



## INTISARI

### ANALISIS SENTIMEN ENTITAS PADA TWITTER BERBAHASA INDONESIA MENGGUNAKAN SUPPORT VECTOR MACHINE

Dimas Tri Novli Ananto

13/356740/PA/15744

*Twitter* adalah salah satu media sosial yang menghasilkan banyak data berupa status, berita, artikel dan lain sebagainya. Data *Twitter* memiliki entitas tertentu, diantaranya nama orang, lokasi dan organisasi. Selain memiliki entitas, status di *Twitter* terkadang juga mengandung sentimen tertentu, diantaranya positif, negatif dan netral. Analisis Sentimen yang terkandung dalam data *twitter* dapat dijadikan acuan oleh masyarakat dalam memberikan penilaian terhadap seseorang, lokasi dan suatu organisasi serta menilai akurasi dari hasil proses klasifikasi yang dilakukan sistem.

Penelitian ini menggunakan metode *Hidden Markov Model* (HMM) untuk melakukan pengenalan entitas dan metode *Support Vector Machine* (SVM) untuk melakukan analisis sentimen hasil dari pengenalan entitas.

Pengujian sistem dengan menggunakan *10-fold cross validation* untuk menentukan entitas dengan HMM menghasilkan rata-rata akurasi 90,09% dan *10-fold cross validation* untuk analisis sentimen dengan SVM menghasilkan rata-rata akurasi tertinggi 71,64% pada kernel Gaussian RBF Gamma 5 dengan 100 atribut. Sedangkan pengujian dengan data uji menggunakan HMM menghasilkan akurasi tertinggi sebesar 96,75%. Untuk analisis sentimen menggunakan SVM menunjukkan bahwa penggunaan kernel dan jumlah atribut yang berbeda tidak memberikan pengaruh yang besar terhadap hasil. Jumlah atribut sebanyak 500 menghasilkan nilai akurasi tertinggi dengan nilai 79,26% pada algoritma *binary occurrence* kernel Gaussian Gamma 0.5 dengan 500 atribut.

**Kata kunci** : pengenalan entitas, analisis sentimen, *hidden markov model*, *support vector machine*, *10-fold cross validation*.



## ABSTRACT

### ***SENTIMENT ANALYSIS OF ENTITIES IN TWITTER INDONESIAN LANGUAGE USING SUPPORT VECTOR MACHINE***

Dimas Tri Novli Ananto

13/356740/PA/15744

Twitter is a on of social media that contain a certain data including news, article, and status. That kind of data contain information which can be send a positive, negative, or neutral sentiment. Sentiment analysis in twitter data can be used as a guide by society to give a rating to people, location, or organization. This rating can also evaluating the accuracy of the result of classification process performed by the system.

This research used Hidden Markov Model (HMM) method to named-entity recognition and Support Vector Machine (SVM) method to analyze the sentiment on its entity.

The examination system process are using 10-fold cross validation to determine the information with HMM, in which produce 90,0% accuration and 10-fold cross validation to analysis information with SVM produce 71,46% accuration in kernel Gaussian RBF Gamma 5 with 100 attributes. However, by using HMM, the highest accuration arround 96,74%. For information analysis with SVM shows that kernel using and with different atribute does not give impact to the result. With 100 attributes, produce 79,26% accuracy in binary occurence algorithm Gaussian RBF Gamma 0.5 kernel shows with 500 attributes.

**Keyword:** named-entity recognition, sentiment analysis, *hidden markov model*, *support vector machine*, *10-fold cross validation*.