

BAB I

PENDAHULUAN

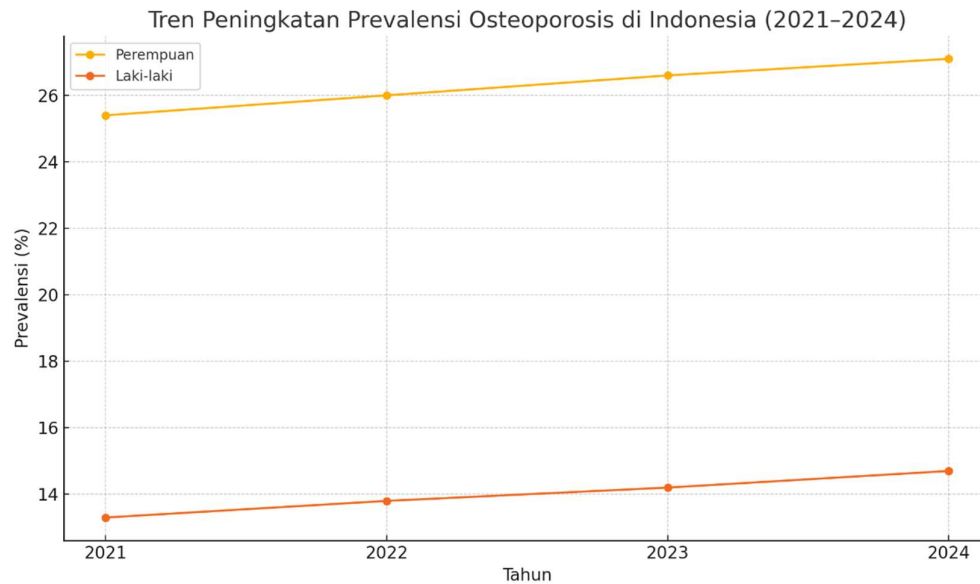
1.1 Latar Belakang

Perkembangan teknologi informasi yang pesat telah membawa perubahan signifikan dalam berbagai bidang, termasuk sektor kesehatan. Salah satu tantangan utama dalam bidang medis saat ini adalah penanganan penyakit degeneratif seperti *Osteoporosis*.

Osteoporosis merupakan salah satu penyakit tulang degeneratif yang paling umum secara global, ditandai oleh penurunan kepadatan tulang dan meningkatnya risiko fraktur. Menurut studi terbaru oleh Shashidhara et al. (2025), lebih dari 18% populasi global mengalami *Osteoporosis*, dengan prevalensi yang lebih tinggi pada wanita pascamenopause dan lansia [1]. *Osteoporosis* adalah penyakit degeneratif yang ditandai dengan menurunnya kepadatan tulang dan struktur tulang menjadi rapuh. Ini menyebabkan peningkatan risiko patah tulang, terutama pada tulang panggul, pergelangan tangan, dan tulang belakang. Penyakit ini lebih umum terjadi pada wanita pascamenopause karena penurunan kadar estrogen yang mempercepat hilangnya massa tulang. Namun, pria lansia juga berisiko, meskipun tingkat kesadaran dan diagnosis pada pria sering kali lebih rendah. Kondisi ini tidak hanya meningkatkan morbiditas dan mortalitas, tetapi juga membebani sistem kesehatan dengan biaya perawatan yang tinggi.

Di Indonesia, prevalensi *Osteoporosis* menunjukkan tren peningkatan yang mengkhawatirkan. Berdasarkan data *Riskesdas* tahun 2018, sekitar 23% wanita

dan 12% pria Indonesia diketahui memiliki risiko tinggi mengalami osteoporosis. Dalam tiga tahun terakhir, prevalensi ini menunjukkan tren peningkatan yang signifikan. Estimasi terkini untuk tahun 2021 dan 2024 memperlihatkan peningkatan risiko menjadi 25,4% dan 27,1% pada wanita, serta 13,3% dan 14,7% pada pria. Kenaikan ini mencerminkan urgensi deteksi dini dan perlunya intervensi berbasis teknologi untuk mengurangi dampak penyakit ini terhadap kualitas hidup masyarakat serta beban biaya pada sistem layanan kesehatan. [2]. Faktor-faktor seperti pola hidup sedentari, kurangnya asupan kalsium dan vitamin D, serta bertambahnya usia populasi berkontribusi terhadap tingginya angka ini. Dengan beban kesehatan yang terus meningkat, deteksi dini *Osteoporosis* menjadi suatu kebutuhan mendesak dalam upaya pencegahan komplikasi lebih lanjut.



