

ABSTRAK

Perkembangan teknologi telah membawa kemudahan, kecepatan, dan praktisitas dalam mencari dan mendapatkan apa yang dibutuhkan. Kemajuan teknologi memberikan kemudahan pada orang Bali yang berada di luar Pulau Bali, terkhusus bagi yang beragama Hindu untuk mendapatkan kalender Bali. Aplikasi tersebut mempermudah mereka untuk dapat mengetahui hari raya suci yang ada pada hari maupun bulan tertentu. Aplikasi Kalender Bali mendapatkan beragam respon dari penggunanya, baik itu pendapat positif seperti pujian maupun pendapat negatif berupa kritik dan keluhan. Adapun penelitian ini akan melakukan klasifikasi sentimen ke dalam dua jenis yaitu positif dan negatif menggunakan metode *K-Nearest Neighbor* (KNN) untuk menentukan sentimen publik mengenai kepuasan pengguna terhadap aplikasi Kalender Bali. KNN merupakan metode klasifikasi yang mengelompokkan data baru berdasarkan jarak data baru ke beberapa data atau tetangga terdekat. Penelitian ini juga menerapkan metode *Tomek Links* untuk meningkatkan kinerja kode program klasifikasi KNN dalam analisis ulasan aplikasi Kalender Bali. Metode *Tomek Links* digunakan untuk mengatasi ketidakseimbangan data (*imbalance dataset*) dengan cara menurunkan jumlah sampel mayoritas yang posisinya berdekatan dengan sampel minoritas di sekitarnya. Data yang tidak seimbang menyulitkan metode klasifikasi dalam melakukan fungsi generalisasi pada proses *machine learning*. Dalam penelitian juga dilakukan pembagian rentang waktu data menjadi sebelum dan setelah pandemi covid, dikarenakan ulasan negatif memiliki jumlah yang lebih tinggi setelah pembaruan aplikasi di masa pandemi covid. Hasil penelitian pada data sebelum covid didapat bahwa akurasi tertinggi pada metode KNN tanpa atau dengan *Tomek Links* dengan parameter $K=5$ dan $K=3$, masing-masing adalah sebesar 93.7% dan 94.3%. Hasil penelitian pada data setelah covid didapat bahwa akurasi tertinggi pada metode KNN tanpa atau dengan *Tomek Links* dengan parameter $K=9$, masing-masing adalah sebesar 86.0% dan 87.2%. Pada data setelah covid, peningkatan akurasi akibat penggunaan *Tomek Links* bernilai lebih tinggi yakni sebesar 1.2% dibandingkan penggunaan pada data sebelum covid yang hanya menaikkan akurasi sebesar 0.6%. Penelitian ini menunjukkan bahwa metode *Tomek Links* mempengaruhi nilai akurasi yang dilakukan oleh KNN saat data yang digunakan tidak seimbang.

Kata kunci : Kalender Bali, Analisis Sentimen, *K-Nearest Neighbor*, *Imbalance Dataset*, *Tomek Links*

ABSTRACT

Technological developments have brought convenience, speed and practicality in finding and getting what is needed. Technological advances make it easier for Balinese people outside the island of Bali, especially Hindus, to get the Balinese calendar. This application makes it easier for them to find out which holy holidays are on certain days or months. The Bali Calendar application has received various responses from its users, both positive opinions such as praise and negative opinions in the form of criticism and complaints. This research will classify sentiment into two types, namely positive and negative using the K-Nearest Neighbor (KNN) method to determine public sentiment regarding user satisfaction with the Bali Calendar application. KNN is a classification method that groups new data based on the distance of the new data to several data or nearest neighbors. This research also applies the Tomek Links method to improve the performance of the KNN classification program code in analyzing reviews of the Bali Calendar application. The Tomek Links method is used to overcome imbalance dataset by reducing the number of majority samples whose positions are close to the surrounding minority samples. Imbalance dataset makes it difficult for classification methods to carry out generalization functions in the machine learning process. In the research, the data time period was also divided into before and after the Covid pandemic, because negative reviews had a higher number after the application was updated during the Covid pandemic. The results of research on pre-covid data showed that the highest accuracy in the KNN method without or with Tomek Links with parameter $K=5$ and $K=3$ was 93.7% and 94.3%, respectively. The results of research on data after Covid showed that the highest accuracy in the KNN method without or with Tomek Links with parameter $K=9$ was 86.0% and 87.2%, respectively. In post-covid data, the increase in accuracy due to the use of Tomek Links is higher, namely 1.2% compared to use of pre-covid data which only increased accuracy by 0.6%. This research shows that the Tomek Links method affects the accuracy values carried out by KNN when the data used is unbalanced.

Keywords: *Balinese Calendar, Sentiment Analysis, K-Nearest Neighbor, Imbalance Dataset, Tomek Links*