INTRODUCTION

Currently, online social networks play a crucial role in managing information by facilitating interaction and information sharing [1]. This has made social networks a recognized source of information [2]. However, the risk of inaccurate information is a significant drawback that must be addressed [3], especially on the X platform, which allow the spread of erroneous information due to inadequate content oversight [4], [5], [6]. As a social media and information source, the X platform enables users to easily access information through tweets, including the latest news, trends, and opinions on current global topics. However, it is important to note that not all information on the X platform can be considered accurate. The credibility of information on this platform is a primary concern in current research, for example after the 2010 earthquake in Chile, when many rumors and unofficial information on the X platform caused anxiety and insecurity among the local population [7].

A study [8] found that users' assessments of information accuracy on the X platform tend to be low. Several studies have been carried out to classify the credibility of information on social media. One of the first studies was done in 2011 by Castillo et al., using machine learning approaches such as Support Vector Machine (SVM), Decision Tree, Bayes Networks and Decision Rules. The focus was to classify the trustworthiness of information on social media X as either credible or not credible. The best results were obtained using the J48 Decision Tree algorithm, achieving an accuracy of 86% [7]. In 2020, Erwin et al. performed similar research to Castillo et al., but they enhanced the analysis by adding combined features from User Profile and Message Content. This study used a dataset from social media X, involving 115 accounts and 19,401 tweets in Indonesian. By incorporating combined features from User Profile and Message Content, Erwin et al. achieved the best results using the J48 Decision Tree algorithm, with an accuracy of 88.42% [9]. The following year, Marina et al. compared algorithms such as SVM, Naïve Bayes, KNN, Random Forest and Logistic Regression to detect information credibility on the X platform. This study used two feature extraction approaches: Content-Based and User-Based. According to the results, the highest accuracy of 83.4% was attained by the Random Forest algorithm [2].

In the next iteration, more studies used machine learning approaches to classify information credibility on social media, particularly the X platform. In 2020, Vyas et al. implemented a deep learning approach using Long Short-Term Memory (LSTM) to assess the performance of the LSTM model in differentiating between trustworthy and untrustworthy news on the X platform. The LSTM approach was found helpful in understanding the continuous representation of microblog events to assess the credibility of information in tweets, classifying them as credible or not. This study used GloVe (Global Vectors) for feature extraction to create word vector representations that include tagging like URLs, hashtags (#), and the "@" symbol. These word vector representations were then used as features in the LSTM model to assess news credibility on the X platform. By using GloVe, this study achieved more accurate word vector representations for tagging features, improving the LSTM model's performance in evaluating news credibility on the X platform. The results showed an accuracy of 81% [10]. In 2023, Fadhli et al. implemented a hybrid deep learning model to detect conversation credibility on the X platform. The employed method was known as CreCDA, a combination of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) models by incorporating post and user features to improve the performance of credibility detection. This approach is aimed to recognize credibility levels in online social interactions efficiently. Additionally, this study used Word2Vec for feature extraction, a neural network model used to learn word vector representations in documents. The accuracy of the hybrid deep learning model reached 81% [11].

In a sentiment analysis study on social media X, Diaz et al. found a 1.1% increase in accuracy by adding semantic features using the Robustly optimized BERT Approach (RoBERTa) [12]. Another study by Anindika et al. focused on detecting hoaxes on social media X, using TF-IDF for feature extraction and GloVe for feature expansion. GloVe was used to find similar words and link them to existing expanded features using a built-in corpus. The results showed that the baseline (TF-IDF N-gram + BERT) combined with the GloVe-expanded corpus achieved a higher accuracy of 98.57%, with a 4.69% improvement over the previous baseline [13]. Previous studies have shown that using methods like TF-IDF, RoBERTa, and GloVe results in better accuracy, even though deep learning approaches are rarely applied. In a study [10], LSTM was used for its ability to evaluate the credibility of information in tweets, classifying them as credible or not. LSTM's advantage lies in its memory cells, outperforming conventional recurrent neural networks [14]. Additionally, because optimization methods have not been applied in research on information credibility classification, optimization methods like Particle Swarm Optimization (PSO) can be used, as they are known to achieve optimal solutions more quickly and easy to implement [15][16]. PSO can enhance the model's accuracy and performance by finding optimal hyperparameters. This was demonstrated in a study by Regina et al., who applied CNN-PSO for sentiment analysis on the social media X. PSO was used to find the optimal hyperparameters for the CNN model, resulting in a 10.07% increase in accuracy [17].

Based on the issues outlined, this research aims to detect the credibility of information on social media X by classifying tweets as credible or non-credible. The approach combines methods proven effective in previous studies to achieve optimal accuracy. The combination includes LSTM as the classification model, TF-IDF as the feature extraction and a baseline, RoBERTa for semantic features, GloVe for feature expansion, and PSO as the optimization method.