

NIM : A11.2021.13254
Nama : YOHANES DIMAS PRATAMA
Program Studi : Teknik Informatika - S1
JUDUL (Bhs. Indonesia) : Perbandingan Teknik Normalisasi Data pada Kinerja Klasifikasi KNN untuk Dataset Pima Indians Diabetes
JUDUL (Bhs. Inggris) : Comparison of Data Normalization Techniques on KNN Classification Performance for Pima Indians Diabetes Dataset

Abstrak (Bhs. Indonesia) :

Studi ini menganalisis perbandingan teknik normalisasi data dalam model K-Nearest Neighbors (KNN) untuk klasifikasi diabetes menggunakan dataset Pima Indians Diabetes. Tiga teknik normalisasi yang dievaluasi adalah Min-Max Scaling, Z-Score Scaling, dan Decimal Scaling. Setelah prapemrosesan, seperti penanganan nilai yang hilang dan penghapusan duplikat, serta seleksi fitur menggunakan metode Random Forest, fitur yang dihapus meliputi SkinThickness, Insulin, Pregnancies, dan BloodPressure. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, F1-Score, spesifisitas, dan ROC AUC. Hasil menunjukkan bahwa Min-Max Scaling memberikan peningkatan signifikan pada semua metrik, dengan akurasi tertinggi 0,8117 dan ROC AUC 0,8050. Z-Score Scaling memberikan hasil yang baik, tetapi tidak sebaik Min-Max Scaling. Decimal Scaling menunjukkan kinerja terendah. Uji statistik menggunakan Paired T-Test menunjukkan perbedaan signifikan antara Min-Max Scaling dan tanpa normalisasi pada semua metrik (P-Value <0,05), sementara Z-Score Scaling dan Decimal Scaling hanya signifikan pada beberapa metrik, dengan P-Value 0,08363 dan 0,43839 masing-masing untuk akurasi dan ROC AUC. Secara keseluruhan, Min-Max Scaling terbukti sebagai metode normalisasi terbaik untuk meningkatkan kinerja KNN dalam klasifikasi diabetes

Abstrak (Bhs. Inggris) :

This study analyzes the comparison of data normalization techniques in the KNearest Neighbors (KNN) model for diabetes classification using the Pima Indians Diabetes dataset. The three normalization techniques evaluated are Min-Max Scaling, Z-Score Scaling, and Decimal Scaling. After preprocessing, such as handling missing values and removing duplicates, as well as feature selection using the Random Forest method, the features removed include SkinThickness, Insulin, Pregnancies, and BloodPressure. The evaluation was carried out using accuracy, precision, recall, F1-Score, specificity, and ROC AUC metrics. The results show that Min-Max Scaling provides a significant improvement in all metrics, with the highest accuracy of 0.8117 and ROC AUC of 0.8050. Z-Score Scaling provides good results, but not as good as Min-Max Scaling. Decimal Scaling shows the lowest performance. Statistical tests using Paired T-Test show significant differences between Min-Max Scaling and without normalization on all metrics (P-Value <0.05), while Z-Score Scaling and Decimal Scaling are only significant on some metrics, with P-Values of 0.08363 and 0.43839 respectively for accuracy and ROC AUC. Overall, Min-Max Scaling proved to be the best normalization method for improving KNN performance in diabetes classification.

Dekan Fakultas Ilmu Komputer

Sri Winarno M.Kom., Ph.D.
NPP 0686.11.1998.142

Verifikator

Nama: Hanny Haryanto
NPP: 0686.11.2003.371

*) lembar ini harus diverifikasi oleh Bp.Hanny Haryanto (ruang Dosen DKV)

*) lembar ini harus 1 halaman

Perbandingan Teknik Normalisasi Data pada Kinerja Klasifikasi KNN untuk Dataset Pima Indians Diabetes

YOHANES DIMAS PRATAMA

*Program Studi Teknik Informatika - S1 - S-1, Fakultas Ilmu
Komputer, Universitas Dian Nuswantoro Semarang*

URL : <http://dinus.ac.id/>

Email : 111202113254@mhs.dinus.ac.id

ABSTRAK

Studi ini menganalisis perbandingan teknik normalisasi data dalam model K-Nearest Neighbors (KNN) untuk klasifikasi diabetes menggunakan dataset Pima Indians Diabetes. Tiga teknik normalisasi yang dievaluasi adalah Min-Max Scaling, Z-Score Scaling, dan Decimal Scaling. Setelah prapemrosesan, seperti penanganan nilai yang hilang dan penghapusan duplikat, serta seleksi fitur menggunakan metode Random Forest, fitur yang dihapus meliputi SkinThickness, Insulin, Pregnancies, dan BloodPressure. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, F1-Score, spesifisitas, dan ROC AUC. Hasil menunjukkan bahwa Min-Max Scaling memberikan peningkatan signifikan pada semua metrik, dengan akurasi tertinggi 0,8117 dan ROC AUC 0,8050. Z-Score Scaling memberikan hasil yang baik, tetapi tidak sebaik Min-Max Scaling. Decimal Scaling menunjukkan kinerja terendah. Uji statistik menggunakan Paired T-Test menunjukkan perbedaan signifikan antara Min-Max Scaling dan tanpa normalisasi pada semua metrik (P-Value <0,05), sementara Z-Score Scaling dan Decimal Scaling hanya signifikan pada beberapa metrik, dengan P-Value 0,08363 dan 0,43839 masing-masing untuk akurasi dan ROC AUC. Secara keseluruhan, Min-Max Scaling terbukti sebagai metode normalisasi terbaik untuk meningkatkan kinerja KNN dalam klasifikasi diabetes.

Kata Kunci : Normalisasi Data, K-Nearest Neighbors, Klasifikasi Diabetes, Skalasi Min-Max, Skalasi Z-Score

Comparison of Data Normalization Techniques on KNN Classification Performance for Pima Indians Diabetes Dataset

YOHANES DIMAS PRATAMA

*Program Studi Teknik Informatika - S1 - S-1, Fakultas Ilmu
Komputer, Universitas Dian Nuswantoro Semarang*

URL : <http://dinus.ac.id/>

Email : 111202113254@mhs.dinus.ac.id

ABSTRACT

This study analyzes the comparison of data normalization techniques in the KNearest Neighbors (KNN) model for diabetes classification using the Pima Indians Diabetes dataset. The three normalization techniques evaluated are Min-Max Scaling, Z-Score Scaling, and Decimal Scaling. After preprocessing, such as handling missing values and removing duplicates, as well as feature selection using the Random Forest method, the features removed include SkinThickness, Insulin, Pregnancies, and BloodPressure. The evaluation was carried out using accuracy, precision, recall, F1-Score, specificity, and ROC AUC metrics. The results show that Min-Max Scaling provides a significant improvement in all metrics, with the highest accuracy of 0.8117 and ROC AUC of 0.8050. Z-Score Scaling provides good results, but not as good as Min-Max Scaling. Decimal Scaling shows the lowest performance. Statistical tests using Paired T-Test show significant differences between Min-Max Scaling and without normalization on all metrics (P-Value <0.05), while Z-Score Scaling and Decimal Scaling are only significant on some metrics, with P-Values of 0.08363 and 0.43839 respectively for accuracy and ROC AUC. Overall, Min-Max Scaling proved to be the best normalization method for improving KNN performance in diabetes classification.

Keyword : Data Normalization, K-Nearest Neighbors, Diabetes Classification, Min-Max Scaling, Z-Score Scaling