, sedangkan mereka interpretasi alami terkait erat dengan biologis kita keberadaan [5]. Diberikan pula fakta bahwa suatu penyakit jarang dan konsekuensi dari kelainan pada satu gen, tetapi itu mencerminkan gangguan jaringan intraseluler dan antar sel yang kompleks yang menghubungkan jaringan dan sistem organ, itu dibenarkan pergeseran komunitas riset untuk menggunakan alat tingkat sistem dan tingkat jaringan untuk mengungkap kompleksitas mekanisme dan penyakit biologis [6]. Sebagai consequence, bidang penelitian baru, yang dikenal sebagai Jaringan Medicine [7], yang merupakan kombinasi dari System Medicine dan Bidang Ilmu Jaringan, mengikuti perspektif ini. Itu masih di dalamnya bayi yang bertujuan untuk mengatasi kompleksitas penyakit secara holistik, dengan lebih terintegrasi. Makalah ini Singkatnya, metode klasifikasi berbasis jalur menggunakan algoritma XGBoost untuk menganalisis data ekspresi gen dijelaskan. Metode yang diusulkan mengidentifikasi jalur penting yang membedakan dua kasus yang diteliti, sementara itu mengisolasi jalur bawah tanah

	<p>yang paling penting yang memainkan peran penting dalam seluruh topologi jalur, gangguan yang memunculkan penyakit yang diteliti. Mengingat fakta bahwa keberadaan biologis kita berakar pada sistem yang kompleks dalam sel kita dengan ribuan interaksi antara gen dan metabolit, komunitas riset di bidang biologi dan medis telah mengalihkan minat mereka ke pendekatan berbasis jaringan. Berdasarkan ini, mengusulkan skema klasifikasi berdasarkan XGBoost, algoritma klasifikasi berbasis pohon baru-baru ini, untuk mendeteksi jalur yang paling diskriminatif terkait dengan suatu penyakit. Metode didemonstrasikan pada dataset ekspresi gen yang menua yang memberikan bukti bahwa XGBoost mengungguli metode klasifikasi terkenal lainnya dalam data biologis, sementara hasil yang disediakan oleh metode mencakup beberapa jalur lama yang terkait serta yang telah lama dilaporkan. Awalnya, jalur bawah tanah linier diekstraksi untuk setiap jalur yang diinginkan dan diukur dengan menjumlahkan semua skor VI dari setiap anggota gen dibagi dengan jumlah semua anggota. Sebagai hasilnya, daftar sub-jalur linear untuk setiap jalur diekspor dan dievaluasi melalui skor VI. Jalur bawah tanah dengan skor terbaik untuk setiap jalur minat dianalisis lebih lanjut berdasarkan penjelasan mereka dalam ontologi proses gen biologis dan istilah jalur. menggunakan Fungsional Anotasi Alat David, yang menyediakan anotasi bets khas dan analisis pengayaan istilah gen-GO untuk menyoroti istilah yang paling relevan, seperti GO dan jalur, yang terkait dengan daftar gen yang diberikan.</p>
--	--

No	6
Judul Artikel	How context affects implementation of the Primary Health Care approach: an analysis of what happened to primary health centres in India.[6]
Topik	Data Mining
Data	-

Metode / Algoritma	Examining macro contexts, Examining the micro context dan Interpreting the macro and micro contexts together.
Abstrak	Dalam makalah ini, menjelaskan tantangan yang diajukan oleh konteks untuk penerapan pendekatan Perawatan Kesehatan Utama, menggunakan contoh pusat kesehatan primer di India. Untuk memeriksa konteks mikro, bekerja dengan data kualitatif empiris dari sebuah distrik pedesaan di Maharashtra — dikumpulkan melalui 12 diskusi kelompok fokus masyarakat, 12 wawancara pasien dan 34 wawancara dengan staf sistem kesehatan. Pada tingkat mikro, penyediaan layanan di pusat kesehatan primer dipengaruhi oleh ketidaktertarikan dokter dalam peran perawatan primer dan konteks kelembagaan yang mempromosikan keengganan risiko dan mengabaikan perawatan rawat jalan. Makalah ini menyoroti beberapa kompleksitas kontekstual dari penerapan PHC - mempertimbangkan masalah tingkat makro dan mikro.
Hasil	Layanan Kesehatan di Inggris. Dalam rencana jangka pendek komite, satu pusat kesehatan primer divisualisasikan untuk setiap 40.000 penduduk. Setiap pusat kesehatan primer dalam program jangka pendek adalah memiliki 2 dokter, 4 Perawat Kesehatan Masyarakat, 4 Bidan, 4 Dais terlatih. Ini membenarkan pengeluaran besar untuk membangun sistem seperti itu sebagai 'investasi yang baik' untuk kemajuan negara. Pusat kesehatan primer adalah bagian dari cetak biru asli yang dimaksudkan sebagai akses pertama fasilitas kesehatan, dalam sistem tiga tingkat.
Kesimpulan	Makalah ini menyoroti beberapa kontekstual kompleksitas penerapan PHC — mempertimbangkan makro (berkaitan dengan ideologi dan prioritas fiskal) dan mikro (berkaitan dengan perilaku sehari-hari dan praktik aktor) masalah level. Ketika kita berkomitmen kembali kepada Alma-Ata, kita harus berhati-hati dari adopsi intervensi upacara, yang terlihat seperti PHC — tetapi tidak bisa mewujudkan cita-citanya.
Penulis	Sudha Ramani, Muthusamy Siva, Lucy Gilson.
Nama jurnal, Volume, Nomor, Tahun	BMJ Glob Health, 10, 136, 2020.

<p>Ulasan artikel</p>	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk menjelaskan tantangan yang diajukan oleh konteks untuk penerapan pendekatan Perawatan Kesehatan Utama, menggunakan contoh pusat kesehatan primer di India. Untuk memeriksa konteks mikro, bekerja dengan data kualitatif empiris dari sebuah distrik pedesaan di Maharashtra — dikumpulkan melalui 12 diskusi kelompok fokus masyarakat, 12 wawancara pasien dan 34 wawancara dengan staf sistem kesehatan. Pada tingkat mikro, penyediaan layanan di pusat kesehatan primer dipengaruhi oleh ketidaktertarikan dokter dalam peran perawatan primer dan konteks kelembagaan yang mempromosikan keengganan risiko dan mengabaikan perawatan rawat jalan. Makalah ini menyoroti beberapa kompleksitas kontekstual dari penerapan PHC — mempertimbangkan masalah tingkat makro dan mikro. Makalah ini menyoroti beberapa kontekstual kompleksitas penerapan PHC — mempertimbangkan makro (berkaitan dengan ideologi dan prioritas fiskal) dan mikro (berkaitan dengan perilaku sehari-hari dan praktik aktor) masalah level. Ketika kita berkomitmen kembali kepada Alma-Ata, kita harus berhati-hati dari adopsi intervensi upacara, yang terlihat seperti PHC — tetapi tidak bisa mewujudkan cita-citanya. Makalah ini menyoroti beberapa kontekstual kompleksitas penerapan PHC — mempertimbangkan makro (berkaitan dengan ideologi dan prioritas fiskal) dan mikro (berkaitan dengan perilaku sehari-hari dan praktik aktor) masalah level.</p>
-----------------------	--

No	7
Judul Artikel	Predicting opioid dependence from electronic health records with machine learning.[7]
Topik	Data Mining
Data	MSMC-EHR
Metode / Algoritma	Machine Learning

<p>Abstrak</p>	<p>Makalah ini melatih model pembelajaran mesin untuk mengklasifikasikan pasien dengan kemungkinan memiliki diagnosis ketergantungan zat menggunakan data EHR dari pasien yang didiagnosis dengan ketergantungan zat, bersama dengan pasien kontrol yang tidak memiliki riwayat kondisi terkait zat, disesuaikan dengan usia, jenis kelamin, dan status HIV, hepatitis C, dan penyakit sel sabit. Model prediktif dapat digunakan untuk mengidentifikasi pasien yang berisiko mengembangkan ketergantungan, risiko overdosis, dan pasien yang mencari opioid yang melaporkan gejala lain dalam kunjungan mereka ke ruang gawat darurat.</p>
<p>Hasil</p>	<p>Dari penelitian ini mempertahankan 11.573 kasus yang diagnosis ketergantungan zat pertama kali dibuat pada usia 20 tahun atau lebih. Usia rata-rata pasien ini pada diagnosis ketergantungan zat pertama mereka adalah 45,6 tahun, dengan pasien termuda berusia 20 tahun, dan yang tertua 89,4 tahun. 9528 dari pasien ini memiliki 1.525.293 rekaman tes laboratorium dan pengukuran tanda-tanda vital selama periode 10 hari sebelum dan 10 hari setelah diagnosis ketergantungan zat. Populasi kasus terakhir diperoleh setelah mengeluarkan pasien dengan kurang dari 17 tes laboratorium dan tanda-tanda vital, meninggalkan 7797 pasien. Dari 880.605 pasien yang memiliki setidaknya satu resep dalam MSMC-EHR, 356.734 pasien memiliki setidaknya satu resep opioid. Median adalah satu hari, kemungkinan karena pasien yang diresepkan metadon untuk pengobatan ketergantungan opioid mereka sebelumnya. Rata-rata 26,7 resep opioid diamati untuk pasien dengan diagnosis gangguan penggunaan opioid, dan rata-rata 5,27 untuk pasien tanpa diagnosis gangguan penggunaan opioid. Dalam hal kasus dan populasi kontrol yang cocok, persentase serupa memiliki resep opioid non-metadon sebelumnya. Namun, untuk pasien yang memiliki resep opioid sebelumnya, pasien kasus memiliki rata-rata 13,07 resep, dan kontrol memiliki rata-rata 4,25 resep. Melihat 100 hari sebelum diagnosis ketergantungan zat, peringkat skor nyeri secara signifikan meningkat dalam</p>

	<p>kasus dibandingkan dengan kontrol pada ~ 80 hari sebelum diagnosis. Peningkatan sebelumnya pada skor nyeri ini dapat menunjukkan waktu khas dari resep opioid awal hingga diagnosis ketergantungan zat. Namun, ditetapkan bahwa perkembangan dari resep opioid ke diagnosis ketergantungan umumnya memakan waktu berbulan-bulan hingga bertahun-tahun. Kemungkinan lain, skor nyeri awal yang tinggi menunjukkan toleransi opioid, dan ambang batas nyeri yang berkurang.</p>
Kesimpulan	<p>Melalui menganalisis catatan kesehatan dari ratusan ribu orang di MSMC-EHR dengan kerangka pembelajaran mesin, selanjutnya meneliti pasien ketergantungan opioid dengan menggunakan pengukuran fisiologis. menemukan bahwa pasien yang tergantung opioid memiliki WBC dan gangguan pernapasan yang secara signifikan lebih tinggi. Pasien yang tergantung opioid juga biasanya kekurangan gizi yang ditandai dengan RCDW dan albumin darah yang rendah dibandingkan dengan kontrol. Analisis fenotip klinis menemukan bahwa pasien yang tergantung opioid lebih mungkin untuk menderita gangguan kejiwaan dan menunjukkan gejala yang berhubungan dengan nyeri. Model prediktif mungkin memiliki kegunaan untuk mengidentifikasi pasien yang berisiko mengembangkan ketergantungan, risiko overdosis, dan mencari pasien opioid bahwa Tabel 5 Top 10 prosedur yang diwakili secara berbeda selama 5 tahun sebelum diagnosis penyalahgunaan zat</p> <p>Prosedur Odds ratio p-value Kelompok lain terapi 19.8 1.23E-69 Wawancara & Evaluasi NEC 11.578 4.06E-42 Penentuan Mental Psikiatri 10.597 1.24E-29 Psikoterapi verbal eksplorator 10.371 2.41E-57 Wawancara & evaluasi singkat 6.224 9.00E-201 Wawancara terbatas / evaluasi 5.804 2.71E-279 Wawancara & evaluasi NOS 5.745 6.16E-102 Wawancara / evaluasi komprehensif 5.065 1.71E-147 Konseling lainnya 4.3 3.93E-36 Pemantauan janin lainnya 0.218 1.14E-30 Ellis et al. Perlu dicatat bahwa menandai pasien dengan kode ICD ketergantungan opioid.</p>

Penulis	Randall J. Ellis, Zichen Wang, Nicholas Genes2 dan Avi Ma'ayan.
Nama jurnal, Volume, Nomor, Tahun	BIO Data Mining, 12, 3, 2019
Ulasan artikel	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk melatih model pembelajaran mesin untuk mengklasifikasikan pasien dengan kemungkinan memiliki diagnosis ketergantungan zat menggunakan data EHR dari pasien yang didiagnosis dengan ketergantungan zat, bersama dengan pasien kontrol yang tidak memiliki riwayat kondisi terkait zat, disesuaikan dengan usia, jenis kelamin, dan status HIV, hepatitis C, dan penyakit sel sabit. Model prediktif dapat digunakan untuk mengidentifikasi pasien yang berisiko mengembangkan ketergantungan, risiko overdosis, dan pasien yang mencari opioid yang melaporkan gejala lain dalam kunjungan mereka ke ruang gawat darurat. Dari penelitian ini mempertahankan 11.573 kasus yang diagnosis ketergantungan zat pertama kali dibuat pada usia 20 tahun atau lebih. Usia rata-rata pasien ini pada diagnosis ketergantungan zat pertama mereka adalah 45,6 tahun, dengan pasien termuda berusia 20 tahun, dan yang tertua 89,4 tahun. 9528 dari pasien ini memiliki 1.525.293 rekaman tes laboratorium dan pengukuran tanda-tanda vital selama periode 10 hari sebelum dan 10 hari setelah diagnosis ketergantungan zat. Populasi kasus terakhir diperoleh setelah mengeluarkan pasien dengan kurang dari 17 tes laboratorium dan tanda-tanda vital, meninggalkan 7797 pasien. Dari 880.605 pasien yang memiliki setidaknya satu resep dalam MSMC-EHR, 356.734 pasien memiliki setidaknya satu resep opioid. Median adalah satu hari, kemungkinan karena pasien yang diresepkan metadon untuk pengobatan ketergantungan opioid mereka sebelumnya. Rata-rata 26,7 resep opioid diamati untuk pasien dengan diagnosis gangguan penggunaan opioid, dan rata-rata 5,27 untuk pasien tanpa diagnosis gangguan penggunaan opioid. Dalam hal kasus dan populasi kontrol yang cocok, persentase serupa memiliki resep opioid non-metadon sebelumnya. Namun, untuk pasien yang memiliki resep opioid sebelumnya, pasien kasus</p>

memiliki rata-rata 13,07 resep, dan kontrol memiliki rata-rata 4,25 resep. Melihat 100 hari sebelum diagnosis ketergantungan zat, peringkat skor nyeri secara signifikan meningkat dalam kasus dibandingkan dengan kontrol pada ~ 80 hari sebelum diagnosis. Peningkatan sebelumnya pada skor nyeri ini dapat menunjukkan waktu khas dari resep opioid awal hingga diagnosis ketergantungan zat. Namun, ditetapkan bahwa perkembangan dari resep opioid ke diagnosis ketergantungan umumnya memakan waktu berbulan-bulan hingga bertahun-tahun. Kemungkinan lain, skor nyeri awal yang tinggi menunjukkan toleransi opioid, dan ambang batas nyeri yang berkurang. Melalui menganalisis catatan kesehatan dari ratusan ribu orang di MSMC-EHR dengan kerangka pembelajaran mesin, selanjutnya meneliti pasien ketergantungan opioid dengan menggunakan pengukuran fisiologis. menemukan bahwa pasien yang tergantung opioid memiliki WBC dan gangguan pernapasan yang secara signifikan lebih tinggi. Pasien yang tergantung opioid juga biasanya kekurangan gizi yang ditandai dengan RCDW dan albumin darah yang rendah dibandingkan dengan kontrol. Analisis fenotip klinis menemukan bahwa pasien yang tergantung opioid lebih mungkin untuk menderita gangguan kejiwaan dan menunjukkan gejala yang berhubungan dengan nyeri. Model prediktif mungkin memiliki kegunaan untuk mengidentifikasi pasien yang berisiko mengembangkan ketergantungan, risiko overdosis, dan mencari pasien opioid bahwa Tabel 5 Top 10 prosedur yang diwakili secara berbeda selama 5 tahun sebelum diagnosis penyalahgunaan zat

Prosedur	Odds ratio	p-value
Kelompok lain terapi	19.8	1.23E-69
Wawancara & Evaluasi NEC	11.578	4.06E-42
Penentuan Mental Psikiatri	10.597	1.24E-29
Psikoterapi verbal eksplorator	10.371	2.41E-57
Wawancara & evaluasi singkat	6.224	9.00E-201
Wawancara terbatas / evaluasi	5.804	2.71E-279
Wawancara & evaluasi NOS	5.745	6.16E-102
Wawancara / evaluasi komprehensif	5.065	1.71E-147
Konseling lainnya	4.3	3.93E-36
Pemantauan janin		

	lainnya 0.218 1.14E-30 Ellis et al. Perlu dicatat bahwa menandai pasien dengan kode ICD ketergantungan opioid.
--	--

