



PERANCANGAN MODEL *ARTIFICIAL INTELLIGENCE* UNTUK
MENGHITUNG PREDIKSI NILAI LDL (*LOW-DENSITY LIPOPROTEIN*)
MENGGUNAKAN XGBOOST



Oleh:

Aurellio Indra Dewanta

22410200010

FAKULTAS TEKNOLOGI DAN INFORMATIKA

UNIVERSITAS DINAMIKA

2025

**Perancangan Model *Artificial Intelligence* Untuk Menghitung Prediksi Nilai
LDL (*Low-Density Lipoprotein*) Menggunakan XGBoost**

Diajukan sebagai salah satu syarat untuk menyelesaikan

Program Strata Satu (S1)

Disusun Oleh :



Nama : AURELLIO INDRA DEWANTA
Nim : 22410200010
Program : S1 (Strata Satu)
Jurusan : Teknik Komputer

UNIVERSITAS
Dinamika

FAKULTAS TEKNOLOGI DAN INFORMATIKA

UNIVERSITAS DINAMIKA

2025

LEMBAR PENGESAHAN

Perancangan Model *Artificial Intelligence* Untuk Menghitung Prediksi Nilai LDL (*Low-Density Lipoprotein*) Menggunakan XGBoost

Laporan Kerja Praktik oleh
AURELLIO INDRA DEWANTA

NIM. 22410200010

Telah diperiksa, diuji, dan disetujui

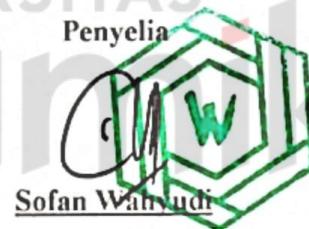


Pembimbing

Heri Pratikno, M.T.
NIDN. 0716117302

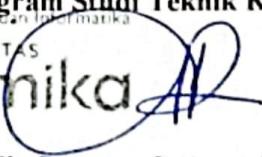
Disetujui:

Surabaya, 30 Juli 2025



Penyelia

Sofan Wahyudi

Mengetahui,
Ketua Program Studi Teknik Komputer
Fakultas Teknologi dan Informatika
UNIVERSITAS
Dinamika 
Pauladie Susanto, S.Kom., M.T.
NIDN. 0729047501

**PERNYATAAN
PERSETUJUAN PUBLIKASI DAN KEASLIAN KARYA ILMIAH**

Sebagai mahasiswa **Universitas Dinamika**, Saya :

Nama : **Aurellio Indra Dewanta**
NIM : **22410200010**
Program Studi : **S1 Teknik Komputer**
Fakultas : **Fakultas Teknologi dan Informatika**
Jenis Karya : **Laporan Kerja Praktik**
Judul Karya : **PERANCANGAN MODEL ARTIFICIAL INTELLIGENCE UNTUK MENGHITUNG PREDIKSI NILAI LDL (*LOW-DENSITY LIPOPROTEIN*) MENGGUNAKAN XGBOOST**

Menyatakan dengan sesungguhnya bahwa :

1. Demi pengembangan Ilmu Pengetahuan, Teknologi dan Seni, Saya menyetujui memberikan kepada **Universitas Dinamika** Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty Free Right*) atas seluruh isi/sebagian karya ilmiah Saya tersebut diatas untuk disimpan, dialihmediakan, dan dikelola dalam bentuk pangkalan data (*database*) untuk selanjutnya didistribusikan atau dipublikasikan demi kepentingan akademis dengan tetap mencantumkan nama Saya sebagai penulis atau pencipta dan sebagai pemilik Hak Cipta.
2. Karya tersebut diatas adalah hasil karya asli Saya, bukan plagiat baik sebagian maupun keseluruhan. Kutipan, karya, atau pendapat orang lain yang ada dalam karya ilmiah ini semata-mata hanya sebagai rujukan yang dicantumkan dalam Daftar Pustaka Saya.
3. Apabila dikemudian hari ditemukan dan terbukti terdapat tindakan plagiasi pada karya ilmiah ini, maka Saya bersedia untuk menerima pencabutan terhadap gelar kesarjanaan yang telah diberikan kepada Saya.

Demikan surat pernyataan ini Saya buat dengan sebenar-benarnya.

Surabaya, 15 Juli 2025



Aurellio Indra Dewanta
NIM : 22410200010

ABSTRAK

Kerja praktek ini dilaksanakan di salah satu anak perusahaan dari PT. Wahana Mediaek Indonesia yaitu, Adamlabs yang berlokasi di Surabaya dengan tujuan untuk merancang dan mengembangkan sistem perhitungan nilai parameter dengan memanfaatkan salah satu cabang dari *artificial intelligence* yaitu, *Machine Learning* (ML), sebagai media bantu dalam memprediksi nilai dari LDL (*Low-Density Lipoprotein*). Pengolahan data ini memanfaatkan parameter laboratorium seperti Total Kolesterol, HDL, dan Triglicerida sebagai input. Tujuan dari proyek ini adalah menciptakan sistem laboratorium berbasis AI yang dapat memprediksi nilai LDL (*Low-Density Lipoprotein*) lalu dibandingkan dengan perhitungan manual seperti rumus Friedewald, dan rumus manual lainnya. Hasil implementasi menunjukkan bahwa model XGBoost mampu memberikan hasil prediksi yang mendekati nilai aktual, dengan nilai koefisien determinasi (R²) yang cukup tinggi dengan nilai R² sebesar 0.98 dan MAE sebesar 1.0 setelah penambahan metode turunan sebagai pola awal bagi *machine learning* belajar dan metode *Grid search* sebagai hyperparameter untuk mencari kombinasi / pola yang akan digunakan oleh model. Proyek ini mendemonstrasikan bahwa teknologi *machine learning* dapat diintegrasikan secara efektif dalam sistem informasi laboratorium untuk mendukung pengambilan keputusan klinis.

Kata Kunci : LDL (*Low-Density Lipoprotein*), XGBoost, *Artificial Intelligence*, *Machine Learning*, Friedewald,..



