

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Penyalahgunaan narkoba merupakan masalah Kesehatan global yang semakin kompleks dengan dampak mortalitas signifikan. Berdasarkan laporan United Nations Office on Drugs and Crime (UNODC), pada tahun 2025 terdapat sekitar 316 juta pengguna narkoba di seluruh dunia dengan kesenjangan besar dalam akses pengobatan (*treatment gap*) yang mencapai lebih dari 80% di banyak negara [1]. Selain itu, World Health Organization (WHO) mencatat lebih dari 3 juta kematian tahunan akibat penggunaan alkohol dan narkoba, dengan mayoritas korban adalah laki-laki [2]. Dampak ini tidak hanya menimbulkan kerugian ekonomi tetapi juga membebani sistem Kesehatan, terutama di negara berkembang, di mana prevalensi penggunaan narkoba terus meningkat dan berpotensi memicu masalah sosial lebih lanjut.

Berbagai algoritma telah terbukti efektif dalam memprediksi risiko Opioid Use Disorder (OUD) dan penggunaan narkoba. Studi fokus pada *logistic regression* (LR) dan *support vector machine* (SVM) menunjukkan performa tinggi dalam identifikasi factor risiko, seperti W.Gao *et al.* [3], Hasan *et al.* [4], serta Yaseliani *et al.* [5] yang mengintegrasikan *fairness-aware* untuk mengurangi bias sisiodemografis. Selain itu pendekatan *fairness* dan bias mitigasi menjadi sorotan, seperti Colakovic dan Karakatic dengan *adaptive boosting* [6] serta Tai *et al.* yang mereview *machine learning* untuk intervensi dini krisis opioid [7]. Pada dataset UCI Drug Consumption, beberapa penelitian fokus pada prediksi konsumsi *psychoactive drugs*, termasuk Almahood *et al* [8]. Akhter dan Miller [9], Redhi *et al* [10]. dengan *explainable AI* , Ghasemkhani *et al.* dengan MMLMT [11], serta Elhachimi *et al.* [12]. Sementara itu, studi *imbalance learning* mendukung penanganan data tidak seimbang, seperti Rezvani *et al.* untuk SVM [13], Ding *et al.* dengan LGGNet [14], Alnaimy *et al.* dengan varian SMOTE [15], dan Branco *et al.* dengan metrik MCC [16].

Penelitian-penelitian tersebut masih memiliki celah, seperti kurangnya perbandingan langsung kinerja LR dan SVM pada dataset UCI Drug Consumption dengan penanganan *imbalance data* yang optimal. Di samping itu, keterbatasan dalam mengintegrasikan *explainable AI* (XAI) untuk interpretasi hasil prediksi membuat aplikasi klinis kurang praktis [10]. Penelitian ini diperlukan untuk mengisi celah tersebut dengan membandingkan kinerja

LR dan SVM dalam prediksi risiko penggunaan narkoba, di mana “risiko” didefinisikan sebagai probabilitas tinggi menjadi *user (high-risk individuals)* berdasarkan klasifikasi biner. Hasil ini dapat mendukung pengembangan *screening tool* untuk *early intervention* dan *decision support system* dalam manajemen risiko overdosis [5], sehingga berkontribusi pada kebijakan kesehatan Masyarakat yang lebih efektif secara global.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana kinerja algoritma *Logistic Regression* dan *Support Vector Machine* dalam memprediksi risiko penggunaan narkoba pada dataset UCI Drug Consumption (*Quantified*)?
2. Fitur apa saja yang paling berpengaruh dalam prediksi risiko penggunaan narkoba menggunakan kedua algoritma tersebut?
3. Bagaimana penanganan *imbalance data* memengaruhi akurasi prediksi pada kedua algoritma?

1.3 Tujuan Penelitian

Tujuan utama penelitian ini adalah untuk membandingkan kinerja algoritma *Logistic Regression* dan *Support Vector Machine* dalam prediksi risiko penggunaan narkoba pada dataset UCI Drug Consumption (*Quantified*), dengan mempertimbangkan penanganan *imbalance data* dan analisis fitur penting. Prediksi ini bertujuan sebagai basis untuk *screening tool* dan *early intervention* dalam manajemen risiko OUD [7].

1.4 Batasan Penelitian

Ruang lingkup penelitian ini dibatasi sebagai berikut:

1. Penggunaan dataset UCI Drug Consumption (*Quantified*) sebagai sumber data utama [17].
2. Penerapan dua algoritma klasifikasi saja, yaitu *Logistic Regression* dan *Support Vector Machine*.
3. Fokus pada prediksi risiko penggunaan narkoba tanpa analisis mendalam terhadap jenis narkoba spesifik.
4. Penggunaan teknik *resampling* seperti SMOTE dan ADASYN untuk penanganan *imbalance data* [15].

1.5 Keterbaruan

Penelitian ini memiliki keterbaruan dibandingkan penelitian terdahulu, sebagaimana berikut:

1. Akhter dan Miller (2025) mengidentifikasi fitur kunci untuk prediksi *Opioid Use Disorder* (UOD) menggunakan *machine learning*, termasuk *Logistic Regression* (LR), yang memberikan dasar untuk analisis fitur penting dalam prediksi risiko narkoba pada dataset UCI Drug Consumption [9].
2. Redhi *et al.* (2024) menerapkan *explainable AI* (XAI) dengan LR untuk prediksi konsumsi narkoba di Bangladesh, menekankan interpretabilitas model yang relevan untuk meningkatkan transparansi prediksi dalam penelitian ini [10].
3. Ghasemkhani *et al.* (2024) mengembangkan *Multi-Class Multi-Label Logistic Model Tree* (MMLMT), sebuah metode prediksi *hybrid* LR untuk tugas klasifikasi *multi-label* pada dataset UCI, yang mendukung pendekatan *multi-label* dalam prediksi konsumsi berbagai jenis narkoba [11].
4. Elhachimi *et al.* (2025) menggunakan *machine learning* untuk prediksi kecanduan *cannabis* berdasarkan performa kognitif dan kualitas tidur, memberikan wawasan tentang integrasi fitur psikometrik yang dapat diterapkan pada prediksi risiko narkoba secara umum [12].
5. Rezvani *et al.* (2024) mereview metode penanganan *imbalance data* khusus untuk *Support Vector Machine* (SVM), termasuk *resampling* dan *fusion methods*, yang menjadi dasar untuk optimasi SVM dalam dataset *imbalance* seperti UCI Drug Consumption [13].
6. Ding *et al.* (2024) mengusulkan LGGNet, sebuah model *graph neural network* untuk representasi *lokal-global* dalam *Brain-Computer Interface* (BCI), yang menginspirasi penggunaan representasi *graph* untuk analisis hubungan fitur demografi dan psikometrik dalam prediksi narkoba [14].
7. Alnaimy *et al.* (2022) mereview varian SMOTE untuk *handling imbalanced datasets*, termasuk *Geometric SMOTE* dan ADASYN, yang mendukung pemilihan teknik *resampling* optimal dalam penelitian ini untuk meningkatkan performa model [13].
8. Branco *et al.* (2021) mereview *predictive modeling* pada domain *imbalance*, menyoroti metrik evaluasi seperti *Matthews Correlation Coefficient* (MCC)

dan *G-Mean*, yang menjadi acuan untuk evaluasi *fairness* dan performa model LR serta SVM di penelitian ini [16].