

PREDIKSI PENYAKIT PARU-PARU MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN ADABOOST

Muhammad bani sadr^{1*}, M. Said Hasibuan²

^{1,2}Magister Teknik Informatika, Informatics & Business Institute Darmajaya

Abstract. *The lungs are the human respiratory system which plays an important role in meeting the body's oxygen needs. Apart from that, the lungs have a function as a place to exchange oxygen from the air with carbon dioxide from the blood. In some conditions, the lungs can experience problems which have a negative impact on the performance of the respiratory system. If the lungs do not function properly it will cause disease. The dataset used uses a dataset from Kaggle a total of 30,000 data using rapid miner tools. The only attributes used are age, gender, work, household, activity of staying up late, sports activity, insurance, congenital disease. The methods used in this research are only Adaboost and Naïve Bayes. Based on the research that has been carried out, it can be concluded that research showing the use of the Adaboost and Naïve Bayes algorithms in predicting lung disease produces a better level of accuracy compared to using only Naïve Bayes. The analysis results show that the accuracy level of this research reached 94.66%, with a precision of 90.71% and a recall of 100.00%. In this experiment, using a combination of Naïve Bayes and adaboost succeeded in increasing the accuracy rate by 7.44%.*

Keywords: Predictions, Lungs, Naïve Bayes, Adaboost

Received Juni 2024 / Revised Juni 2024 / Accepted Juni 2024

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).



PENDAHULUAN

Perkembangan teknologi informasi pada saat ini sangat pesat dan telah digunakan dalam banyak bidang seperti perbankan, pemerintahan, industri, pendidikan, bahkan kesehatan. Dalam bidang kesehatan[1] perkembangan teknologi informasi memiliki peran signifikan dalam penanganan berbagai penyakit salah satunya adalah paru-paru.

Paru-paru merupakan sistem pernapasan pada manusia yang berperan penting untuk memenuhi kebutuhan oksigen dalam tubuh. Selain itu, paru-paru memiliki fungsi sebagai tempat bertukarnya oksigen dari udara dengan karbon dioksida dari darah. Pada beberapa kondisi, paru-paru dapat mengalami gangguan yang berakibat buruk pada kinerja sistem pernapasan, jika paru-paru tidak berfungsi dengan baik maka akan menyebabkan timbulnya suatu penyakit [2].

Badan Kesehatan Dunia (WHO) menyebut Penyakit Paru Obstruktif Kronis (PPOK) merupakan penyebab ketiga terbanyak di dunia. Sebanyak 3,23juta kematian di tahun 2019 dengan merokok sebagai penyebab utamanya. Tahun 2020, Global initiative for Chronic Obstructive Lung Disease memperkirakan secara epidemiologi di tahun 2030 angka prevalensi Penyakit Paru Obstruktif Kronik akan terus meningkat karena meningkatnya jumlah angka orang yang merokok. Di Indonesia berdasarkan data riset kesehatan dasar 2013 prevalensi Penyakit Paru Obstruktif Kronik mencapai 3,7% atau sekitar 9,2 juta jiwa yang mengalami PPOK [3]. Riset Kesehatan Kementerian Kesehatan memperlihatkan jumlah perokok di Indonesia masih sangat tinggi, kira-kira 33,8% atau 1 dari 3 orang di Indonesia merokok. Hal ini memberikan kontribusi pada kejadian Penyakit Paru Obstruktif Kronik yang besar. Angka merokok dengan perokok pria mempunyai proporsi yang besar sekitar 63% atau 2 dari 3 pria di Indonesia saat ini merokok. Selain itu peningkatan prevalensi merokok cenderung lebih tinggi pada kelompok remaja usia 10 sampai 18 tahun, yakni sekitar 7,2% naik menjadi 9,1% di tahun 2018 atau hampir 1 dari 10 anak di Indonesia merokok. Hasil Riskesdas 2013 menunjukkan prevalensi asma, Penyakit Paru Obstruktif Kronik , dan kanker di Indonesia masing-masing 4,5 persen, 3,7 persen, dan 1,4 per mil. Prevalensi asma dan kanker lebih tinggi pada perempuan, prevalensi Penyakit Paru Obstruktif

¹*Korespondensi author.

Alamat Email: banisadr.2221210041@mail.darmajaya.ac.id (bani), msaid@darmajaya.ac.id (Said)

Kronik lebih tinggi pada laki-laki. Di Jawa Barat sendiri prevalensi PPOK Penyakit Paru Obstruktif Kronik menempati peringkat tertinggi kedua setelah asma (5,0%), PPOK (4,0%) dan Kanker (0,1%). Banyak yang tidak mengetahui tanda-tanda penyakit ini sehingga terlambat untuk datang mencari pengobatan, karakteristik penyakit pun penting untuk diketahui masyarakat.