UNIVERSITAS
Dinamika

KATA PENGANTAR

Puji syukur ke hadirat Tuhan Yang Maha Esa atas rahmat-Nya, sehingga laporan kerja praktik di Adamlabs ini dapat diselesaikan dengan baik. Laporan ini mendokumentasikan pengalaman dan pembelajaran saya selama menjalani kerja praktik sebagai Datalogger, khususnya dalam pengembangan sistem prediksi data laboratorium berbasis kecerdasan buatan (AI) di bidang teknologi informasi kesehatan. Penulis menyampaikan terima kasih sebesar-besarnya kepada:

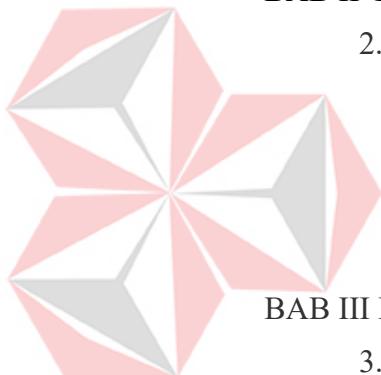
1. Tuhan Yang Maha Esa yang memberikan rahmat, hidayah, dan kesempatan yang diberikan sehingga penulis dapat menyelesaikan program kerja praktik dengan baik.
2. Bapak Heri Pratikno, M.T., selaku dosen pembimbing yang telah memberikan arahan, dukungan, serta masukan selama proses pelaksanaan kerja praktik hingga penyusunan laporan ini.
3. Bapak Wigananda Firdaus Putra Aditya, S.Kom., sebagai penyelenggara program Dinamika Industrial internship (DII) yang mana dapat pada akhirnya saya dapat melakukan konversi untuk kerja praktik ini.
4. Bapak Sofan Wahyudi, selaku mentor dalam program ini dan bagian R&D serta *Datalogger*, atas bimbingan dan kesempatan yang sudah diberikan
5. Rekan-rekan mahasiswa serta seluruh pihak yang telah memberikan semangat dan bantuan selama masa magang.

Diharapkan laporan ini dapat memberikan gambaran yang jelas mengenai pelaksanaan kerja praktik serta menjadi referensi yang bermanfaat bagi pembaca. Semoga pengalaman ini menjadi landasan yang kuat untuk pengembangan diri dan kontribusi lebih lanjut di dunia profesional.

Surabaya, 15 Juli 2025

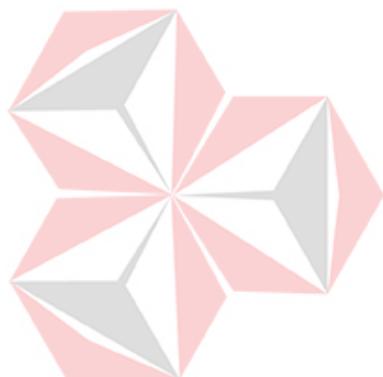
Aurellio Indra Dewanta

DAFTAR ISI



ABSTRAK.....	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	ix
BAB I PENDAHULUAN	10
1.1 Latar Belakang.....	10
1.2 Rumusan Masalah.....	10
1.3 Batasan Masalah	11
1.4 Tujuan	12
1.5 Manfaat.....	12
BAB II GAMBARAN UMUM PERUSAHAAN	13
2.1. Sejarah Singkat Perusahaan.....	13
2.1.1. Visi dan Misi.....	13
2.1.2. Tujuan	14
2.1.3. Struktur Organisasi	14
2.1.4. Informasi Kontak	14
BAB III LANDASAN TEORI	15
3.1 <i>Artificial Intelligence</i> (AI).....	15
3.1.1. Jenis-Jenis <i>Artificial Intelligence</i> (AI)	15
3.2 <i>Machine Learning</i> (ML).....	16
3.3 Algoritma XGBoost (<i>Extreme Gradient Boosting</i>).....	17
3.4 Library Pendukung Pengembangan Sistem.....	17
3.5 Rumus Konvesional Perhitungan LDL (<i>Low-Density Lipoprotein</i>)	18
3.6 Rumus Evaluasi Model Yang Digunakan Untuk Validasi	18
BAB IV Hasil Pekerjaan dan Pembahasan.....	21
4.1 Deskripsi Kerja Pratik	21
4.2 Alur Kerja Sistem	21
4.3 Uraian Perancangan Sistem	24
4.4 Perancangan <i>Software</i>	25
4.4.1 Pustaka Utama yang Digunakan	25

4.4.2 Struktur dan Modularisasi.....	26
4.5 Hasil Pengujian Dengan Program Test	27
BAB V Kesimpulan dan Saran.....	29
5.1 Kesimpulan.....	29
5.2 Saran.....	29
DAFTAR PUSTAKA.....	31



UNIVERSITAS
Dinamika

DAFTAR GAMBAR

Gambar 2. 1. Struktur Organisasi	14
Gambar 4. 1. Tampilan hasil dari proses training.....	24
Gambar 4. 2. Struktur yang digunakan.....	27
Gambar 4. 3. Hasil pengujian dengan program test	28



BAB I

PENDAHULUAN

1.1 Latar Belakang

Model *Artificial Intelligence* (AI) kini menjadi salah satu pendekatan unggulan dalam dunia teknik dan kesehatan, terutama untuk menangani masalah prediksi dan klasifikasi berdasarkan data medis. Salah satu contohnya adalah prediksi nilai LDL (*Low-Density Lipoprotein*), yang merupakan indikator penting dalam menilai risiko penyakit jantung. Namun, dalam proses pembelajaran, mahasiswa sering kali hanya memahami konsep AI secara teoritis, tanpa adanya praktik langsung yang melibatkan perancangan model, pengolahan data, serta evaluasi performa prediksi. Kondisi ini menyebabkan pemahaman terhadap *machine learning* menjadi kurang aplikatif.

Sebagai solusi dari permasalahan tersebut, telah dirancang dan dibangun sebuah model prediksi nilai LDL berbasis algoritma XGBoost. Model ini menggunakan data laboratorium seperti total kolesterol, HDL (*High-Density Lipoprotein*), dan Trigliserida sebagai parameter input untuk menghasilkan estimasi nilai LDL. Pembuatan model ini melibatkan tahapan preprocessing data, pemisahan data pelatihan dan pengujian, pelatihan model menggunakan pustaka XGBoost, serta evaluasi akurasi model dengan metrik regresi seperti RMSE (*Root Mean Squared Error*). Dengan algoritma XGBoost yang dikenal efektif dalam menangani data tabular dan kompleksitas non-linear, model ini mampu memberikan hasil prediksi yang cukup akurat dan cepat.

