

## ABSTRAK

Sepak bola, sebagai olahraga paling populer di dunia, menarik minat jutaan penggemar yang tersebar di berbagai negara. Komunikasi antara klub sepak bola dan penggemar sering terjadi melalui media sosial, terutama Instagram, sebuah platform berbagi foto dan video. Studi ini berfokus pada Chelsea Football Club (CFC), yang memiliki basis penggemar besar, khususnya di Indonesia. Penelitian ini bertujuan untuk menganalisis sentimen komentar pada akun Instagram resmi klub (@chelseaindo) menggunakan metode K-Nearest Neighbor (KNN). Melalui analisis sentimen ini, diharapkan dapat terungkap pola pendapat penggemar terhadap performa CFC. Hasil analisis ini memberikan wawasan berharga mengenai persepsi dan emosi penggemar terhadap klub, serta dapat berkontribusi dalam pengelolaan hubungan klub dengan penggemar di Indonesia. Data diperoleh melalui web scraping dari Instagram, dengan total 21.456 komentar. Proses penelitian mencakup preprocessing data, labeling menggunakan Vader dan TextBlob, dan pengelompokan sentimen menjadi positif, negatif, dan netral. Untuk penyeimbangan data digunakan metode SMOTE, dan seleksi fitur dilakukan menggunakan chi-square. Dalam perbandingan performa, hasil paling optimal diperoleh dari Labeling TextBlob yang menggunakan algoritma KNN. Seleksi fitur menggunakan skor chi-square menunjukkan bahwa akurasi tertinggi dari model KNN adalah 87,82% dengan precision 87,90%, recall 87,82%, dan F1 score 87,77%. Kondisi optimal ini dicapai dengan nilai k-tetangga sebesar 3, nilai k-fold sebesar 9, serta menggunakan 360 fitur, serta nilai *k-best* 360.

Kata Kunci : *Chelsea Football Club(CFC), Analisis Sentimen, Instagram. @chelseaindo, Web Scraping, VADER, TEXTBLOB, K-Nearest Neighbor, K-FOLD, Akurasi, Precision, Recall, dan F1 Score, Chi-Square.*

## ABSTRACT

Football, as the most popular sport in the world, attracts millions of fans spread across various countries. Communication between football clubs and fans often occurs via social media, especially Instagram, a photo and video sharing platform. This study focuses on Chelsea Football Club (CFC), which has a large fan base, especially in Indonesia. This research aims to analyze the sentiment of comments on the club's official Instagram account (@chelseaindo) using the K-Nearest Neighbor (KNN) method. Through this sentiment analysis, it is hoped that patterns of fan opinion regarding CFC's performance can be revealed. The results of this analysis provide valuable insight into fans' perceptions and emotions towards clubs, and can contribute to the management of club relationships with fans in Indonesia. Data was obtained through web scraping from Instagram, with a total of 21,456 comments. The research process includes data preprocessing, labeling using Vader and TextBlob, and grouping sentiment into positive, negative, and neutral. To balance the data, the SMOTE method was used, and feature selection was carried out using chi-square. In performance comparison, the most optimal results were obtained from TextBlob Labeling which uses the KNN algorithm. Feature selection using the chi-square score shows that the highest accuracy of the KNN model is 87.82% with precision 87.90%, recall 87.82%, and F1 score 87.77%. This optimal condition is achieved with a k-neighbor value of 3, a k-fold value of 9, and using 360 features, as well as k-best 360 value.

*Keywords: Chelsea Football Club (CFC), Sentiment Analysis, Instagram. @chelseaindo, Web Scraping, VADER, TEXTBLOB, K-Nearest Neighbor, K-FOLD, Accuracy, Precision, Recall, and F1 Score, Chi-Square*