



INTISARI

KLASTERING DATA TWITTER KEJAHATAN PENCURIAN MENGGUNAKAN METODE k-MODES

Oleh

Yant Parama Mulya
12/331346/PA/14610

Survei merupakan metode pengumpulan data yang umum digunakan dalam proses penelitian. Namun, apabila data yang dikumpulkan berupa kejahatan pencurian di Indonesia, akan memakan banyak biaya dan waktu yang lama. Pengumpulan data akan lebih cepat dan murah jika dilakukan dengan *crawling* dan menggunakan *text mining* untuk memperoleh atribut yang berkaitan dengan pencurian. Selain itu, untuk memperoleh informasi yang maksimal metode klastering harus disesuaikan dengan bentuk data dari hasil *text mining*.

Penelitian ini menggunakan data kategorik. Data didapat dengan memanfaatkan REST API twitter untuk mencari tweet yang berkaitan dengan kejadian pencurian di Indonesia. Data melalui tahap preproses untuk mendapatkan 4 atribut kategori yaitu kategori waktu dengan 4 sub kategori, hari dengan 7 sub kategori, objek pencurian dengan 54 sub kategori, dan lokasi dengan 34 sub kategori. Metode klastering yang digunakan adalah metode *k-modes*. Metode *k-modes* merupakan metode modifikasi dari *k-mean* yang khusus menangani data kategorik. Selanjutnya, metode *silhouette coefficient* digunakan untuk menentukan jumlah klaster yang optimal.

Hasil yang diperoleh dari penelitian ini adalah *k*-klaster optimal ketika *k* bernilai 6 dengan data *input* sejumlah 3265. Jumlah *k* tersebut ditentukan dari nilai akhir *silhouette coefficient* terbaik dengan nilai 0,1175 dan nilai *silhouette coefficient* perklaster yang bernilai positif. Selanjutnya, jumlah iterasi yang dilakukan dalam proses klastering *k-modes* sebanyak 3 kali.

Kata kunci: *k-modes*, *silhouette coefficient*, kejahatan pencurian, twitter



ABSTRACT

CLUSTERING DATA THEFT CRIMES BASED ON TWITTER USING k-MODES METHOD

By

Yant Parama Mulya
12/331346/PA/14610

Surveys are data collection methods commonly used in the research process. However, if the data collected in the form of theft crimes in Indonesia, it will take a lot of time and money. Data collection will be faster and cheaper if done by crawling and using text mining to obtain attributes related to theft. In addition, to obtain maximum information, the clustering method must be adjusted to the data form from the results of text mining.

This study uses categorical data. Data is obtained by utilizing the twitter REST API to search for tweets related to theft in Indonesia. Data through the preprocessing stage to get 4 category attributes, namely the time category with 4 sub-categories, days with 7 sub-categories, object theft with 54 sub-categories, and locations with 34 sub-categories. The clustering analysis used is the k-modes method. The k-modes method is a modified method of k-mean that specifically handles categorical data. Then the silhouette coefficient method is used to determine the optimal number of clusters.

The results obtained from this study are optimal k-cluster when k is 6 with input data of 3265. The number of k is determined from the final value of the best silhouette coefficient with a value of 0.1175 and the value of the silhouette coefficient of the cluster is positive. Furthermore, the number of iterations performed in the clustering process k-modes 3 times.

Keywords: k-modes, silhouette coefficient, theft crime, twitter