No	8
Judul Artikel	Machine learning to predict rapid progression of carotid atherosclerosis in patients with impaired glucose tolerance.[8]
Topik	Data Mining
Data	ACT NOW dataset
Metode / Algoritma	Naive Bayes, Multilayer Perceptron dan Random Forest Methods.
Abstrak	<p>Dalam makalah ini, menyelidiki faktor-faktor penting yang mempengaruhi prediksi, menggunakan beberapa metode pembelajaran mesin, perkembangan cepat ketebalan intima-media karotid pada peserta toleransi glukosa yang terganggu. Dalam studi Actos Now for Prevention of Diabetes, 382 peserta dengan IGT menjalani evaluasi ultrasonografi intima-media karotid pada awal dan pada 15-18 bulan, dan dibagi menjadi pelanjut cepat dan pelambat tidak cepat. Perbandingan metode dan faktor dilakukan dengan menggunakan area di bawah analisis kurva karakteristik operasi penerima dan skor Brier. Hasil eksperimen menunjukkan bahwa metode pembelajaran yang diusulkan bekerja dengan baik dalam mengidentifikasi atau memprediksi RP. Di antara metode, kinerja Naïve Bayes adalah yang terbaik dibandingkan dengan perceptron multilayer dan hutan acak. Dengan berurusan dengan data multi-modal, metode pembelajaran yang diusulkan menunjukkan efektivitas dalam memprediksi prediabetik yang berisiko untuk perkembangan aterosklerosis yang cepat.</p>
Hasil	<p>Dari makalah ini ada perbedaan yang signifikan dalam situs pendaftaran, proporsi dengan ras Hispanik, mikroalbumin urin, kreatinin plasma dan serum inhibitor aktivator plasminogen serum-1 antara RP dan NRP. Secara khusus, data secara acak dibagi menjadi satu set pelatihan yang datanya digunakan untuk membangun model, dan set tes yang datanya digunakan untuk memvalidasi model yang dibangun. Sementara masing-masing</p>

	<p>metode memiliki kinerja yang baik secara keseluruhan, Naïve Bayes dengan pemilihan fitur mencapai kinerja terbaik, yang menghasilkan klasifikasi yang benar pada 340 dari 382 subjek, AUC 0,797 dan skor Brier 0,086. Selain itu, menyelidiki efektivitas memperkenalkan metode pemilihan fitur dalam kerangka analisis data. Hasil percobaan menunjukkan bahwa ketiga metode mencapai hasil yang jauh lebih baik dengan menggunakan pemilihan fitur, dan metode Naïve Bayes mencapai AUC sebesar 0,797 dan 0,745.</p>
Kesimpulan	<p>Metode Naïve Bayes menunjukkan kinerja yang unggul di atas multilayer perceptron dan metode hutan acak dan pemilihan fitur meningkatkan kinerja prediksi.</p>
Penulis	<p>Xia Hu, Peter D. Reaven, Aramesh Saremi, Ninghao Liu, Mohammad Ali Abbasi, Huan Liu dan Raymond Q. Migrino</p>
Nama jurnal, Volume, Nomor, Tahun	<p>EUARASIP Journal on Bioinformatics and Systems Biology, 14, 2016.</p>
Ulasan artikel	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk menyelidiki faktor-faktor penting yang mempengaruhi prediksi, menggunakan beberapa metode pembelajaran mesin, perkembangan cepat ketebalan intima-media karotid pada peserta toleransi glukosa yang terganggu. Dalam studi Actos Now for Prevention of Diabetes, 382 peserta dengan IGT menjalani evaluasi ultrasonografi intima-media karotid pada awal dan pada 15-18 bulan, dan dibagi menjadi pelanjut cepat dan pelambat tidak cepat. Perbandingan metode dan faktor dilakukan dengan menggunakan area di bawah analisis kurva karakteristik operasi penerima dan skor Brier. Hasil eksperimen menunjukkan bahwa metode pembelajaran yang diusulkan bekerja dengan baik dalam mengidentifikasi atau memprediksi RP. Di antara metode, kinerja Naïve Bayes adalah yang terbaik dibandingkan dengan perceptron multilayer dan hutan acak. Dengan berurusan dengan data multi-modal, metode pembelajaran yang diusulkan menunjukkan efektivitas dalam memprediksi prediabetik yang berisiko untuk perkembangan aterosklerosis yang cepat. makalah ini ada perbedaan yang</p>

	<p>signifikan dalam situs pendaftaran, proporsi dengan ras Hispanik, mikroalbumin urin, kreatinin plasma dan serum inhibitor aktivator plasminogen serum-1 antara RP dan NRP. Secara khusus, data secara acak dibagi menjadi satu set pelatihan yang datanya digunakan untuk membangun model, dan set tes yang datanya digunakan untuk memvalidasi model yang dibangun. Sementara masing-masing metode memiliki kinerja yang baik secara keseluruhan, Naïve Bayes dengan pemilihan fitur mencapai kinerja terbaik, yang menghasilkan klasifikasi yang benar pada 340 dari 382 subjek, AUC 0,797 dan skor Brier 0,086. Selain itu, menyelidiki efektivitas memperkenalkan metode pemilihan fitur dalam kerangka analisis data. Hasil percobaan menunjukkan bahwa ketiga metode mencapai hasil yang jauh lebih baik dengan menggunakan pemilihan fitur, dan metode Naïve Bayes mencapai AUC sebesar 0,797 dan 0,745. Metode Naïve Bayes menunjukkan kinerja yang unggul di atas multilayer perceptron dan metode hutan acak dan pemilihan fitur meningkatkan kinerja prediksi.</p>
--	--

No	9
Judul Artikel	Using a machine learning approach to predict mortality in critically ill influenza patients: a cross-sectional retrospective multicentre study in Taiwan.[9]
Topik	Data Mining
Data	-
Metode / Algoritma	XGBoost, Decision Tree.
Abstrak	Partisipan Sebanyak 336 pasien yang membutuhkan penerimaan ICU untuk influenza yang terbukti secara virologi di delapan rumah sakit selama epidemi influenza antara Oktober 2015 dan Maret 2016. Hasil Kumpulan data berisi 76 fitur dari 336 pasien dengan influenza berat. Tingkat keparahannya tampaknya tinggi, seperti yang ditunjukkan oleh skor Fisiologi Akut dan Kesehatan Kronis II yang tinggi dan skor indeks keparahan pneumonia. Area model XGBoost di bawah kurva mengungguli

	<p>RF dan LR untuk memprediksi mortalitas 30 hari. Untuk memberi dokter pemahaman intuitif tentang eksploitasi fitur, membuat stratifikasi fitur berdasarkan domain klinis. Kepentingan fitur kumulatif dalam domain keseimbangan cairan, domain ventilasi, domain data laboratorium, domain demografi dan gejala, domain manajemen dan domain skor keparahan masing-masing adalah 0,253, 0,113, 0,177, 0,140, 0,140, 0,152 dan 0,165. selanjutnya menggunakan plot SHAP untuk menggambarkan hubungan antara fitur dan mortalitas 30 hari pada pasien influenza yang sakit kritis.</p>
Hasil	<p>Sebanyak 336 pasien dengan influenza yang terbukti secara virologi terdaftar, dan 76 variabel dengan data lengkap dari 336 pasien ini dianalisis. Usia rata-rata pasien adalah 61 tahun, dan 62,8% adalah laki-laki. Untuk menyelidiki faktor-faktor yang terkait dengan kematian di rumah sakit, membagi 336 subyek menjadi kelompok yang selamat dan tidak selamat berdasarkan mortalitas pada 30 hari.</p>
Kesimpulan	<p>menggunakan fitur-fitur penting berbasis domain dan plot SHAP untuk realisasi yang divisualisasikan dan pendekatan ini setidaknya harus mengurangi sebagian kekhawatiran masalah kotak hitam. Penelitian prospektif di masa depan dijamin untuk memvalidasi model yang diusulkan dan menerjemahkan keunggulan model ML ke dalam hasil pasien yang ditingkatkan melalui DSS otomatis dan real-time.</p>
Penulis	<p>Chien-An Hu, Chia-Ming Chen, Yen-Chun Fang, Shinn-Jye Liang, Hao-Chien Wang, Wen-Feng Fang, Chau-Chyun Sheu, Wann-Cherng Perng, Kuang-Yao Yang, Kuo-Chin Kao, Chieh-Liang Wu, Chwei-Shyong Tsai, Ming-Yen Lin dan Wen-Cheng Chao.</p>
Nama jurnal, Volume, Nomor, Tahun	<p>BMJ Open, 10, 1136, 2020.</p>
Ulasan artikel	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk membutuhkan penerimaan ICU untuk influenza yang terbukti secara virologi di delapan rumah sakit selama epidemi influenza antara Oktober 2015 dan Maret 2016. Hasil Kumpulan data berisi 76 fitur dari 336 pasien dengan influenza berat. Tingkat keparahannya tampaknya tinggi, seperti yang ditunjukkan oleh</p>

	<p>skor Fisiologi Akut dan Kesehatan Kronis II yang tinggi dan skor indeks keparahan pneumonia. Area model XGBoost di bawah kurva mengungguli RF dan LR untuk memprediksi mortalitas 30 hari. Untuk memberi dokter pemahaman intuitif tentang eksploitasi fitur, membuat stratifikasi fitur berdasarkan domain klinis. Kepentingan fitur kumulatif dalam domain keseimbangan cairan, domain ventilasi, domain data laboratorium, domain demografi dan gejala, domain manajemen dan domain skor keparahan masing-masing adalah 0,253, 0,113, 0,177, 0,140, 0,140, 0,152 dan 0,165. selanjutnya menggunakan plot SHAP untuk menggambarkan hubungan antara fitur dan mortalitas 30 hari pada pasien influenza yang sakit kritis. Sebanyak 336 pasien dengan influenza yang terbukti secara virologi terdaftar, dan 76 variabel dengan data lengkap dari 336 pasien ini dianalisis. Usia rata-rata pasien adalah 61 tahun, dan 62,8% adalah laki-laki. Untuk menyelidiki faktor-faktor yang terkait dengan kematian di rumah sakit, membagi 336 subyek menjadi kelompok yang selamat dan tidak selamat berdasarkan mortalitas pada 30 hari. Dengan menggunakan fitur-fitur penting berbasis domain dan plot SHAP untuk realisasi yang divisualisasikan dan pendekatan ini setidaknya harus mengurangi sebagian kekhawatiran masalah kotak hitam. Penelitian prospektif di masa depan dijamin untuk memvalidasi model yang diusulkan dan menerjemahkan keunggulan model ML ke dalam hasil pasien yang ditingkatkan melalui DSS otomatis dan real-time.</p>
--	---

No	10
Judul Artikel	Diagnosis of Cervical cancer using CLAHE and SGLDM on RGB Pap smear Images through ANN.[10]
Topik	Data Mining
Data	Management and Decision Engineering Laboratory (MDE-LAB) database.

Metode / Algoritma	SGLDM- Spatial Gray Level Difference Method, RDM- Run Difference Method, LBP- Local Binary Pattern dan HOG- Histogram of Oriented Gradients.
Abstrak	<p>Dalam makalah ini, teknik yang diusulkan digunakan untuk meningkatkan gambar Pap smear dengan membandingkan Histogram Equalization dalam algoritma Contrast Stretching, Transformasi Hukum Daya untuk Koreksi Gamma, Koreksi Shading, Kontras Terbatas Adaptive Histogram Equalization. Klasifikasi JST digunakan untuk setiap algoritma ekstraksi fitur untuk mengevaluasi tingkat akurasi. Dengan demikian, CLAHE mencapai hasil yang baik untuk peningkatan dan algoritma ekstraksi fitur SGLDM mencapai akurasi 93% menggunakan ANN.</p>
Hasil	<p>Pada fase pertama membandingkan tiga algoritma peningkatan transformasi Power law untuk koreksi gamma, Contrast stretching dan Contrast Limited Adaptive Histogram Equalization untuk mencapai penilaian kualitas. Dengan demikian, CLAHE akan mempertahankan kualitas gambar yang baik dalam Pra-pemrosesan. Ada 13 fitur yang telah diekstraksi dari setiap algoritma ekstraksi fitur menggunakan Enhanced CLAHE Pap smear image. SGLDM memberikan ekstraksi fitur yang lebih baik dan akurasi 93,30% yang diperoleh dibandingkan dengan algoritma dan klasifikasi lain yang dilakukan menggunakan Neural Network. Operator LBP bekerja dengan delapan tetangga piksel, menggunakan nilai piksel tengah sebagai ambang. Nilai ambang batas adalah 90 telah diambil dari gambar CLAHE yang diusulkan. Penggunaan dengan set Enhanced CLAHE Pap smear Gambar telah diekstraksi dan dimuat dalam Neural Network dan melatih jaringan menggunakan propagasi gradien konjugasi skala belakang.</p>
Kesimpulan	<p>Gambar Pap smear mikroskopis kanker serviks untuk menentukan fakta sebenarnya dari keadaan kanker, yang membantu dokter untuk mengambil keputusan yang tepat dan dapat memberikan perawatan yang tepat kepada pasien</p>

	<p>sebelumnya. Dalam bidang medis sulit untuk mendeteksi kanker serviks pada gambar pap karena nuklei berukuran sangat kecil. Pada fase kedua, digunakan empat algoritma ekstraksi fitur untuk mengekstraksi fitur dari gambar CLAHE Pap smear yang diusulkan. SGLDM adalah ekstraksi fitur terbaik untuk Gambar Pap mikroskopis jika dibandingkan dengan LBP, RDM, HOG dan mencapai akurasi 93,30% menggunakan matriks kebingungan.</p>
Penulis	S. Jaya dan M. Latha.
Nama jurnal, Volume, Nomor, Tahun	International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9, 1, 2019.
Ulasan artikel	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk meningkatkan gambar Pap smear dengan membandingkan Histogram Equalization dalam algoritma Contrast Stretching, Transformasi Hukum Daya untuk Koreksi Gamma, Koreksi Shading, Kontras Terbatas Adaptive Histogram Equalization. Klasifikasi JST digunakan untuk setiap algoritma ekstraksi fitur untuk mengevaluasi tingkat akurasi. Dengan demikian, CLAHE mencapai hasil yang baik untuk peningkatan dan algoritma ekstraksi fitur SGLDM mencapai akurasi 93% menggunakan ANN. Pada penelitian ini fase pertama membandingkan tiga algoritma peningkatan transformasi Power law untuk koreksi gamma, Contrast stretching dan Contrast Limited Adaptive Histogram Equalization untuk mencapai penilaian kualitas. Dengan demikian, CLAHE akan mempertahankan kualitas gambar yang baik dalam Pra-pemrosesan. Ada 13 fitur yang telah diekstraksi dari setiap algoritma ekstraksi fitur menggunakan Enhanced CLAHE Pap smear image. SGLDM memberikan ekstraksi fitur yang lebih baik dan akurasi 93,30% yang diperoleh dibandingkan dengan algoritma dan klasifikasi lain yang dilakukan menggunakan Neural Network. Operator LBP bekerja dengan delapan tetangga piksel, menggunakan nilai piksel tengah sebagai ambang. Nilai ambang batas adalah 90 telah diambil dari gambar CLAHE yang diusulkan.</p>

	<p>Penggunaan dengan set Enhanced CLAHE Pap smear Gambar telah diekstraksi dan dimuat dalam Neural Network dan melatih jaringan menggunakan propagasi gradien konjugasi skala belakang. Gambar Pap smear mikroskopis kanker serviks untuk menentukan fakta sebenarnya dari keadaan kanker, yang membantu dokter untuk mengambil keputusan yang tepat dan dapat memberikan perawatan yang tepat kepada pasien sebelumnya. Dalam bidang medis sulit untuk mendeteksi kanker serviks pada gambar pap karena nuklei berukuran sangat kecil. Pada fase kedua, digunakan empat algoritma ekstraksi fitur untuk mengekstraksi fitur dari gambar CLAHE Pap smear yang diusulkan. SGLDM adalah ekstraksi fitur terbaik untuk Gambar Pap mikroskopis jika dibandingkan dengan LBP, RDM, HOG dan mencapai akurasi 93,30% menggunakan matriks kebingungan.</p>
--	--

No	11
Judul Artikel	Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning.[11]
Topik	Data Mining
Data	
Metode / Algoritma	
Abstrak	<p>Penerapan algoritma pendukung keputusan klinis untuk pencitraan medis menghadapi tantangan dengan keandalan dan interpretabilitas. Kerangka kerja menggunakan transfer learning, yang melatih jaringan saraf dengan sebagian kecil dari data pendekatan konvensional. Dengan menerapkan pendekatan ini pada kumpulan data gambar tomografi koherensi optik, mendemonstrasikan kinerja yang sebanding dengan kinerja ahli manusia dalam mengklasifikasikan degenerasi makula terkait usia dan edema makula diabetik. juga memberikan diagnosis yang lebih transparan dan dapat ditafsirkan dengan menyoroti daerah yang dikenali oleh jaringan saraf. selanjutnya menunjukkan penerapan umum sistem AI untuk diagnosis pneumonia pediatrik menggunakan gambar X-ray.</p>

