

INTISARI

KLASIFIKASI DATA STREAM BERBASIS PEMBELAJARAN BERTAHAP MENGGUNAKAN GAUSSIAN NAIVE BAYES

Oleh

Annisa Sekartierra Mulyanto

22/494177/PA/21257

Kemajuan teknologi digital telah mendorong peningkatan volume data global secara signifikan dan memungkinkan data dihasilkan secara kontinu dengan laju yang sangat tinggi, yang dikenal sebagai *data stream*. Dalam konteks ini, penyimpanan data berskala besar serta penerapan metode *machine learning* klasik berbasis *retraining* sering kali tidak efisien karena keterbatasan memori dan beban komputasi yang tinggi. Penelitian ini mengkaji penerapan skema *incremental learning* dan *mini-batch learning* dalam kerangka *online learning* menggunakan algoritma *Gaussian Naive Bayes* sebagai model klasifikasi yang ekonomis secara komputasi untuk pemrosesan *data stream*. Pada pendekatan ini, parameter model diperbarui secara bertahap seiring kedatangan data baru tanpa memerlukan pemrosesan ulang terhadap data historis. Eksperimen dilakukan pada delapan dataset sekunder dengan mengevaluasi performa model menggunakan metrik akurasi, *F1-macro*, dan *log-loss*. Hasil pengujian menunjukkan bahwa skema pembelajaran bertahap mampu mencapai performa yang setara dengan metode *machine learning* klasik. Selain itu, pada beberapa dataset dengan karakteristik ketidakseimbangan kelas (*class imbalance*), skema pembelajaran bertahap menunjukkan performa yang relatif lebih baik, khususnya pada metrik yang sensitif terhadap kelas minoritas. Temuan ini menunjukkan bahwa pembelajaran bertahap berbasis *Gaussian Naive Bayes* merupakan alternatif yang efisien dan kompetitif untuk pemrosesan data berskala besar dan mengalir.

Kata Kunci: *data stream, incremental learning, mini-batch, Gaussian Naive Bayes, test-then-train.*

ABSTRACT

DATA STREAM CLASSIFICATION WITH INCREMENTAL LEARNING USING GAUSSIAN NAIVE BAYES

By

Annisa Sekartiertha Mulyanto

22/494177/PA/21257

Advances in digital technology have driven a significant increase in global data volume and enabled data to be generated continuously at very high rates, commonly referred to as data streams. In this context, large-scale data storage and the application of classical machine learning methods based on repeated retraining are often inefficient due to memory limitations and high computational costs. This study investigates the application of incremental learning and mini-batch learning schemes within an online learning framework using the Gaussian Naive Bayes algorithm as a computationally efficient classification model for data stream processing. Under this approach, model parameters are updated incrementally as new data arrive, without requiring repeated processing of historical data. Experiments are conducted on eight secondary datasets, and model performance is evaluated using accuracy, F1-macro, and log-loss metrics. The experimental results show that the proposed incremental learning schemes achieve performance comparable to that of classical machine learning methods. Moreover, for several datasets exhibiting class imbalance, the incremental and mini-batch approaches demonstrate relatively superior performance, particularly on metrics that are sensitive to minority classes. These findings indicate that Gaussian Naive Bayes-based incremental learning provides an efficient and competitive alternative for large-scale and streaming data processing.

Keywords: *data stream, incremental learning, mini-batch, Gaussian Naive Bayes, test-then-train.*