Dengan memanfaatkan pendekatan berbasis data dan model *machine learning*, mahasiswa tidak hanya memahami teori tentang AI dan prediksi medis, tetapi juga mendapatkan pengalaman langsung dalam merancang, menguji, dan mengimplementasikan model prediksi secara nyata. Harapannya, proyek ini dapat mendukung pembelajaran yang lebih kontekstual dan aplikatif, serta memperkuat pemahaman mahasiswa terhadap penerapan kecerdasan buatan dalam bidang kesehatan.

1.2 Rumusan Masalah

Dari latar belakang, maka dapat perumusan masalah sebagai berikut:

1. Bagaimana merancang dan membangun model prediksi nilai LDL menggunakan algoritma XGBoost berbasis data hasil pemeriksaan laboratorium seperti total kolesterol, HDL, dan trigliserida?
2. Sejauh mana tingkat akurasi model XGBoost dalam memprediksi nilai LDL dibandingkan dengan nilai aktual dari data uji?
3. Bagaimana proses preprocessing data dan pemilihan fitur memengaruhi performa model prediksi LDL?
4. Bagaimana cara pengimplementasian model AI ini ke dalam aplikasi sederhana yang dapat digunakan sebagai media pembelajaran atau sistem bantu diagnosa awal?

1.3 Batasan Masalah

Batasan masalah pada pelaksanaan kerja praktik ini adalah sebagai berikut:

1. Data yang digunakan terbatas pada parameter laboratorium, yaitu:
 - Total Kolesterol
 - HDL (*High-Density Lipoprotein*)
 - Trigliserida
2. Data lain seperti usia, jenis kelamin, berat badan, tekanan darah, atau riwayat penyakit tidak digunakan dalam model ini.
3. Model prediksi yang digunakan hanya berbasis algoritma XGBoost, tanpa perbandingan dengan algoritma lain seperti Random Forest, dan *Linear Regression*..
4. Jenis *machine learning* yang digunakan adalah *supervised learning* dengan pendekatan regresi, karena target yang diprediksi (nilai LDL) berupa data numerik kontinu.
5. Evaluasi performa model hanya dilakukan menggunakan metrik regresi, seperti RMSE (*Root Mean Squared Error*) dan MAE (*Mean Absolute Error*), tanpa melibatkan uji statistik lanjutan atau validasi klinis.
6. Implementasi terbatas pada bentuk simulasi atau aplikasi sederhana berbasis Python, tanpa integrasi ke dalam sistem laboratorium atau perangkat medis secara nyata.
7. Dataset yang digunakan bersumber dari data terbuka atau dummy, dan bukan

dari data pasien rumah sakit yang sesungguhnya karena keterbatasan akses dan pertimbangan etika.

1.4 Tujuan

Tujuan Kerja praktik di Adamlabs adalah sebagai berikut:

- 1 Merancang dan membangun model *machine learning* menggunakan algoritma XGBoost untuk memprediksi nilai LDL berdasarkan data laboratorium.
- 2 Melakukan proses preprocessing dan pemilihan fitur yang relevan terhadap prediksi nilai LDL.
- 3 Menguji dan mengevaluasi performa model prediksi LDL menggunakan metrik regresi seperti RMSE dan MAE.
- 4 Mengembangkan media pembelajaran praktis berbasis model AI dalam bidang kesehatan, khususnya untuk pengenalan penerapan AI pada data medis.
- 5 Meningkatkan pemahaman mahasiswa terhadap penerapan nyata algoritma *machine learning*, khususnya XGBoost, dalam memecahkan masalah dunia nyata.

1.5 Manfaat

Model *Artificial Intelligence* untuk Menghitung Prediksi Nilai Low-Density Lipoprotein (LDL) Menggunakan XGBoost dengan Python memiliki berbagai manfaat, baik bagi pengguna maupun lingkungan laboratorium itu sendiri, antara lain:

- 1 Memberikan solusi alternatif untuk estimasi nilai LDL tanpa perlu pengukuran langsung melalui tes laboratorium.
- 2 Menjadi media pembelajaran interaktif dalam praktik penerapan AI dan data science bagi mahasiswa, khususnya di bidang teknik komputer dan kesehatan digital.
- 3 Membantu pengguna atau mahasiswa memahami cara kerja model prediktif mulai dari proses pelatihan hingga interpretasi hasil.
- 4 Menjadi dasar pengembangan sistem bantu diagnosa awal berbasis AI, yang dapat dikembangkan lebih lanjut di masa depan.

BAB II

GAMBARAN UMUM PERUSAHAAN

2.1. Sejarah Singkat Perusahaan

Adamlabs, atau dikenal juga dengan nama ADAMLIS, merupakan sebuah perusahaan teknologi kesehatan (*healthtech*) yang berbasis di Surabaya, Indonesia. Perusahaan ini dikelola oleh PT. Wahana Meditek Indonesia dan resmi didirikan pada tahun 2019, meskipun pengembangan awal sistem mereka, yaitu ADAMLIS versi *Alpha*, telah dimulai sejak tahun 2018. Fokus utama Adamlabs adalah menyediakan solusi digital berbasis web-based SaaS (*Software as a Service*) untuk kebutuhan laboratorium, klinik, rumah sakit, dan layanan radiologi. Seiring waktu, Adamlabs terus mengembangkan produknya hingga berhasil merilis ADAMLIS versi 1.0 pada tahun 2022, dan melanjutkan dengan pengembangan versi 2.0 yang lebih canggih.

Produk-produk yang ditawarkan Adamlabs mencakup berbagai sistem informasi terintegrasi, di antaranya ADAMLIS (*Laboratory Information System*), ADAMEDS (*Clinic Information System*), ADAMEDSpro (*Hospital Information System*), ADAMPACS (sistem teleradiologi dan PACS), ADAMQUE (antrian digital), hingga AdamExpertise (platform monitoring pasien untuk dokter). Selain menyediakan *software*, Adamlabs juga aktif dalam berbagai kegiatan edukatif dan kolaboratif, seperti mengikuti event SURAMADENUSRA 2024 dan mengadakan *roadshow Health Connect* di berbagai kota termasuk Madiun dan Surabaya, guna mendukung digitalisasi pelayanan kesehatan di Indonesia.

