

INTISARI

PENGEMBANGAN PROSEDUR *ROBUST* UNTUK ANALISIS DISKRIMINAN BERBASIS PROYEKSI ORTOGONAL TERHADAP STRUKTUR LATEN

Oleh

NOVIANA PRATIWI

21/485359/SPA/00807

Orthogonal Projection to Latent Structures Discriminant Analysis (OPLS-DA) adalah metode klasifikasi multivariat yang banyak digunakan karena kemampuannya dalam memisahkan informasi yang relevan untuk diskriminasi kelas dari variabel prediktor yang berkorelasi. Metode ini juga mampu mengatasi multikolinearitas dan menghasilkan model yang stabil dan mudah diinterpretasikan. Namun, kelemahan dari OPLS-DA klasik adalah sensitivitasnya terhadap pencilan, yang dapat mengganggu struktur data, menurunkan keakuratan klasifikasi, dan mengurangi keandalan generalisasi model.

Penelitian ini memfokuskan pada pendekatan *Robust* OPLS-DA dengan dua strategi utama untuk mengatasi masalah sensitivitas terhadap pencilan. Pertama, pendekatan pra-pemrosesan data dengan skala *robust* juga eksplorasinya sebagai pengembangannya. Pendekatan ini dapat memperbaiki struktur data sebelum dilakukan pemodelan dengan OPLS-DA. Kedua, pendekatan pembobotan *robust* menggunakan fungsi Huber dan Tukey diintegrasikan ke dalam proses iterasi model. Fungsi Huber memberikan kompromi antara error kuadrat dan deviasi absolut sehingga efektif mengurangi pengaruh pencilan. Sementara itu, fungsi Tukey menekan pengaruh pencilan ekstrem dengan memberi bobot mendekati nol pada observasi yang sangat menyimpang, sehingga meningkatkan ketahanan model dalam kondisi kontaminasi tinggi.

Evaluasi metode dilakukan melalui simulasi menyeluruh dengan berbagai skenario kontaminasi data (proporsi, jarak pencilan yang bervariasi serta 3 jenis pengulangan), serta pengujian pada data nyata. Hasil menunjukkan bahwa kedua strategi *robust* secara signifikan meningkatkan akurasi klasifikasi, stabilitas model, dan ketahanan terhadap pencilan, dibandingkan dengan OPLS-DA klasik. Model yang dikembangkan juga menunjukkan kemampuan generalisasi yang lebih baik pada dataset berdimensi tinggi dan mengandung pencilan. Selain itu, pengembangan metode ini juga menunjukkan peningkatan spesifisitas, yaitu kemampuan model

dalam mengidentifikasi kelas negatif dengan benar. Hal ini sangat penting untuk aplikasi-aplikasi kritis seperti deteksi penyakit, keamanan siber, dan deteksi penipuan, di mana kesalahan positif (*false positive*) dapat menimbulkan dampak yang serius.

Secara keseluruhan, pengembangan *Robust* OPLS-DA ini memberikan kontribusi penting dalam analisis diskriminan multivariat, terutama pada bidang-bidang seperti biomedis, kemometrik, dan metabolomik, di mana keberadaan pencilan sering tak terhindarkan dan kualitas prediksi sangat krusial.

Kata-kata kunci: Klasifikasi, ortogonal, OPLS-DA, penskalaan robust, pembobotan robust

ABSTRACT

ROBUST ORTHOGONAL PROJECTION TO LATENT STRUCTURES DISCRIMINANT ANALYSIS

By

NOVIANA PRATIWI

21/485359/SPA/00807

Orthogonal Projection to Latent Structures Discriminant Analysis (OPLS-DA) is a widely used multivariate classification method due to its ability to separate relevant class-discriminative information from correlated predictor variables. This method also effectively handles multicollinearity and produces stable, interpretable models. However, a key limitation of classical OPLS-DA is its sensitivity to outliers, which can disrupt data structure, reduce classification accuracy, and impair the generalizability of the model.

This study proposes a Robust OPLS-DA approach with two main strategies to address the issue of outlier sensitivity. First, a pre-processing approach with robust scaling. This strategy aims to improve the data structure before modeling with OPLS-DA. In this approach several robust scaling strategies are also carried out to determine the effect of robust scaling on robustness. Second, a robust weighting approach using Huber and Tukey functions is integrated into the iterative modeling process. The Huber function provides a compromise between squared error and absolute deviation, effectively reducing the influence of moderate outliers. In contrast, the Tukey function suppresses the effect of extreme outliers by assigning near-zero weights to highly deviating observations, thereby enhancing model robustness under high contamination conditions. The method was evaluated through comprehensive simulations across various data contamination scenarios (with varying outlier proportions and distances), as well as real data applications. Results show that both robust strategies significantly improve classification accuracy, model stability, and resilience to outliers compared to classical OPLS-DA. The developed model also demonstrates better generalization on high-dimensional datasets containing outliers. Furthermore, this method shows improved specificity—the ability to correctly identify negative classes—which is particularly important for critical applications such as disease detection, cybersecurity, and fraud detection, where false positives can have serious consequences.

Overall, the development of Robust OPLS-DA provides a meaningful contri-

bution to multivariate discriminant analysis, especially in fields such as biomedicine, chemometrics, and metabolomics, where the presence of outliers is often unavoidable and prediction quality is crucial.

Keywords: Classification, orthogonal, OPLS-DA, robust scaling, robust weighting