<p>Hasil</p>	<p>Aplikasi utama dari algoritma pembelajaran transfer adalah dalam diagnosis gambar OCT retina. Spectral-domain OCT menggunakan cahaya untuk menangkap penampang retina optik resolusi tinggi in vivo yang dapat dirakit menjadi gambar tiga dimensi volume jaringan retina hidup. Selain itu, hampir 750.000 orang yang berusia 40 atau lebih menderita edema makula diabetik, suatu bentuk retinopati diabetik yang mengancam penglihatan yang melibatkan penumpukan cairan di retina sentral. OCT sangat penting untuk memandu administrasi terapi anti-VEGF dengan memberikan representasi penampang yang jelas dari patologi retina dalam kondisi ini, memungkinkan visualisasi lapisan retina individu, yang tidak mungkin dilakukan dengan pemeriksaan klinis oleh mata manusia atau dengan fotografi fundus berwarna.</p>
<p>Kesimpulan</p>	<p>Ketika model dilatih dengan jumlah gambar yang jauh lebih kecil, ia mempertahankan kinerja tinggi dalam akurasi, sensitivitas, spesifisitas, dan area di bawah kurva ROC untuk mencapai diagnosis dan rujukan yang benar, dengan demikian menggambarkan kekuatan sistem pembelajaran transfer untuk membuat sangat klasifikasi yang efektif, bahkan dengan set data pelatihan yang sangat terbatas. Meskipun platform AI dilatih dan divalidasi menggunakan sistem pencitraan Heidelberg Spectralis, standar Digital Imaging and Communications in Medicine membuat gambar OCT dari produsen yang berbeda cukup konsisten. Tujuan dari pendekatan pendahuluan ini adalah untuk mengembangkan sistem dan mendemonstrasikan metode yang tepat. Manfaat terbesar dari tes oklusi adalah bahwa ia mengungkapkan wawasan ke dalam keputusan jaringan saraf, yang dikenal sebagai "kotak hitam" tanpa transparansi. Semua area yang mengandung drusen dikenali dengan benar pada semua gambar yang digunakan untuk pengujian, sementara edema makula diabetik dan tes oklusi neovaskularisasi koroid kadang-kadang tidak menunjukkan hal yang menarik. Hal ini kemungkinan karena lesi dan kantong cairan neovaskularisasi koroid dan edema makula diabetik</p>

	kadang-kadang muncul jauh lebih besar daripada jendela oklusi, sementara drusen cenderung berukuran lebih kecil.
Penulis	Daniel S. Kermany, Michael Goldbaum, Wenjia Cai, M. Anthony Lewis, Huimin Xia, Kang Zhang.
Nama jurnal, Volume, Nomor, Tahun	Cell, 172, 1122-1131, 2018.
Ulasan artikel	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk pencitraan medis menghadapi tantangan dengan keandalan dan interpretabilitas. Kerangka kerja menggunakan transfer learning, yang melatih jaringan saraf dengan sebagian kecil dari data pendekatan konvensional. Dengan menerapkan pendekatan ini pada kumpulan data gambar tomografi koherensi optik, mendemonstrasikan kinerja yang sebanding dengan kinerja ahli manusia dalam mengklasifikasikan degenerasi makula terkait usia dan edema makula diabetik. juga memberikan diagnosis yang lebih transparan dan dapat ditafsirkan dengan menyoroti daerah yang dikenali oleh jaringan saraf. selanjutnya menunjukkan penerapan umum sistem AI untuk diagnosis pneumonia pediatrik menggunakan gambar X-ray. Ketika model dilatih dengan jumlah gambar yang jauh lebih kecil, ia mempertahankan kinerja tinggi dalam akurasi, sensitivitas, spesifisitas, dan area di bawah kurva ROC untuk mencapai diagnosis dan rujukan yang benar, dengan demikian menggambarkan kekuatan sistem pembelajaran transfer untuk membuat sangat klasifikasi yang efektif, bahkan dengan set data pelatihan yang sangat terbatas. Meskipun platform AI dilatih dan divalidasi menggunakan sistem pencitraan Heidelberg Spectralis, standar Digital Imaging and Communications in Medicine membuat gambar OCT dari produsen yang berbeda cukup konsisten. Tujuan dari pendekatan pendahuluan ini adalah untuk mengembangkan sistem dan mendemonstrasikan metode yang tepat. Manfaat terbesar dari tes oklusi adalah bahwa ia mengungkapkan wawasan ke dalam keputusan jaringan saraf, yang dikenal sebagai "kotak hitam" tanpa transparansi. Semua area yang mengandung drusen dikenali</p>

	<p>dengan benar pada semua gambar yang digunakan untuk pengujian, sementara edema makula diabetik dan tes oklusi neovaskularisasi koroid kadang-kadang tidak menunjukkan hal yang menarik. Hal ini kemungkinan karena lesi dan kantong cairan neovaskularisasi koroid dan edema makula diabetik kadang-kadang muncul jauh lebih besar daripada jendela oklusi, sementara drusen cenderung berukuran lebih kecil. Hasil dari aplikasi utama dari algoritma pembelajaran transfer adalah dalam diagnosis gambar OCT retina. Spectral-domain OCT menggunakan cahaya untuk menangkap penampang retina optik resolusi tinggi in vivo yang dapat dirakit menjadi gambar tiga dimensi volume jaringan retina hidup. Selain itu, hampir 750.000 orang yang berusia 40 atau lebih menderita edema makula diabetik, suatu bentuk retinopati diabetik yang mengancam penglihatan yang melibatkan penumpukan cairan di retina sentral. OCT sangat penting untuk memandu administrasi terapi anti-VEGF dengan memberikan representasi penampang yang jelas dari patologi retina dalam kondisi ini, memungkinkan visualisasi lapisan retina individu, yang tidak mungkin dilakukan dengan pemeriksaan klinis oleh mata manusia atau dengan fotografi fundus berwarna.</p>
--	---

No	12
Judul Artikel	An infrastructure for precision medicine through analysis of big data.[12]
Topik	Data Mining
Data	(i.e., laboratory exams, pathological anatomy exams, biopsy exams)
Metode / Algoritma	Collection layer, Anonymization layer, Machine learning and aggregation layer, Import and aggregate data.
Abstrak	Saat ini, semakin meningkatnya ketersediaan data omics, karena kemajuan dalam perolehan hasil biologi molekuler dan dalam teknologi simulasi biologi sistem, memberikan dasar untuk pengobatan presisi. Keberhasilan dalam kedokteran presisi tergantung pada akses ke layanan kesehatan dan data biomedis.

	<p>Digitalisasi sangat penting untuk mengumpulkan, berbagi, dan mengumpulkan volume besar data heterogen untuk mendukung penemuan pola tersembunyi dengan tujuan untuk mendefinisikan model prediksi untuk keperluan biomedis. Berbagi data pasien adalah proses penting. Data dari tiga platform berbeda dikumpulkan. Infrastruktur telah dirancang untuk memungkinkan ekstraksi dan agregasi data tidak terstruktur dan semi-terstruktur. Data diperlakukan dengan baik untuk memastikan keamanan dan privasi data. Infrastruktur ini memungkinkan integrasi berbagai sumber data yang anonim dari berbagai platform klinis yang beragam dan heterogen. Data tidak terstruktur dan semi-terstruktur diproses untuk mendapatkan analisis historis yang tepat dari aktivitas klinis satu atau lebih pasien. Agregasi data memungkinkan dilakukannya serangkaian penilaian statistik yang diperlukan untuk menjawab pertanyaan kompleks yang dapat digunakan dalam berbagai bidang, seperti kedokteran prediktif dan presisi.</p>
<p>Hasil</p>	<p>hasil yang diperoleh dengan menganalisis data dari laboratorium, biopsi dan ujian anatomi patologis dengan perhatian khusus pada mekanisme impor dan agregasi dari berbagai jenis nilai. Data laboratorium sebagian besar bersifat numerik dengan kemungkinan adanya komentar teks bebas. Data yang diambil dari laboratorium diekspor ke file Format File Nilai Dipisahkan Koma. Itu mungkin untuk mengamati terjadinya pasien di pusat serta untuk mengevaluasi berapa banyak pasien mengulangi ujian beberapa kali. Setelah perubahan dalam sistem manajemen ujian, sejak 2014 data dari ujian biopsi disimpan dalam dua sumber data yang berbeda. Adapun sumber sebelumnya, data yang diekspor ke format Format File Microsoft Excel terdiri dari 3587 baris untuk 18 bidang yang berbeda, kebanyakan dari mereka teks gratis. Sedangkan untuk sumber terakhir, data yang diekspor dalam format Format File Nilai Dipisahkan Koma terdiri dari 3130 baris. Bidang utama seperti kuis diagnostik, situs nyeri, administrasi farmakologis, dan deskripsi ujian disediakan</p>

sebagai teks gratis. Data ujian bioptik dari kedua sumber data terkait dengan 6716 kunjungan dibagi menjadi 5.186 pasien. Data terkait dengan 264.339 kunjungan dibagi menjadi 15.442 pasien. Metode pembelajaran mesin yang diawasi telah digunakan untuk menganalisis ujian yang dilaporkan sebagai teks bebas. Sedangkan untuk pemeriksaan anatomi, bidang yang relevan terkait dengan lokasi sampel dan diagnosis. Dalam Tabel 4 dilaporkan beberapa contoh kalimat yang digunakan untuk melatih dan menguji classifier. 3 menggunakan peta panas untuk merepresentasikan akurasi rata-rata dari simulasi yang berbeda. Seperti yang dilakukan sebelumnya, beberapa penilaian diekspor, diklasifikasikan secara manual, dan digunakan untuk melatih dan menguji pengklasifikasi. Akurasi rata-rata pengklasifikasi adalah 0,8 menunjukkan bahwa sistem mampu membuat katalog jenis penyakit dengan keandalan yang baik. Akhirnya, pengklasifikasi diterapkan untuk semua data yang tidak terklasifikasi yang dikumpulkan untuk periode yang dipertimbangkan dan dimungkinkan untuk mengklasifikasikan pemeriksaan yang berbeda sesuai dengan lokasi sampel dan klasifikasi diagnosis. Sebagai contoh, untuk tahun 2015 ditemukan 509 kasus «Kontrol / Monitoring / Tanpa Neoplastik», 779 kasus «Preneoplastik» dan 91 kasus «Neoplastik» dilaporkan terjadi pada sistem pencernaan. Pada saat yang sama, analisis lebih lanjut dilakukan pada data uji dalam biopsi dengan menentukan pengklasifikasi yang terkait dengan kepositifan atau negativitas yang terkait dengan kanker usus besar. 5 menunjukkan akurasi rata-rata dari 500 simulasi yang dilakukan untuk pengklasifikasi ini. Eksperimen yang dilakukan pada tes biopsi memungkinkan untuk mengamati keadaan patologi dengan menganalisis riwayat klinis pasien. 6 memungkinkan untuk mengamati berapa banyak perubahan keadaan dikaitkan dengan pasien yang berbeda. Informasi ini memungkinkan untuk mengidentifikasi pasien yang telah memulai jalur klinis dalam keadaan negatif dan kemudian menunjukkan penyakitnya. Catatan klinis agregat

	<p>memungkinkan untuk mengevaluasi riwayat klinis pasien dan untuk mengevaluasi berbagai pemeriksaan laboratorium yang dilakukan dari waktu ke waktu dan untuk menghubungkannya dengan departemen pemeriksaan lainnya. 7 memperhitungkan 130 pasien yang telah memiliki penyakit sejak mereka telah dirawat di Pusat Diag-nostik Italia dan berapa banyak ujian yang dilakukan masing-masing selama 30, 180 dan 365 hari. Seperti yang bisa diamati, pemeriksaan yang dilakukan oleh banyak pasien adalah pemeriksaan hitung darah lengkap.</p>
Kesimpulan	<p>Infrastruktur yang dikembangkan memungkinkan integrasi berbagai sumber data anonim dan heterogen dari berbagai platform klinis yang digunakan oleh Pusat Diag-nostik Italia. Melalui anonimisasi, menjaga pembagian data tanpa alasan privasi dan keamanan. Algoritme yang diimplementasikan memungkinkan ekstraksi dan penggunaan data tidak terstruktur atau semi-terstruktur, memperoleh analisis historis yang tepat dari aktivitas klinis satu atau lebih pasien. Secara khusus, mempelajari riwayat klinis pasien yang telah mengembangkan patologi serupa, dapat memungkinkan untuk memprediksi atau membedakan tanda yang memungkinkan diagnosis dini dari kemungkinan penyakit.</p>
Penulis	<p>Marco Moscatelli, Andrea Manconi, Mauro Pessina, Giovanni Fellegara, Stefano Rampoldi, Luciano Milanese, Andrea Casasco dan Matteo Gnocchi.</p>
Nama jurnal, Volume, Nomor, Tahun	<p>BMC Bioinformatics, 19:351, 2018.</p>
Ulasan artikel	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk kemajuan dalam perolehan hasil biologi molekuler dan dalam teknologi simulasi biologi sistem, memberikan dasar untuk pengobatan presisi. Keberhasilan dalam kedokteran presisi tergantung pada akses ke layanan kesehatan dan data biomedis. Digitalisasi sangat penting untuk mengumpulkan, berbagi, dan mengumpulkan volume besar data heterogen untuk mendukung penemuan pola tersembunyi dengan tujuan untuk mendefinisikan model prediksi untuk keperluan biomedis.</p>

Berbagai data pasien adalah proses penting. Data dari tiga platform berbeda dikumpulkan. Infrastruktur telah dirancang untuk memungkinkan ekstraksi dan agregasi data tidak terstruktur dan semi-terstruktur. Data diperlakukan dengan baik untuk memastikan keamanan dan privasi data. Infrastruktur ini memungkinkan integrasi berbagai sumber data yang anonim dari berbagai platform klinis yang beragam dan heterogen. Data tidak terstruktur dan semi-terstruktur diproses untuk mendapatkan analisis historis yang tepat dari aktivitas klinis satu atau lebih pasien. Agregasi data memungkinkan dilakukannya serangkaian penilaian statistik yang diperlukan untuk menjawab pertanyaan kompleks yang dapat digunakan dalam berbagai bidang, seperti kedokteran prediktif dan presisi. Artikel ini memiliki infrastruktur yang dikembangkan memungkinkan integrasi berbagai sumber data anonim dan heterogen dari berbagai platform klinis yang digunakan oleh Pusat Diag-nostik Italia. Melalui anonimisasi, menjaga pembagian data tanpa alasan privasi dan keamanan. Algoritme yang diimplementasikan memungkinkan ekstraksi dan penggunaan data tidak terstruktur atau semi-terstruktur, memperoleh analisis historis yang tepat dari aktivitas klinis satu atau lebih pasien. Secara khusus, mempelajari riwayat klinis pasien yang telah mengembangkan patologi serupa, dapat memungkinkan untuk memprediksi atau membedakan tanda yang memungkinkan diagnosis dini dari kemungkinan penyakit. Hasil yang diperoleh dengan menganalisis data dari laboratorium, biopsi dan ujian anatomi patologis dengan perhatian khusus pada mekanisme impor dan agregasi dari berbagai jenis nilai. Data laboratorium sebagian besar bersifat numerik dengan kemungkinan adanya komentar teks bebas. Data yang diambil dari laboratorium diekspor ke file Format File Nilai Dipisahkan Koma. Itu mungkin untuk mengamati terjadinya pasien di pusat serta untuk mengevaluasi berapa banyak pasien mengulangi ujian beberapa kali. Setelah perubahan dalam sistem manajemen ujian, sejak 2014 data dari

ujian biopsi disimpan dalam dua sumber data yang berbeda. Adapun sumber sebelumnya, data yang diekspor ke format Format File Microsoft Excel terdiri dari 3587 baris untuk 18 bidang yang berbeda, kebanyakan dari mereka teks gratis. Sedangkan untuk sumber terakhir, data yang diekspor dalam format Format File Nilai Dipisahkan Koma terdiri dari 3130 baris. Bidang utama seperti kuis diagnostik, situs nyeri, administrasi farmakologis, dan deskripsi ujian disediakan sebagai teks gratis. Data ujian bioptik dari kedua sumber data terkait dengan 6716 kunjungan dibagi menjadi 5.186 pasien. Data terkait dengan 264.339 kunjungan dibagi menjadi 15.442 pasien. Metode pembelajaran mesin yang diawasi telah digunakan untuk menganalisis ujian yang dilaporkan sebagai teks bebas. Sedangkan untuk pemeriksaan anatomi, bidang yang relevan terkait dengan lokasi sampel dan diagnosis. Dalam Tabel 4 dilaporkan beberapa contoh kalimat yang digunakan untuk melatih dan menguji classifier. 3 menggunakan peta panas untuk merepresentasikan akurasi rata-rata dari simulasi yang berbeda. Seperti yang dilakukan sebelumnya, beberapa penilaian diekspor, diklasifikasikan secara manual, dan digunakan untuk melatih dan menguji pengklasifikasi. Akurasi rata-rata pengklasifikasi adalah 0,8 menunjukkan bahwa sistem mampu membuat katalog jenis penyakit dengan keandalan yang baik. Akhirnya, pengklasifikasi diterapkan untuk semua data yang tidak terklasifikasi yang dikumpulkan untuk periode yang dipertimbangkan dan dimungkinkan untuk mengklasifikasikan pemeriksaan yang berbeda sesuai dengan lokasi sampel dan klasifikasi diagnosis. Sebagai contoh, untuk tahun 2015 ditemukan 509 kasus «Kontrol / Monitoring / Tanpa Neoplastik», 779 kasus «Preneoplastik» dan 91 kasus «Neoplastik» dilaporkan terjadi pada sistem pencernaan. Pada saat yang sama, analisis lebih lanjut dilakukan pada data uji dalam biopsi dengan menentukan pengklasifikasi yang terkait dengan kepositifan atau negativitas yang terkait dengan kanker usus besar. 5 menunjukkan akurasi rata-rata dari 500 simulasi

	<p>yang dilakukan untuk pengklasifikasi ini. Eksperimen yang dilakukan pada tes biopsi memungkinkan untuk mengamati keadaan patologi dengan menganalisis riwayat klinis pasien. 6 memungkinkan untuk mengamati berapa banyak perubahan keadaan dikaitkan dengan pasien yang berbeda. Informasi ini memungkinkan untuk mengidentifikasi pasien yang telah memulai jalur klinis dalam keadaan negatif dan kemudian menunjukkan penyakitnya. Catatan klinis agregat memungkinkan untuk mengevaluasi riwayat klinis pasien dan untuk mengevaluasi berbagai pemeriksaan laboratorium yang dilakukan dari waktu ke waktu dan untuk menghubungkannya dengan departemen pemeriksaan lainnya. 7 memperhitungkan 130 pasien yang telah memiliki penyakit sejak mereka telah dirawat di Pusat Diag-nostik Italia dan berapa banyak ujian yang dilakukan masing-masing selama 30, 180 dan 365 hari. Seperti yang bisa diamati, pemeriksaan yang dilakukan oleh banyak pasien adalah pemeriksaan hitung darah lengkap.</p>
--	---