2.1.1. Visi dan Misi

Visi

Dengan dukungan tim yang berpengalaman, kami berupaya memenuhi semua kebutuhan industri kesehatan dengan pendekatan Ekosistem Teknologi kesehatan *Healthtech*, sehingga pelayanan kesehatan menjadi lebih baik. Integrasi adalah kata yang menggambarkan visi kita untuk masa depan.

Misi

Sebagai penyedia layanan Sistem Informasi kesehatan untuk Rumah Sakit, Klinik, Laboratorium, Radiologi, yang terintegrasi, kami berharap dapat bekerja

sama dengan Anda untuk meningkatkan kualitas layanan kesehatan di Indonesia dengan pengembangan yang berkelanjutan.

2.1.2. Tujuan

1. Menyediakan layanan kesehatan yang lebih baik.

2.1.3. Struktur Organisasi



2.1.4. Informasi Kontak

Tempat	: Eastern Park AB, Jl. Raya Sukolilo Mulia No.Ruko B23, Keputih, Kec. Sukolilo, Surabaya, Jawa Timur 60111
Email	: support.marketing@adamlabs.id
Website	: adamlabs.id
No telfon	: 0882-0104-53808
Sosial Media	
Linkedin	: PT Wahana Meditek Indonesia
Facebook	: Adamlabs
TikTok	: adamlabs.id
Instagram	: @adamlabs.id

BAB III

LANDASAN TEORI

3.1 *Artificial Intelligence (AI)*

Kecerdasan Buatan atau *Artificial Intelligence (AI)* merupakan bidang dalam ilmu komputer yang mempelajari bagaimana menciptakan mesin atau sistem komputer yang mampu meniru kecerdasan manusia. AI dapat mengerjakan berbagai tugas seperti pengambilan keputusan, pemecahan masalah, pengenalan pola, dan pembelajaran dari pengalaman (Russell & Norvig, 2020). Dalam konteks teknologi modern, AI telah menjadi fondasi dari berbagai inovasi, mulai dari asisten virtual, sistem rekomendasi, hingga sistem prediksi dalam bidang medis.

Dalam bidang kesehatan, AI berperan penting dalam membantu proses diagnostik, pengolahan data medis, dan pengambilan keputusan klinis. AI dapat mengolah data dalam jumlah besar secara cepat dan akurat, sehingga sangat potensial digunakan dalam mendeteksi penyakit, memperkirakan risiko kesehatan, serta memberikan prediksi hasil laboratorium, termasuk estimasi nilai LDL dalam darah (Topol, 2019).

3.1.1. Jenis-Jenis *Artificial Intelligence (AI)*

a. *Rule-Based System*

Sistem ini beroperasi berdasarkan aturan-aturan logika yang telah ditentukan secara manual oleh pengembang atau pakar. *Rule-based system* bersifat deterministik, artinya setiap input akan menghasilkan output yang pasti sesuai dengan aturan yang telah didefinisikan sebelumnya. Meskipun tergolong sederhana, sistem ini sangat efektif pada lingkungan yang memiliki parameter dan kondisi tetap, namun kurang adaptif terhadap variasi atau ketidakpastian data (Luger, 2005).

b. *Machine Learning*

Machine Learning adalah pendekatan AI yang memungkinkan sistem belajar dari data, tanpa perlu pemrograman eksplisit. Dalam *machine learning*, sistem dilatih menggunakan data historis untuk menemukan pola dan membuat prediksi terhadap data baru. Model *machine learning* seperti

decision tree, *support vector machine*, dan XGBoost sangat populer digunakan untuk tugas prediksi dan klasifikasi (Mitchell, 1997). Machine learning sangat cocok digunakan dalam permasalahan medis yang melibatkan data numerik seperti kadar kolesterol.

c. ***Natural Language Processing***

Natural Language Processing adalah cabang AI yang bertujuan untuk memahami, menganalisis, dan memproses bahasa manusia dalam bentuk teks atau suara. NLP memungkinkan sistem komputer untuk menjalankan fungsi seperti memahami pertanyaan, menjawab dalam bahasa alami, menganalisis dokumen medis, hingga mengekstraksi informasi penting dari catatan medis elektronik (Jurafsky & Martin, 2021). Ini sangat bermanfaat dalam manajemen data rekam medis pasien.

d. ***Computer Vision***

Computer vision merupakan cabang AI yang berfokus pada pengolahan dan pemahaman data visual seperti gambar atau video. Dalam dunia medis, *computer vision* banyak digunakan untuk mendeteksi kelainan dari citra diagnostik seperti MRI, CT-Scan, atau rontgen, serta untuk klasifikasi sel kanker atau identifikasi tumor (Litjens et al., 2017).

e. ***Robotics***

Robotics adalah penggabungan AI dengan sistem fisik yang dirancang untuk melakukan tugas-tugas secara otomatis. Dalam bidang kesehatan, robot digunakan untuk membantu proses pembedahan, perawatan pasien, pengantaran obat, dan bahkan layanan konsultasi berbasis AI. Robot cerdas dilengkapi sensor dan pengambilan keputusan otomatis yang didukung oleh algoritma AI.

3.2 ***Machine Learning (ML)***

Machine Learning (ML) adalah salah satu cabang dari AI yang berfokus pada pengembangan algoritma yang memungkinkan komputer belajar dari data dan membuat prediksi atau keputusan tanpa diprogram secara eksplisit. ML mengandalkan proses pelatihan model menggunakan dataset untuk mengidentifikasi pola atau relasi antar data, dan kemudian menerapkan hasil pelatihan tersebut untuk menganalisis data baru (Goodfellow et al., 2016).

