

## ABSTRAK

Manuaba, Ida Bagus Nyoman Wijana (2020), *Analisis Sentimen Data Provider Layanan Internet Pada Twitter Menggunakan Support Vector Machine Dengan Penambahan Algoritma Levenshtein Distance*, Tesis, Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha

Tesis ini sudah disetujui dan diperiksa oleh Pembimbing I: Dr. Gede Rasben Dantes, S.T, M.T.I. dan Pembimbing II: Dr. Gede Indrawan, S.T., M.T.

*Kata-kata kunci:* Analisis Sentimen, *Support Vector Machine*, *Levenshtein Distance*, *Text Preprocessing*

Twitter menjadi salah satu media sosial yang banyak digunakan di Indonesia. Pengguna twitter dapat dengan bebas memberikan berbagai komentar yang mengandung opini terkait suatu objek atau topik. Selain itu twitter juga menyediakan *API* yang dapat digunakan para *developer* untuk memperoleh data dan mengolahnya untuk kepentingan pengembangan atau penelitian. Banyaknya dokumentasi dan kemudahan penggunaan *API* jika dibandingkan dengan media sosial lainnya, menjadi salah satu pertimbangan digunakannya twitter sebagai data penelitian. Analisis sentimen dilakukan terkait provider atau penyedia layanan internet yang ada di Indonesia menggunakan *Support Vector Machine (SVM)* dan penambahan algoritma *Levenshtein Distance* dengan tujuan memperbaiki kesalahan kata pada tahap *text preprocessing* untuk mendapatkan hasil klasifikasi dengan kategori positif atau negatif. Proses klasifikasi melalui serangkaian tahapan meliputi, pengumpulan data twitter, penghapusan *duplicate* data, pemberian label data, tahap *text preprocessing (convert emoticon, cleansing, case folding, stemming, stopword removal, and tokenizing)*, penerapan algoritma *Levenshtein Distance*, *stopword removal lanjutan convert negation*, *feature extraction (TF-IDF)*, serta proses klasifikasi menggunakan *SVM*. Hasil pengujian dengan menggunakan *confusion matrix*, menunjukkan peningkatan hasil klasifikasi yang lebih baik setelah menggunakan *algoritma Levenshtein Distance* pada tahap *text preprocessing*. Nilai *accuracy* mengalami peningkatan sebesar 2%, *recall* positif 3%, *recall negatif* 1%, *precision* positif 1%, dan *precision negatif* 2%. Tetapi kecepatan waktu proses pada tahap *text preprocessing* dengan penambahan *algoritma Levenshtein Distance* lebih lambat sebesar 295,606 detik, jika dibandingkan tanpa adanya penambahan *algoritma Levenshtein Distance*. Peningkatan waktu proses karena adanya perbaikan kata dengan membandingkan antara kata yang tidak ada dalam kamus (*undefined word*) dengan kata pada kamus yang digunakan, dan mencari nilai minimum operasi perubahan jarak antara kata yang dibandingkan. Untuk pengembangan penelitian selanjutnya, dapat melakukan pengecekan kembali kata (*term*) hasil perbaikan algoritma *Levenshtein Distance* yang menghasilkan lebih dari satu *term* dengan bobot yang sama pada seluruh *term* pada komentar, jadi *term* yang merupakan *term* dengan *frekuensi* atau kemunculan yang paling banyak sehingga diharapkan dapat meningkatkan hasil klasifikasi. Selain itu penggunaan *kernel trick* yang ada pada SVM dapat dibandingkan untuk mengetahui perbedaan hasil klasifikasi.

## ABSTRACT

Ida Bagus Nyoman Wijana Manuaba (2020), Sentiment Analysis Data Provider Of Internet Services On Twitter Using Support Vector Machine With Additional Levenshtein Distance Algorithm. Thesis, Computer Science, Post Graduate Study Program, Ganesha University of Education

This thesis has been supervised and approved by Supervisor I: Dr. Gede Rasben Dantes, S.T, M.T.I. and Supervisor II: Dr. Gede Indrawan, S.T., M.T.

*Key Words:* Sentiment Analysis, Support Vector Machine, Levenshtein Distance, Text Preprocessing

Twitter has become one of social media used by many people in Indonesia. Twitter users can freely give any comments including opinions related to an object or topic. Twitter also provides an API that developers can use to obtain data and process it for development or research purposes. The amount of documentation and ease of use of the API when compared to other social media are one of the considerations for using Twitter as research data. Sentiment analysis is carried out regarding internet service providers in Indonesia using Support Vector Machine (SVM) and the addition of the Levenshtein Distance algorithm with the aim of correcting word errors at the text preprocessing stage to obtain classification results in positive or negative categories. The classification process includes a sequence of phases like crawling using twitter's APIs, erasing duplicated data, data labeling, text processing (convert emoticon, cleansing, case folding, stemming, stopword removal, and tokenizing, Levenshtein Distance algorithms implementation, stopword removal continuation, convert negation), feature extraction (*TF-IDF*), and also the classification process using SVM. The test results using confusion matrix, show an increase in better classification results after using the Levenshtein Distance algorithm at the text preprocessing stage. The accuracy value has increased by 2%, positive recall 3%, negative recall 1%, positive precision 1%, and negative precision 2%. But the processing time at the text preprocessing stage with the addition of the Levenshtein Distance algorithm is slower by 295,606 seconds, when compared without the addition of the Levenshtein Distance algorithm. Increased processing time due to word improvements by comparing the words that are not in the dictionary (undefined words) with the words in the dictionary used, and looking for the minimum value of the operation of changing the distance between the words being compared. For the development of research, it can re-check the word results of the improvement of the Levenshtein Distance algorithm which produces more than one term with the same cost or weight value on all terms in all documents, so the term used as a result of improvement is a term with frequency or the most appearances and are expected to improve classification results. In addition, the use of kernel tricks in the SVM can be compared to determine the differences in classification results. In addition, the use of kernel tricks in SVM can be compared to determine the differences in classification results.