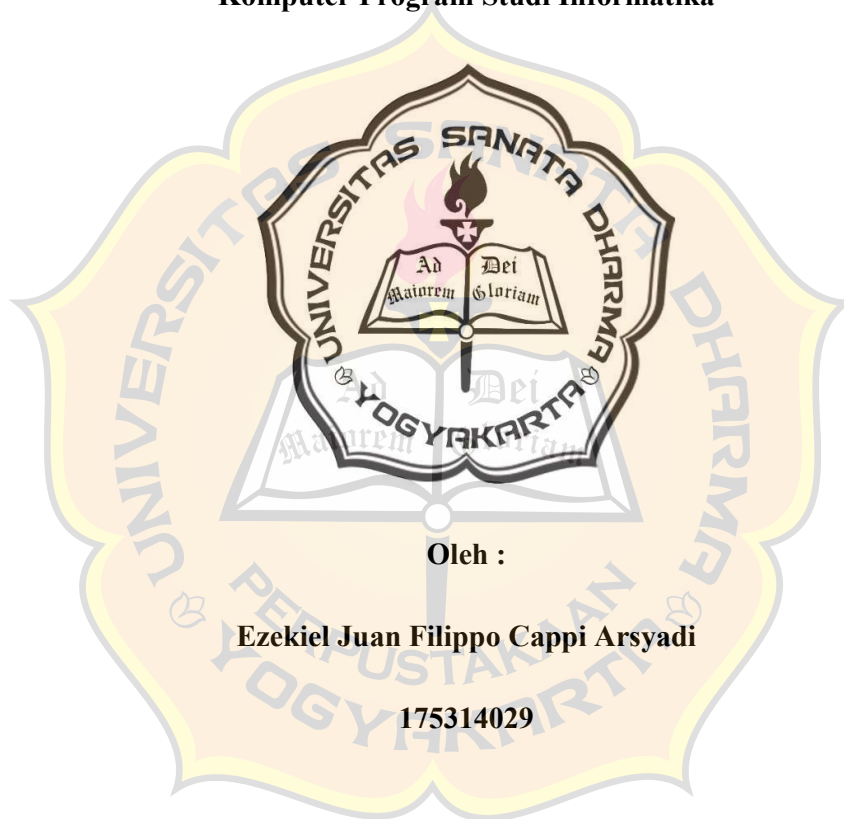


**ANALISIS SENTIMEN KOMENTAR *NETIZEN TWITTER* PADA MASA
PANDEMI *COVID-19* TERHADAP AKUN *@TELKOMSEL*
MENGUNAKAN METODE *K-NEAREST NEIGHBOR***

SKRIPSI

**Diajukan Untuk Memenuhi Salah Satu Syarat Memperoleh Gelar Sarjana
Komputer Program Studi Informatika**



Oleh :

Ezekiel Juan Filippo Cappi Arsyadi

175314029

PROGRAM STUDI INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS SANATA DHARMA

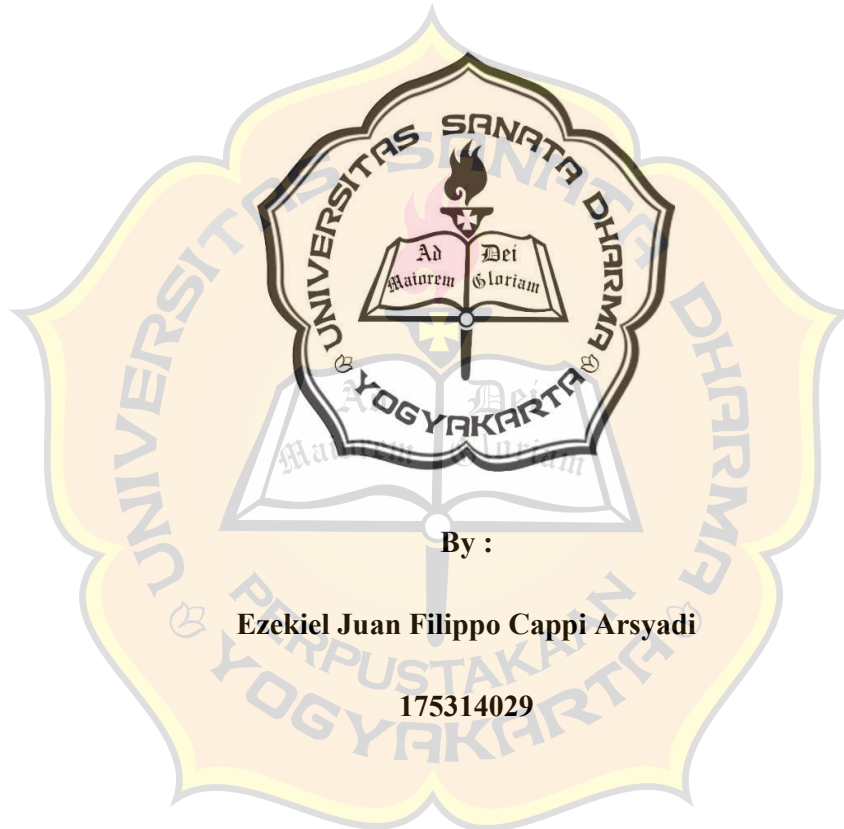
YOGYAKARTA

2022

***SENTIMENT ANALYSIS OF TWITTER NETIZEN'S COMMENTS DURING
THE COVID-19 PANDEMIC ON @TELKOMSEL ACCOUNT USING K-
NEAREST NEIGHBOR METHOD***