Jenis-jenis pembelajaran mesin dibagi menjadi tiga kategori utama:

1. *supervised learning*, yang memerlukan data dengan label sebagai acuan
2. *unsupervised learning*, yang tidak menggunakan label dan lebih menekankan pada pengelompokan atau eksplorasi data
3. *reinforcement learning*, di mana agen belajar berdasarkan imbalan dari setiap aksi yang diambil (Mitchell, 1997).

Dalam penelitian ini, jenis *supervised learning* digunakan karena nilai LDL sebagai label target diketahui dan dijadikan referensi untuk melatih model prediksi. Model akan belajar dari input seperti total kolesterol, HDL, dan Triglicerida untuk mengestimasi nilai LDL sebagai output.

3.3 Algoritma XGBoost (*Extreme Gradient Boosting*)

XGBoost (*Extreme Gradient Boosting*) merupakan algoritma berbasis *decision tree* yang dikembangkan untuk meningkatkan performa prediksi dalam tugas klasifikasi maupun regresi. Algoritma ini mengadopsi teknik *gradient boosting*, yang merupakan metode ensemble learning, di mana beberapa pohon keputusan dibangun secara bertahap, dan masing-masing pohon bertugas memperbaiki kesalahan prediksi dari pohon sebelumnya (Chen & Guestrin, 2016).

Keunggulan utama XGBoost adalah kecepatannya dalam pelatihan model, efisiensi memori, serta kemampuan menangani data yang hilang (*missing values*). Selain itu, XGBoost juga dilengkapi dengan regularisasi (*L1* dan *L2*) untuk menghindari *overfitting*, menjadikannya pilihan populer dalam kompetisi *data science* seperti Kaggle (Brownlee, 2019).

Dalam konteks prediksi nilai LDL, XGBoost sangat cocok karena mampu mengolah data tabular medis yang memiliki hubungan non-linear antar variabel. Dengan fitur *feature importance*, model ini juga mampu memberikan wawasan mengenai seberapa besar kontribusi masing-masing fitur terhadap hasil prediksi.

3.4 Library Pendukung Pengembangan Sistem

Dalam pengembangan sistem prediksi nilai LDL berbasis algoritma XGBoost, digunakan beberapa library Python yang memiliki peran penting dalam proses pengolahan data, pelatihan model, hingga evaluasi performa. Berikut ini adalah penjelasan mengenai library-library tersebut:

1. NumPy

NumPy adalah library untuk komputasi numerik dengan Python. NumPy menyediakan dukungan untuk array multidimensi dan berbagai fungsi matematis. Dalam proyek ini, NumPy digunakan untuk perhitungan numerik dasar seperti rata-rata dan standar deviasi (Harris et al., 2020).

2. Pandas

Pandas digunakan untuk manipulasi data berbasis tabel (*DataFrame*). Library ini mempermudah proses pembersihan data, pemfilteran nilai, dan transformasi dataset input (McKinney, 2012).

3. Scikit-learn

Library ini menyediakan berbagai metode pembelajaran mesin, termasuk *preprocessing*, pembagian data (*train-test split*), dan evaluasi model (Pedregosa et al., 2011).

3.5 Rumus Konvesional Perhitungan LDL (*Low-Density Lipoprotein*)

Salah satu metode paling umum yang digunakan untuk menghitung kadar LDL (*Low-Density Lipoprotein*) dalam darah adalah rumus Friedewald, yaitu suatu pendekatan matematis yang didasarkan pada pengukuran laboratorium terhadap total kolesterol, HDL (*High-Density Lipoprotein*), dan trigliserida.

$$LDL = \text{Total Kolesterol} - HDL - \frac{\text{Trigliserida}}{5}$$

Rumus ini digunakan secara luas karena efisien dan tidak memerlukan prosedur pengukuran langsung LDL yang mahal dan memakan waktu. Namun, rumus ini memiliki keterbatasan, yakni hanya berlaku akurat ketika kadar Trigliserida di bawah 400 mg/dL dan tidak digunakan pada pasien dengan kondisi seperti dislipidemia atau diabetes (Friedewald et al., 1972).

Rumus Friedewald digunakan sebagai referensi dasar dan validasi hasil prediksi AI terhadap nilai LDL yang diperkirakan oleh model XGBoost.

3.6 Rumus Evaluasi Model Yang Digunakan Untuk Validasi

Berikut ini adalah rumus evaluasi yang digunakan dalam sistem untuk nilai hasil prediksi program:

1. R2 (*R-Squared*)

R² atau *coefficient of determination*, adalah ukuran seberapa baik model regresi mampu menjelaskan variasi data aktual.

$$R^2 = 1 - \frac{\sum_{i=1}^n (\gamma_i - \hat{y}_i)^2}{\sum_{i=1}^n (\gamma_i - \bar{y})^2}$$

- n = Jumlah data (banyaknya sampel atau observasi)
- γ_i = Nilai aktual dari data ke-i
- \hat{y}_i = Nilai prediksi dari model untuk data ke-i
- \bar{y} = Rata-rata dari seluruh nilai aktual

2. Mean Absolute Error (MAE)

Mengukur rata-rata kesalahan absolut antara nilai aktual (observasi) dan nilai hasil prediksi:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\gamma_i - \hat{y}_i|$$

- n = Jumlah data (banyaknya sampel atau observasi)

MAE memberikan gambaran rata-rata seberapa besar nilai prediksi menyimpang dari nilai sebenarnya.

3. Mean Squared Error (MSE)

Merupakan rata-rata kuadrat dari selisih antara nilai aktual dan prediksi:

$$MSE = \frac{1}{n} (\gamma_i - \hat{y}_i)^2$$

- n = Jumlah data (banyaknya sampel atau observasi)
- γ_i = Nilai aktual dari data ke-i
- \hat{y}_i = Nilai prediksi dari model untuk data ke-i

MSE sangat sensitif terhadap error besar karena kuadrat dari selisih memperbesar efek outlier.

4. Root Mean Squared Error (RMSE)

Merupakan akar dari MSE, dan sering digunakan karena memiliki satuan yang sama dengan data asli:

$$RMSE = \sqrt{\frac{1}{n}(\gamma_i - \hat{y}_i)^2}$$

- n = Jumlah data (banyaknya sampel atau observasi)
- γ_i = Nilai aktual dari data ke-i
- \hat{y}_i = Nilai prediksi dari model untuk data ke-i

RMSE memberikan informasi yang mudah diinterpretasikan untuk melihat performa model dalam satuan yang relevan (misalnya mg/dL untuk LDL).



