

Implementasi Data Mining dalam Klasifikasi Diagnosa Kanker Payudara Menggunakan Algoritma Logistic Regression

Oleh:

Yahya Anugerah Dwi Khurrota A'yunan

Uce Indahyanti, M.Kom

Informatika

Universitas Muhammadiyah Sidoarjo

Agustus, 2023



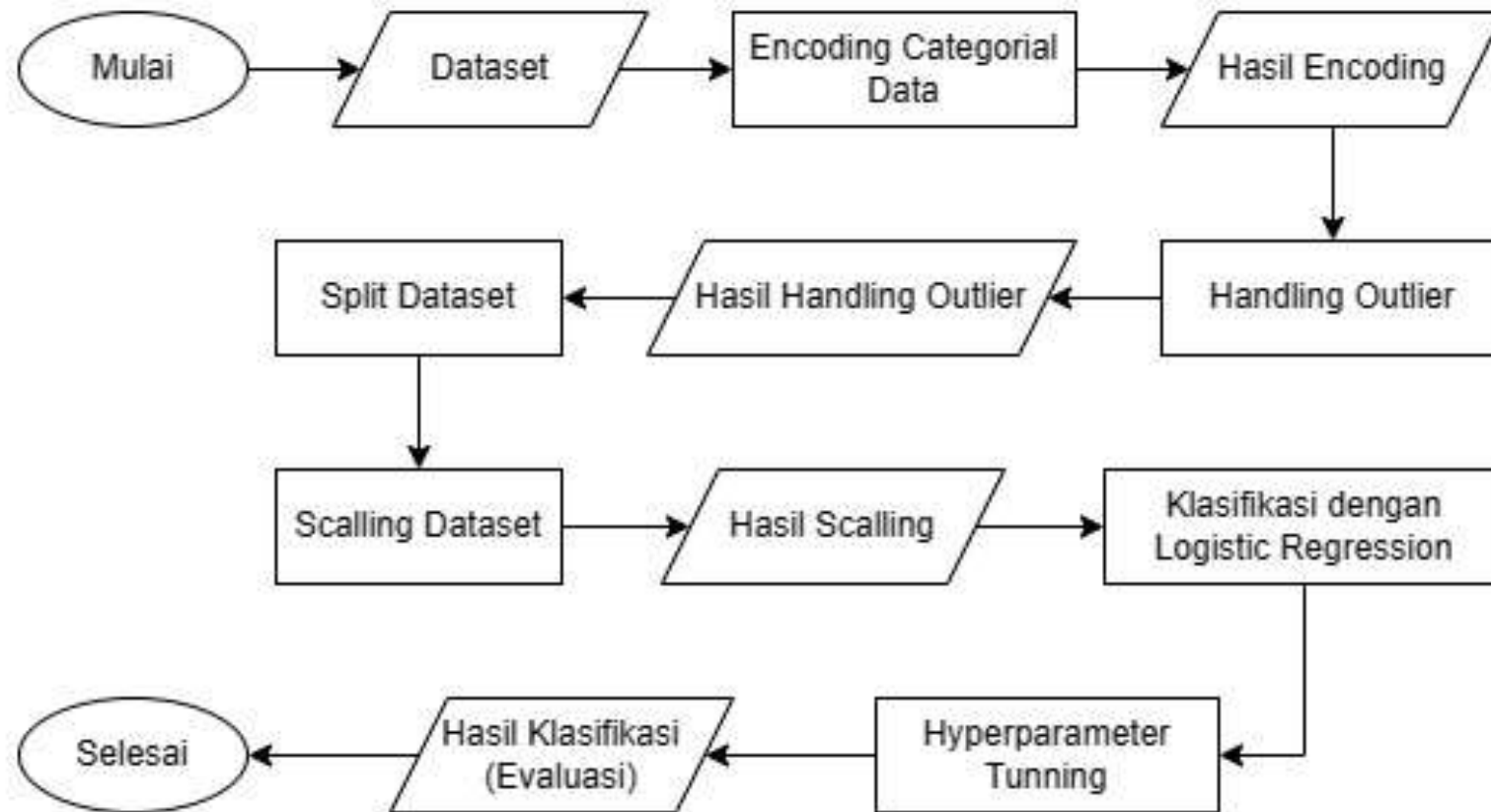
Pendahuluan

- Kanker payudara merupakan kanker yang sangat umum di kalangan wanita di dunia.
- Diagnosis dini penyakit kanker payudara dapat dilakukan melalui metode data mining.
- Algoritma Regresi Logistik adalah salah satu konsep yang krusial dalam dunia statistik, analisis data, dan Machine Learning.

Rumusan Masalah

1. Bagaimana cara melakukan klasifikasi kanker payudara menggunakan metode *Logistic Regression* ?
2. Bagaimana mengukur kinerja klasifikasi menggunakan metode *confusion matrix*?

Metode



Hasil

	Precision	Recall	F1-score	support
0	0.98	0.99	0.98	93
1	0.98	0.96	0.97	56
Accuracy			0.98	149
Macro avg	0.98	0.98	0.98	149
Weighted avg	0.98	0.98	0.98	149

Pembahasan

1. Pengambilan Dataset
2. Preprocessing
3. Klasifikasi dan Evaluasi

Temuan Penting Penelitian

Dari hasil dan proses evaluasi yang menggunakan teknik pengukuran kinerja confusion matrix, model klasifikasi diperoleh tingkat akurasi yang sangat baik yaitu 98%.

Manfaat Penelitian

Penelitian ini dapat dikembangkan menggunakan data privat yang lebih banyak dan menambahkan algoritma klasifikasi lainnya untuk membandingkan tingkat akurasi yang diperoleh.

