Min-Max Scaling, Z-Score, Decimal Scaling

Penelitian tentang pengaruh teknik normalisasi terhadap algoritma KNN telah banyak dilakukan untuk mengeksplorasi cara meningkatkan kinerja klasifikasi. Salah satu penelitian penting yang dilakukan oleh Muasir Pagan et al. pada tahun 2023, yang menyelidiki dampak teknik skala data pada kinerja algoritma K-Nearest Neighbor (K-NN) menggunakan sepuluh dataset dari berbagai domain. Tiga teknik skala yang dievaluasi dalam penelitian ini adalah min-max normalization, Z-score, dan decimal scaling. Hasil penelitian menunjukkan bahwa pilihan teknik skala memiliki dampak signifikan terhadap kinerja algoritma k-NN, dengan Z-score dan decimal scaling umumnya memberikan hasil yang lebih baik dibandingkan min-max normalization.

Secara khusus, penelitian ini menemukan bahwa Z-score secara konsisten memberikan akurasi, presisi, recall, dan F1-score yang lebih tinggi di sebagian besar dataset yang diuji. Meskipun min-max normalization menunjukkan performa yang baik pada beberapa dataset tertentu, seperti dataset Airfoil self-noise dan Concrete compressive strength, Z-score dan decimal scaling lebih unggul dalam konteks yang lebih luas. Temuan ini menekankan pentingnya pemilihan teknik skala yang tepat berdasarkan karakteristik spesifik dari dataset, menjadikan dasar penting untuk mengeksplorasi pengaruh normalisasi pada algoritma KNN.

SITASI: Investigating the impact of data scaling on the KNN algorithm

Min-Max Scaling, Z-Score, Decimal Scaling

Penelitian yang dilakukan oleh Alshdaifat et al. (2021) mengevaluasi dampak berbagai teknik normalisasi, termasuk Min-Max Normalization, Z-Score Normalization, dan Decimal Scaling Normalization, terhadap kinerja algoritma klasifikasi seperti Support Vector Machines (SVM) dan Artificial Neural Networks (ANN). Meskipun penelitian ini berfokus pada SVM dan ANN, temuan yang menunjukkan pengaruh signifikan teknik normalisasi terhadap performa algoritma klasifikasi juga relevan untuk algoritma K-Nearest Neighbor (KNN), mengingat KNN sangat sensitif terhadap skala fitur dalam perhitungan jaraknya.

Penelitian ini dilakukan pada 18 dataset benchmark dari berbagai domain, termasuk kesehatan, bisnis, dan fisika. Hasilnya menunjukkan bahwa Z-Score Normalization sering kali memberikan hasil terbaik, khususnya karena kemampuannya menangani outlier, sedangkan Decimal Scaling Normalization dianggap kurang efektif dalam meningkatkan akurasi model secara keseluruhan.

SITASI: The Effect of Preprocessing Techniques, Applied to Numeric Features, on Classification Algorithms’ Performance

Min-Max Scaling, Decimal Scaling

Penelitian yang dilakukan oleh Saichon Sinsomboonthong (2022) bertujuan untuk membandingkan kinerja berbagai teknik normalisasi dalam proses klasifikasi menggunakan *Artificial Neural Network (ANN)*. Dalam penelitian ini, delapan metode normalisasi diuji, yaitu *min-max normalization*, *decimal scaling normalization*, *adjusted decimal scaling normalization*, *statistical column normalization*, serta empat varian *adjusted min-max normalization* yang meliputi *Adjusted-1 min-max*, *Adjusted-2 min-max*, *Adjusted-3 min-max*, dan *Adjusted-4 min-max*. Penelitian ini menggunakan enam dataset yang mencakup berbagai domain, antara lain *White Wine Quality* dari UCI, *Pima Indians Diabetes* dari Kaggle, *Vertebral Column* dari Kaggle, *Indian Liver Patient* dari MLDAT, *Working Hour* dari UCI, dan *Avocado* dari Kaggle.

Hasil penelitian menunjukkan bahwa *min-max normalization* secara umum memberikan akurasi tertinggi dan *mean square error* (MSE) terendah di sebagian besar dataset yang diuji. Namun, dalam beberapa kasus, seperti pada dataset *White Wine Quality*, *Pima Indians Diabetes*, *Vertebral Column*, dan *Indian Liver Disease*, *Adjusted-2 min-max normalization* memberikan hasil yang lebih baik, dengan akurasi lebih tinggi dan MSE yang lebih rendah. Untuk dataset seperti *Working Hour* dan *Avocado*, normalisasi *Statistical Column* dan *Decimal Scaling* juga menunjukkan hasil yang kompetitif, meskipun tidak lebih unggul daripada *min-max normalization*. Secara keseluruhan, penelitian ini menyoroti pentingnya memilih teknik normalisasi yang tepat sesuai dengan karakteristik dataset yang digunakan.

SITASI: Performance Comparison of New Adjusted Min-Max with Decimal Scaling and Statistical Column Normalization Methods for Artificial Neural Network Classification

Min-Max Scaling, Z-Score, Robust Scaler

Penelitian lainnya dilakukan oleh **Md Manjurul Ahsan et al.** pada tahun **2021**, yang mengeksplorasi pengaruh metode skala data terhadap algoritma pembelajaran mesin dan kinerja model. Penelitian ini menganalisis beberapa teknik skala, termasuk **min-max normalization**, **Z-score normalization**, dan **robust scaling**, dengan tujuan untuk memahami bagaimana masing-masing metode mempengaruhi akurasi dan efisiensi berbagai algoritma pembelajaran mesin, seperti k-NN, Decision Tree, dan Support Vector Machine (SVM). Dataset yang digunakan mencakup berbagai domain, memberikan gambaran yang komprehensif tentang dampak teknik skala pada kinerja model.

Hasil penelitian menunjukkan bahwa metode skala data memiliki pengaruh signifikan terhadap kinerja algoritma pembelajaran mesin. Secara umum, Z-score normalization dan robust scaling menunjukkan hasil yang lebih baik dibandingkan dengan min-max normalization, terutama pada dataset yang memiliki outlier. Penelitian ini juga menemukan bahwa pemilihan teknik skala yang tepat dapat meningkatkan akurasi model secara signifikan, serta mengurangi waktu pelatihan. Temuan ini memberikan wawasan penting bagi praktisi dan peneliti dalam memilih metode skala yang sesuai untuk meningkatkan kinerja model pembelajaran mesin mereka.

SITASI: Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance