

ABSTRAK

Sapi Madura sebagai salah satu ras asli dari Pulau Madura telah lama menjadi bagian dari kehidupan masyarakat lokal baik sebagai sumber pangan maupun penggerak ekonomi. Dalam pengelompokan data terkait kerapian sapi, terjadi ketidakseimbangan data (data imbalance) sehingga menyebabkan model cenderung mengklasifikasikan data ke dalam kelas mayoritas dan mengabaikan kelas minoritas. Tujuan dari penelitian ini adalah mengetahui hasil klasifikasi dengan menggunakan algoritma K-Nearest Neighbor (KNN) dan teknik penyeimbang data Synthetic Minority Over-sampling Technique (SMOTE). Algoritma KNN dipilih karena merupakan algoritma klasifikasi berbasis jarak yang bekerja dengan membandingkan data baru terhadap sejumlah k terdekat dalam data latih, sedangkan SMOTE digunakan untuk mengatasi ketidakseimbangan data yang dapat menyebabkan bias model terhadap kelas mayoritas. Metode penelitian ini meliputi pengumpulan dataset sapi Madura, *preprocessing* data, penerapan SMOTE, pelatihan model dengan KNN, serta evaluasi kinerja menggunakan metrik akurasi, presisi, recall, dan F1-score. Berdasarkan hasil penelitian dapat disimpulkan bahwa performa model KNN dan KNN-SMOTE dipengaruhi oleh nilai K , penggunaan SMOTE, serta pembagian data latih dan uji. Pada KNN, akurasi tertinggi pada pembagian 80:20 dicapai pada $K = 5$ dan $K = 7$ sebesar 0,850, sedangkan pada pembagian 60:40 akurasi terbaik diperoleh pada $K = 3$ sebesar 0,857. Sedangkan pada KNN-SMOTE menunjukkan peningkatan akurasi, di mana pada pembagian 80:20 nilai tertinggi diperoleh pada $K = 9$ sebesar 0,857, dan pada pembagian 60:40 akurasi terbaik dicapai pada $K = 3$ sebesar 0,871, yang merupakan akurasi tertinggi dari seluruh percobaan yang dilakukan. Dengan demikian, model KNN-SMOTE dengan pembagian data 60:40 dan $K = 3$ menunjukkan performa terbaik.

KATA KUNCI : Data *Imbalance*, Kerapian Sapi, Klasifikasi, K-Nearest Neighbor, SMOTE

ABSTRACT

Madura cattle, as one of the native breeds of Madura Island, have long been part of the local community's life as a source of food and an economic driver. In grouping data related to cattle racing, data imbalance occurs, causing the model to tend to classify data into the majority class and ignore the minority class. The purpose of this study is to determine the classification results using the K-Nearest Neighbor (KNN) algorithm and the Synthetic Minority Over-sampling Technique (SMOTE) data balancing technique. The KNN algorithm was chosen because it is a distance-based classification algorithm that works by comparing new data to the k nearest data in the training data, while SMOTE was used to overcome data imbalance that could cause model bias towards the majority class. The research method included collecting a Madura cattle dataset, preprocessing the data, applying SMOTE, training the model with KNN, and evaluating performance using accuracy, precision, recall, and F1-score metrics. Based on the results, it can be concluded that the performance of the KNN and KNN-SMOTE models is influenced by the K value, the use of SMOTE, and the division of training and test data. In KNN, the highest accuracy in the 80:20 division was achieved at $K = 5$ and $K = 7$ at 0.850, while in the 60:40 division, the best accuracy was obtained at $K = 3$ at 0.857. Meanwhile, KNN-SMOTE showed an increase in accuracy, where in the 80:20 division the highest value was obtained at $K = 9$ at 0.857, and in the 60:40 division the best accuracy was achieved at $K = 3$ at 0.871, which was the highest accuracy of all the experiments conducted. Thus, the KNN-SMOTE model with a 60:40 data split and $K = 3$ showed the best performance.

KEYWORDS: *Bull Racing, Classification, Data Imbalance, K-Nearest Neighbor, SMOTE.*