No	13
Judul Artikel	Imputation and characterization of uncoded self-harm in major mental illness using machine learning.[13]
Topik	Data Mining
Data	The IBM MarketScan database (2003-2016).
Metode / Algoritma	Natural Language Processing (NLP), Regression Methods, Random Forests, Bayesian Models.
Abstrak	<p>Penelitian ini bertujuan untuk memperhitungkan tindakan menyakiti diri sendiri yang tidak dikodekan dalam data klaim administratif individu dengan penyakit mental utama, mencirikan insiden tindakan menyakiti diri sendiri, dan mengidentifikasi faktor yang terkait dengan bias pengkodean. Lima pengklasifikasi pembelajaran mesin diuji pada subset data yang seimbang, dengan XGBoost dipilih untuk set data lengkap. Kinerja klasifikasi divalidasi melalui kesalahan label data acak dan perbandingan dengan «standar emas» yang diturunkan dari dokter. Insiden selfharm yang dikodekan dan diperhitungkan</p>

	<p>ditandai dengan tahun, usia pasien, jenis kelamin, U. Undercoding menyakiti diri sendiri lebih tinggi pada pria daripada pada individu wanita dan meningkat seiring bertambahnya usia. Hanya 1 dari 19 peristiwa melukai diri sendiri yang dikodekan untuk individu dengan MMI. ML dapat secara efektif memulihkan tindakan menyakiti diri yang tidak tercatat dalam data klaim dan menginformasikan studi epidemiologis dan observasi psikiatris.</p>
<p>Hasil</p>	<p>Dari 20.783.244 kunjungan-meta yang dicatat selama 29.799.203 tahun pengamatan pasien, probabilitas model data-lengkap XGboost dari selfharm dijumlahkan menjadi 1.592.703, sesuai dengan insiden tahunan yang diperhitungkan secara keseluruhan sebesar 5,34%. Dari semua 20 juta kunjungan meta, 842263 memiliki probabilitas kelas $1 > 0,5$ dan 246 511 memiliki probabilitas kelas $1 > 0,95$. Juga, dari 83.113 kunjungan-meta yang dikodekan untuk melukai diri sendiri, 79.882 memiliki probabilitas kelas $1 > 0,5$ dan 62929 memiliki probabilitas kelas $1 > 0,95$. Performa model ML berbasis XGboost yang dilatih dan diuji pada kumpulan data berbeda ditunjukkan pada Tabel 1. Kinerja dari 5 algoritma klasifikasi ML yang berbeda yang diterapkan pada dataset yang seimbang ditunjukkan pada Tabel 2. Kesepakatan berpasangan antara Full-data-model, individu ahli klinis, dan standar emas konsensus ditunjukkan pada Tabel 3. Dari 200 meta -kunjungan, 79 dikategorikan sebagai kelas 1 oleh standar emas. Di antara 100 kunjungan meta dengan tindakan menyakiti diri sendiri, para ahli klinis memberi label 52 seperti itu, sedangkan dalam 100 kunjungan meta tanpa tindakan menyakiti diri sendiri yang terdokumentasi, dokter mengklasifikasikan hanya 27 yang memiliki tindakan menyakiti diri sendiri. Kesepakatan keseluruhan antara ML dan standar emas adalah 84%. 15 kovariat dengan skor perolehan tertinggi ditunjukkan pada Tabel 4. Model yang dibangun secara eksklusif dengan kovariat kondisi hanya memiliki kinerja yang sedikit lebih buruk dibandingkan dengan model yang dibangun dengan semua kelas</p>

	<p>kovariat. Menambahkan konsep leluhur tingkat tinggi memiliki efek yang dapat diabaikan pada kinerja model, kecuali untuk prosedur, di mana skor AUC meningkat dari 0,800 menjadi 0,828 setelah menambahkan istilah leluhur ICD-10-PCS. Ketika perangkat kovariat tidak umum dan prediktif buruk digunakan sendiri, AUC dan akurasi 0,51, dan PKS 0,1. 100 kovariat yang berkontribusi paling besar pada Coding-biasmodel ada di Tabel Tambahan S2.</p>
Kesimpulan	<p>Insiden melukai diri sendiri memuncak lebih cepat dan turun lebih awal pada pasien wanita vs pria muda dengan MMI. Namun, pasien laki-laki lebih cenderung melukai diri sendiri dengan undercode pada semua usia, menunjukkan kemungkinan bias pelabelan penyedia terkait dengan stereotip jenis kelamin. Indikator psikiatris tentang bunuh diri dan kejadian somatik dikaitkan dengan pengkodean dengan melukai diri sendiri yang lebih tinggi pada pasien dengan MMI.</p>
Penulis	<p>Praveen Kumar, Anastasiya Nestsiarovich, Stuart J. Nelson, Berit Kerner, Douglas J. Perkins dan Christophe G. Lambert</p>
Nama jurnal, Volume, Nomor, Tahun	<p>American Medical Informatics Association (AMIA), 27, 136-146, 2020.</p>
Ulasan artikel	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk memperhitungkan tindakan menyakiti diri sendiri yang tidak dikodekan dalam data klaim administratif individu dengan penyakit mental utama, mencirikan insiden tindakan menyakiti diri sendiri, dan mengidentifikasi faktor yang terkait dengan bias pengkodean. Lima pengklasifikasi pembelajaran mesin diuji pada subset data yang seimbang, dengan XGBoost dipilih untuk set data lengkap. Kinerja klasifikasi divalidasi melalui kesalahan label data acak dan perbandingan dengan «standar emas» yang diturunkan dari dokter. Insiden selfharm yang dikodekan dan diperhitungkan ditandai dengan tahun, usia pasien, jenis kelamin, U. Undercoding menyakiti diri sendiri lebih tinggi pada pria daripada pada individu wanita dan meningkat seiring bertambahnya usia. Hanya 1 dari 19 peristiwa</p>

melukai diri sendiri yang dikodekan untuk individu dengan MMI. ML dapat secara efektif memulihkan tindakan menyakiti diri yang tidak tercatat dalam data klaim dan menginformasikan studi epidemiologis dan observasi psikiatris. Dalam artikel ini insiden melukai diri sendiri memuncak lebih cepat dan turun lebih awal pada pasien wanita vs pria muda dengan MMI. Namun, pasien laki-laki lebih cenderung melukai diri sendiri dengan undercode pada semua usia, menunjukkan kemungkinan bias pelabelan penyedia terkait dengan stereotip jenis kelamin. Indikator psikiatris tentang bunuh diri dan kejadian somatik dikaitkan dengan pengkodean dengan melukai diri sendiri yang lebih tinggi pada pasien dengan MMI. Dari 20.783.244 kunjungan-meta yang dicatat selama 29.799.203 tahun pengamatan pasien, probabilitas model data-lengkap XGboost dari selfharm dijumlahkan menjadi 1.592.703, sesuai dengan insiden tahunan yang diperhitungkan secara keseluruhan sebesar 5,34%. Dari semua 20 juta kunjungan meta, 842263 memiliki probabilitas kelas $1 > 0,5$ dan 246 511 memiliki probabilitas kelas $1 > 0,95$. Juga, dari 83.113 kunjungan-meta yang dikodekan untuk melukai diri sendiri, 79.882 memiliki probabilitas kelas $1 > 0,5$ dan 62929 memiliki probabilitas kelas $1 > 0,95$. Performa model ML berbasis XGboost yang dilatih dan diuji pada kumpulan data berbeda ditunjukkan pada Tabel 1. Kinerja dari 5 algoritma klasifikasi ML yang berbeda yang diterapkan pada dataset yang seimbang ditunjukkan pada Tabel 2. Kesepakatan berpasangan antara Full-data-model, individu ahli klinis, dan standar emas konsensus ditunjukkan pada Tabel 3. Dari 200 meta -kunjungan, 79 dikategorikan sebagai kelas 1 oleh standar emas. Di antara 100 kunjungan meta dengan tindakan menyakiti diri sendiri, para ahli klinis memberi label 52 seperti itu, sedangkan dalam 100 kunjungan meta tanpa tindakan menyakiti diri sendiri yang terdokumentasi, dokter mengklasifikasikan hanya 27 yang memiliki tindakan menyakiti diri sendiri. Kesepakatan keseluruhan antara ML dan standar emas adalah 84%. 15 kovariat dengan skor perolehan tertinggi

	<p>ditunjukkan pada Tabel 4. Model yang dibangun secara eksklusif dengan kovariat kondisi hanya memiliki kinerja yang sedikit lebih buruk dibandingkan dengan model yang dibangun dengan semua kelas kovariat. Menambahkan konsep leluhur tingkat tinggi memiliki efek yang dapat diabaikan pada kinerja model, kecuali untuk prosedur, di mana skor AUC meningkat dari 0,800 menjadi 0,828 setelah menambahkan istilah leluhur ICD-10-PCS. Ketika perangkat kovariat tidak umum dan prediktif buruk digunakan sendiri, AUC dan akurasi 0,51, dan PKS 0,1. 100 kovariat yang berkontribusi paling besar pada Coding-biasmodel ada di Tabel Tambahan S2.</p>
--	---

No	14
Judul Artikel	API design for machine learning software: experiences from the scikit-learn project.[14]
Topik	Data Mining
Data	
Metode / Algoritma	
Abstrak	<p>scikit-learn adalah library machine learning yang semakin populer. Ditulis dengan Python, ia dirancang untuk menjadi sederhana dan efisien, dapat diakses oleh non-ahli, dan dapat digunakan kembali dalam berbagai konteks. Dalam tulisan ini, menyajikan dan mendiskusikan pilihan desain untuk antarmuka pemrograman aplikasi proyek. Secara khusus, menjelaskan antarmuka sederhana dan elegan yang dimiliki oleh semua unit pembelajaran dan pemrosesan di perpustakaan dan kemudian membahas kelebihanannya dalam hal komposisi dan kegunaan kembali. Makalah ini juga mengomentari detail implementasi khusus untuk ekosistem Python dan menganalisis hambatan yang dihadapi oleh pengguna dan pengembang perpustakaan.</p>
Hasil	<p>Ada beberapa arah yang menjadi fokus proyek scikit-learn dalam pengembangan masa depan. Pada tingkat yang lebih rendah, pemrosesan paralel merupakan titik peningkatan yang potensial. Beberapa estimator di scikit-learn sudah dapat memanfaatkan prosesor multicore, tetapi hanya dengan cara</p>

	<p>kasar. Oleh karena itu, setiap dekomposisi tugas paralel harus dilakukan di dalam modul Cython, atau pada tingkat yang cukup tinggi untuk menjamin overhead pembuatan beberapa proses tingkat OS, dan komunikasi antar-proses berikutnya. Pencarian grid paralel adalah contoh dari pendekatan terakhir yang telah diterapkan. Versi terbaru Cython menyertakan dukungan untuk standar OpenMP, yang merupakan kandidat teknologi yang layak untuk dukungan multicore yang lebih halus di scikit-learn. Masalah API ini akan diatasi di masa mendatang sebagai persiapan untuk rilis 1.0 scikit-learn.</p>
Kesimpulan	<p>Scikit-learn API dan caranya memetakan konsep dan tugas machine learning ke objek dan operasi dalam bahasa pemrograman Python. Melalui pengetikan bebek, API yang konsisten mengarah ke pustaka yang mudah diperluas, dan memungkinkan estimator yang ditentukan pengguna untuk digabungkan ke dalam alur kerja scikit-learn tanpa pewarisan objek eksplisit apa pun. Meskipun bagian dari scikit-learn API harus khusus Python, konsep inti mungkin berlaku untuk aplikasi pembelajaran mesin dan toolkit yang ditulis dalam bahasa pemrograman lain. Kekuatan, dan ekstensibilitas dari scikit-learn API dibuktikan oleh basis pengguna yang besar dan terus bertambah, penggunaannya untuk memecahkan masalah nyata di berbagai bidang, serta tampilan paket pihak ketiga yang mengikuti scikit- pelajari konvensi. Dirilis sebagai bagian dari paket lightning8, juga mengikuti konvensi scikit-learn.</p>
Penulis	<p>Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas C. Müller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton¹⁰, Jake Vanderplas, Arnaud Joly, Brian Holt dan Gaël Varoquaux.</p>
Nama jurnal, Volume, Nomor, Tahun	-
Ulasan artikel	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk menjadi sederhana dan efisien, dapat diakses oleh non-ahli, dan dapat digunakan kembali dalam berbagai konteks. Dalam</p>

tulisan ini, menyajikan dan mendiskusikan pilihan desain untuk antarmuka pemrograman aplikasi proyek. Secara khusus, menjelaskan antarmuka sederhana dan elegan yang dimiliki oleh semua unit pembelajaran dan pemrosesan di perpustakaan dan kemudian membahas kelebihanannya dalam hal komposisi dan kegunaan kembali. Makalah ini juga mengomentari detail implementasi khusus untuk ekosistem Python dan menganalisis hambatan yang dihadapi oleh pengguna dan pengembang perpustakaan. Scikit-learn API dan caranya memetakan konsep dan tugas machine learning ke objek dan operasi dalam bahasa pemrograman Python. Melalui pengetikan bebek, API yang konsisten mengarah ke pustaka yang mudah diperluas, dan memungkinkan estimator yang ditentukan pengguna untuk digabungkan ke dalam alur kerja scikit-learn tanpa pewarisan objek eksplisit apa pun. Meskipun bagian dari scikit-learn API harus khusus Python, konsep inti mungkin berlaku untuk aplikasi pembelajaran mesin dan toolkit yang ditulis dalam bahasa pemrograman lain. Kekuatan, dan ekstensibilitas dari scikit-learn API dibuktikan oleh basis pengguna yang besar dan terus bertambah, penggunaannya untuk memecahkan masalah nyata di berbagai bidang, serta tampilan paket pihak ketiga yang mengikuti scikit-learn konvensi. Dirilis sebagai bagian dari paket lightning8, juga mengikuti konvensi scikit-learn. Hasil dari artikel ini beberapa arah yang menjadi fokus proyek scikit-learn dalam pengembangan masa depan. Pada tingkat yang lebih rendah, pemrosesan paralel merupakan titik peningkatan yang potensial. Beberapa estimator di scikit-learn sudah dapat memanfaatkan prosesor multicore, tetapi hanya dengan cara kasar. Oleh karena itu, setiap dekomposisi tugas paralel harus dilakukan di dalam modul Cython, atau pada tingkat yang cukup tinggi untuk menjamin overhead pembuatan beberapa proses tingkat OS, dan komunikasi antar-proses berikutnya. Pencarian grid paralel adalah contoh dari pendekatan terakhir yang telah diterapkan. Versi terbaru Cython menyertakan dukungan untuk standar OpenMP, yang merupakan kandidat teknologi yang

	layak untuk dukungan multicore yang lebih halus di scikit-learn. Masalah API ini akan diatasi di masa mendatang sebagai persiapan untuk rilis 1.0 scikit-learn.
--	---

No	15
Judul Artikel	Aerial Mapping of Forests Affected by Pathogens Using UAVs, Hyperspectral Sensors, and Artificial Intelligence.[15]
Topik	Data Mining
Data	
Metode / Algoritma	
Abstrak	<p>Pendekatan penginderaan jauh baru-baru ini telah menawarkan survei yang cepat, berskala luas, dan terjangkau serta indikator tambahan yang dapat melengkapi pengujian di lapangan. Makalah ini mengusulkan kerangka kerja yang menggabungkan wawasan berbasis situs dan kemampuan penginderaan jauh untuk mendeteksi dan mensegmentasi kerusakan oleh patogen jamur di hutan alam dan hutan tanaman. Pendekatan ini diilustrasikan dengan kasus percobaan karat myrtle pada pohon teh paperbark di New South Wales, Australia. Wawasan mengungkapkan tingkat deteksi individu sebesar 95% untuk pohon yang sehat, 97% untuk pohon yang rusak, dan tingkat deteksi multikelas global sebesar 97%.</p>
Hasil	<p>Untuk memvisualisasikan manfaat dari memasukkan skema optimasi di Langkah 8 dari Algoritma 1, tingkat deteksi dilacak dengan melatih dan menjalankan classifier beberapa kali dengan hanya satu set fitur yang difilter per instance. Fitur-fitur tersebut diberi peringkat dengan relevansinya oleh pengklasifikasi XGBoost dan diurutkan secara konsekuen, seperti yang diilustrasikan pada Gambar 7. Pengklasifikasi dapat mencapai tingkat akurasi yang tinggi melebihi 97% dari akurasi global ketika memproses data menggunakan dari 10 hingga 40 fitur saja, dengan jumlah yang optimal fitur dari 24. Di sisi lain, pengklasifikasi hanya meningkatkan register mereka ketika jumlah fitur yang diproses lebih banyak. Ciri-ciri yang paling relevan dari studi kasus ini digambarkan pada Tabel 1 dan</p>

	<p>Gambar 8. Meskipun ilustrasi mereka menunjukkan wawasan tentang intensitas yang dapat dibedakan antara kawasan pohon yang sehat dan yang terkena dampak dari Gambar 5, rangkaian fitur ini tidak cukup untuk mensegmentasi kawasan objek lain. Selain itu, fitur yang diproses dengan kernel 2D memperoleh skor relevansi yang lebih baik daripada rekan yang tidak diproses. Perbedaan itu bahkan lebih besar untuk fitur yang diproses menggunakan kernel jendela besar dengan mempertimbangkan bahwa jumlah noise yang tinggi, yang umum dalam citra hiperspektral mentah, mengubah performa pendekatan.</p>
Kesimpulan	<p>Makalah ini menjelaskan metodologi pipeline untuk deteksi dan pemetaan yang efektif dari indikator kesehatan yang buruk di hutan dan pohon perkebunan yang mengintegrasikan pendekatan teknologi UAS dan kecerdasan buatan. Teknik diilustrasikan dengan klasifikasi yang akurat dan tugas segmentasi pohon teh paperbark yang rusak oleh karat myrtle dari percobaan eksklusi di NSW, Australia. Di sini, sistem mencapai tingkat deteksi 97,24% untuk pohon sehat dan 94,72% untuk pohon yang terkena dampak. Algoritma tersebut memperoleh tingkat deteksi multikelas sebesar 97,35%. Pendekatan ini dapat digunakan untuk melatih berbagai kumpulan data dari berbagai sensor untuk meningkatkan tingkat deteksi yang ditawarkan solusi tunggal serta kemampuan memproses kumpulan data besar menggunakan perangkat lunak freeware. Studi kasus mendemonstrasikan pendekatan efektif yang memungkinkan indikator cepat dan akurat, dan untuk perubahan area yang terpapar pada tahap awal. Namun, pemahaman epidemiologi penyakit dan interaksi antara patogen dan inang masih diperlukan untuk penggunaan teknologi ini secara efektif.</p>
Penulis	Juan Sandino, Geoff Pegg, Felipe Gonzalez dan Grant Smith.
Nama jurnal, Volume, Nomor, Tahun	Sensors, 18, 944, 2018.
Ulasan artikel	Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk

mengusulkan kerangka kerja yang menggabungkan wawasan berbasis situs dan kemampuan penginderaan jauh untuk mendeteksi dan mensegmentasi kerusakan oleh patogen jamur di hutan alam dan hutan tanaman. Pendekatan ini diilustrasikan dengan kasus percobaan karat myrtle pada pohon teh paperbark di New South Wales, Australia. Wawasan mengungkapkan tingkat deteksi individu sebesar 95% untuk pohon yang sehat, 97% untuk pohon yang rusak, dan tingkat deteksi multikelas global sebesar 97%. Makalah ini menjelaskan metodologi pipeline untuk deteksi dan pemetaan yang efektif dari indikator kesehatan yang buruk di hutan dan pohon perkebunan yang mengintegrasikan pendekatan teknologi UAS dan kecerdasan buatan. Teknik diilustrasikan dengan klasifikasi yang akurat dan tugas segmentasi pohon teh paperbark yang rusak oleh karat myrtle dari percobaan eksklusif di NSW, Australia. Di sini, sistem mencapai tingkat deteksi 97,24% untuk pohon sehat dan 94,72% untuk pohon yang terkena dampak. Algoritma tersebut memperoleh tingkat deteksi multikelas sebesar 97,35%. Pendekatan ini dapat digunakan untuk melatih berbagai kumpulan data dari berbagai sensor untuk meningkatkan tingkat deteksi yang ditawarkan solusi tunggal serta kemampuan memproses kumpulan data besar menggunakan perangkat lunak freeware. Studi kasus mendemonstrasikan pendekatan efektif yang memungkinkan indikator cepat dan akurat, dan untuk perubahan area yang terpapar pada tahap awal. Namun, pemahaman epidemiologi penyakit dan interaksi antara patogen dan inang masih diperlukan untuk penggunaan teknologi ini secara efektif. Untuk memvisualisasikan manfaat dari memasukkan skema optimasi di Langkah 8 dari Algoritma 1, tingkat deteksi dilacak dengan melatih dan menjalankan classifier beberapa kali dengan hanya satu set fitur yang difilter per instance. Fitur-fitur tersebut diberi peringkat dengan relevansinya oleh pengklasifikasi XGBoost dan diurutkan secara konsekuen, seperti yang diilustrasikan pada Gambar 7. Pengklasifikasi dapat mencapai tingkat akurasi yang tinggi

	<p>melebihi 97% dari akurasi global ketika memproses data menggunakan dari 10 hingga 40 fitur saja, dengan jumlah yang optimal fitur dari 24. Di sisi lain, pengklasifikasi hanya meningkatkan register mereka ketika jumlah fitur yang diproses lebih banyak. Ciri-ciri yang paling relevan dari studi kasus ini digambarkan pada Tabel 1 dan Gambar 8. Meskipun ilustrasi mereka menunjukkan wawasan tentang intensitas yang dapat dibedakan antara kawasan pohon yang sehat dan yang terkena dampak dari Gambar 5, rangkaian fitur ini tidak cukup untuk mensegmentasi kawasan objek lain. Selain itu, fitur yang diproses dengan kernel 2D memperoleh skor relevansi yang lebih baik daripada rekan yang tidak diproses. Perbedaan itu bahkan lebih besar untuk fitur yang diproses menggunakan kernel jendela besar dengan mempertimbangkan bahwa jumlah noise yang tinggi, yang umum dalam citra hiperspektral mentah, mengubah performa pendekatan.</p>
--	--

No	16
Judul Artikel	A metabolite-based machine learning approach to diagnose Alzheimer-type dementia in blood: Results from the European Medical Information Framework for Alzheimer disease biomarker discovery cohort.[16]
Topik	Data Mining
Data	Analyzed metabolite 357 participants
Metode / Algoritma	K-nearest neighbor (RF and DL), XGBoost, Nested Cross Validation (NCV).
Abstrak	Di sini menguji kinerja metabolit dalam darah untuk mengategorikan DA jika dibandingkan dengan biomarker CSF. Demensia tipe AD memanfaatkan metabolit plasma. Studi ini menunjukkan bahwa metabolit plasma memiliki potensi untuk menyamai AUC biomarker AD CSF yang telah terbentuk dengan baik dalam kohort yang relatif kecil. Studi lebih lanjut dalam kohort independen diperlukan untuk memvalidasi apakah panel metabolit darah tertentu ini dapat memisahkan DA dari

	kontrol, dan seberapa spesifiknya untuk DA dibandingkan dengan gangguan neurodegeneratif lainnya.
Hasil	<p>Dalam studi ini, menganalisis data metabolit yang berasal dari sampel darah dari 357 partisipan yang sebelumnya dilaporkan dalam Kim et al. Pada data pengujian, model DL menghasilkan nilai Karakteristik Operasi Receiver Under the Curve sebesar 0,85 dengan interval kepercayaan 95% berkisar antara 0,8038 dan 0,8895. Model XGBoost menghasilkan nilai AUC 0,88. Simulasi MC dilakukan dengan XGBoost, yang merupakan model prediktif superior dalam analisis, menghasilkan distribusi Gaussian dari nilai AUC menurut dan seperti yang dikonfirmasi oleh uji Shapiro-Wilk. Uji-t menunjukkan bahwa rata-rata AUC yang sebenarnya untuk XGBoost yang diterapkan pada metabolit plasma tidak lebih rendah dari 0,87. Sebagai perbandingan, juga menyelidiki tingkat amiloid, p-tau dan t-tau, yang tambahkan juga usia dan jenis kelamin, dan prediksi mereka untuk AD klinis versus CN. Model XGBoost dibangun dengan cara yang sama seperti untuk prediktor metabolit. Bersama dengan usia dan jenis kelamin, amiloid menyebabkan AUC 0,78; p-tau menyebabkan AUC 0,83; dan t-tau menyebabkan AUC 0,87. 20 prediktor peringkat teratas dari 347 yang dipilih dengan metode yang disajikan pada bagian sebelumnya ditunjukkan pada Gambar. Analisis jalur mengungkapkan bahwa jalur Nitrogen terwakili secara berlebihan dalam panel.</p>
Kesimpulan	<p>Dalam studi ini, menggunakan dua algoritma canggih, DL dan XGBoost, dan algoritma yang lebih konvensional, RF, untuk mendapatkan model akurasi tinggi untuk memprediksi AD versus CN dengan metabolit sebagai prediktor. Studi menunjukkan bahwa model terbaik didasarkan pada XGBoost, yang merupakan bentuk penyempurnaan dari metode Gradient Boosting Machines berdasarkan pohon keputusan. Data inisiatif untuk prediksi AD, XGBoost menunjukkan hasil yang unggul AUC 0,97 ketika menyertakan parameter pencitraan sebagai prediktor dan jika dibandingkan dengan RF, Support Vector</p>

	<p>Machines, Proses Gaussian, dan Stochastic Gradient Boosting. Dalam studi lain di mana kognisi dan MRI digunakan sebagai prediktor, Regresi Kernel Ridge dilakukan ke R² 5 0.87 ketika prediktor kognisi dan MRI dimasukkan. Analisis jalur menggunakan 20 AD teratas yang memprediksi metabolit yang berasal dari metode Relief menunjukkan bahwa jalur nitrogen terlalu terwakili. Metabolit baru yang mungkin menarik dan belum pernah dilaporkan sebelumnya terkait dengan DA adalah fitanat dan furoylglycine. Tujuan dari makalah ini adalah untuk membandingkan kinerja algoritma ML yang berbeda untuk mengidentifikasi orang dengan DA dari individu yang tidak mengalami gangguan kognitif. Di sini menunjukkan pertama bahwa ketiga pendekatan yang digunakan menunjukkan kekuatan diskriminatif yang baik, kedua bahwa XGBoost agak lebih efektif dalam kumpulan data khusus ini daripada RF dan DL dan ketiga, bahwa akurasi untuk diagnosis klinis ini secara luas mirip dengan yang dicapai oleh penanda CSF patologi AD. Kurangnya replikasi dan validasi dataset membatasi interpretasi dari temuan ini, namun demikian, prediksi kuat kategori diagnostik dari kumpulan biomarker metabolit berbasis darah merupakan bukti lebih lanjut dari potensi pendekatan tersebut untuk melengkapi biomarker lain dalam mengidentifikasi orang dengan kemungkinan AD.</p>
<p>Penulis</p>	<p>Daniel Stamatea,b,c, Min Kimd, Petroula Proitsie, Sarah Westwoodf, Alison Bairdf, Alejo NevadoHolgadof, Abdul Hyee, Isabelle Bosg,h, Stephanie J. B. Vosg, Rik Vandenbergheh, Charlotte E. Teunissen, Mara Ten Kateh,i, Philip Scheltensh, Silvy Gabelj,k, Karen Meersmansk,Olivier Blinm, Jill Richardsson, Ellen De Roecko,p,q, Sebastiaan Engelborghsp,q,r, Kristel Slegersq,s, Regis Bordett, Lorena Ramitu, Petronella Kettunenv, Magda Tsolakiw, Frans Verheyg, Daniel Alcoleax, Alberto Leox, Gwendoline Peyratouty, Mikel Taintaz, Peter Johannsenaa, Yvonne Freund-Levie,bb, Lutz Fr€olichcc, Valerija Dobricicdd, Giovanni B. Frisoniee,ff, Jose L. Molinuevot,gg, Anders Wallinhh, Julius</p>

	Poppy,ii, Pablo MartinezLagez, Lars Bertramdd,jj, Kaj Blennowk, Henrik Zetterbergk,mm,nn, Johannes Strefferoo, Pieter J. Visserg,h, Simon Lovestonef,pp, Cristina Legido-Quigley.
Nama jurnal, Volume, Nomor, Tahun	ELSEVIER, 5, 933-938, 2019.
Ulasan artikel	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk menguji kinerja metabolit dalam darah untuk mengkategorikan DA jika dibandingkan dengan biomarker CSF. Demensia tipe AD memanfaatkan metabolit plasma. Studi ini menunjukkan bahwa metabolit plasma memiliki potensi untuk menyamai AUC biomarker AD CSF yang telah terbentuk dengan baik dalam kohort yang relatif kecil. Studi lebih lanjut dalam kohort independen diperlukan untuk memvalidasi apakah panel metabolit darah tertentu ini dapat memisahkan DA dari kontrol, dan seberapa spesifiknya untuk DA dibandingkan dengan gangguan neurodegeneratif lainnya. Dalam studi ini, menggunakan dua algoritma canggih, DL dan XGBoost, dan algoritma yang lebih konvensional, RF, untuk mendapatkan model akurasi tinggi untuk memprediksi AD versus CN dengan metabolit sebagai prediktor. Studi menunjukkan bahwa model terbaik didasarkan pada XGBoost, yang merupakan bentuk penyempurnaan dari metode Gradient Boosting Machines berdasarkan pohon keputusan. Data inisiatif untuk prediksi AD, XGBoost menunjukkan hasil yang unggul AUC 0,97 ketika menyertakan parameter pencitraan sebagai prediktor dan jika dibandingkan dengan RF, Support Vector Machines, Proses Gaussian, dan Stochastic Gradient Boosting. Dalam studi lain di mana kognisi dan MRI digunakan sebagai prediktor, Regresi Kernel Ridge dilakukan ke R² 0,87 ketika prediktor kognisi dan MRI dimasukkan. Analisis jalur menggunakan 20 AD teratas yang memprediksi metabolit yang berasal dari metode Relief menunjukkan bahwa jalur nitrogen terlalu terwakili. Metabolit baru yang mungkin menarik dan belum pernah dilaporkan sebelumnya terkait dengan DA adalah fitanat dan</p>

furoylglycine. Tujuan dari makalah ini adalah untuk membandingkan kinerja algoritma ML yang berbeda untuk mengidentifikasi orang dengan DA dari individu yang tidak mengalami gangguan kognitif. Di sini menunjukkan pertama bahwa ketiga pendekatan yang digunakan menunjukkan kekuatan diskriminatif yang baik, kedua bahwa XGBoost agak lebih efektif dalam kumpulan data khusus ini daripada RF dan DL dan ketiga, bahwa akurasi untuk diagnosis klinis ini secara luas mirip dengan yang dicapai oleh penanda CSF patologi AD. Kurangnya replikasi dan validasi dataset membatasi interpretasi dari temuan ini, namun demikian, prediksi kuat kategori diagnostik dari kumpulan biomarker metabolit berbasis darah merupakan bukti lebih lanjut dari potensi pendekatan tersebut untuk melengkapi biomarker lain dalam mengidentifikasi orang dengan kemungkinan AD. Dalam studi ini, menganalisis data metabolit yang berasal dari sampel darah dari 357 partisipan yang sebelumnya dilaporkan dalam Kim et al. Pada data pengujian, model DL menghasilkan nilai Karakteristik Operasi Receiver Under the Curve sebesar 0,85 dengan interval kepercayaan 95% berkisar antara 0,8038 dan 0,8895. Model XGBoost menghasilkan nilai AUC 0,88. Simulasi MC dilakukan dengan XGBoost, yang merupakan model prediktif superior dalam analisis, menghasilkan distribusi Gaussian dari nilai AUC menurut dan seperti yang dikonfirmasi oleh uji Shapiro-Wilk. Uji-t menunjukkan bahwa rata-rata AUC yang sebenarnya untuk XGBoost yang diterapkan pada metabolit plasma tidak lebih rendah dari 0,87. Sebagai perbandingan, juga menyelidiki tingkat amiloid, p-tau dan t-tau, yang tambahkan juga usia dan jenis kelamin, dan prediksi mereka untuk AD klinis versus CN. Model XGBoost dibangun dengan cara yang sama seperti untuk prediktor metabolit. Bersama dengan usia dan jenis kelamin, amiloid menyebabkan AUC 0,78; p-tau menyebabkan AUC 0,83; dan t-tau menyebabkan AUC 0,87. 20 prediktor peringkat teratas dari 347 yang dipilih dengan metode yang disajikan pada bagian sebelumnya ditunjukkan pada

	Gambar. Analisis jalur mengungkapkan bahwa jalur Nitrogen terwakili secara berlebihan dalam panel.
--	--

No	17
Judul Artikel	Prediction of future gastric cancer risk using a machine learning algorithm and comprehensive medical check-up data: A casecontrol study.[17]
Topik	Data Mining
Data	25.942 peserta endoskopi dari tahun 2006 hingga 2017
Metode / Algoritma	XGBoost, classification and regression trees (CARTs).
Abstrak	Metode skrining komprehensif menggunakan pembelajaran mesin dan banyak faktor, yang dikumpulkan setiap hari sebagai data di rumah sakit, dapat meningkatkan keakuratan skrining untuk mengklasifikasikan pasien berisiko tinggi atau rendah terkena kanker lambung. menggunakan XGBoost, metode klasifikasi yang dikenal untuk mencapai berbagai solusi pemenang dalam kompetisi analisis data, untuk menangkap hubungan nonlinier di antara banyak variabel input dan hasil menggunakan pendekatan peningkatan pembelajaran mesin. Data pemeriksaan kesehatan longitudinal dan komprehensif dikumpulkan dari 25.942 peserta yang menjalani banyak endoskopi dari tahun 2006 hingga 2017 di satu fasilitas di Jepang.
Hasil	Mempertimbangkan masalah klasifikasi mengenai apakah subjek akan memiliki risiko kanker lambung di masa depan dengan memprediksi apakah ia akan didiagnosis menderita kanker lambung dalam 122 bulan ke depan. mengklasifikasikan peserta ke dalam kelompok kasus atau kelompok kontrol jika wasor kanker lambung tidak terdeteksi, masing-masing, selama periode 122 bulan. Dari peserta, 1.144 dipilih secara acak untuk digunakan dalam model klasifikasi pelatihan, dan 287 sisanya digunakan untuk mengevaluasi akurasi prediksi model yang dibangun. Kinerja klasifikasi diukur dengan kurva karakteristik operasi penerima dan area mereka di bawah nilai kurva. membangun 10 model klasifikasi untuk menjawab dua

pertanyaan penelitian berikut. Tabel 1 menunjukkan daftar 10 model klasifikasi yang dibangun menggunakan XGBoost dan regresi logistik, sambil secara bertahap menambahkan variabel input yang terkait dengan faktor risiko kanker lambung. Meningkatkan variabel input untuk memprediksi kanker lambung di masa depan. Pertanyaan pertama adalah apakah hanya kedua faktor ini yang cukup untuk memprediksi kanker lambung di masa depan. Gambar 1 menunjukkan kurva KOP lima yang dihasilkan untuk model A – E yang diperoleh dengan menggunakan teknik XGBoost. Model D, yang menambahkan faktor latar belakang biologis ke model C, ditunjukkan dengan garis merah. melihat bahwa kinerja klasifikasi ditingkatkan dengan menambahkan variabel input. Nilai ABK umumnya meningkat dengan meningkatnya jumlah variabel masukan. Model E, yang memanfaatkan semua informasi untuk variabel input, menunjukkan nilai AUC terbaik untuk data uji yang tidak diketahui, dengan akurasi = 0,777, sensitivitas = 0,933 dan spesifitas = 0,768. Pengaturan hyperparameter yang berbeda dapat menghasilkan hasil klasifikasi yang berbeda. XGBoost memiliki beberapa hiperparameter yang harus dioptimalkan menggunakan data pelatihan. Untuk mengoptimalkan hyperparameter, memanfaatkan metode pengoptimalan Bayesian^{18,19}, yang memungkinkan pemilihan otomatis dari kombinasi hyperparameter yang disukai menggunakan regresi proses Gaussian. Gambar Tambahan S1 hingga S5 menunjukkan perjalanan waktu nilai AUC pada CV dan data uji sebagai fungsi pembaruan hyperparameter XGBoost oleh BO. Terakhir, Tabel Tambahan S2 menyajikan skor penting untuk setiap variabel input yang disediakan oleh XGBoost, yang berkontribusi pada prediksi kanker lambung di masa depan. Dalam model E, HbA1c, rata-rata volume korpuskuler, rasio limfosit, usia, BMI dan pasca gastrektomi ditemukan sebagai variabel yang lebih penting. Faktor risiko yang dihitung secara otomatis untuk kanker lambung adalah wajar, seperti yang dijelaskan di bagian Diskusi di bawah. Pembelajaran mesin

	<p>nonlinier versus regresi logistik linier. Pertanyaan kedua adalah apakah metode pembelajaran mesin nonlinear yang maju dan sukses bisa efektif untuk memprediksi kanker lambung di masa depan dibandingkan dengan metode linier tradisional. bertujuan untuk menghitung keuntungan yang bisa diperoleh dengan menggunakan XGBoost dibandingkan dengan regresi logistik linier. Pada Tabel 2, model F – J menunjukkan hasil regresi logistik linier, yang menggunakan variabel input yang sama dengan XGBoost. Menemukan bahwa XGBoost mengungguli regresi logistik ketika variabel input meningkat. Sedangkan model A, B, C, F, G dan H hanya memasukkan variabel biner, model D, E, I dan J memasukkan banyak variabel kontinu. XGBoost mampu menangkap hubungan nonlinear antara hasil dan banyak variabel input dengan belajar, dan menunjukkan kinerja yang unggul dibandingkan dengan regresi logistik tradisional. Gambar Tambahan S1 hingga S5 menunjukkan perbandingan antara XGBoost dan regresi logistik dalam hal waktu perjalanan nilai AUC di bawah pembaruan hyperparameter oleh BO. Meskipun XGBoost dan regresi logistik menunjukkan hasil yang hampir sama pada beberapa variabel masukan, namun menunjukkan perbedaan yang signifikan ketika meningkatkan jumlah variabel masukan. Meningkatkan jumlah faktor masukan yang terkait dengan kanker lambung dan penggunaan metode pembelajaran mesin nonlinier lanjutan tampaknya efektif untuk memprediksi risiko kanker lambung di masa mendatang secara akurat.</p>
Kesimpulan	<p>Dalam makalah ini 21 menyarankan bahwa skrining kanker lambung di U. harus dikelompokkan berdasarkan wilayah, usia, riwayat keluarga kanker lambung, infeksi H. pylori dan kondisi lambung. Penilaian komprehensif dari beberapa faktor risiko dapat berkontribusi pada peningkatan akurasi skrining kanker lambung, dan laporan sebelumnya menunjukkan faktor lain yang mungkin terkait dengan risiko pengembangan kanker lambung, seperti diabetes dan obesitas^{22,23}. Anemia pernisiiosa juga telah diduga terkait dengan adanya kanker lambung²⁴.</p>

Selain itu, jumlah leukosit diferensial, termasuk jumlah limfosit, mungkin berguna dalam memprediksi prognosis atau adanya kanker lambung^{25,26}. Risiko kanker lambung dapat meningkat setelah operasi lambung²⁷. Hasil dapat memberikan informasi untuk mengklasifikasikan «pasien berisiko tinggi» yang harus direkomendasikan untuk skrining endoskopi yang sering untuk kanker lambung, dan «pasien berisiko rendah» yang tidak boleh. Beberapa penelitian telah dilakukan pada interval optimal untuk skrining kanker lambung endoskopi, dan saat ini tidak ada pedoman. Skrining kanker lambung direkomendasikan setiap 1–2 tahun untuk pasien berisiko tinggi di banyak negara²¹.²¹ merekomendasikan skrining kanker lambung endoskopi setiap 3–5 tahun untuk pasien berisiko rendah di U. Dengan mengacu pada laporan sebelumnya ini, jika tes skrining yang akurat diperoleh setelah memperluas dan meningkatkan penelitian ini, dapat merekomendasikan skrining kanker lambung setiap 1–2 tahun untuk pasien risiko tinggi dan setiap 3–5 tahun untuk pasien risiko rendah. menetapkan pasien yang tidak ada kanker lambung yang dapat dideteksi selama 122 bulan atau lebih sebagai kelompok kontrol. menetapkan 122 bulan sebagai batas waktu karena ini adalah periode terlama di mana kanker lambung dapat dideteksi dalam kelompok kasus. Secara umum diperkirakan bahwa permulaan kanker dimulai sekitar 20 tahun sebelum deteksi. Tus, kelompok kontrol mungkin termasuk pasien yang berpotensi mengembangkan kanker lambung di masa depan. Kelompok kontrol dalam penelitian tidak terdiri dari «pasien yang tidak akan menderita kanker lambung», melainkan «pasien di mana kanker lambung tidak akan terdeteksi dalam beberapa tahun». Selain itu, berbagai faktor risiko lingkungan dan faktor yang berhubungan dengan inang telah diduga terkait dengan kanker lambung. Karena penelitian menggunakan data pemeriksaan awal, pasien dengan temuan lambung awal yang tidak diidentifikasi mungkin telah dimasukkan dalam kelompok kasus. Prediksi dihitung berdasarkan data dari 1.431 pasien. Diperlukan waktu kurang

	lebih 10 tahun untuk mendapatkan satu sampel untuk prediksi jangka panjang perkembangan kanker lambung.
Penulis	Junichi Taninaga, Yu Nishiyama, Kazutoshi Fujibayashi, Toshiaki Gunji, Noriko Sasabe, Kimiko Iijima & Toshio Naito.
Nama jurnal, Volume, Nomor, Tahun	SCIENTIFIC REPORTS, 9, 12384, 2019.
Ulasan artikel	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk pembelajaran mesin dan banyak faktor, yang dikumpulkan setiap hari sebagai data di rumah sakit, dapat meningkatkan keakuratan skrining untuk mengklasifikasikan pasien berisiko tinggi atau rendah terkena kanker lambung. menggunakan XGBoost, metode klasifikasi yang dikenal untuk mencapai berbagai solusi pemenang dalam kompetisi analisis data, untuk menangkap hubungan nonlinier di antara banyak variabel input dan hasil menggunakan pendekatan peningkatan pembelajaran mesin. Data pemeriksaan kesehatan longitudinal dan komprehensif dikumpulkan dari 25.942 peserta yang menjalani banyak endoskopi dari tahun 2006 hingga 2017 di satu fasilitas di Jepang. Dalam makalah ini 21 menyarankan bahwa skrining kanker lambung di U. harus dikelompokkan berdasarkan wilayah, usia, riwayat keluarga kanker lambung, infeksi H. pylori dan kondisi lambung. Penilaian komprehensif dari beberapa faktor risiko dapat berkontribusi pada peningkatan akurasi skrining kanker lambung, dan laporan sebelumnya menunjukkan faktor lain yang mungkin terkait dengan risiko pengembangan kanker lambung, seperti diabetes dan obesitas^{22,23}. Anemia pernisiiosa juga telah diduga terkait dengan adanya kanker lambung²⁴. Selain itu, jumlah leukosit diferensial, termasuk jumlah limfosit, mungkin berguna dalam memprediksi prognosis atau adanya kanker lambung^{25,26}. Risiko kanker lambung dapat meningkat setelah operasi lambung²⁷. Hasil dapat memberikan informasi untuk mengklasifikasikan «pasien berisiko tinggi» yang harus direkomendasikan untuk skrining endoskopi yang sering untuk</p>

kanker lambung, dan «pasien berisiko rendah» yang tidak boleh. Beberapa penelitian telah dilakukan pada interval optimal untuk skrining kanker lambung endoskopi, dan saat ini tidak ada pedoman. Skrining kanker lambung direkomendasikan setiap 1–2 tahun untuk pasien berisiko tinggi di banyak negara²¹.²¹ merekomendasikan skrining kanker lambung endoskopi setiap 3–5 tahun untuk pasien berisiko rendah di U. Dengan mengacu pada laporan sebelumnya ini, jika tes skrining yang akurat diperoleh setelah memperluas dan meningkatkan penelitian ini, dapat merekomendasikan skrining kanker lambung setiap 1–2 tahun untuk pasien risiko tinggi dan setiap 3–5 tahun untuk pasien risiko rendah. menetapkan pasien yang tidak ada kanker lambung yang dapat dideteksi selama 122 bulan atau lebih sebagai kelompok kontrol. menetapkan 122 bulan sebagai batas waktu karena ini adalah periode terlama di mana kanker lambung dapat dideteksi dalam kelompok kasus. Secara umum diperkirakan bahwa permulaan kanker dimulai sekitar 20 tahun sebelum deteksi. Tus, kelompok kontrol mungkin termasuk pasien yang berpotensi mengembangkan kanker lambung di masa depan. Kelompok kontrol dalam penelitian tidak terdiri dari «pasien yang tidak akan menderita kanker lambung», melainkan «pasien di mana kanker lambung tidak akan terdeteksi dalam beberapa tahun». Selain itu, berbagai faktor risiko lingkungan dan faktor yang berhubungan dengan inang telah diduga terkait dengan kanker lambung. Karena penelitian menggunakan data pemeriksaan awal, pasien dengan temuan lambung awal yang tidak diidentifikasi mungkin telah dimasukkan dalam kelompok kasus. Prediksi dihitung berdasarkan data dari 1.431 pasien. Diperlukan waktu kurang lebih 10 tahun untuk mendapatkan satu sampel untuk prediksi jangka panjang perkembangan kanker lambung. Hasil dari artikel ini Mempertimbangkan masalah klasifikasi mengenai apakah subjek akan memiliki risiko kanker lambung di masa depan dengan memprediksi apakah ia akan didiagnosis menderita kanker lambung dalam 122 bulan ke depan.

mengklasifikasikan peserta ke dalam kelompok kasus atau kelompok kontrol jika wasor kanker lambung tidak terdeteksi, masing-masing, selama periode 122 bulan. Dari peserta, 1.144 dipilih secara acak untuk digunakan dalam model klasifikasi pelatihan, dan 287 sisanya digunakan untuk mengevaluasi akurasi prediksi model yang dibangun. Kinerja klasifikasi diukur dengan kurva karakteristik operasi penerima dan area mereka di bawah nilai kurva. membangun 10 model klasifikasi untuk menjawab dua pertanyaan penelitian berikut. Tabel 1 menunjukkan daftar 10 model klasifikasi yang dibangun menggunakan XGBoost dan regresi logistik, sambil secara bertahap menambahkan variabel input yang terkait dengan faktor risiko kanker lambung. Meningkatkan variabel input untuk memprediksi kanker lambung di masa depan. Pertanyaan pertama adalah apakah hanya kedua faktor ini yang cukup untuk memprediksi kanker lambung di masa depan. Gambar 1 menunjukkan kurva KOP lima yang dihasilkan untuk model A – E yang diperoleh dengan menggunakan teknik XGBoost. Model D, yang menambahkan faktor latar belakang biologis ke model C, ditunjukkan dengan garis merah. melihat bahwa kinerja klasifikasi ditingkatkan dengan menambahkan variabel input. Nilai ABK umumnya meningkat dengan meningkatnya jumlah variabel masukan. Model E, yang memanfaatkan semua informasi untuk variabel input, menunjukkan nilai AUC terbaik untuk data uji yang tidak diketahui, dengan akurasi = 0,777, sensitivitas = 0,933 dan spesifitas = 0,768. Pengaturan hyperparameter yang berbeda dapat menghasilkan hasil klasifikasi yang berbeda. XGBoost memiliki beberapa hiperparameter yang harus dioptimalkan menggunakan data pelatihan. Untuk mengoptimalkan hyperparameter, memanfaatkan metode pengoptimalan Bayesian^{18,19}, yang memungkinkan pemilihan otomatis dari kombinasi hyperparameter yang disukai menggunakan regresi proses Gaussian. Gambar Tambahan S1 hingga S5 menunjukkan perjalanan waktu nilai AUC pada CV dan data uji sebagai fungsi

pembaruan hyperparameter XGBoost oleh BO. Terakhir, Tabel Tambahan S2 menyajikan skor penting untuk setiap variabel input yang disediakan oleh XGBoost, yang berkontribusi pada prediksi kanker lambung di masa depan. Dalam model E, HbA1c, rata-rata volume korpuskuler, rasio limfosit, usia, BMI dan pasca gastrektomi ditemukan sebagai variabel yang lebih penting. Faktor risiko yang dihitung secara otomatis untuk kanker lambung adalah wajar, seperti yang dijelaskan di bagian Diskusi di bawah. Pembelajaran mesin nonlinier versus regresi logistik linier. Pertanyaan kedua adalah apakah metode pembelajaran mesin nonlinear yang maju dan sukses bisa efektif untuk memprediksi kanker lambung di masa depan dibandingkan dengan metode linier tradisional. bertujuan untuk menghitung keuntungan yang bisa diperoleh dengan menggunakan XGBoost dibandingkan dengan regresi logistik linier. Pada Tabel 2, model F – J menunjukkan hasil regresi logistik linier, yang menggunakan variabel input yang sama dengan XGBoost. Menemukan bahwa XGBoost mengungguli regresi logistik ketika variabel input meningkat. Sedangkan model A, B, C, F, G dan H hanya memasukkan variabel biner, model D, E, I dan J memasukkan banyak variabel kontinu. XGBoost mampu menangkap hubungan nonlinear antara hasil dan banyak variabel input dengan belajar, dan menunjukkan kinerja yang unggul dibandingkan dengan regresi logistik tradisional. Gambar Tambahan S1 hingga S5 menunjukkan perbandingan antara XGBoost dan regresi logistik dalam hal waktu perjalanan nilai AUC di bawah pembaruan hyperparameter oleh BO. Meskipun XGBoost dan regresi logistik menunjukkan hasil yang hampir sama pada beberapa variabel masukan, namun menunjukkan perbedaan yang signifikan ketika meningkatkan jumlah variabel masukan. Meningkatkan jumlah faktor masukan yang terkait dengan kanker lambung dan penggunaan metode pembelajaran mesin nonlinier lanjutan tampaknya efektif untuk memprediksi risiko kanker lambung di masa mendatang secara akurat.

No	18
Judul Artikel	Predicting urinary tract infections in the emergency department with machine learning.[18]
Topik	Data Mining
Data	Electronic Health Record (EHR) dataset.
Metode / Algoritma	Random Forest, Extreme Gradient Boosting, Adaptive Boosting, Support Vector machine, Elastic Net, Neural Network dan Logistic Regression.
Abstrak	<p>Infeksi saluran kemih adalah diagnosis gawat darurat umum dengan tingkat kesalahan diagnostik yang dilaporkan tinggi. Karena kultur urin, bagian dari standar emas untuk diagnosis ISK, biasanya tidak tersedia selama 24-48 jam setelah kunjungan UGD, keputusan diagnosis dan pengobatan didasarkan pada gejala, temuan fisik, dan hasil laboratorium lainnya, yang berpotensi menyebabkan penggunaan berlebihan, resistensi antibiotik, dan pengobatan tertunda. Penelitian sebelumnya telah menunjukkan kinerja diagnostik yang tidak memadai untuk tes laboratorium individu dan alat prediksi. Tujuan, adalah untuk melatih, memvalidasi, dan membandingkan model prediktif berbasis pembelajaran mesin untuk ISK pada pasien DE yang sangat beragam. Analisis kohort retrospektif satu pusat, multi-tempat dari 80.387 kunjungan DE dewasa dengan hasil kultur urin dan gejala ISK. mengembangkan model untuk prediksi ISK dengan enam algoritme pembelajaran mesin menggunakan informasi demografis, tanda vital, hasil laboratorium, obat-obatan, riwayat kesehatan masa lalu, keluhan utama, dan temuan pemeriksaan fisik dan riwayat terstruktur. Prediksi ISK dibandingkan antara model dan proksi penilaian penyedia.</p>
Hasil	<p>Sebanyak 80.387 kunjungan UGD memiliki hasil kultur urin, gejala yang berpotensi disebabkan ISK, dan akhirnya dimasukkan dalam analisis akhir. Ada 18.284 kultur urin positif, 14.335 pada wanita, dan 3.755 pada pria. Sementara menetapkan sensitivitas model berkinerja terbaik ke nilai yang sama dengan kombinasi antibiotik ATAU dokumentasi</p>

	<p>diagnosis ISK, model lengkap dan tereduksi dengan kinerja terbaik menunjukkan spesifisitas yang jauh lebih unggul dengan perbedaan 33,3 dan 29,6, masing-masing. Dibingkai dalam perspektif yang lebih klinis, dalam menerapkan model untuk keseluruhan validasi kohort yang diterima / dipulangkan, sekitar 1 dari 4 pasien akan dikategorikan ulang dari positif palsu menjadi negatif benar jika dibandingkan dengan penilaian penyedia yang ditentukan oleh diagnosis ISK dan resep antibiotik . Membandingkan hanya diagnosis ISK dengan model berkinerja terbaik yang ditetapkan pada spesifisitas yang sama, model lengkap dan tereduksi dengan kinerja terbaik juga menunjukkan sensitivitas yang jauh lebih unggul dengan perbedaan 38,7 dan 33,2. Dalam kohort validasi keseluruhan yang diterima / dipulangkan sekitar 1 dari 11 pasien akan dikategorikan ulang dari negatif palsu menjadi positif benar jika dibandingkan dengan penilaian penyedia yang ditentukan oleh diagnosis ISK saja. Di antara kunjungan yang menerima antibiotik, ada 156 kunjungan dengan diagnosis infeksi alternatif yang jelas pada mereka dengan kultur urin positif dan 529 pada mereka dengan kultur urin negatif. Di antara kunjungan pulang yang menerima antibiotik, ada 52 kunjungan dengan diagnosis infeksi alternatif yang jelas dan 200 kunjungan dengan kultur urin negatif.</p>
Kesimpulan	<p>Dalam penelitian ini mengembangkan dan memvalidasi model untuk prediksi infeksi saluran kemih pada kunjungan gawat darurat pada dataset EHR yang besar, algoritma pembelajaran mesin yang berkinerja terbaik, XGBoost, hasil kultur urin positif yang terdiagnosis secara akurat dan berkinerja lebih baik. model yang dikembangkan sebelumnya dalam literatur dan beberapa proxy untuk penilaian penyedia. Studi implementasi masa depan harus secara prospektif memeriksa dampak model pada hasil dan kesalahan diagnostik.</p>
Penulis	<p>R. Andrew Taylor, Christopher L. Moore, Kei-Hoi Cheung, Cynthia Brandt.</p>
Nama jurnal, Volume, Nomor, Tahun	<p>PLOS ONE, 10, 1371, 2018.</p>

Ulasan artikel

Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk diagnosis gawat darurat umum dengan tingkat kesalahan diagnostik yang dilaporkan tinggi. Karena kultur urin, bagian dari standar emas untuk diagnosis ISK, biasanya tidak tersedia selama 24-48 jam setelah kunjungan UGD, keputusan diagnosis dan pengobatan didasarkan pada gejala, temuan fisik, dan hasil laboratorium lainnya, yang berpotensi menyebabkan penggunaan berlebihan, resistensi antibiotik, dan pengobatan tertunda. Penelitian sebelumnya telah menunjukkan kinerja diagnostik yang tidak memadai untuk tes laboratorium individu dan alat prediksi. Tujuan, adalah untuk melatih, memvalidasi, dan membandingkan model prediktif berbasis pembelajaran mesin untuk ISK pada pasien DE yang sangat beragam. Analisis kohort retrospektif satu pusat, multi-tempat dari 80.387 kunjungan DE dewasa dengan hasil kultur urin dan gejala ISK. mengembangkan model untuk prediksi ISK dengan enam algoritme pembelajaran mesin menggunakan informasi demografis, tanda vital, hasil laboratorium, obat-obatan, riwayat kesehatan masa lalu, keluhan utama, dan temuan pemeriksaan fisik dan riwayat terstruktur. Prediksi ISK dibandingkan antara model dan proksi penilaian penyedia. Dalam penelitian ini mengembangkan dan memvalidasi model untuk prediksi infeksi saluran kemih pada kunjungan gawat darurat pada dataset EHR yang besar, algoritma pembelajaran mesin yang berkinerja terbaik, XGBoost, hasil kultur urin positif yang terdiagnosis secara akurat dan berkinerja lebih baik. model yang dikembangkan sebelumnya dalam literatur dan beberapa proxy untuk penilaian penyedia. Studi implementasi masa depan harus secara prospektif memeriksa dampak model pada hasil dan kesalahan diagnostik. Hasil dari artikel ini Sebanyak 80.387 kunjungan UGD memiliki hasil kultur urin, gejala yang berpotensi disebabkan ISK, dan akhirnya dimasukkan dalam analisis akhir. Ada 18.284 kultur urin positif, 14.335 pada wanita, dan 3.755 pada pria. Sementara menetapkan sensitivitas

	<p>model berkinerja terbaik ke nilai yang sama dengan kombinasi antibiotik ATAU dokumentasi diagnosis ISK, model lengkap dan tereduksi dengan kinerja terbaik menunjukkan spesifisitas yang jauh lebih unggul dengan perbedaan 33,3 dan 29,6, masing-masing. Dibingkai dalam perspektif yang lebih klinis, dalam menerapkan model untuk keseluruhan validasi kohort yang diterima / dipulangkan, sekitar 1 dari 4 pasien akan dikategorikan ulang dari positif palsu menjadi negatif benar jika dibandingkan dengan penilaian penyedia yang ditentukan oleh diagnosis ISK dan resep antibiotik . Membandingkan hanya diagnosis ISK dengan model berkinerja terbaik yang ditetapkan pada spesifisitas yang sama, model lengkap dan tereduksi dengan kinerja terbaik juga menunjukkan sensitivitas yang jauh lebih unggul dengan perbedaan 38,7 dan 33,2. Dalam kohort validasi keseluruhan yang diterima / dipulangkan sekitar 1 dari 11 pasien akan dikategorikan ulang dari negatif palsu menjadi positif benar jika dibandingkan dengan penilaian penyedia yang ditentukan oleh diagnosis ISK saja. Di antara kunjungan yang menerima antibiotik, ada 156 kunjungan dengan diagnosis infeksi alternatif yang jelas pada mereka dengan kultur urin positif dan 529 pada mereka dengan kultur urin negatif. Di antara kunjungan pulang yang menerima antibiotik, ada 52 kunjungan dengan diagnosis infeksi alternatif yang jelas dan 200 kunjungan dengan kultur urin negatif.</p>
--	--

No	19
Judul Artikel	Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: machine-learning algorithms and validation using national health data from Kuwait—a cohort study.[19]
Topik	Data Mining
Data	Kuwait Health Network
Metode / Algoritma	Cross Validation (CV), Logistic regression (LR), k-Nearest neighbours, Support Vector Machine (SVM) dan Multifactor Dimensionality Reduction (MDR).

<p>Abstrak</p>	<p>Membangun model klasifikasi dan alat penilaian risiko untuk diabetes, hipertensi, dan komorbiditas menggunakan algoritme machine learning pada data dari Kuwait. Insiden diabetes tipe 2, hipertensi dan komorbiditas. Alat penilaian risiko berdasarkan model klasifikasi k-NN mampu menetapkan risiko 'tinggi' untuk 75% pasien diabetes dan 94% pasien hipertensi. Kondisi patologis diabetes pada populasi umum atau pada populasi hipertensi dan hipertensi dimodelkan. Model klasifikasi agregat dua tahap dan alat penilaian risiko, dibangun dengan menggabungkan kedua model komponen pada diabetes, berkinerja lebih baik daripada model individual. Data tentang diabetes, hipertensi dan komorbiditas dari negara kosmopolitan Kuwait tersedia untuk pertama kalinya. Alat-alat ini membantu dalam penilaian non-intrusif awal dari populasi. Penilaian risiko perlu dikembangkan menggunakan data regional karena menunjukkan penerapan kalkulator online Asosiasi Diabetes Amerika pada data dari Kuwait.</p>
<p>Hasil</p>	<p>Penerapan teknik pembelajaran mesin untuk membedakan penderita diabetes tipe 2 dari populasi non-diabetes dan pasien hipertensi dari pasien non-hipertensi diperiksa. Model-model tersebut dilatih dengan data tentang parameter dasar non-intrusif dari Jaringan Kesehatan Kuwait nasional tentang diabetes dan hipertensi. Nilai akurasi > 85% untuk mengklasifikasikan penderita diabetes dari non-diabetes dengan benar, dan > 90% untuk mengklasifikasikan hipertensi dari populasi non-hipertensi dengan benar dapat dilakukan dengan model klasifikasi yang dibangun menggunakan SVM dan k-NN. Hingga 75% penderita diabetes dikelompokkan ke dalam risiko 'tinggi', dan sedikitnya 5% pasien nondiabetes dikelompokkan ke dalam kategori risiko 'tinggi'. Situasi patologi yang berbeda dimodelkan, yaitu diabetes pada populasi umum, diabetes pada populasi hipertensi, hipertensi pada populasi umum dan hipertensi pada populasi diabetes. Model klasifikasi agregat dua tahap, yang dibuat dengan menggabungkan kedua model pada diabetes atau kedua model pada hipertensi, berkinerja jauh lebih</p>

	<p>baik daripada model individual. Pemeriksaan kinerja alat penilaian risiko online ADA pada data dari Kuwait menunjukkan bahwa alat ADA bekerja hampir secara acak dalam membedakan penderita diabetes dari non-diabetes di Kuwait. Ini sesuai dengan gagasan bahwa disposisi terhadap diabetes meningkatkan kecenderungan terhadap hipertensi dan sebaliknya.</p>
Kesimpulan	<p>Dengan menggunakan parameter non-invasif dasar yang tidak berbasis laboratorium, berhasil memprediksi, dengan tingkat keakuratan yang tinggi, timbulnya diabetes dan hipertensi pada pasien di Kuwait, serupa dengan yang terlihat dalam penelitian lain. 7 Kedua, dapat memodelkan peningkatan kerentanan pada pasien diabetes untuk mengembangkan hipertensi dan sebaliknya. 26 27 28 Saat mengembangkan model klasifikasi untuk pasien di Kuwait, menghapus bidang etnis dari data menyebabkan penurunan akurasi minimal 6%. Karena yang terakhir dibuat menggunakan pasien di AS, yang secara alami memiliki demografi etnis yang berbeda dengan Kuwait, melihat perbedaan besar dalam hasil.</p>
Penulis	<p>Bassam Farran, Arshad Mohamed Channanath, Kazem Behbehani, Thangavel Alphonse Thanaraj.</p>
Nama jurnal, Volume, Nomor, Tahun	<p>BMJ Open, 10, 1136, 2013.</p>
Ulasan artikel	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk membangun model klasifikasi dan alat penilaian risiko untuk diabetes, hipertensi, dan komorbiditas menggunakan algoritme machine learning pada data dari Kuwait. Insiden diabetes tipe 2, hipertensi dan komorbiditas. Alat penilaian risiko berdasarkan model klasifikasi k-NN mampu menetapkan risiko 'tinggi' untuk 75% pasien diabetes dan 94% pasien hipertensi. Kondisi patologis diabetes pada populasi umum atau pada populasi hipertensi dan hipertensi dimodelkan. Model klasifikasi agregat dua tahap dan alat penilaian risiko, dibangun dengan menggabungkan kedua model komponen pada diabetes, berkinerja lebih baik daripada model individual. Data tentang</p>

diabetes, hipertensi dan komorbiditas dari negara kosmopolitan Kuwait tersedia untuk pertama kalinya. Alat-alat ini membantu dalam penilaian non-intrusif awal dari populasi. Penilaian risiko perlu dikembangkan menggunakan data regional karena menunjukkan penerapan kalkulator online Asosiasi Diabetes Amerika pada data dari Kuwait. Dengan menggunakan parameter non-invasif dasar yang tidak berbasis laboratorium, berhasil memprediksi, dengan tingkat keakuratan yang tinggi, timbulnya diabetes dan hipertensi pada pasien di Kuwait, serupa dengan yang terlihat dalam penelitian lain. 7 Kedua, dapat memodelkan peningkatan kerentanan pada pasien diabetes untuk mengembangkan hipertensi dan sebaliknya. 26 27 28 Saat mengembangkan model klasifikasi untuk pasien di Kuwait, menghapus bidang etnis dari data menyebabkan penurunan akurasi minimal 6%. Karena yang terakhir dibuat menggunakan pasien di AS, yang secara alami memiliki demografi etnis yang berbeda dengan Kuwait, melihat perbedaan besar dalam hasil. Penerapan teknik pembelajaran mesin untuk membedakan penderita diabetes tipe 2 dari populasi non-diabetes dan pasien hipertensi dari pasien non-hipertensi diperiksa. Model-model tersebut dilatih dengan data tentang parameter dasar non-intrusif dari Jaringan Kesehatan Kuwait nasional tentang diabetes dan hipertensi. Nilai akurasi > 85% untuk mengklasifikasikan penderita diabetes dari non-diabetes dengan benar, dan > 90% untuk mengklasifikasikan hipertensi dari populasi non-hipertensi dengan benar dapat dilakukan dengan model klasifikasi yang dibangun menggunakan SVM dan k-NN. Hingga 75% penderita diabetes dikelompokkan ke dalam risiko 'tinggi', dan sedikitnya 5% pasien nondiabetes dikelompokkan ke dalam kategori risiko 'tinggi'. Situasi patologi yang berbeda dimodelkan, yaitu diabetes pada populasi umum, diabetes pada populasi hipertensi, hipertensi pada populasi umum dan hipertensi pada populasi diabetes. Model klasifikasi agregat dua tahap, yang dibuat dengan menggabungkan kedua model pada diabetes atau kedua model pada hipertensi, berkinerja jauh lebih

	<p>baik daripada model individual. Pemeriksaan kinerja alat penilaian risiko online ADA pada data dari Kuwait menunjukkan bahwa alat ADA bekerja hampir secara acak dalam membedakan penderita diabetes dari non-diabetes di Kuwait. Ini sesuai dengan gagasan bahwa disposisi terhadap diabetes meningkatkan kecenderungan terhadap hipertensi dan sebaliknya.</p>
--	---

No	20
Judul Artikel	Prediction of Acute Kidney Injury With a Machine Learning Algorithm Using Electronic Health Record Data.[20]
Topik	Data Mining
Data	300 000 inpatient encounters
Metode / Algoritma	Sequential Organ Failure Assessment (SOFA).
Abstrak	<p>Dokter dapat melakukan intervensi selama apa yang mungkin merupakan tahap penting untuk mencegah cedera ginjal permanen jika pasien dengan AKI yang baru jadi dan mereka yang berisiko tinggi mengembangkan AKI dapat diidentifikasi. Dalam studi ini, mengevaluasi algoritma pembelajaran mesin untuk deteksi dini dan prediksi AKI. menggunakan teknik pembelajaran mesin, meningkatkan ansambel pohon keputusan, untuk melatih alat prediksi AKI pada data retrospektif yang diambil dari lebih dari 300.000 pertemuan pasien rawat inap. Data dikumpulkan dari bangsal rawat inap di Stanford Medical Center dan pasien unit perawatan intensif di Beth Israel Deaconess Medical Center. Pasien yang berusia lebih dari 18 tahun yang dirawat di rumah sakit berlangsung antara 5 dan 1000 jam dan yang memiliki setidaknya satu pengukuran detak jantung, laju pernapasan, suhu, kreatinin serum, dan Skala Koma Glasgow yang terdokumentasi. menguji kemampuan algoritme untuk mendeteksi AKI saat onset dan untuk memprediksi AKI 12, 24, 48, dan 72 jam sebelum onset. menguji deteksi dan prediksi AKI menggunakan Algoritma AKI National Health Service Inggris sebagai standar emas. Karena sifat retrospektif dari penelitian ini, tidak dapat menarik</p>

	<p>kesimpulan tentang dampak prediksi algoritma terhadap hasil pasien dalam pengaturan klinis. Hasil eksperimen ini menunjukkan bahwa alat prediksi AKI berbasis pembelajaran mesin mungkin menawarkan kemampuan prognostik penting untuk menentukan pasien mana yang kemungkinan menderita AKI, yang berpotensi memungkinkan dokter untuk campur tangan sebelum kerusakan ginjal terjadi.</p>
<p>Hasil</p>	<p>Pendekatan machine learning yang dijelaskan di sini menghasilkan alat prediksi yang menunjukkan kinerja prediksi yang kuat, dalam hal AUROC, hingga 72 jam sebelum onset AKI tahap 2 atau tahap 3, di bawah kriteria NHS dan KDIGO untuk AKI. Dengan hanya membutuhkan 5 pengukuran ini untuk prediksi AKI, algoritma ini dirancang untuk dapat memprediksi risiko AKI pada sebagian besar populasi rumah sakit dalam pekerjaan klinis di masa depan. Berdasarkan hasil ini, percaya MLA ini dapat memberikan dokter kesempatan untuk meningkatkan hasil pasien melalui deteksi AKI lebih dini dan intervensi berikutnya, yang mungkin termasuk resusitasi volume atau menghindari obat nefrotoksik untuk meminimalkan cedera ginjal lebih lanjut. 28 menekankan bahwa MLA berkinerja sama baiknya pada kumpulan data BIDMC dan Stanford. Data BIDMC hanya memasukkan pasien yang dirawat di ICU, sedangkan kumpulan data Stanford berisi informasi tentang rawat inap dari semua bangsal rumah sakit. Kedua kumpulan data ini mewakili pengaturan rumah sakit dengan demografi yang berbeda, frekuensi pengumpulan pengukuran pasien, tingkat penyediaan perawatan, dan tingkat keparahan penyakit pada pasien. Kemampuan prediksi algoritme di seluruh kumpulan data ini menunjukkan bahwa algoritme dapat mengidentifikasi pasien yang berisiko terkena AKI di berbagai pengaturan rumah sakit. Karena AKI adalah komplikasi umum dari rawat inap di rumah sakit dengan sifat yang beragam, kemampuan ini sangat penting dalam alat prediksi AKI. Dalam studi sebelumnya, menilai MLA untuk deteksi sepsis di kedua pengaturan retrospektif²⁹⁻³¹ dan prospektif. Dalam penelitian</p>

