**Prediksi Risiko Penyakit Diabetes Menggunakan Algoritma Regresi Logistik**

**riview**

1. **Judul Prediksi Risiko Penyakit Diabetes Menggunakan Algoritma Regresi Logistik**

**2. Penulis**

* Qatrunnada Refa Cahyani
* Mochammad Januar Finandi
* Jathu Rianti
* Devi Lestari Arianti
* Arya Dwi Pratama Putra

**3. Afiliasi dan Korespondensi**

* **Afiliasi**:
  + Qatrunnada Refa Cahyani, Jathu Rianti – Universitas Diponegoro, Semarang
  + Mochammad Januar Finandi, Arya Dwi Pratama Putra – Universitas Nasional, Jakarta
  + Devi Lestari Arianti – Politeknik Elektronika Negeri Surabaya
* **Penulis korespondensi**:  
  Qatrunnada Refa Cahyani  
  Email: refa220801@gmail.com

**4. Jurnal / Seminar (Proceeding)**

Jurnal: **JOMLAI (Journal of Machine Learning and Artificial Intelligence)**  
Volume 1, Nomor 2, Juni 2022  
ISSN: 2828-9102 (Print) | 2828-9099 (Online)  
DOI: 10.55123/jomlai.v1i2.598

**5. Abstrak**

* **Latar belakang**: Penyakit diabetes dipengaruhi oleh banyak faktor seperti tekanan darah tinggi, kadar gula, dan riwayat keluarga.
* **Tujuan**: Memprediksi risiko seseorang menderita diabetes menggunakan algoritma regresi logistik.
* **Metode dan Data**: Menggunakan dataset Pima Indians dari Kaggle, dilakukan preprocessing data (normalisasi dan pembersihan), kemudian model dibangun dengan algoritma regresi logistik.
* **Hasil**: Recall tanpa normalisasi adalah 43%, sedangkan setelah normalisasi meningkat menjadi 55%.
* **Kesimpulan**: Normalisasi data dapat meningkatkan performa prediksi risiko diabetes menggunakan regresi logistik.

**Kata Kunci**: Diabetes, Regresi Logistik, Recall, Confusion Matrix, Normalisasi

**6. Bab 1 – Pendahuluan**

* **Masalah umum**: Diabetes adalah penyakit tidak menular yang sangat berisiko bagi siapa saja.
* **Masalah utama**: Kebutuhan akan sistem cerdas untuk memprediksi risiko diabetes secara akurat.
* **Penelitian lain**: Beberapa penelitian sebelumnya telah menggunakan regresi logistik, seperti oleh Gunawan et al (akurasi 83.33% dengan grid search).
* **Kontribusi**: Mengembangkan model prediksi diabetes menggunakan regresi logistik dan membandingkan performanya dengan dan tanpa normalisasi.

**7. Sitasi (Referensi)**

Penulisan sitasi menggunakan format **penomoran** dan disusun secara **berurutan** dalam teks. Referensi di akhir juga dituliskan sesuai urutan kemunculan, contoh:

* [1] Y. Safitri, I. K. A. Nurhayati, dst.
* [2] F. Fatmawati, dst.

**8. Bab 2 – Tinjauan Pustaka / Penelitian Terkait**

* **Teori Kasus**: Diabetes dipengaruhi oleh banyak variabel fisiologis seperti tekanan darah, insulin, indeks massa tubuh, dan faktor keturunan.
* **Teori Metode**:
  + **Regresi logistik** digunakan sebagai metode klasifikasi dengan output biner (diabetes/tidak).
  + **Preprocessing** seperti normalisasi diperlukan untuk meningkatkan kinerja model karena perbedaan skala antar variabel input.

**9. Bab 3 – Metode**

* **Data**: Dataset Pima Indians Diabetes (768 data), dengan 8 fitur dan 1 label (Outcome).
* **Alur Sistem**:
  1. **Data Acquisition**
  2. **Data Preprocessing** (penanganan missing value, outlier, normalisasi)
  3. **Data Exploration** (korelasi variabel)
  4. **Modeling** menggunakan regresi logistik
  5. **Evaluation** menggunakan Confusion Matrix
* **Diskusi**:
  1. Data dibersihkan dari duplikat, missing value diganti dengan rata-rata, dan outlier dihapus.
  2. Dua eksperimen dilakukan: dengan dan tanpa normalisasi.
  3. Hasil menunjukkan recall meningkat setelah normalisasi.

**10. Bab 4 – Kesimpulan dan Saran**

* **Kesimpulan**:
  + Model dengan normalisasi menghasilkan recall **55%** (lebih baik dari tanpa normalisasi yaitu **43%**).
  + Recall dianggap lebih penting karena lebih baik menghindari kesalahan negatif (false negative).
* **Saran**:
  + Disarankan untuk mencoba algoritma machine learning lain (seperti Random Forest, SVM, KNN) agar performa model meningkat.
  + Perlu validasi lebih lanjut pada data yang lebih besar dan bervariasi.

Link jurnal : <https://www.journal.literasisains.id/index.php/jomlai/article/view/598/470>