Dari tingginya kasus penderita penyakit paru-paru, untuk dapat mengatasi masalah tersebut banyak dilakukan dalam penelitian bidang ilmu computer, penelitian terkait untuk prediksi penyakit paru-paru dilakukan oleh Toni Arifin, dkk, "Prediksi Keberhasilan Immunotherapy Pada Penyakit Kutil Dengan Menggunakan Algoritma Naïve Bayes Pada penelitian ini akurasi tertinggi adalah 81,11%"[4], Pada penelitian kedua yang dilakukan oleh Siska, dkk, "Implementasi Metode Naïve Bayes Pada Prediksi Penyakit Seliak pada penelitian ini mendapatkan nilai akurasinya sebesar 93,03%" [5].

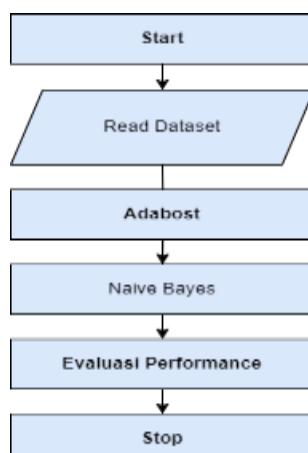
Berdasarkan masalah di atas, maka peneliti dalam hal ini mengambil judul "Prediksi Penyakit Paru-Paru Menggunakan Algoritma Naïve Bayes Dan Adaboost "Dalam penelitian ini akan dilakukan penerapan algoritma Naïve Bayes Dan Adaboost dengan mengoptimalkan atribut-atribut yang berasal dari Dataset untuk memprediksi penyakit Paru-paru, dan tools yang digunakan adalah rapid miner sehingga dapat mengetahui performa yang baik dari algoritma tersebut untuk mengetahui penyakit paru-paru.

METODE

Adaboost merupakan akronim dari Adaptive Boosting termasuk kedalam Ensemble Methods /Boosting Methods yang sering dipakai. Secara garis besar proses yang dilakukan dalam Adaboost ialah membangun sejumlah weak learners yang tidak memiliki korelasi satu sama lain, lalu kemudian menggabungkan prediksinya. Dalam penerapannya Adaboost dikombinasikan dengan algoritma lain dengan tujuan untuk mengoptimalkan performa yang dihasilkan [6]. Sedangkan Algoritma Naïve Bayes menggunakan teknik percabangan matematika dengan mencari peluang terbesar dari kemungkinan dalam klasifikasi berdasarkan frekuensi tiap klasifikasi terhadap data training yang sering disebut dengan teori pro [7]. Pada hipotesis Naïve Bayes adalah merupakan label identitas kelas yang telah menjadi target pada pemetaan klasifikasi dengan telah terkait pada korelasi hipotesis, dan bukti tersebut dapat berupa fitur-fitur yang telah menjadi input pada model klasifikasi [8].

Pada penelitian ini menggunakan Matriks konfigurasi, yaitu tabel yang terdiri dari jumlah baris data uji yang diprediksi benar dan salah dengan model klasifikasi yang digunakan. Tabel Confusion Matrix diperlukan untuk memilih kinerja terbaik dari sebuah model klasifikasi [9]. Confusion matrix adalah matrix 2×2 yang merepresentasikan hasil klasifikasi biner pada suatu dataset. Terdapat beberapa rumus umum yang dapat digunakan untuk menghitung performa klasifikasi. Hasil dari nilai accuracy, precision dan recall bisa ditampilkan dalam persentase [10].

Dataset yang digunakan menggunakan dataset dari kaggle dengan jumlah sebanyak 30.000 data yang menggunakan tools rapid miner. Atribut yang digunakan hanya Usia, jenis kelamin, bekerja, rumah tangga, Aktivitas Begadang, Aktivitas Olahraga, Asuransi, Penyakit Bawaan. Metode yang digunakan dalam penelitian ini hanya Adaboost dan Naïve Bayes. Untuk mencapai hal tersebut, maka dirancang langkah-langkah berikut sebagai alur dalam tahapan penelitian yang akan dilakukan:



Gambar Alur Dalam Tahapan Penelitian

Berikut ini merupakan penjelasan dari gambar 1 yang merupakan alur dari penelitian ini

