

## ABSTRAK

**Ardiansyah, Dede** (2024), *Analisis Karakteristik Resampling Smote, Smote-Enn, dan Borderline-Smote Berdasarkan Nilai Silhouette Coefficient.*

Tesis, Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Tesis ini sudah disetujui dan diperiksa oleh Pembimbing I: Dr. I Gede Aris Gunadi, S.Si., M.Kom., dan Pembimbing II: Dr. Putu Hendra Suputra, S.Kom.

*Kata Kunci:* Pembelajaran mesin, Ketidakeimbangan data, *Resampling* data, SMOTE, SMOTE-ENN, *Borderline-SMOTE*, *Silhouette Coefficient*.

Penelitian ini berfokus pada permasalahan ketidakseimbangan data dalam pembelajaran mesin, yang dapat menyebabkan model bias terhadap kelas yang lebih umum dan menurunkan akurasi untuk kelas yang kurang umum. Hal ini dapat mengakibatkan hasil yang tidak akurat dan tidak dapat diandalkan. Tujuan penelitian ini adalah untuk menganalisis karakteristik metode resampling SMOTE, SMOTE-ENN, dan *Borderline-SMOTE* berdasarkan nilai Silhouette Coefficient. Silhouette Coefficient adalah metrik yang digunakan untuk mengevaluasi kualitas pembentukan kelas yang sering digunakan untuk mengevaluasi hasil klusterisasi, menentukan jumlah kluster optimal, dan membandingkan algoritma klusterisasi. Penelitian ini menggunakan dataset yang telah dimodifikasi menjadi distribusi kelas yang tidak seimbang, yang kemudian di-resampling menggunakan ketiga metode tersebut. Resampling yang baik adalah meningkatkan kelas minoritas, menjaga Silhouette Coefficient tinggi, mengurangi *outlier*, dan memastikan akurasi model stabil. Hasil penelitian menunjukkan bahwa metode SMOTE-ENN, yang menggabungkan teknik SMOTE dengan Edited Nearest Neighbor (ENN), menghasilkan kelas yang lebih seimbang berdasarkan jumlah *outlier*, nilai silhouette yang dihasilkan. Metode ini lebih unggul dibandingkan SMOTE, yang meningkatkan jumlah sampel kelas minoritas melalui interpolasi tetapi cenderung menciptakan kelas yang terlalu ketat dan kurang bervariasi, serta *Borderline-SMOTE*, yang efektif dalam menangani kasus di perbatasan kelas namun kadang meningkatkan risiko overfitting pada dataset yang sangat bervariasi.

## ABSTRACT

**Ardiansyah, Dede** (2024), *Analysis of Smote, Smote-Enn, and Borderline-Smote Resampling Characteristics Based on Silhouette Coefficient Values.*

Thesis, Computer Science, Postgraduate Program, Ganesha University of Education.

This thesis has been approved and examined by Supervisor I: Dr. I Gede Aris Gunadi, S.Si., M.Kom., and Supervisor II: Dr. Putu Hendra Suputra, S.Kom.

*Keywords:* Machine learning, Data imbalance, Data resampling, SMOTE, SMOTE-ENN, *Borderline-SMOTE*, *Silhouette Coefficient*.

This study focuses on the issue of data imbalance in machine learning, which can lead to models being biased towards the more common classes and reducing accuracy for less common classes. This can result in inaccurate and unreliable outcomes. The aim of this study is to analyze the characteristics of the resampling methods SMOTE, SMOTE-ENN, and *Borderline-SMOTE* based on the Silhouette Coefficient values. The Silhouette Coefficient is a metric used to evaluate the quality of class formation, often used to assess clustering results, determine the optimal number of clusters, and compare clustering algorithms. This study uses datasets that have been modified to have an imbalanced class distribution, which are then resampled using the three methods. Good resampling improves minority classes, maintains a high Silhouette Coefficient, reduces *outliers*, and ensures model accuracy remains stable. The results show that the SMOTE-ENN method, which combines SMOTE with Edited Nearest Neighbor (ENN), produces more balanced classes based on the number of *outliers* and Silhouette values obtained. This method is superior compared to SMOTE, which increases the number of minority class samples through interpolation but tends to create overly tight and less varied classes, and *Borderline-SMOTE*, which is effective in handling *Borderline* cases but sometimes increases the risk of overfitting in highly varied datasets.