	<p>ini, terdapat penurunan yang signifikan secara statistik pada titik akhir primer rata-rata lama rawat di rumah sakit dan pada titik akhir sekunder dari angka kematian di rumah sakit. 32 Studi ini menunjukkan kelayakan penggunaan samping tempat tidur dari sistem prediksi berbasis pembelajaran mesin, serta potensi sistem tersebut untuk meningkatkan hasil pasien melalui deteksi dini dan lebih akurat dari kondisi pasien. Implementasi klinis di masa mendatang dari sistem yang dijelaskan dalam pekerjaan ini dimaksudkan untuk tidak mengganggu alur kerja klinis atau memerlukan pekerjaan tambahan dari penyedia perawatan. Metode pembelajaran mesin sebelumnya telah diterapkan pada deteksi AKI. Manfaat terukur dari fokus pada prediksi AKI dapat memungkinkan dokter untuk lebih cepat menentukan penyebab kemunduran pasien dan, dengan demikian, memberikan perawatan yang tepat secara lebih tepat waktu.</p>
Kesimpulan	<p>Pendekatan machine learning yang dijelaskan dalam penelitian ini secara akurat memprediksi AKI tahap 2 atau tahap 3 hingga 72 jam masuk kemajuan onset pada saat dilatih dan diuji pada dua yang berbeda set data. Algoritme ini dapat meningkatkan deteksi AKI di pengaturan klinis, memungkinkan untuk intervensi lebih awal dan meningkatkan hasil pasien.</p>
Penulis	<p>Hamid Mohamadlou, Anna Lynn-Palevsky, Christopher Barton, Uli Chettipally, Lisa Shieh, Jacob Calvert, Nicholas R. Saber dan Ritankar Das.</p>
Nama jurnal, Volume, Nomor, Tahun	<p>Canadian Journal of Kidney Health and Disease, 5, 1-9, 2018</p>
Ulasan artikel	<p>Penulisan Artikel ini secara keseluruhannya adalah baik dan mudah dipahami. Ulasan artikel dari Penelitian ini untuk melakukan intervensi selama apa yang mungkin merupakan tahap penting untuk mencegah cedera ginjal permanen jika pasien dengan AKI yang baru jadi dan mereka yang berisiko tinggi mengembangkan AKI dapat diidentifikasi. Dalam studi ini, mengevaluasi algoritma pembelajaran mesin untuk deteksi dini dan prediksi AKI. menggunakan teknik pembelajaran mesin, meningkatkan ansambel pohon keputusan, untuk melatih alat prediksi AKI pada data retrospektif yang diambil dari lebih</p>

dari 300.000 pertemuan pasien rawat inap. Data dikumpulkan dari bangsal rawat inap di Stanford Medical Center dan pasien unit perawatan intensif di Beth Israel Deaconess Medical Center. Pasien yang berusia lebih dari 18 tahun yang dirawat di rumah sakit berlangsung antara 5 dan 1000 jam dan yang memiliki setidaknya satu pengukuran detak jantung, laju pernapasan, suhu, kreatinin serum, dan Skala Koma Glasgow yang terdokumentasi. menguji kemampuan algoritme untuk mendeteksi AKI saat onset dan untuk memprediksi AKI 12, 24, 48, dan 72 jam sebelum onset. menguji deteksi dan prediksi AKI menggunakan Algoritma AKI National Health Service Inggris sebagai standar emas. Karena sifat retrospektif dari penelitian ini tidak dapat menarik kesimpulan tentang dampak prediksi algoritma terhadap hasil pasien dalam pengaturan klinis. Hasil eksperimen ini menunjukkan bahwa alat prediksi AKI berbasis pembelajaran mesin mungkin menawarkan kemampuan prognostik penting untuk menentukan pasien mana yang kemungkinan menderita AKI, yang berpotensi memungkinkan dokter untuk campur tangan sebelum kerusakan ginjal terjadi. Pendekatan machine learning yang dijelaskan dalam penelitian ini secara akurat memprediksi AKI tahap 2 atau tahap 3 hingga 72 jam masuk kemajuan onset pada saat dilatih dan diuji pada dua yang berbeda set data. Algoritme ini dapat meningkatkan deteksi AKI di pengaturan klinis, memungkinkan untuk intervensi lebih awal dan meningkatkan hasil pasien. Pendekatan machine learning yang dijelaskan di sini menghasilkan alat prediksi yang menunjukkan kinerja prediksi yang kuat, dalam hal AUROC, hingga 72 jam sebelum onset AKI tahap 2 atau tahap 3, di bawah kriteria NHS dan KDIGO untuk AKI. Dengan hanya membutuhkan 5 pengukuran ini untuk prediksi AKI, algoritma ini dirancang untuk dapat memprediksi risiko AKI pada sebagian besar populasi rumah sakit dalam pekerjaan klinis di masa depan. Berdasarkan hasil ini, percaya MLA ini dapat memberikan dokter kesempatan untuk meningkatkan hasil pasien melalui deteksi AKI lebih dini

	<p>dan intervensi berikutnya, yang mungkin termasuk resusitasi volume atau menghindari obat nefrotoksik untuk meminimalkan cedera ginjal lebih lanjut. 28 menekankan bahwa MLA berkinerja sama baiknya pada kumpulan data BIDMC dan Stanford. Data BIDMC hanya memasukkan pasien yang dirawat di ICU, sedangkan kumpulan data Stanford berisi informasi tentang rawat inap dari semua bangsal rumah sakit. Kedua kumpulan data ini mewakili pengaturan rumah sakit dengan demografi yang berbeda, frekuensi pengumpulan pengukuran pasien, tingkat penyediaan perawatan, dan tingkat keparahan penyakit pada pasien. Kemampuan prediksi algoritme di seluruh kumpulan data ini menunjukkan bahwa algoritme dapat mengidentifikasi pasien yang berisiko terkena AKI di berbagai pengaturan rumah sakit. Karena AKI adalah komplikasi umum dari rawat inap di rumah sakit dengan sifat yang beragam, kemampuan ini sangat penting dalam alat prediksi AKI. Dalam studi sebelumnya, menilai MLA untuk deteksi sepsis di kedua pengaturan retrospektif²⁹⁻³¹ dan prospektif. Dalam penelitian ini, terdapat penurunan yang signifikan secara statistik pada titik akhir primer rata-rata lama rawat di rumah sakit dan pada titik akhir sekunder dari angka kematian di rumah sakit. 32 Studi ini menunjukkan kelayakan penggunaan samping tempat tidur dari sistem prediksi berbasis pembelajaran mesin, serta potensi sistem tersebut untuk meningkatkan hasil pasien melalui deteksi dini dan lebih akurat dari kondisi pasien. Implementasi klinis di masa mendatang dari sistem yang dijelaskan dalam pekerjaan ini dimaksudkan untuk tidak mengganggu alur kerja klinis atau memerlukan pekerjaan tambahan dari penyedia perawatan. Metode pembelajaran mesin sebelumnya telah diterapkan pada deteksi AKI. Manfaat terukur dari fokus pada prediksi AKI dapat memungkinkan dokter untuk lebih cepat menentukan penyebab kemunduran pasien dan, dengan demikian, memberikan perawatan yang tepat secara lebih tepat waktu.</p>
--	--

Tabel ini terdiri dari minimal 20 Jurnal

IV. ULASAN

Nama Jurnal	Ulasan Penelitian
Prediction of future gastric cancer risk using a machine learning algorithm and comprehensive medical check-up data: A casecontrol study.	Risiko kanker lambung di masa depan dengan memprediksi apakah ia akan didiagnosis menderita kanker lambung dalam 122 bulan ke depan. mengklasifikasikan peserta ke dalam kelompok kasus atau kelompok kontrol jika wasor kanker lambung tidak terdeteksi, masing-masing, selama periode 122 bulan. Dari peserta, 1.144 dipilih secara acak untuk digunakan dalam model klasifikasi pelatihan, dan 287 sisanya digunakan untuk mengevaluasi akurasi prediksi model yang dibangun. Kinerja klasifikasi diukur dengan kurva karakteristik operasi penerima dan area mereka di bawah nilai kurva.
Pathway analysis using XGBoost classification in Biomedical Data.	Makalah ini Singkatnya, metode klasifikasi berbasis jalur menggunakan algoritma XGBoost untuk menganalisis data ekspresi gen dijelaskan. Metode yang diusulkan mengidentifikasi jalur penting yang membedakan dua kasus yang diteliti, sementara itu mengisolasi jalur bawah tanah yang paling penting yang memainkan peran penting dalam seluruh topologi jalur, gangguan yang memunculkan penyakit yang diteliti.
Machine learning to predict rapid progression of carotid atherosclerosis in patients with impaired glucose tolerance	Perbandingan metode dan faktor dilakukan dengan menggunakan area di bawah analisis kurva karakteristik operasi penerima dan skor Brier. Hasil eksperimen menunjukkan bahwa metode pembelajaran yang diusulkan bekerja dengan baik dalam mengidentifikasi atau memprediksi RP. Di antara metode, kinerja Naïve Bayes adalah yang terbaik dibandingkan dengan perceptron multilayer dan hutan acak. Dengan berurusan dengan data multi-modal, metode pembelajaran yang diusulkan menunjukkan efektivitas dalam memprediksi prediabetik yang berisiko untuk perkembangan aterosklerosis yang cepat.

DAFTAR PUSTAKA

- [1] A. Shaham, G. Chodick, V. Shalev, and D. Yamin, "Personal and social patterns predict influenza vaccination decision," *BMC Public Health*, vol. 20, no. 1, pp. 1–12, 2020.
- [2] A. Dinh, S. Miertschin, A. Young, and S. D. Mohanty, "A data-driven approach to predicting diabetes and cardiovascular disease with machine learning," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, pp. 1–15, 2019.
- [3] W. Chang *et al.*, "A machine-learning-based prediction method for hypertension outcomes based on medical data," *Diagnostics*, vol. 9, no. 4, 2019.
- [4] D. J. Albers, N. Elhadad, J. Claassen, R. Perotte, A. Goldstein, and G. Hripcsak, "Estimating summary statistics for electronic health record laboratory data for use in high-throughput phenotyping algorithms," *J. Biomed. Inform.*, vol. 78, no. December 2017, pp. 87–101, 2018.
- [5] G. N. Dimitrakopoulos, A. G. Vrahatis, K. Sgarbas, and V. Plagianakos, "Pathway analysis using xgboost classification in biomedical data," *ACM Int. Conf. Proceeding Ser.*, pp. 1–6, 2018.
- [6] S. Ramani, M. Sivakami, and L. Gilson, "How context affects implementation of the Primary Health Care approach: an analysis of what happened to primary health centres in India," *BMJ Glob. Heal.*, vol. 3, no. Suppl 3, p. e001381, 2019.
- [7] R. J. Ellis, Z. Wang, N. Genes, and A. Ma'Ayan, "Predicting opioid dependence from electronic health records with machine learning," *BioData Min.*, vol. 12, no. 1, pp. 1–19, 2019.
- [8] X. Hu *et al.*, "Machine learning to predict rapid progression of carotid atherosclerosis in patients with impaired glucose tolerance," *Eurasip J. Bioinforma. Syst. Biol.*, vol. 2016, no. 1, 2016.
- [9] C. A. Hu *et al.*, "Using a machine learning approach to predict mortality in critically ill influenza patients: A cross-sectional retrospective multicentre study in Taiwan," *BMJ Open*, vol. 10, no. 2, pp. 1–10, 2020.
- [10] S. Jaya and M. Latha, "Diagnosis of cervical cancer using CLAHE and SGLDM on RGB pap smear images through ANN," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 1, pp. 530–534, 2019.
- [11] D. S. Kermany *et al.*, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, 2018.
- [12] M. Moscatelli *et al.*, "An infrastructure for precision medicine through analysis of big data," *BMC Bioinformatics*, vol. 19, no. Suppl 10, 2018.
- [13] P. Kumar, A. Nestsiarovich, S. J. Nelson, B. Kerner, D. J. Perkins, and C. G. Lambert, "Imputation and characterization of uncoded self-harm in major mental illness using machine learning," *J. Am. Med. Informatics Assoc.*, vol. 27, no. 1, pp. 136–146, 2020.
- [14] L. Buitinck *et al.*, "API design for machine learning software: experiences from the scikit-learn project," no. July 2014, 2013.
- [15] J. Sandino, G. Pegg, F. Gonzalez, and G. Smith, "Aerial mapping of forests affected by pathogens using UAVs, hyperspectral sensors, and artificial intelligence," *Sensors (Switzerland)*, vol. 18, no. 4, pp. 1–17, 2018.
- [16] D. Stamate *et al.*, "A metabolite-based machine learning approach to diagnose Alzheimer-type dementia in blood: Results from the European Medical Information Framework for Alzheimer disease biomarker discovery cohort," *Alzheimer's Dement. Transl. Res. Clin. Interv.*, vol. 5, pp. 933–938, 2019.
- [17] J. Taninaga *et al.*, "Prediction of future gastric cancer risk using a machine learning

- algorithm and comprehensive medical check-up data: A case-control study,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, 2019.
- [18] R. A. Taylor, C. L. Moore, K. H. Cheung, and C. Brandt, “Predicting urinary tract infections in the emergency department with machine learning,” *PLoS One*, vol. 13, no. 3, pp. 1–15, 2018.
- [19] B. Farran, A. M. Channanath, K. Behbehani, and T. A. Thanaraj, “Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: Machine-learning algorithms and validation using national health data from Kuwait-a cohort study,” *BMJ Open*, vol. 3, no. 5, pp. 1–10, 2013.
- [20] H. Mohamadlou *et al.*, “Prediction of Acute Kidney Injury With a Machine Learning Algorithm Using Electronic Health Record Data,” *Can. J. Kidney Heal. Dis.*, vol. 5, 2018.



LAMPIRAN

SURAT PERNYATAAN

Yang bertanda tangan di bawah ini, pemegang hak cipta:

Nama : 1. Muhamad Iqbal 2. DR. Mujiono Sadikin, MT, CISA, CGEIT
Kewarganegaraan : Indonesia
Alamat : Jakarta Barat, DKI Jakarta, Indonesia

Dengan ini menyatakan bahwa:

1. Karya Cipta yang saya mohonkan:
Berupa : Jurnal Ilmiah
Berjudul : Analisa Data Rekam Medis Menggunakan Teknik Data Mining Association Rules Dengan Algoritma Clustering.
 - Tidak meniru dan tidak sama secara esensial dengan Karya Cipta milik pihak lain atau obyek kekayaan intelektual lainnya sebagaimana dimaksud dalam Pasal 68 ayat (2);
 - Bukan merupakan Ekspresi Budaya Tradisional sebagaimana dimaksud dalam Pasal 38;
 - Bukan merupakan Ciptaan yang tidak diketahui penciptanya sebagaimana dimaksud dalam Pasal 39;
 - Bukan merupakan hasil karya yang tidak dilindungi Hak Cipta sebagaimana dimaksud dalam Pasal 41 dan 42;
 - Bukan merupakan Ciptaan seni lukis yang berupa logo atau tanda pembeda yang digunakan sebagai merek dalam perdagangan barang/jasa atau digunakan sebagai lambang organisasi, badan usaha, atau badan hukum sebagaimana dimaksud dalam Pasal 65 dan;
 - Bukan merupakan Ciptaan yang melanggar norma agama, norma susila, ketertiban umum, pertahanan dan keamanan negara atau melanggar peraturan perundang-undangan sebagaimana dimaksud dalam Pasal 74 ayat (1) huruf d Undang-Undang Nomor 28 Tahun 2014 tentang Hak Cipta.
2. Sebagai pemohon mempunyai kewajiban untuk menyimpan asli contoh ciptaan yang dimohonkan dan harus memberikan apabila dibutuhkan untuk kepentingan penyelesaian sengketa perdata maupun pidana sesuai dengan ketentuan perundang-undangan.
3. Karya Cipta yang saya mohonkan pada Angka 1 tersebut di atas tidak pernah dan tidak sedang dalam sengketa pidana dan/atau perdata di Pengadilan.
4. Dalam hal ketentuan sebagaimana dimaksud dalam Angka 1 dan Angka 3 tersebut di atas saya / kami melanggar, maka saya / kami bersedia secara sukarela bahwa:
 - a. permohonan karya cipta yang saya ajukan dianggap ditarik kembali; atau
 - b. Karya Cipta yang telah terdaftar dalam Daftar Umum Ciptaan Direktorat Hak Cipta, Direktorat Jenderal Hak Kekayaan Intelektual, Kementerian Hukum Dan Hak Asasi Manusia RI dihapuskan sesuai dengan ketentuan perundang-undangan yang berlaku.
 - c. Dalam hal kepemilikan Hak Cipta yang dimohonkan secara elektronik sedang dalam berperkara dan/atau sedang dalam gugatan di Pengadilan maka status kepemilikan surat pencatatan elektronik tersebut ditangguhkan menunggu putusan Pengadilan yang berkekuatan hukum tetap.

Demikian Surat pernyataan ini saya/kami buat dengan sebenarnya dan untuk dipergunakan sebagaimana mestinya.

Jakarta, 26 Agustus 2022


METERAI TEMPEL
C 014 JX943834300
(Muhamad Iqbal)

* Semua pemegang hak cipta agar menandatangani di atas materai