

ABSTRACT

IMPROVING DECISION TREE CLASSIFICATION PERFORMANCE ON ALZHEIMER DISEASE WITH DATA BALANCING TECHNIQUES

Recognizing the limitations of deep learning methods regarding transparency and interpretability, there arises a compelling demand for a classification methodology that healthcare professionals can readily adopt and integrate into their clinical decision-making processes. This study aims to find a more transparent and comprehensible classification methodology within the healthcare domain, specifically in Alzheimer's disease (AD) classification.

To accomplish this goal, the author leverages data from the National Alzheimer's Coordinating Center (NACC) Uniform Data Set version 3, collected since 2005 from Alzheimer's Disease Research Centers (ADRCs) across the United States. The author conducts a thorough evaluation and comparison of data balancing techniques for Binary Classification. The imbalance in the data from the NACC Uniform Data Set poses challenges for classification results due to the unequal representation of different classes. This disproportion can skew the model's learning process, leading to biased predictions favoring the majority class and potentially neglecting the minority class. It affects the classifier's ability to generalize and accurately predict outcomes for underrepresented classes, impacting the overall performance and reliability of the classification model.

This comprehensive strategy not only sheds light on critical Alzheimer's disease features but also establishes a transparent framework for efficiently classifying datasets. Ultimately, this work aims to facilitate early detection and enhance the management of this condition. It addresses the pressing need for more interpretable classification methods in healthcare, fostering trust among healthcare professionals. By aiding early diagnosis by minimizing false negative, this study holds the promise of enhancing the quality of life for those impacted by Alzheimer's.

The research methodology encompasses data cleaning, data balancing, and model training. Hyperparameter optimization and model evaluation are also conducted to ensure robust model performance.

Keyword: alzheimer disease, decision-tree, data balancing techniques, National Alzheimer's Coordinating Center (NACC)

ABSTRAK

MENINGKATKAN PERFORMA KLASIFIKASI DECISION TREE PADA PENYAKIT ALZHEIMER DENGAN TEKNIK DATA BALANCING

Mengenali keterbatasan metode pembelajaran mendalam terkait transparansi dan interpretabilitas, muncul kebutuhan mendesak akan metodologi klasifikasi yang dapat dengan mudah diadopsi dan diintegrasikan oleh profesional kesehatan ke dalam proses pengambilan keputusan klinis mereka. Studi ini bertujuan untuk menemukan metodologi klasifikasi yang lebih transparan dan mudah dipahami dalam domain kesehatan, khususnya dalam klasifikasi penyakit Alzheimer (AD).

Untuk mencapai tujuan ini, penulis memanfaatkan data dari National Alzheimer's Coordinating Center (NACC) Uniform Data Set versi 3, yang dikumpulkan sejak tahun 2005 dari Pusat Penelitian Penyakit Alzheimer (ADRC) di seluruh Amerika Serikat. Penulis melakukan evaluasi dan perbandingan menyeluruh terhadap teknik *data balancing* untuk Klasifikasi Biner. Ketidakseimbangan dalam data dari NACC Uniform Data Set menimbulkan tantangan bagi hasil klasifikasi karena representasi yang tidak seimbang dari berbagai kelas. Ketidakseimbangan ini dapat mengaburkan proses pembelajaran model, mengarah pada prediksi yang bias yang lebih menguntungkan kelas mayoritas dan berpotensi mengabaikan kelas minoritas. Hal ini memengaruhi kemampuan *classifier* untuk menggeneralisasi dan memprediksi hasil dengan akurat untuk kelas yang kurang terwakili, berdampak pada kinerja dan keandalan model klasifikasi secara keseluruhan.

Strategi komprehensif ini tidak hanya mengungkap fitur-fitur penting penyakit Alzheimer tetapi juga membangun kerangka kerja transparan untuk mengklasifikasikan dataset secara efisien. Pada akhirnya, karya ini bertujuan untuk memfasilitasi deteksi dini dan meningkatkan manajemen kondisi ini. Ini mengatasi kebutuhan mendesak akan metode klasifikasi yang lebih dapat diinterpretasikan dalam perawatan kesehatan, menumbuhkan kepercayaan di antara profesional kesehatan. Dengan membantu diagnosis dini dengan meminimalkan hasil negatif palsu, studi ini berpotensi meningkatkan kualitas hidup bagi mereka yang terdampak oleh penyakit Alzheimer.

Metodologi penelitian mencakup pembersihan data, penyeimbangan data, dan pelatihan model. Optimisasi hyperparameter dan evaluasi model juga dilakukan untuk memastikan kinerja model yang kuat.

Keyword: penyakit alzheimer, decision-tree, teknik data balancing, National Alzheimer's Coordinating Center (NACC)