BAB IV

Hasil Pekerjaan dan Pembahasan

4.1 Deskripsi Kerja Pratik

Kerja praktik ini dilaksanakan di salah satu anak perusahaan dari PT. Wahana Meditek Indonesia yaitu, AdamLabs yang berlokasi di Surabaya dan berfokus pada pengembangan sistem prediksi nilai LDL (*Low-Density Lipoprotein*) menggunakan model *machine learning* XGBoost (*Extreme Gradient Boosting*). Tujuan dari proyek ini adalah untuk merancang dan mengembangkan sistem prediksi nilai – nilai parameter dengan memanfaatkan salah satu cabang dari *artificial intelligence* yaitu, *Machine Learning* (ML), sebagai media bantu dalam memprediksi nilai dari LDL (*Low-Density Lipoprotein*) dan nilai hasil laboratorium lainnya. Model ini dikembangkan dengan memanfaatkan salah satu model yang sudah ada yaitu XGBoost yang mana menggunakan konsep dasar dari *decision tree* yang dikembangkan untuk meningkatkan performa prediksi dalam tugas klasifikasi maupun regresi yang mana sesuai dengan *machine learning* dalam hal prediksi.

Sistem ini memanfaatkan metode turunan dan grid search yang mana kedua metode ini digunakan sebagai metode untuk meningkatkan akurasi dari nilai hasil prediksi nilai dari LDL (*Low-Density Lipoprotein*) yang mana hasil prediksi akan di validasi lagi dengan cara membandingkan hasil prediksi model XGBoost dengan nilai aktual dari Perhitungan manual dengan menggunakan rumus Friedewald dan akan terlihat tingkat akurasi dan berapa nilai selisihnya.

Rumus Friedewald sendiri memiliki keterbatasan terutama ketika kadar Trigliserida > 400 mg/dL, sehingga metode prediksi berbasis pembelajaran mesin menjadi solusi alternatif yang fleksibel. Model XGBoost yang digunakan dalam kerja praktik ini telah terbukti efektif dalam menangani data tabular dan menghasilkan akurasi tinggi dalam berbagai kasus medis, sebagaimana dilaporkan dalam studi oleh (Weng et al., 2017).

4.2 Alur Kerja Sistem

Alur kerja sistem prediksi LDL (*Low-Density Lipoprotein*) merupakan serangkaian tahapan terstruktur yang mencakup seluruh proses dari input data mentah hingga menghasilkan output prediksi. Setiap tahapan memiliki peran

penting dalam menjamin performa dan akurasi model. Berikut ini penjabaran detailnya:

1. Penyiapan Dataset

Dataset merupakan komponen paling vital dalam proyek *machine learning*. Pada tahapan ini dilakukan:

1. Pengumpulan data dummy (karena keterbatasan akses data medis aktual) yang berisi:
 - a. Total_Kolesterol
 - b. HDL
 - c. Trigliserida
 - d. LDL (sebagai nilai target)
2. Lalu format data disimpan dalam file berformat csv.

Struktur data dalam file.csv:

total_cholesterol,triglycerides, hdl, ldl

220,120,50,146

200,110,55,135

180,100,60,120

.....

2. Preprocessing dan Pembagian Data

Tahapan ini digunakan untuk melakukan pembacaan file csv dan membagi data yang ada didalamnya dengan program seperti dibawah ini:

```
def load_and_preprocess_csv(filepath, input_cols,
target_col=None, scale=True):
    df = pd.read_csv(filepath)
    X = df[input_cols].copy()
    y = df[target_col] if target_col else None

    if scale:
        scaler = StandardScaler()
        X = scaler.fit_transform(X)

    return X, y

def split_data(X, y, test_size=0.3, random_state=42):
    return train_test_split(X, y, test_size=test_size,
random_state=random_state)
```

Pada kode diatas menunjukan bahwa data akan di bagi dengan dengan rentan 30% pengujian melalui dan pembagian data akan selalu konsisten

dalam rentan 42 sesuai dengan yang tertera dalam kode.

3. Pelatihan dan Evaluasi Model

Proses pelatihan model prediksi nilai LDL (*Low-Density Lipoprotein*) dilakukan menggunakan pendekatan *supervised learning* dengan algoritma XGBoost Regressor. Dalam implementasinya melalui skrip train_ldl_xgboost.py, langkah awal dimulai dari pemanggilan dataset menggunakan fungsi yang telah disiapkan, yaitu:

```
df, y = load_and_preprocess_csv(DATA_PATH, INPUT_COLS,
TARGET_COL, return_df=True)
```

Data ini mencakup total kolesterol, HDL, dan Trigliserida sebagai fitur input, serta LDL sebagai target. Setelah data dimuat dan dibersihkan, dilakukan proses *feature engineering* dengan menambahkan rasio kolesterol terhadap HDL dan rasio Trigliserida terhadap HDL yang mana fitur ini kemudian disimpan agar bisa digunakan saat *deployment*:

```
df = add_derived_features(df)
...
save_feature_list(X)
```

Langkah selanjutnya adalah normalisasi data untuk menyeimbangkan skala setiap fitur. Proses ini dilakukan melalui fungsi scale_features, yang menghasilkan data terstandarisasi serta objek *scaler*-nya:

```
scaler, X_scaled = scale_features(X)
```

Setelah itu, dataset dibagi menjadi data latih dan data uji menggunakan train_test_split:

```
X_train, X_test, y_train, y_test = train_test_split(X_scaled,
y, test_size=0.2, random_state=42)
```

Model kemudian dilatih menggunakan teknik *Grid Search Cross Validation* untuk mencari kombinasi parameter terbaik, seperti n_estimators, max_depth, dan learning_rate. Parameter ini diuji dalam beberapa kombinasi menggunakan GridSearchCV:

```

params = {
    'n_estimators': [300, 400],
    'max_depth': [11, 12],
    'learning_rate': [0.03, 0.08],
}
grid = GridSearchCV(XGBRegressor(random_state=42), params,
scoring='r2', cv=3)
grid.fit(X_train, y_train)
model = grid.best_estimator_

