

## INTISARI

### Identifikasi Suporter Tim Sepakbola Menggunakan *Supervised Learning* dan *Keyword Filtering* pada Data Twitter

Oleh

Ersyaf Ikhsanul Fikri  
14/371694/PPA/04597

Sepakbola telah menjadi daya tarik dimana kekuatannya dapat menarik jutaan pasang mata di dunia. Tidak bisa dipungkiri sepakbola menjadi olahraga yang sangat populer dan menarik untuk ditonton. Membahas sepakbola tidak lengkap apabila tidak membahas pendukungnya karena pendukung sepakbola membawa euforia berbeda di dalam suatu pertandingan. Di sisi lain dengan adanya kemajuan teknologi yang pesat, kebiasaan masyarakat untuk mengutarakan dukungan terhadap tim sepakbola ditunjukkan melalui komentar yang diutarakan melalui sosial media, salah satunya adalah Twitter. Indonesia menduduki peringkat kelima pengguna twitter di dunia. Dengan jumlah *tweet* mencapai 29 juta setiap jam-nya. Penelitian ini melakukan pengidentifikasian suporter menggunakan algoritma *Naïve Bayes Classifier* berdasarkan *tweet*.

Penelitian ini menggunakan algoritma *Naïve Bayes Classifier* sebagai model sistem untuk mengklasifikasikan *tweet* positif sebagai bentuk dukungan suporter kepada sebuah tim sepakbola dengan penambahan *keyword filtering* tagline untuk setiap tim sepakbola dalam penentuan *userId tweet* sebagai suporter. Hasil dari proses pengklasifikasian dan *keyword filtering* kemudian di visualisasi menggunakan peta Indonesia untuk menunjukkan banyaknya suporter terhadap tiap tim sepakbola. Seleksi fitur pada teks yang digunakan adalah *term-frequency* dan *term frequency-IDF*.

Hasil penelitian diperoleh jumlah suporter dari tim Manchester United sebanyak 168 ID yang teridentifikasi lokasinya, sedangkan 241 ID yang tidak teridentifikasi lokasinya; tim Barcelona 60 ID yang teridentifikasi lokasinya, sedangkan 138 yang tidak teridentifikasi lokasinya; dan tim Real Madrid 140 ID yang teridentifikasi lokasinya, sedangkan 163 ID yang tidak teridentifikasi lokasinya. Penggunaan algoritma *Naïve Bayes Classifier* dengan ekstraksi fitur *term-frequency* pada penelitian ini diperoleh rata-rata akurasi sebesar 70% sedangkan menggunakan ekstraksi *term frequency-IDF* didapatkan akurasi sebesar 69,7%.

Kata Kunci: *Naïve Bayes Classifier*, *SVM*, *TF-IDF*, *Sentiment*, *Tweet*.

## ABSTRACT

### *Football Team s Identification using Supervised Learning and Keyword Filtering on Twitter Data*

By

Ersyaf Ikhsanul Fikri  
14/371694/PPA/04597

Football has become an attraction where its power can attract millions of pairs of eyes in the world. Then it can not be denied soccer become a very popular sport and interesting to watch. Discussing football is incomplete if it does not address his s. Because football fan euphoria brought different in a game. On the other hand with the rapid advances in technology, people's habits to express support for the football team is shown through the comments expressed via social media, one of them is Twitter. Indonesia ranked fifth in the world of Twitter users. With the amount of 29 million *tweets* every hour of his. In this study, providing identification of s using Naïve Bayes classifier algorithm based on the *tweet*.

In this study modeled the system using Naïve Bayes classifier algorithm to classify the positive *tweets* as a form of support of fans to a football team with the addition of keyword filtering using a tagline to every football team in determining the *userId tweet* as s. The results of the classification and keyword filtering then the visualization using Indonesia map to indicate the number of fans to every football team. Selection feature on the text used is term-frequency and frequency-IDF term.

The results were obtained an average accuracy of methods Naïve Bayes classifier with term-frequency feature extraction by 70% while using frequency-IDF term extraction obtained an accuracy of 69.7%. To test the accuracy of the classification, Naive Bayes classifier algorithm than the algorithm Support Vector Machine. The test results an average accuracy with Support Vector Machine algorithm uses the term-frequency extraction obtained an accuracy of 74%, while using frequency-IDF term extraction obtained the same results.

Keyword: Naïve Bayes Classifier, SVM, TF-IDF, Sentiment, *Tweet*.