

Proposal Proyek KDS - Kelompok 6 (K1)

13521049 Brian Kheng
13522013 Denise Felicia Tiowanni
13522053 Erdianti Wiga Putri Andini

Kategori Topik: 4 - Penerapan Machine Learning dalam Biologi

Judul: Penerapan Model Machine Learning untuk Klasifikasi Risiko Kanker Serviks Berdasarkan Data Klinis dan Perilaku

Deskripsi:

Proyek ini bertujuan untuk mengembangkan model machine learning yang dapat mengklasifikasikan risiko kanker serviks berdasarkan data klinis dan faktor perilaku pasien. Dengan menggunakan teknik klasifikasi, model diharapkan mampu mengidentifikasi tingkat risiko kanker serviks secara lebih akurat dibandingkan metode konvensional.

- **Dataset**

Dataset yang digunakan berasal dari Kaggle dengan judul [Cervical Cancer Risk Classification](#). Dataset ini berisi informasi klinis dan faktor risiko meliputi usia, riwayat medis, kebiasaan seksual, hasil tes medis dengan 858 sampel dan 36 fitur.

- **Metode Komputasi**

Proyek ini akan menggunakan beberapa model machine learning untuk klasifikasi risiko kanker serviks:

1. **Logistic Regression:** Model ini akan digunakan untuk masalah klasifikasi biner dan multi kelas, dengan interpretasi yang jelas mengenai hubungan variabel independen dan risiko kanker serviks (rendah, sedang, tinggi).
2. **Random Forest:** Sebagai ensemble learning method, Random Forest menggunakan banyak decision trees untuk meningkatkan akurasi klasifikasi dan mengurangi overfitting.
3. **Support Vector Machine (SVM):** SVM akan digunakan untuk klasifikasi dengan margin maksimum, diharapkan memberikan klasifikasi yang lebih akurat dalam mendeteksi kelas risiko tinggi.

- **Contoh Input dan Output**

- **Input:** Data pasien berupa variabel klinis dan faktor perilaku.
- **Output:** Prediksi klasifikasi risiko kanker serviks (rendah, sedang, tinggi).

- **Pengukuran Kinerja**

Keberhasilan model akan dievaluasi dengan metrik berikut:

1. **Akurasi:** Persentase prediksi yang benar dari seluruh data.
2. **Precision (Presisi):** Mengukur proporsi prediksi positif yang benar, penting untuk menghindari kesalahan tipe I dalam diagnosis medis.
3. **Recall (Sensitivitas):** Mengukur kemampuan model dalam mendeteksi semua kasus positif (kasus kanker serviks yang berisiko tinggi).
4. **F1-Score:** Rata-rata harmonik dari precision dan recall, digunakan untuk keseimbangan antara keduanya.
5. **Cross-Validation:** Teknik validasi silang untuk memastikan model tidak overfit dan dapat digeneralisasi ke data yang lebih luas.

- **Penggunaan**

Model ini dapat digunakan sebagai alat bantu dalam skrining awal kanker serviks, membantu tenaga medis dalam pengambilan keputusan, serta meningkatkan kesadaran pasien terhadap faktor risiko yang dimilikinya.