Chapter 21
Hyperparameter Optimization of Machine Learning Models Using Grid Search for Twitter Sentiment Analysis

Wesam Ahmed
https://orcid.org/0009-0008-5675-5724
Information Technology Department, Hurghada University, Egypt

Noura A. Semary
Menoufia University, Egypt

Khalid Amin
https://orcid.org/0000-0002-9594-8827
Menoufia University, Egypt

Mohamed Hammad
https://orcid.org/0000-0002-6506-3083
Menoufia University, Egypt

ABSTRACT

Twitter has emerged as a significant social media platform and has garnered significant interest from sentiment analysis researchers. Text mining is an active subfield that includes Twitter sentiment analysis (TSA) research. TSA is the term used to describe the utilization of algorithms to analyze the subjective nature of Twitter data, which includes its sentiments and opinions. The extraction of inferences from user interactions is facilitated by machine learning (ML) approaches. A wide range of machine learning methodologies are employed to analyze emotions. This research compares four supervised machine learning techniques with the term frequency-inverse document frequency (TF-IDF) method and hyperparameter tuning with grid search for the Dell tweets dataset. This classification technique will demonstrate how sentiment analysis is performed using performance metrics.

DOI: 10.4018/979-8-3693-7011-7.ch021

INTRODUCTION

Twitter is a social networking platform that allows users to share their thoughts and opinions, connect with others, and contribute to specific issues through short updates called tweets, limited to 140 characters. Users can engage with each other using various forms of media, such as texts, photographs, and videos (Chaturvedi et al. 2018). They can communicate by utilizing features like liking, commenting, and reposting. As to the information provided by Twitter, the platform currently boasts over 206 million daily active users in 2022. With the increasing number of individuals participating in social media, it is possible to analyze online information to gain insights into the shifts in people's views, behavior, and psychology. Consequently, using Twitter data for sentiment analysis has become more common. Due to the increasing interest in social media analysis, greater emphasis has been placed on the technologies associated with text analysis, such as NLP and Artificial Intelligence (AI) (Valdivia et al. 2018).

Opinion mining is a specialized field of study that concentrates on sentiment analysis within linguistics and natural language processing (NLP). It evaluates the degree of polarity in words and phrases to extract and analyze opinions and emotions from textual data. Sentiment analysis has also been applied to business and social research. Companies such as Microsoft and Google recently developed Sentiment analysis tools to facilitate their business operations. (Alamoodi et al., 2021). The TSA aims to tackle the challenge of assessing the covert connotations of tweets shared on Twitter, which is regarded as a novel domain of sentiment analysis. One of the main obstacles faced by the TSA is the limitation on the size of messages. Several machine learning methods are utilized for sentiment analysis. The efficacy of these categorization algorithms depends on their unique domain (Birjali et al. 2021). This study employed grid search to optimize hyperparameters to enhance the parameters of four supervised machine learning models. This research's primary goal is to ascertain the hyperparameters' most suitable or optimal values to generate the most superior model. It also considers the unpredictability of the computation costs during the training process (Choi et al.2017). Furthermore, the use of varying hyperparameter values facilitates the evaluation of machine learning models. Building machine learning models that are accurate and effective for Twitter sentiment analysis necessitates the adjustment of hyperparameters. A specific group of hyperparameters is explicitly defined in this approach, and the model's performance on a validation set is optimized by determining optimal hyperparameter values through methods such as grid search. Refinement of hyperparameters can enhance machine learning models' generalization and precision capabilities for Twitter sentiment analysis(Cambria et al.2017).

We are primarily interested in enhancing the classification performance of Twitter sentiment analysis in this study. The present analysis aims to make the following contributions in this context:

1. offering a recently collected dataset of Dell tweets that facilitates the examination of customer sentiments and opinions unique to Dell products and services.

2- Improves the accuracy and performance of the models, enabling more precise predictions and enhanced analysis of the Dell tweet dataset through the grid search method.

3. Which machine learning technique achieves the highest level of performance?