Referensi

- [1] A. Suyanto, *Data Mining in Early Diagnosis of Breast Cancer*. Journal of Medical Systems, 2017.
- [2] Kemenkes RI., “Infodatin. Bulan Peduli Kanker Payudara Jakarta Kemenkes RI.,” *Jakarta Selatan, Indones. Kementeri. Kesehat. Republik Indones.*, pp. 1–17, 2016.
- [3] E. Susilowati, A. T. Hapsari, M. Efendi, and P. Edi, “Diagnosa Penyakit Kanker Payudara Menggunakan Metode K - Means Clustering,” *J. Sist. Informasi, Teknol. Inform. dan Komput.*, vol. 10, no. 1, pp. 27–32, 2019.
- [4] I. Mubarog, A. Setyanto, and H. Sismoro, “Sistem Klasifikasi Pada Penyakit Breast Cancer Dengan Menggunakan Metode Naïve Bayes,” *Creat. Inf. Technol. J.*, vol. 6, no. 2, p. 109, 2021, doi: 10.24076/citec.2019v6i2.246.
- [5] Suyanto, *Data mining untuk klasifikasi dan klasterisasi data*. Bandung: Informatika Bandung, 2017.
- [6] N. Meilani and O. Nurdiawan, “Data Mining untuk Klasifikasi Penderita Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor,” *J. Wahana Inform.*, vol. 2, no. 1, pp. 177–187, 2023, [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer>.
- [7] M. I. Gunawan, D. Sugiarto, and I. Mardianto, “Peningkatan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode Grid Search pada Algoritma Logistic Regression,” *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 280, 2020, doi: 10.26418/jp.v6i3.40718.