**Gambar 1. 1 Tren Peningkatan Risiko Osteoporosis Menurut Riskesdas
(2021-2024)**

Grafik di atas menampilkan estimasi tren prevalensi osteoporosis di Indonesia untuk periode 2021 hingga 2024. Terlihat bahwa angka prevalensi terus meningkat, dari 25,4% menjadi 27,1% pada perempuan, dan dari 13,3% menjadi 14,7% pada laki-laki. Kenaikan ini konsisten dengan proyeksi Kementerian Kesehatan berdasarkan pertumbuhan penduduk lansia dan faktor risiko gaya hidup yang belum berubah secara signifikan.

Dalam bidang ilmu komputer, pendekatan berbasis machine learning telah digunakan secara luas untuk meningkatkan kemampuan prediksi berbagai penyakit, termasuk osteoporosis. Dua algoritma yang paling sering digunakan adalah *Logistic Regression* (LR) dan *Support Vector Machine* (SVM). Logistic Regression merupakan metode statistik klasik yang efektif untuk klasifikasi biner,

sedangkan SVM dikenal karena kemampuannya menangani data berdimensi tinggi dan nonlinear [3].

Beberapa penelitian menunjukkan hasil yang menjanjikan dari penggunaan algoritma ini. Misalnya, penelitian oleh Hasanpour et al. (2025) menunjukkan bahwa SVM mampu memberikan hasil prediksi yang lebih akurat dibandingkan dengan LR dalam mendeteksi risiko fraktur vertebra pada pasien osteoporosis [4]. Sementara itu, studi lain oleh Zhao et al. (2024) menemukan bahwa kombinasi antara SVM dan teknik ekstraksi fitur dapat meningkatkan sensitivitas model prediksi dibandingkan metode konvensional [5].

Namun, perbandingan sistematis dan visualisasi performa kedua algoritma tersebut dalam konteks prediksi osteoporosis masih belum banyak dilakukan, khususnya dengan mempertimbangkan variasi dataset dan parameter evaluasi yang beragam. Hal ini menciptakan celah penelitian (research gap) yang penting untuk diisi guna memahami kelebihan dan kekurangan masing-masing algoritma dalam konteks klinis dan teknis yang nyata [6].

Untuk mendukung penelitian prediktif terhadap osteoporosis, data dari *Riskesdas* memang memberikan gambaran makro yang penting. Namun, penggunaan *dataset* tambahan dari platform seperti *Kaggle* menjadi relevan dan strategis. Salah satu alasannya adalah karena *dataset* di *Kaggle* seringkali menyediakan atribut klinis yang lebih lengkap, seperti hasil pemeriksaan *bone mineral density* atau catatan medis rinci yang tidak tersedia dalam data survei nasional. Selain itu, *dataset* tersebut mencakup populasi global yang lebih

beragam, sehingga dapat meningkatkan *generalizability* model prediktif yang dibangun [7].

Keunggulan lainnya, data dari *Kaggle* telah tersedia dalam format terstruktur seperti *CSV* atau *JSON*, yang sangat mendukung proses *data preprocessing* dalam proyek *machine learning*. Bahkan, banyak di antaranya digunakan dalam kompetisi *data science*, sehingga telah disertai dengan model dasar dan *benchmark* performa yang bisa dijadikan acuan pembandingan. Terakhir, karena sifatnya yang terbuka dan dapat diakses publik, penggunaan *dataset Kaggle* mendukung semangat *open research* dan meningkatkan kemungkinan replikasi serta validasi hasil penelitian oleh pihak lain [8].

