

## **ABSTRACT**

### **COMPARISON BETWEEN FASTTEXT, SKIP GRAM, AND CBOW MODEL USING SVM CLASSIFIER ON INDONESIAN HOTEL REVIEW**

Salsabila Saraswati  
16/395997/PA/17338

Hotel reviews become crucial when it comes to deciding which hotel to book, online hotel review is more dependable since fellow hotel customers honestly review it based on their real experience. Hotel management can take advantage of this new trend but still finds it challenging to evaluate the data since the online reviews are real time. The new one will keep frequently appearing in an enormous amount of it.

This research of Indonesian hotel review using supervised learning are done using SVM to classify the online review whether the review is positive or negative. So the hotelier take it as feedback to improve their quality of service.

Feature extraction method used in this research can give additional information, making it easier for classifiers to learn the data by transforming textual data to numerical vector representation. This research uses Word2Vec and FastText to extract text features and SVM to classify Indonesian hotel reviews. Grid Search cross-validation is also used to find the best SVM hyper-parameters for each given feature extraction methods.

Using dataset consisting of 2976 positive hotel review and 2024 negative hotel review, evaluation results on 20% testing data shows that SVM has the best performance when using features generated by Word2Vec CBOW with accuracy of 88%, precision of 87%, recall of 87%, and F1 Score of 87%.

**Keywords:** Hotel Review, FastText, CBOW, Skip-gram, Text Classification, SVM, Feature Extraction, Word Embedding

## ABSTRAK

### PERBANDINGAN FASTTEXT, SKIP GRAM, DAN CBOW MODEL MENGUNAKAN SVM CLASSIFIER PADA REVIEW HOTEL DALAM BAHASA INDONESIA

Salsabila Saraswati  
16/395997/PA/17338

Ulasan hotel menjadi sangat penting ketika harus memutuskan hotel mana yang akan dipesan, ulasan hotel *online* lebih dapat diandalkan karena sesama pelanggan hotel mengulasnya dengan jujur berdasarkan pengalaman nyata mereka. Manajemen hotel dapat memanfaatkan tren baru ini, tetapi masih merasa kesulitan untuk mengevaluasi data karena ulasan *online* dilakukan dalam waktu nyata. Yang baru akan terus muncul dalam jumlah besar.

Penelitian *review* hotel Indonesia dengan menggunakan *supervised learning* ini dilakukan dengan menggunakan SVM untuk mengklasifikasikan *review online* apakah *review* tersebut positif atau negatif. Jadi pihak hotel dapat memanfaatkan *review* tersebut untuk meningkatkan kualitas layanan mereka.

Metode ekstraksi fitur yang digunakan dalam penelitian ini dapat memberikan informasi tambahan, sehingga memudahkan pengklasifikasi untuk mempelajari data dengan mengubah data tekstual menjadi representasi vektor numerik. Penelitian ini menggunakan Word2Vec dan FastText untuk mengekstraksi fitur teks dan SVM untuk mengklasifikasikan ulasan hotel Indonesia. Validasi silang Grid Search juga digunakan untuk menemukan *hyperparameter* SVM terbaik untuk setiap metode ekstraksi fitur yang diberikan.

Menggunakan set data yang terdiri dari 2976 *review* hotel positif dan 2024 *review* hotel negatif, hasil evaluasi pada 20% data pengujian menunjukkan bahwa SVM memiliki kinerja terbaik ketika menggunakan fitur yang dihasilkan oleh Word2Vec CBOW dengan akurasi 88%, presisi 87%, recall 87% , dan Skor F1 sebesar 87%.

**Kata kunci:** *Hotel Review, FastText, CBOW, Skip-gram, Text Classification, SVM, Feature Extraction, Word Embedding*