Referensi

- [8] A. Bimantara and T. A. Dina, “Klasifikasi Web Berbahaya Menggunakan Metode Logistic Regression,” *Annu. Res. Semin.*, vol. 4, no. 1, pp. 173–177, 2019, [Online]. Available: <https://seminar.ilkom.unsri.ac.id/index.php/ars/article/view/1932>.
- [9] G. P. PB, “Klasifikasi Persetujuan Permohonan Pinjaman Pada Koperasi Simpan Pinjam Menggunakan Algoritma Logistic Regression,” *J. Ilmu Data*, vol. 2, no. 12, pp. 1–12, 2022, [Online]. Available: <http://ilmudata.org/index.php/ilmudata/article/view/281%0Ahttp://ilmudata.org/index.php/ilmudata/article/download/281/270>.
- [10] F. M. Faruk, F. M. Faruk, F. S. Doven, and B. Budyanra, “Penerapan Metode Regresi Logistik Biner Untuk Mengetahui Determinan Kesiapsiagaan Rumah Tangga Dalam Menghadapi Bencana Alam,” *Semin. Nas. Off. Stat.*, vol. 2019, no. 1, pp. 379–389, 2020, doi: 10.34123/semnasoffstat.v2019i1.146.
- [11] N. G. Ramadhan, F. D. Adhinata, A. J. T. Segara, and D. P. Rakhmadani, “Deteksi Berita Palsu Menggunakan Metode Random Forest dan Logistic Regression,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, p. 251, 2022, doi: 10.30865/jurikom.v9i2.3979.
- [12] A. K. A. I, F. Nurhadi, I. K. O. Setiawan, I. A. Rizky, and R. B. Manurung, “Pengaruh Normalisasi Data pada Klasifikasi Harga Ponsel Berdasarkan Spesifikasi Menggunakan Klasifikasi Naive Bayes dan Multinomial Logistic Regression,” *J. Rekayasa Elektro Sriwij.*, vol. 3, no. 1, pp. 8–16, 2022.
- [13] A. D. Achmad, “KLASIFIKASI BREAST CANCER MENGGUNAKAN METODE LOGISTIC REGRESSION,” vol. 9, no. 1, 2022.
- [14] I. N. Atthalla, A. Jovandy, and H. Habibie, “Klasifikasi Penyakit Kanker Payudara Menggunakan Metode K Nearest Neighbor,” *Pros. Annu. Res. Semin.*, vol. 4, no. 1, pp. 148–151, 2018.

Referensi

- [15] A. K. Santoso, A. Noviriandini, A. Kurniasih, B. D. Wicaksono, and A. Nuryanto, “Klasifikasi Persepsi Pengguna Twitter Terhadap Kasus Covid-19 Menggunakan Metode Logistic Regression,” *JIK (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 234–241, 2021.
- [16] R. Nofitri and N. Irawati, “Analisis Data Hasil Keuntungan Menggunakan Software Rapidminer,” *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 5, no. 2, pp. 199–204, 2019, doi: 10.33330/jurteks.v5i2.365.
- [17] A. Saleh and F. Nasari, “Penerapan Equal-Width Interval Discretization Dalam Metode Naive Bayes Untuk Meningkatkan Akurasi Prediksi Pemilihan Jurusan Siswa (Studi Kasus: Mas Pab 2 Helvetia, Medan),” *Masy. Telemat. Dan Inf. J. Penelit. Teknol. Inf. dan Komun.*, vol. 8, no. 1, p. 1, 2018, doi: 10.17933/mti.v8i1.98.
- [18] N. Barkah, E. Sutinah, and N. Agustina, “Metode Asosiasi Data Mining Untuk Analisa Persediaan Fiber Optik Menggunakan Algoritma Apriori,” *J. Kaji. Ilm.*, vol. 20, no. 3, pp. 237–248, 2020, doi: 10.31599/jki.v20i3.288.
- [19] O. I. Desanti, I. Sunarsih, and Supriyati, “Persepsi Wanita Berisiko Kanker Payudara Tentang Pemeriksaan Payudara Sendiri Di Kota Semarang, Jawa Tengah,” *Ber. Kedokt. Masy.*, vol. 26, no. 3, pp. 152–161, 2010.
- [20] A. Alharthi, Abdulrahman ; Al-Mutairi, “Performance evaluation of classification models using confusion matrix,” *Int. J. Adv. Comput. Sci. Appl.*, pp. 427–432, 2020