```

Model yang telah dilatih kemudian digunakan untuk memprediksi nilai LDL dari data uji, dan hasilnya dievaluasi menggunakan metrik performa:

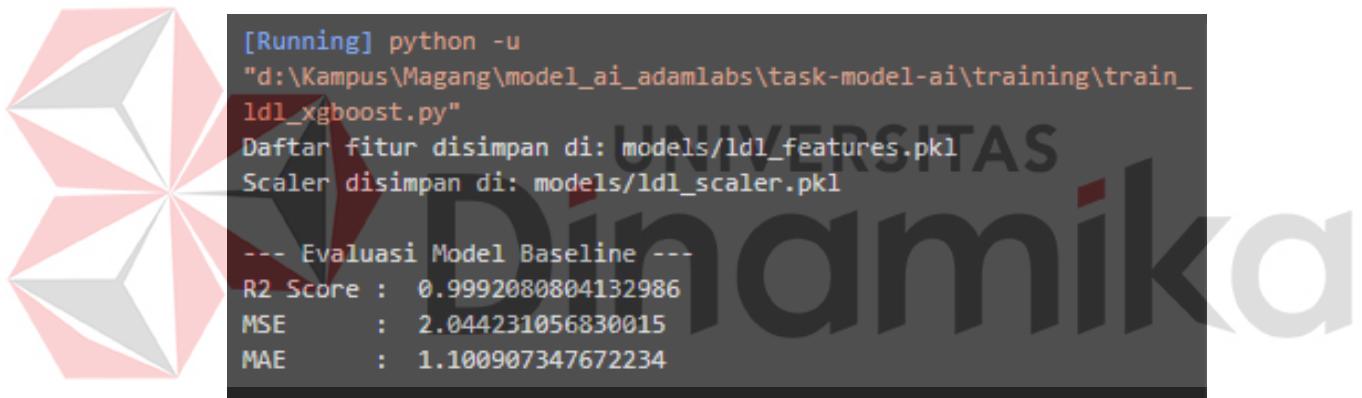
```

y_pred = model.predict(X_test)
evaluate_model(y_test, y_pred)

```

Jika performa model dianggap cukup baik, model akan disimpan dalam bentuk file .pkl agar dapat dipanggil kembali pada sistem:

```
save_model(model, MODEL_OUTPUT)
```



Gambar 4. 1. Tampilan hasil dari proses training

4.3 Uraian Perancangan Sistem

Perancangan sistem merupakan tahapan fundamental dalam proses pengembangan aplikasi prediksi nilai LDL (*Low-Density Lipoprotein*). Tahapan ini bertujuan untuk merumuskan dan membentuk kerangka sistem yang tidak hanya mampu menjalankan tugas analitik secara efisien, tetapi juga mampu berkembang dan beradaptasi sesuai kebutuhan di masa depan. Sistem yang dirancang harus memiliki alur kerja yang logis, modular, dan mudah dipelihara.

Sistem ini dirancang dalam konteks pemrosesan data laboratorium klinis yang umumnya terdiri dari parameter biokimia dasar, seperti kadar kolesterol total, trigliserida, HDL, gula darah, dan lain-lain. Data tersebut kemudian diolah dan dianalisis menggunakan pendekatan *machine learning*, dengan algoritma utama

yang digunakan adalah XGBoost (*Extreme Gradient Boosting*).

Perancangan sistem secara umum mencakup dua aspek utama:

1. Perancangan Arsitektur *Software*

- a. Menentukan struktur direktori proyek.
- b. Menerapkan pendekatan *separation of concerns* agar setiap bagian sistem memiliki tanggung jawab tunggal, seperti *preprocessing*, model *training*, *deployment*, dan evaluasi.

2. Struktur *Input-Output* Data

- a. Data *input* berupa dataset hasil uji laboratorium yang diformat dalam file csv atau melalui API.
- b. *Output* sistem berupa prediksi nilai LDL yang ditampilkan dalam format numerik dan dapat juga diintegrasikan dengan antarmuka visual seperti *dashboard* web.
- c. Proses input-output ini dirancang dalam *pipeline* bertahap yang meliputi: pengambilan data → *preprocessing* → pelatihan model → prediksi → evaluasi dan visualisasi hasil.
- d. Alur kerja sistem dimulai dari proses ingest data (pengambilan dan pembersihan), dilanjutkan dengan eksplorasi data dan transformasi fitur, pelatihan model dengan algoritma XGBoost, evaluasi hasil model, dan integrasi model ke dalam sistem yang siap digunakan oleh *end-user*.

4.4 Perancangan *Software*

Perancangan software dalam sistem ini dilakukan menggunakan bahasa pemrograman Python, yang dipilih karena bersifat open-source, memiliki komunitas pengguna yang sangat besar, serta dukungan pustaka yang sangat lengkap, khususnya dalam bidang data science dan *machine learning*.

4.4.1 Pustaka Utama yang Digunakan

1. Pandas

Berfungsi sebagai alat utama dalam manipulasi data tabular. *Library* ini memudahkan proses pembersihan data (*data cleaning*), transformasi fitur (*feature engineering*), serta eksplorasi data (EDA – *Exploratory Data*

Analysis).

2. NumPy

Digunakan dalam operasi numerik tingkat dasar dan lanjutan. Pustaka ini mendukung berbagai macam perhitungan matematis dan sangat efisien dalam pemrosesan array dan matriks, yang merupakan format utama dalam *machine learning*.

3. Scikit-learn

Merupakan *library machine learning* klasik yang digunakan dalam proses *preprocessing* data, seperti normalisasi, *encoding*, *splitting dataset* menjadi data latih dan data uji, serta menyediakan berbagai metrik evaluasi seperti MSE (*Mean Squared Error*), MAE (*Mean Absolute Error*), dan R2 (*coefficient of determination*).

4. XGBoost

Merupakan pustaka utama yang digunakan dalam pelatihan model prediksi LDL. Algoritma ini terkenal karena kemampuannya dalam melakukan *boosting decision tree* dengan efisiensi tinggi dan akurasi yang kompetitif. XGBoost juga mendukung teknik regularisasi untuk menghindari *overfitting*.

