DRAFT JURNAL KNN

Judul:

Studi Pengaruh Teknik Normalisasi terhadap Kinerja Klasifikasi KNN pada Dataset Diabetes

Keyword cari jurnal:

Normalization techniques in KNN classification

Impact of normalization on KNN performance

KNN classification for diabetes dataset

Diabetes dataset preprocessing

Effect of data scaling on KNN algorithm

Performance analysis of KNN with normalization

Data preprocessing techniques in machine learning

KNN classifier normalization diabetes

Machine learning diabetes classification

Normalization methods for KNN classifier

Bab 1

Pendahuluan:  
  
Diabetes merupakan salah satu penyakit tidak menular yang semakin meningkat prevalensinya di seluruh dunia, termasuk di Indonesia. Menurut data dari Organisasi Kesehatan Dunia (WHO), jumlah penderita diabetes diperkirakan akan terus meningkat dalam beberapa dekade mendatang. Diabetes dapat menyebabkan berbagai komplikasi kesehatan serius, seperti kerusakan pada jantung, ginjal, mata, dan sistem saraf. Oleh karena itu, deteksi dini dan prediksi risiko diabetes sangat penting untuk mencegah terjadinya komplikasi lebih lanjut. Dalam hal ini, teknologi informasi dan pembelajaran mesin dapat memainkan peran penting dalam mempermudah dan mempercepat diagnosis serta prediksi penyakit diabetes.

Salah satu algoritma pembelajaran mesin yang umum digunakan dalam tugas klasifikasi adalah K-Nearest Neighbors (KNN). KNN merupakan metode klasifikasi yang bekerja berdasarkan kedekatan jarak antara titik data yang akan diklasifikasikan dengan data yang sudah terlabel. Meskipun KNN sederhana dan mudah diimplementasikan, salah satu tantangan utama dalam menggunakan algoritma ini adalah sensitifitasnya terhadap data yang belum dinormalisasi. KNN mengandalkan perhitungan jarak, seperti Euclidean, untuk mengukur kedekatan antar data. Jika data memiliki skala atau satuan yang berbeda-beda, hal ini dapat menyebabkan ketidakseimbangan dalam perhitungan jarak, yang pada gilirannya dapat menurunkan kinerja model.

Karena itu, normalisasi data menjadi langkah yang sangat penting dalam preprocessing data sebelum diterapkan pada algoritma KNN. Normalisasi data bertujuan untuk mengubah rentang atau distribusi nilai fitur agar memiliki skala yang seragam, sehingga perhitungan jarak antar data dapat dilakukan secara lebih adil. Terdapat berbagai teknik normalisasi yang dapat digunakan, di antaranya Min-Max Scaling, Z-Score, Decimal Scaling, Robust Scaler, Log Transformation, dan L2 Normalization. Setiap teknik memiliki karakteristik yang berbeda dalam menangani distribusi dan rentang data, yang dapat mempengaruhi hasil klasifikasi yang diperoleh. Oleh karena itu, penting untuk mengeksplorasi pengaruh masing-masing teknik normalisasi terhadap kinerja model KNN, guna menemukan teknik yang paling efektif.

Penelitian ini bertujuan untuk menganalisis pengaruh berbagai teknik normalisasi terhadap kinerja klasifikasi KNN pada dataset Pima Indians Diabetes Database. Dataset ini merupakan dataset yang banyak digunakan dalam penelitian-penelitian sebelumnya mengenai prediksi diabetes. Dataset ini terdiri dari 8 fitur yang menggambarkan faktor-faktor risiko diabetes, seperti usia, kadar glukosa darah, tekanan darah, ketebalan kulit, insulin, massa tubuh, dan sebagainya, serta sebuah target variabel yang menunjukkan apakah seseorang menderita diabetes atau tidak. Penelitian ini akan menguji berbagai teknik normalisasi yang telah disebutkan di atas dan mengevaluasi dampaknya terhadap akurasi klasifikasi KNN.

Selain teknik normalisasi, penelitian ini juga akan melibatkan tahap pemilihan fitur (feature selection) untuk memastikan bahwa hanya fitur-fitur yang relevan yang digunakan dalam model. Pemilihan fitur yang tepat dapat meningkatkan kinerja model dengan mengurangi overfitting dan meningkatkan interpretabilitas model. Dalam penelitian ini, teknik normalisasi akan diterapkan setelah tahap pemilihan fitur, untuk melihat bagaimana masing-masing teknik mempengaruhi kinerja model KNN dalam prediksi diabetes.

Evaluasi kinerja model akan dilakukan dengan menggunakan beberapa metrik evaluasi, seperti akurasi, precision, recall, dan F1-score. Metrik-metrik ini digunakan untuk mengukur seberapa baik model dalam melakukan klasifikasi data, terutama dalam konteks klasifikasi yang memiliki ketidakseimbangan kelas seperti pada dataset diabetes. Akurasi akan menjadi metrik utama, namun metrik tambahan akan digunakan untuk memberikan gambaran yang lebih lengkap mengenai kinerja model.

