

**IMPLEMENTASI ALGORITMA K-MEANS DAN *k*-NEAREST  
NEIGHBOR UNTUK TOPIC DETECTION AND TRACKING PADA  
MICROBLOG TWITTER**

**ABSTRAK**

Twitter merupakan salah satu situs *microblog* yang dioperasikan oleh Twitter, Inc. Twitter memungkinkan penggunanya mengirim dan membaca pesan dengan panjang maksimal 140 karakter (*tweet*). Bandung merupakan salah satu kota tujuan wisata di Indonesia. Untuk mencari informasi tentang Bandung dapat diperoleh dari mana saja, salah satunya dari Twitter. Pada penelitian ini penulis memfokuskan penelitian untuk mencari topik-topik apa saja yang dibicarakan pengguna twitter mengenai Bandung. Teknik yang digunakan adalah *Topic detection and tracking* (TDT). *Topic detection* digunakan untuk mendeteksi topic apa saja yang terkumpul, tahapan ini menggunakan metode *clustering* dengan algoritma K-Means. Sedangkan tahapan *topic tracking* digunakan untuk mengklasifikasikan *tweet* baru ke dalam topik-topik yang telah dihasilkan pada tahap *topic detection*, metode yang digunakan adalah klasifikasi dengan menggunakan algoritma k-Nearest Neighbors (k-NN). K-Means memberikan hasil yang cukup baik, terbukti dengan *overall similarity* sebesar 0,65, selain itu k-NN juga memberikan hasil yang cukup baik dengan rata-rata *overall similarity* sebesar 0,70.

**Kata Kunci:** TDT, K-Means, kNN, Twitter, Visualisasi.

**IMPLEMENTATION OF THE ALGORITHM K-MEANS AND K-  
NEAREST NEIGHBOR FOR *TOPIC DETECTION AND TRACKING ON*  
TWITTER**

**ABSTRACT**

Twitter is a microblog sites operated by Twitter, Inc. Twitter allows users to send and read messages with a maximum length of 140 characters (tweet). Bandung is one of the tourist destinations in Indonesia. To find information about Bandung can be obtained from anywhere, one of Twitter. In this study, the authors focused research to find the topics discussed anything about Bandung by twitter users. This study used technique Topic detection and Tracking (TDT). Topic detection is used to detect any topic are collected, this task using clustering with K-Means algorithm. Topic tracking is used to classify of new tweets into topics that have been produced at the task of topic detection, this task using classification with k-Nearest Neighbors (k-NN) algorithms. K-Means give good results, the overall similarity score is 0.65, k-NN also gives good results with an average overall similarity score is 0.70.

**Keywords:** TDT, K-Means, kNN, Twitter, Visualization.