1. INTRODUCTION

The relocation of the Indonesian capital from Jakarta to Kalimantan as the Ibu Kota Nusantara (IKN) is carried out to accelerate Indonesia's economic equality, reduce development inequality, and create jobs [1]. The move is also expected to ease the burden on Jakarta, including welfare, health, and environmental issues such as congestion, air pollution, flooding, clean water shortages, crime, and land subsidence[2] [3]. The relocation plan has been under discussed the era of since Indonesia's first president Soekarno's era and continued until President Susilo Bambang Yudhoyono's era, and finally received serious attention under the leadership of President Joko Widodo's [4]. The process of relocating the capital also involves several stages of good planning, such as feasibility studies, infrastructure planning, public consultation, and stakeholder involvement [5]. While the relocation of the capital city was well intentioned, it still triggered many public responses and discussions on support and opposition, as the project was underway. This indicates the necessity of comprehending public sentiment, as it exerts a substantial influence on the public's perspective regarding the decision to relocate the capital city[6]. With this understanding, the government can gain valuable insights to predict public views, craft more appropriate responses, prevent potential conflicts, and make wiser decisions in the future [7]. Therefore, this research is conducted to analyze the public sentiment related to the policy of relocating Indonesia's capital policy using sentiment analysis based on machine learning.

In the digital era, social media has become the primary means of communication for individuals to express their views and opinions [8][9]. One such platform is X, formerly known as Twitter. X is a social media that allows its users to write and express opinions or actions through tweets [10][11]. The hashtag and keyword features help users find the latest information on trending topics [12][13]. This research will therefore utilise tweet data on the X platform to explore information, understand opinions, and public reactions to IKN policies.

The background of the problem, to find out and analyze tweets related to IKN, sentiment analysis research is carried out. Sentiment analysis is a technique in Natural Language Processing (NLP) that employs machine learning to identify and categorize emotions, opinions, or perspectives in text as negative [14][15][16]. This research uses Logistic Regression (LR), a common machine learning method for text classification and sentiment analysis [17]. Logistic Regression is a classification algorithm used for sentiment analysis and produces positive or negative classes [18][19]. Logistic Regression produces The most effective sentiment prediction evaluates accuracy, recall, precision, and F1-score using the N-Gram and TF-IDF methods [20]. The classification approach that results in consistent test results is Logistic Regression [21]. In addition, Logistic Regression provides class probabilities that can help in strengthening sentiment analysis results [22][23].

There are several studies that have conducted sentiment analysis using the Logistic Regression method. Previous research [17] shows that Logistic Regression with TF-IDF using VADER labeling has an accuracy of 85.22%, superior to KNN Textblob 73.64% in sentiment analysis of the Covid-19 vaccine. The study found that the highest accuracy of the three classification models, LR, SVM and MNB, was 86.23% for bigram [23]. Research by Ramadhon et al. [24] sentiment analysis of capital city relocation using N-grams with bigram and K-Nearest Neighbor reached accuracy of 66%. Twitter data on capital city relocation used TF-IDF and three algorithms [25]. The findings revealed that the SVM algorithm exhibited the highest accuracy, at 97.72%, in comparison to Logistic Regression and KNN. Sinha et al. [26] discussed sentiment analysis regarding the conflict between Russia and Ukraine by comparing methods between KNN, Decision Tree and Logistic Regression with the results obtained accuracy of 88.89%, 90.45%, and 94.58%. Research by Wahyuningsih et al. [27] compared LR, NB, and random forest algorithms to predict students' reasoning using data sources from Kaggle, LR has the highest accuracy of 94.34%.

Previous research has discussed sentiment analysis, but no one has specifically focused on the topic of IKN using the Logistic Regression method. Previous studies focused on comparisons between methods or on a single type of simple feature extraction. This study differs from previous studies because this study uses logistic regression by applying three labeling techniques and testing different combinations of data preprocessing and feature extraction methods. This combination is done to explore and identify the most effective approach for generating sentiment classification using Logistic Regression related to IKN topics.

This research is aims to analyze the sentiment of the public over the capital relocation policy of the Indonesian government through the application of machine learning. The present study utilizes a machine learning algorithm, specifically Logistic Regression. During the analysis process, a combination of preprocessing and feature extraction is employed to ascertain the most effective approach for conducting sentiment analysis of public opinion concerning the relocation of the National Capital. It is from this approach that this research is expected to contribute to an understanding of the perception and dynamics of public opinion towards the policy.