Sejumlah penelitian sebelumnya telah membuktikan efektivitas penggunaan *machine learning* dalam prediksi osteoporosis. Chung et al. berhasil mengembangkan model prediksi menggunakan algoritma non-DXA dan mencapai akurasi hingga 93%, dengan variabel penting seperti usia dan berat badan [9]. Lim et al. juga menunjukkan bahwa model berbasis *machine learning* dengan data radiomik dan CT abdomen-pelvis mampu mencapai akurasi dan spesifisitas lebih dari 93% dalam memprediksi risiko fraktur osteoporosis [10]. Selain itu, penelitian Zhang et al. menggunakan *deep learning* pada citra X-ray tulang belakang wanita pascamenopause, dan menghasilkan AUC sebesar 0.767 dengan sensitivitas 73.7% [11].

Studi oleh Qiu et al. membandingkan performa berbagai algoritma, dan menemukan bahwa *deep neural network* (DNN) memberikan performa terbaik

dengan AUC 0.848, menjadikannya salah satu pendekatan paling andal dalam prediksi osteoporosis [12]. Di tingkat nasional, penelitian oleh Wiryawan dan Agastya menunjukkan bahwa algoritma Gradient Boosting mencapai akurasi 91.07% dalam mendeteksi osteoporosis pada data berbasis one-hot encoding [13].

Jika dibandingkan dengan data dari platform terbuka seperti Kaggle, dataset dari institusi kesehatan memiliki keunggulan dari sisi validitas dan kedalaman fitur klinis yang lebih tinggi, sehingga lebih representatif untuk konteks medis. Namun demikian, Kaggle tetap menjadi sumber yang sangat berguna dalam penelitian karena formatnya yang terstruktur, ketersediaan publik, serta keberagaman atribut yang mendukung eksperimen awal dan pengembangan model. Dengan demikian, penggunaan dataset Kaggle dapat menjadi langkah awal yang strategis, namun tetap memerlukan validasi lebih lanjut menggunakan data klinis nyata agar model yang dikembangkan dapat diimplementasikan secara efektif dalam praktik medis.

Sebagai solusi, penelitian ini difokuskan pada perbandingan performa dua algoritma *Machine Learning*, yaitu *Logistic Regression (LR)* dan *Support Vector Machine (SVM)*, dalam mengklasifikasikan risiko penyakit Osteoporosis berdasarkan data pasien. Setelah perbandingan dilakukan, model prediksi akan dibangun menggunakan algoritma dengan performa terbaik di antara keduanya. Pemilihan *LR* dan *SVM* didasarkan pada karakteristik unik masing-masing algoritma: *Logistic Regression* dikenal dengan model yang sederhana, transparan, dan mudah diinterpretasikan oleh praktisi medis, sedangkan *SVM* memiliki

keunggulan dalam menangani data dengan pola kompleks dan *non-linear*, serta sering menunjukkan performa yang tinggi dalam berbagai studi prediksi penyakit.

Urgensi dari penelitian ini adalah untuk mengidentifikasi algoritma yang tidak hanya akurat, tetapi juga praktis digunakan dalam konteks medis. Berdasarkan studi terdahulu, berbagai algoritma telah digunakan untuk prediksi Osteoporosis, seperti *Decision Tree*, *Random Forest*, dan *K-Nearest Neighbors (KNN)*. Namun, hasil penelitian sebelumnya menunjukkan bahwa algoritma-algoritma tersebut memiliki kelemahan masing-masing, seperti overfitting pada *Decision Tree* atau kebutuhan tuning parameter yang tinggi pada *KNN*. Oleh karena itu, fokus pada perbandingan *LR* dan *SVM* dianggap relevan karena keduanya cenderung lebih stabil dan banyak digunakan dalam skenario klinis nyata.

Selain membandingkan performa berdasarkan metrik evaluasi standar seperti akurasi, *precision*, *recall*, dan *AUC-ROC*, penelitian ini juga akan menampilkan visualisasi hasil prediksi. Visualisasi tersebut bertujuan untuk menyampaikan informasi model dengan cara yang intuitif, sehingga dapat dimanfaatkan secara langsung oleh tenaga medis dalam pengambilan keputusan. Dengan pendekatan ini, diharapkan penelitian dapat memberikan kontribusi nyata terhadap pengembangan sistem pendukung keputusan medis berbasis *Machine Learning*, khususnya dalam upaya deteksi dini dan penanganan Osteoporosis secara lebih efektif dan efisien.