4.4.2 Struktur dan Modularisasi

Struktur proyek dirancang secara modular agar setiap bagian dari sistem dapat dikembangkan, diuji, dan dipelihara secara terpisah. Berikut adalah penjelasan struktur folder yang digunakan:

- ***training/*:** Menyimpan *script* dan *notebook* untuk pelatihan model, termasuk konfigurasi parameter.
- ***utils/*:** Berisi fungsi bantu seperti *cleaning* data, normalisasi, *encoding*, dan lain-lain.
- ***models/*:** Menyimpan file model terlatih (*serialized model*) serta class-model untuk prediksi.
- ***setup/*:** Menyediakan file konfigurasi lingkungan dan dependensi
- ***tests/*:** Folder ini menyimpan script pengujian sistem baik untuk unit *test* maupun *integration test*.
- ***schemas/*:** Menyimpan struktur data untuk validasi input-output

- **routers/**: Berisi endpoint API jika sistem diintegrasikan dengan backend web

Modularisasi ini bertujuan agar sistem dapat lebih mudah dikembangkan untuk kebutuhan lanjutan seperti integrasi dengan sistem rumah sakit, penambahan fitur visualisasi, maupun pengembangan sistem rekomendasi berdasarkan hasil prediksi LDL.

```
task-model-ai/
├── models/
├── routers/
├── schemas/
├── setup/
├── tests/
└── training/
    └── utils/
```

Gambar 4. 2. Struktur yang digunakan

4.5 Hasil Pengujian Dengan Program Test

Pengujian dilakukan dengan menjalankan program test yang telah dibuat dengan *software* VS Code. Setelah program selesai berjalan maka tampilan yang ada pada jendela output VS Code adalah seperti gambar dibawah ini. Hasil pengujian dengan *dataset* sebanyak 10.000 data *dummy* dan parameter yang digunakan adalah total_cholesterol = 180 mg/dL, triglycerides = 150 mg/dL, dan hdl = 45 mg/dL adalah 105,999 yang mana jika LDL (*Low-Density Lipoprotein*) dihitung dengan rumus Friedewald 105 mg/dL.

```
[Running] python -u
"d:\Kampus\Magang\model_ai_adamlabs\task-model-ai\tests\test_mode
ls.py"
.....
Prediksi LDL -- XGBoost: 105.999
...
-----
Ran 15 tests in 38.252s
OK
[Done] exited with code=0 in 42.391 seconds
```

Gambar 4. 3. Hasil pengujian dengan program test



BAB V

Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan pelaksanaan kerja praktik yang dilakukan di AdamLabs, serta seluruh tahapan mulai dari perancangan hingga implementasi dan evaluasi model, maka dapat disimpulkan hal-hal berikut:

1. Model prediksi nilai LDL (*Low-Density Lipoprotein*) berhasil dirancang dan dibangun menggunakan algoritma *XGBoost* dengan memanfaatkan data hasil pemeriksaan laboratorium seperti total kolesterol, HDL (*High-Density Lipoprotein*), dan trigliserida sebagai fitur utama.
2. Model *XGBoost* menunjukkan performa yang cukup baik dalam memprediksi nilai LDL (*Low-Density Lipoprotein*), dengan nilai akurasi yang diukur melalui metrik RMSE (*Root Mean Square Error*) dan MAE (*Mean Absolute Error*) mendekati nilai aktual dari data uji.
3. Proses *preprocessing* data, seperti penanganan nilai kosong dan normalisasi, serta pemilihan fitur yang relevan terbukti signifikan dalam meningkatkan akurasi dan stabilitas model prediksi LDL (*Low-Density Lipoprotein*).
4. Model yang telah dibangun berhasil diimplementasikan ke dalam sebuah aplikasi sederhana yang dapat digunakan sebagai media pembelajaran maupun sistem bantu diagnosa awal dalam bidang kesehatan.

5.2 Saran

Untuk pengembangan lebih lanjut, sistem prediksi nilai LDL ini sebaiknya divalidasi menggunakan data klinis yang lebih representatif agar hasil prediksi dapat mencerminkan kondisi nyata di lapangan. Selain itu, perlu dilakukan perbandingan performa antara algoritma XGBoost dengan metode lain seperti Random Forest, Support Vector Machine, dan Neural Network guna memperoleh gambaran menyeluruh terkait keunggulan dan keterbatasan masing-masing pendekatan dalam konteks data laboratorium. Dari sisi implementasi teknis, pengembangan *web application* dengan dukungan *framework* seperti *Flask* atau

FastAPI akan sangat membantu dalam menyediakan antarmuka yang mudah digunakan baik untuk keperluan pembelajaran maupun operasional laboratorium. Di samping itu, penerapan teknik *feature selection* dan *dimensionality reduction* seperti PCA dapat meningkatkan efisiensi komputasi sekaligus meminimalkan risiko *overfitting*, khususnya jika jumlah fitur meningkat. Pengujian lanjutan dengan skema *cross-validation* serta pemanfaatan *real-world data* dari berbagai sumber dan populasi juga diharapkan dapat memperkuat reliabilitas sistem dan membuka peluang penerapan lebih luas, baik untuk edukasi, penelitian, maupun sistem bantu diagnosis awal berbasis kecerdasan buatan.



DAFTAR PUSTAKA

- Brownlee, J. (2019). *XGBoost with Python: Gradient Boosted Trees with XGBoost and scikit-learn*. Machine Learning Mastery.
- Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794).
- Friedewald, W. T., Levy, R. I., & Fredrickson, D. S. (1972). *Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge*. Clinical Chemistry, 18(6), 499–502.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). *Array programming with NumPy*. Nature, 585(7825), 357–362.
- Jurafsky, D., & Martin, J. H. (2021). *Speech and language processing* (3rd ed.). Draft manuscript.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). *A survey on deep learning in medical image analysis*. Medical image analysis, 42, 60–88.
- Luger, G. F. (2005). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (5th ed.). Pearson Education.
- McKinney, W. (2012). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media, Inc.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw Hill.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825–2830.
- Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
- Topol, E. J. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books.
- Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). *Can machine-learning improve cardiovascular risk prediction using routine clinical data?*. PLoS ONE, 12(4), e0174944