1. Tahap pertama hal yang dilakukan adalah pengumpulan data, data yang digunakan menggunakan data Publik dari kaggle.
2. Tahap selanjutnya dalam pengelolaan data , Pada tahap ini fokus utama adalah memastikan bahwa data yang terkumpul dapat diolah secara efisien, relevan, dan aman. Dataset yang digunakan menggunakan dataset dari kaggle dengan jumlah sebanyak 30.000 data yang menggunakan tools rapid miner. Atribut yang digunakan hanya Usia, jenis kelamin, bekerja, rumah tangga, Aktivitas Begadang, Aktivitas Olahraga, Asuransi, Penyakit Bawaan. Metode yang digunakan dalam penelitian ini hanya Adaboost dan Naïve Bayes.
3. Tahap ketiga penerapan metode dan pengujian metode yang akan digunakan untuk melakukan untuk menemukan hasil terbaik.
4. Tahap yang terakhir yaitu evaluation, dimana tahap ini adalah keberhasilan dari model yang digunakan seperti menghasilkan nilai akurasi pada data diuji.

HASIL DAN DISKUSI

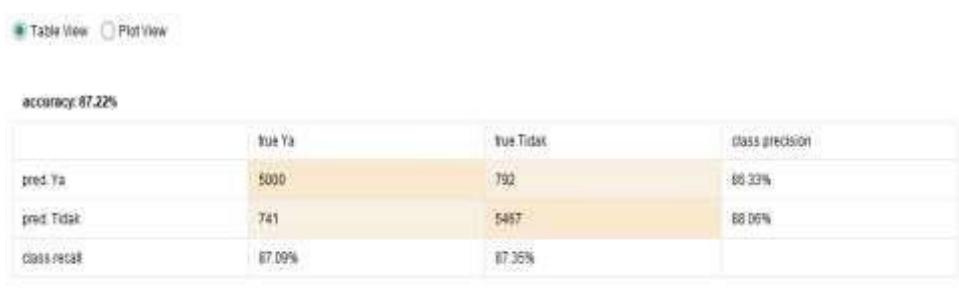
Dataset yang digunakan menggunakan dataset dari kaggle dengan jumlah sebanyak 30.000 data yang menggunakan tools rapid miner. Atribut yang digunakan hanya Usia, jenis kelamin, bekerja, rumah tangga, Aktivitas Begadang, Aktivitas Olahraga, Asuransi, Penyakit Bawaan. Metode yang digunakan dalam penelitian ini hanya Adaboost dan Naïve Bayes. Data penyakit paru terdapat pada gambar dibawah ini:

	Aktivitas_Begadang	Aktivitas_Olahraga	A_Asuransi	A_Penyakit_Bawaan
5%	Ya	Jarang	60%	Ada
49%	Tidak	Sering	40%	Tidak
	Ya	Sering	Ada	Tidak
	Ya	Jarang	Ada	Ada
	Tidak	Jarang	Ada	Tidak
	Tidak	Sering	Tidak	Ada

Gambar Dataset Penyakit Paru-Paru

1. Penelitian Menggunakan Algoritma Naïve Bayes

Penerapan data pada Rapidminer untuk prediksi penyakit paru-paru menggunakan algoritma Naïve Bayes ditunjukkan pada gambar dibawah ini:



Gambar Penerapan Data Prediksi Penyakit Paru-Paru Menggunakan Algoritma Naïve Bayes Pada Rapidminer

Gambar diatas confusion matrix yang menunjukan hasil eksperimen, didalam confusion matrix kita dapat melihat hasil akurasi, class presisi, dan class recall. Akurasi yang dihasilkan adalah 87,22 % class recall 87,35 % dan class presisi 88,06 %.

2. Penelitian Menggunakan Adabost dan Algoritma Naïve Bayes

Metode yang digunakan menggunakan metode adabost dengan algoritma Naïve Bayes. Penerapan Metode dan algoritma pada rapidminer ditunjukkan pada gambar di bawah ini:

accuracy: 94.66%			
	true Ya	true Tidak	class presisi
pred Ya	12750	0	100,00%
pred Tidak	1582	15648	90,71%
class recall	88,84%	100,00%	

Gambar Confusion Matrix Hasil Prediksi Paru Menggunakan Algoritma Naïve Bayes dan Adabost pada Rapidminer

Gambar diatas adalah confusion matrix yang menunjukan hasil eksperimen lanjutan, didalam confusion matrix kita dapat melihat hasil akurasi, class presisi, dan class recall. Akurasi yang dihasilkan adalah 94,66 % presisi 90,71%. dan recall 100,00 %.

$$\frac{12750}{12750+1582} \times 100\% = 0,9466$$

$$\frac{12750}{12750+1582} \times 100\% = 94,66\% \text{ Precision} = = = 88,84\%$$

$$\text{Recall} = = =$$

3. Perbandingan Penelitian Tidak Menggunakan Adabost dan Menggunakan Adabost

Perbandingan hasil eksperimen yang tidak menggunakan metode Adabost dan eksperimen yang menggunakan Adabost ditunjukkan pada tabel dibawah ini:

Tabel 1 Perbandingan Hasil Akurasi Penggunaan Adabost

Penelitian	Tingkat Akurasi
Naïve Bayes	87,22 %
Naïve Bayes dan Adabost	94,66 %

Hasil dari eksperimen ini untuk menghasilkan tingkat akurasi yang tinggi menggunakan naïve bayes dan adabost sebesar 94,66% dengan kenaikan akurasi sebesar 7,44%

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dapat diambil kesimpulan bahwa penelitian yang menunjukkan penggunaan algoritma adaboost dan Naïve Bayes dalam prediksi penyakit paru-paru menghasilkan tingkat akurasi yang lebih baik dibandingkan dengan hanya menggunakan naïve bayes.

Hasil analisis menunjukkan bahwa tingkat akurasi dari penelitian ini mencapai 94,66%, dengan presisi sebesar 90,71% dan recall mencapai 100,00%. Dalam eksperimen ini, penggunaan kombinasi Naïve Bayes dan adaboost berhasil meningkatkan tingkat akurasi sebesar 7,44%. Akurasi sebesar 94,66% menunjukkan kemampuan model untuk secara tepat mengidentifikasi kasus penyakit paru-paru, sementara presisi sebesar 90,71% menunjukkan sejauh mana model dapat mengidentifikasi kasus positif dengan benar. Recall yang mencapai 100,00% menunjukkan kemampuan model untuk menemukan semua kasus positif yang sebenarnya.

REFERENSI

- [1] F. Ramadhana, F. Fauziah, and W. Winarsih, “Aplikasi Sistem Pakar untuk Mendiagnosa Penyakit ISPA menggunakan Metode Naive Bayes Berbasis Website,” *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 4, no. 3, p. 320, 2020, doi: 10.30998/string.v4i3.5441.
- [2] F. Meila Azzahra Sofyan, A. Voutama, and Y. Umaidah, “Penerapan Algoritma C4.5 Untuk Prediksi Penyakit Paru-Paru Menggunakan Rapidminer,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 2, pp. 1409–1415, 2023, doi: 10.36040/jati.v7i2.6810.
- [3] A. M. Anas, L. Agustin, and B. T. Wahyudi, “Pengaruh Latihan Batuk Efektif Dan Fisioterapi Dada Terhadap Pengeluaran Sputum Pada Pasien Penyakit Paru Obstruksi Kronik Di Rs Khusus Paru Karawang,” *J. Kesehat. dan Fisioter.*, vol. 0, no. 0, pp. 118–124, 2023, [Online]. Available: <https://ejournal.insightpower.org/index.php/KeFis/article/view/214>.
- [4] T. Arifin and S. Syalwah, “Prediksi Keberhasilan Immunotherapy Pada Penyakit Kutil Dengan Menggunakan Algoritma Naïve Bayes,” *J. Responsif Ris. Sains dan Inform.*, vol. 2, no. 1, pp. 38–43, 2020, doi: 10.51977/jti.v2i1.177.
- [5] S. Siska, G. A. Saputra, C. L. Rohmat, and F. Sidik, “Implementasi Metode Naive Bayes pada Prediksi Penyakit Seliaik,” *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 7, no. 1, pp. 8–13, 2023, doi: 10.32485/kopertip.v7i1.325.
- [6] R. B. Sinaga, D. Widiyanto, B. T. Wahyono, K. Kunci, K. Paru, and R. Forest, “Deteksi Dini Penyakit Kanker Paru dengan Gabungan Algoritma Adaboost dan Random Forest,” pp. 682–691, 2022.
- [7] R. Dahlia *et al.*, “Penerapan Data Mining Terhadap Data Covid - 19,” *J. Inform.*, vol. 21, no. 1, pp. 44–52, 2021.
- [8] S. H. F. Hakim, I. Cholissodin, and A. W. Widodo, “Seleksi Fitur Dengan Particle Swarm Optimization Untuk Pengenalan Pola Wajah Menggunakan Naive Bayes (Studi Kasus Pada Mahasiswa Universitas Brawijaya Fakultas Ilmu Komputer Gedung A),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 10, pp. 1045–1057, 2017.
- [9] M. R. Romadhon and F. Kurniawan, “A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia,” *3rd 2021 East Indones. Conf. Comput. Inf. Technol. EIconCIT 2021*, pp. 41–44, 2021, doi: 10.1109/EIconCIT50028.2021.9431845.
- [10] L. A. Andika, P. A. N. Azizah, and R. Respatiwulan, “Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier,” *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019, doi: 10.13057/ijas.v2i1.29998.