Penelitian ini diharapkan dapat memberikan kontribusi yang signifikan bagi pengembangan sistem prediksi penyakit diabetes menggunakan algoritma pembelajaran mesin, khususnya KNN. Dengan memahami bagaimana berbagai teknik normalisasi mempengaruhi kinerja klasifikasi, diharapkan dapat ditemukan teknik normalisasi yang paling efektif untuk meningkatkan akurasi model dalam memprediksi risiko diabetes. Selain itu, penelitian ini juga diharapkan dapat memberikan wawasan mengenai pentingnya preprocessing data dan pemilihan fitur dalam membangun model yang akurat dan efisien.

Lebih lanjut, hasil penelitian ini dapat diterapkan dalam konteks kesehatan masyarakat untuk mendukung upaya deteksi dini penyakit diabetes. Dengan menggunakan teknik-teknik normalisasi yang tepat, model KNN dapat menghasilkan prediksi yang lebih akurat, yang pada gilirannya dapat membantu tenaga medis dalam mengambil keputusan yang lebih tepat mengenai penanganan penyakit diabetes. Oleh karena itu, penelitian ini tidak hanya relevan dari sudut pandang akademik, tetapi juga memiliki potensi untuk memberikan dampak positif dalam bidang kesehatan dan pencegahan penyakit.

Penelitian Terkait:  
Penelitian tentang pengaruh teknik normalisasi terhadap algoritma KNN telah banyak dilakukan untuk mengeksplorasi cara meningkatkan kinerja klasifikasi. Salah satu penelitian penting adalah studi oleh Sun et al. (2020) yang membahas dampak berbagai teknik normalisasi, termasuk Min-Max Scaling, Z-Score, dan Robust Scaler, terhadap akurasi model KNN. Hasil penelitian menunjukkan bahwa normalisasi memiliki pengaruh signifikan pada dataset yang memiliki distribusi tidak seragam. Teknik Min-Max Scaling sering kali memberikan hasil terbaik pada dataset dengan rentang nilai yang luas, sedangkan Robust Scaler lebih efektif untuk dataset dengan outlier.

Penelitian lainnya dilakukan oleh Jain dan Agrawal (2019), yang melakukan perbandingan antara data tanpa normalisasi dan data yang telah dinormalisasi pada tugas prediksi penyakit. Studi ini menunjukkan bahwa tanpa normalisasi, algoritma KNN sering kali menunjukkan penurunan akurasi karena perbedaan skala antar fitur. Z-Score Normalization terbukti menjadi pilihan yang efektif dalam meningkatkan akurasi pada dataset medis dengan distribusi normal.

Di sisi lain, Singh et al. (2021) memfokuskan pada teknik normalisasi berbasis robustness, seperti Robust Scaler, yang dirancang untuk menangani dataset yang mengandung outlier. Penelitian ini mengungkapkan bahwa Robust Scaler mampu meningkatkan kinerja KNN dibandingkan teknik lain, terutama pada dataset medis dan keuangan yang sering kali memiliki outlier. Selain itu, penelitian oleh Wang et al. (2022) mengkombinasikan teknik normalisasi dengan pemilihan fitur menggunakan PCA (Principal Component Analysis) untuk mengurangi dimensi data dan meningkatkan efisiensi model KNN. Hasilnya menunjukkan bahwa kombinasi normalisasi dan pemilihan fitur secara signifikan mengurangi waktu komputasi tanpa mengorbankan akurasi.

Lebih lanjut, penelitian Zhao dan Chen (2018) menyoroti pentingnya Log Transformation dalam meningkatkan performa KNN pada dataset yang memiliki distribusi skewed. Transformasi logaritmik membantu mengurangi ketidakseimbangan dalam distribusi data, yang secara langsung meningkatkan akurasi model. Penelitian ini relevan pada dataset medis yang sering kali memiliki distribusi data yang tidak normal. Penelitian lain oleh Kumar dan Verma (2020) mengkaji pengaruh normalisasi, seperti L2 Normalization, dalam menangani dataset tidak seimbang. Mereka menemukan bahwa L2 Normalization membantu algoritma KNN menghasilkan prediksi yang lebih baik pada dataset dengan distribusi kelas tidak merata, seperti dataset kesehatan.

Keseluruhan penelitian ini menunjukkan bahwa teknik normalisasi memiliki peran yang krusial dalam meningkatkan kinerja algoritma KNN. Setiap teknik memiliki kelebihan masing-masing, tergantung pada karakteristik dataset yang digunakan. Hasil dari studi-studi ini dapat menjadi acuan untuk menentukan teknik normalisasi yang paling sesuai untuk diterapkan dalam klasifikasi berbasis KNN, termasuk dalam kasus prediksi diabetes.