THESIS

**Presented as Partial Fulfillment of Requirements For the Degree of Sarjana
Komputer in Informatics Study Program**



By :

Ezekiel Juan Filippo Cappi Arsyadi

175314029

INFORMATICS STUDY PROGRAM

FACULTY OF SCIENCE AND TECHNOLOGY

SANATA DHARMA UNIVERSITY

YOGYAKARTA

2022

ABSTRAK

Pada bulan Maret 2020, Indonesia dilanda pandemi *covid-19*. Untuk mencegah penyebaran virus ini, maka hampir seluruh aktivitas masyarakat dilakukan secara daring atau via internet. *Telkomsel* merupakan salah satu provider *internet* yang paling banyak digunakan oleh masyarakat Indonesia. Provider telekomunikasi ini tengah dihujani banyak komentar dari *customer telkomsel* ditengah pandemi *covid-19* melalui media sosial *twitter*. Oleh karena itu akan sangat membantu apabila dilakukan suatu *sentiment analysis* sebagai salah satu cara untuk membantu mengevaluasi layanan dari *telkomsel* dengan cara mengklasifikasi *tweet customer* menjadi positif dan negatif. Penelitian ini membandingkan metode *cosine similarity* dan *euclidean distance* pada metode klasifikasi *K-Nearest Neighbor* untuk mendapatkan akurasi yang terbaik. Proses *sentiment analysis* pada penelitian ini meliputi *labeling* data secara manual, *pre-processing*, perhitungan bobot *term frequency-inverse document frequency* dan pembagian data menggunakan *k-fold cross validation*. Hasil akurasi paling optimal yang dapat dihasilkan oleh algoritma *K-Nearest Neighbor* menggunakan metode *cosine similarity* adalah 83,456% dengan *precision* 94,217%, *recall* 87,658%, dan *f1-score* 90,819% pada kondisi 3 *k-fold* dan jumlah tetangga $k=9$. Sedangkan hasil akurasi paling optimal yang dapat dihasilkan menggunakan metode *euclidean distance* adalah 76,6% dengan *precision* 90,548%, *recall* 82,271%, dan *f1-score* 86,211% pada kondisi 5 *k-fold* dan jumlah tetangga $k=3$. Berdasarkan hasil perbandingan diatas maka dapat disimpulkan bahwa pada percobaan dengan jumlah tetangga $k=1$ sampai dengan $k=9$, metode perhitungan similaritas *cosine similarity* lebih baik dibandingkan dengan metode perhitungan jarak *euclidean distance* pada model *K-Nearest Neighbor* dalam mengklasifikasi komentar *netizen twitter* pada akun *@Telkomsel* pada masa *covid-19*.

Kata kunci : *Tweet*, Klasifikasi, *K-Nearest Neighbor*, *Sentiment analysis*.

ABSTRACT

In March 2020, Indonesia was hit by the COVID-19 pandemic. To prevent the spread of this virus, most of people activities are carried out online. Telkomsel is one of the most internet providers that frequently used by the people of Indonesia. This telecommunications provider gets a lot of comments from their customers in the midst of the COVID-19 pandemic through twitter social media. Therefore, it will be very helpful if a sentiment analysis is carried out as a way to help evaluate Telkomsel's services by classifying customer tweets into positive and negative. This study compares the cosine similarity and euclidean distance methods on the K-Nearest Neighbor classification method to get the best accuracy.. The sentiment analysis process in this study are manual data labeling, pre-processing, calculation of term frequency-inverse document frequency weights and data split using k-fold cross validation. The highest accuracy results that can be produced by the cosine similarity method are 83.456% with precision 94.217%, recall 87.658%, and f1-score 90.819% on condition 3 fold and the number of neighbors $k=9$. The highest accuracy results that can be produced using the euclidean distance method are 76.6% with precision 90.548%, recall 82.271%, and f1-score 86.211% on condition 5 fold and the number of neighbors $k = 3$. The conclusion is that for $k=1$ until $k=9$, the cosine similarity calculation method is better than the Euclidean distance calculation method on the K-Nearest Neighbor model in classifying Twitter netizen comments on the @Telkomsel account during the covid-19 period.

Keywords : Tweet, Classification, K-Nearest Neighbor, Sentiment analysis.