1. INTRODUCTION

Identifying user behavior is a research activity from data or information showing a person's behavior. Identification of user behavior aims to determine the various problems to be studied. The user's behavior can be seen from the attitude shown towards something. For example, the attitude shown in tweet uploads can be positive (compliments or good sentences) or negative (blasphemy against someone or the user's state is anger). With online social media, users can increase their activities, especially regarding various kinds of information [1].

Twitter is a platform for getting information from anywhere and in any field [2]. The platform is considered a micro-blogging system so users can share what they think in 140 characters [3], [4]. Sometimes, the user creates a status message (tweet) with various information. Usually, users on the platform always share various kinds of things that are positive (praise) or negative (blasphemy) from some circulating information [5]. Sometimes users also spam on the Twitter platform to give their opinion regarding cases that are trending at the time [6], [7]. Therefore, the identification ofuser behavior requires a dataset.

One exciting piece of research related to this Twitter dataset is user behavior analysis. The problems that exist on the Twitter platform are very diverse. This study chose the political field that had been trending some time ago. The problem started when the President of Indonesia and his staff's performances was not sufficiently able to control several cases in Indonesia, so the Indonesian people staged a demonstration to ask the President of Indonesia to resign during the current period. The political field was chosen because it is still infrequent in other studies related to user behavior identification. Various tweet comments, such as positive (praise) or negative (blasphemous) tweets, appeared on Twitter until it became a trending topic in the hope that the president heard their voices. These positive and negative tweets can be interpreted as user behavior. From the mood of the user's tweet, we can determine whether the user's behavior is positive or negative [8] so that the user's mood is happy or angry.

Datasets can be obtained through data crawling. This way, we can find information from a large dataset needed to identify user behavior. After getting a dataset, such as research [3], [5] performs data preprocessing so that the data obtained is clean of missing values and becomes efficient data to get accurate results or good performance values. Furthermore, several studies on the identification of user behavior on the Twitter platform have been carried out with various methods related to this research. As in research [1], [9], [10], [11] to identify user behavior with the clustering method can use the Birch algorithm, K-Means, Agglomerative Hierarchical, and others as needed. In addition to these methods, for identification, classification methods can be used [3], [5], [12] with the Naive Bayes algorithm, or other methods such as J48, SVM, Fuzzy Analytic Hierarchy, and Logistic Regression [13], [14]. Process Identification of user behavior can use the calculation of centrality or similarity [1], [15], [16]. With these calculations, analyzing user behavior from several groups (groups can be obtained in the clustering method) will be more accessible and display network visualizations. However, research [17] explains that user behavior identification can be made manually by viewing or collecting data only. This will take a very long time because to group the analyzed discussion, you have to check each tweet and user manual.

The drawbacks of previous research are several steps in identifying incomplete user behavior or using the same method, many of which use the clustering method with the K-Means algorithm. This study uses a clustering

method with a different algorithm from previous studies, namely Mean Shift Clustering. The algorithm was chosen because it is still rarely used in previous research. In addition, what distinguishes Mean Shift from other algorithms is that the number of clusters is determined automatically using bandwidth, but this does not guarantee that the number of clusters produced will be optimal [18]. In addition to the clustering method, the centrality calculation is also carried out so that the identification process of user behavior becomes more effortless.

This research phase begins with collecting data, data preprocessing to become efficient data, and calculating word weights using the TF-IDF Vectorizer. Then use the clustering method with Mean Shift algorithm. In addition, network analysis was also carried out, consisting of checking the number of users (nodes), relationships between users (edges), and calculating centrality. Then perform network visualization to identify each cluster obtained. The last stage is analyzing the user's behavior [1]. This study purpose to analyze the types of user behavior on Twitter data in the political field and to analyze the study results using the Mean Shift Clustering algorithm.