1.2 PERUMUSAN MASALAH

Berdasarkan latar belakang tersebut, maka perumusan masalah dalam penelitian ini adalah:

1. Bagaimana melakukan perbandingan pada prediksi data svm & LR
2. Bagaimana membangun sebuah model prediksi dari hasil komparasi svm & LR

1.3 BATASAN MASALAH

Agar penelitian lebih terarah, batasan masalah dalam penelitian ini ditetapkan sebagai berikut:

1. Pada penelitian ini memfokuskan kepada data penyakit osteoporosis dengan data set dan data ini mencakup informasi demografis, gaya hidup, serta faktor-faktor risiko lainnya yang relevan dalam memprediksi kemungkinan seseorang mengalami Osteoporosis.
2. Pada penelitian ini memfokuskan kepada dataset penyakit osteoporosis dengan berisi 1.000 entri data pasien.
3. Model prediksi yang digunakan berdasarkan dari hasil perbandingan akurasi, presisi, recall, dan auc roc pada algoritma svm & LR

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

1. Membandingkan performa algoritma Logistic Regression dan Support Vector Machine dalam memprediksi penyakit Osteoporosis menggunakan metrik akurasi, precision, recall, dan AUC-ROC.

2. Menampilkan hasil prediksi dari kedua algoritma untuk mendukung analisis, interpretasi hasil serta mendukung pemahaman terhadap kelebihan dan kelemahan masing-masing metode.
3. Menentukan algoritma mana yang lebih optimal dan aplikatif dalam mendukung sistem pendukung keputusan medis, khususnya dalam deteksi dini *Osteoporosis*.

1.4.2 Manfaat Penelitian

1. Penelitian ini diharapkan dapat memperkaya literatur mengenai penerapan machine learning dalam bidang medis, khususnya dalam prediksi *Osteoporosis*.
2. Hasil dari penelitian ini dapat menjadi acuan bagi praktisi medis dalam memilih metode prediksi yang lebih akurat dan mudah diinterpretasikan.

1.5 Sistematika Penulisan

Untuk memberikan gambaran umum mengenai keseluruhan penulisan ilmiah, dapat dilihat melalui sistematika penulisan yang meliputi :

BAB I : PENDAHULUAN

Latar Belakang Masalah, Perumusan Masalah, Pembatasan Masalah Tujuan & Manfaat Penelitian dan Sistematika Penulisan

BAB II : LANDASAN TEORI

Bab ini berisi tentang teori-teori dasar yang mendukung penelitian yang dikutip dari internet, buku, jurnal, dan juga pendapat atau ide para pakar yang berhubungan dengan permasalahan yang akan diangkat seperti *Machine Learning*, *Algoritma SVM*, *Logistic Reggresion* dan Penyakit *Osteoporosis*.

BAB III : METODOLOGI PENELITIAN

Uraikan secara rinci desain, metoda atau pendekatan yang akan digunakan dalam menjawab permasalahan penelitian untuk mencapai tujuan penelitian. Uraian dapat meliputi parameter penelitian, model yang digunakan, rancangan penelitian, teknikteknik pengumpulan data, teknik analisis data, cara penafsiran dan pengumpulan data bila menggunakan metode kualitatif. Perlu juga dijelaskan pendekatan yang digunakan dalam penelitian, serta tools (alat bantu) yang digunakan jika ada.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian yang meliputi evaluasi performa dari masing-masing algoritma berdasarkan metrik yang telah ditentukan. Di bagian ini juga ditampilkan hasil visualisasi data prediksi untuk masing-masing algoritma. Selanjutnya dilakukan analisis terhadap hasil yang diperoleh, membandingkan keunggulan dan kekurangan Logistic Regression dan SVM dalam

konteks prediksi *Osteoporosis*, serta membahas temuan penelitian secara kritis.

BAB V : PENUTUP

Bab terakhir berisi kesimpulan dari hasil penelitian yang telah dilakukan, yang merangkum jawaban atas perumusan masalah. Selain itu, bab ini juga memberikan saran untuk pengembangan penelitian lebih lanjut, baik dari segi peningkatan metode, penggunaan dataset yang lebih besar, ataupun eksplorasi algoritma lain yang lebih kompleks untuk prediksi penyakit *Osteoporosis*.