

## ABSTRAK

Aplikasi TIX ID merupakan salah satu penyedia layanan penjualan tiket bioskop online yang memiliki jaringan kerjasama dengan bioskop bioskop yang tersebar di Indonesia. Namun aplikasi ini juga cukup banyak menuai komentar komentar yang berisi komentar positif, neutral maupun negatif, sehingga ada kemungkinan data pada ulasan aplikasi TIX ID memiliki ketidak seimbangan. Oleh karena itu, diperlukan analisis sentimen yang mampu mengklasifikasikan ulasan pengguna aplikasi tix id menjadi sentimen positif, negatif dan netral berdasarkan isi ulasan yang diberikan oleh pengguna.

Penggunaan metode klasifikasi ini dapat secara akurat mengklasifikasikan sebuah kalimat kedalam sentimen positif, negatif, ataupun netral. Metode K-nearest neighbor atau yang biasa disebut dengan KNN merupakan Kode Program klasifikasi yang sering digunakan untuk proses analisis sentimen yang dimana prosesnya dengan cara mengelompokan data baru berdasarkan jarak data baru ke beberapa data atau tetangga terdekat. Tujuan penelitian ini adalah untuk mengetahui tingkat akurasi dari penerapan metode SMOTE untuk meningkatkan kinerja Kode Program klasifikasi K-Nearest Neighbor dalam analisis ulasan aplikasi TIX ID. Metode SMOTE digunakan untuk mengatasi imbalance dataset. Data yang tidak seimbang tersebut menyulitkan metode klasifikasi dalam melakukan fungsi generalisasi pada proses machine learning.

Hasil penelitian ini didapat akurasi pada metode K-Nearest Neighbor sebelum menggunakan SMOTE dengan parameter  $K=7$  akurasi yang didapat sebesar 76,65% sedangkan dengan menggunakan SMOTE didapat akurasi tertinggi pada parameter  $K=3$  sebesar 87,05%. Sehingga pada penelitian ini menunjukkan bahwa metode SMOTE sangat mempengaruhi nilai akurasi yang dilakukan oleh K-Nearest neighbor pada saat data tidak seimbang.

**Kata Kunci :** *Imbalance Dataset, K-Nearest Neighbor, TIX ID, SMOTE*

## **ABSTRACT**

*The TIX ID application is an online cinema ticket sales service provider that has a network of collaborations with cinema theaters spread across Indonesia. However, this application also reaps quite a lot of comments containing positive, neutral or negative comments, so there is a possibility that the data in the TIX ID application review has an imbalance. Therefore, a sentiment analysis is needed that is able to classify user reviews of the tix id application into positive, negative and neutral sentiments based on the contents of the reviews provided by users.*

*The use of this classification method can accurately classify a sentence into positive, negative, or neutral sentiments. The K-nearest neighbor method or commonly referred to as KNN is a classification program code that is often used for the sentiment analysis process where the process is by grouping new data based on the distance of the new data to some data or the nearest neighbors. The purpose of this study was to determine the level of accuracy of applying the SMOTE method to improve the performance of the K-Nearest Neighbor classification Program Code in the analysis of TIX ID application reviews. The SMOTE method is used to overcome imbalance dataset. The unbalanced data makes it difficult for the classification method to perform generalization functions in the machine learning process.*

*The results of this study obtained accuracy on the K-Nearest Neighbor method before using SMOTE with parameter  $K = 7$ . The accuracy obtained was 76.65%, while using SMOTE the highest accuracy was obtained on parameter  $K = 3$  of 87.05%. So this study shows that the SMOTE method greatly affects the accuracy value performed by K-Nearest neighbors when the data is unbalanced.*

**Keywords:** *imbalance dataset, K-Nearest Neighbor, TIX ID, SMOTE*