Sentiment Analysis of User Comments on Machine Learning Posts: A Data-driven Approach Using R*

Joseph Chung

Apr 2, 2023

Abstract

In this study, we present a comprehensive data analysis paper that utilize a sentiment model in R to analyze user comments from a dataset containing post_ids and comments related to the topic of Machine Learning. We aim to investigate sentiment polarity of user comments and its distribution across different posts, identifying trends and patterns that may provide valuable insights into the Machine Learning community's perception and engagement.

Table of contents

*Code and data are available at: https://github.com/UtopianYoungChung/Sentiment-Analysis-Using-R.git

List of Figures

List of Tables

1 1. Introduction

The rapid growth of online forums and communities discussing various topics, including Machine Learning(ML), has led to an abundance of user-generated content in the form of text data. Analyzing this data can reveal valuable insights into user perceptions, opinions, and sentiment towards specific subjects. Sentiment analysis, or opinion mining, is a natural language processing technique used to determine the sentiment expressed in a piece of text, such as positive, negative, or neutral. This paper focuses on employing sentiment analysis techniques to understand the sentiment polarity of user comments related to the topic of Machine Learning.

1.1 1.1 Background and motivation

Machine Learning has become a popular field of study and research due to its vast applications in various industries. As a result, online communities have emerged where people discuss ML topics, share knowledge, and express their opinions. Analyzing the sentiment of these user comments can provide insights into the perception of the ML community, identify common issues or concerns, and uncover emerging trends. This understanding could be valuable for researchers, educators, and industry professionals who aim to improve their work based on community feedback.

1.2 1.2 Research questions

This study aims to address the following research questions:

1. What is the overall sentiment polarity (positive, negative, or neutral) of user comments related to Machine Learning?
2. How is the sentiment polarity distributed across different posts in the dataset?
3. Are there any correlations between post attributes (e.g., post length, engagement) and sentiment polarity?
4. Can any trends or patterns in sentiment polarity be identified over time or in in relation to specific subtopics within Machine Learning?

1.3 1.3 Scope and limitations

The scope of this study is limited to the analysis of a dataset containing post IDs and comments related to the topic of Machine Learning. The results and interpretations are based solely on the data provided and may not be generalizable to other online communities or topics. The sentiment analysis model employed in this study is subject to the limitations inherent in natural language processing techniques, such as context ambiguity and handling of sarcasm or irony. Additionally, the study does not account for demographic information or biases present in the dataset, which could potentially influence the sentiment analysis results.

2. Literature Review {#sec-literature review} ## 2.1 Sentiment analysis in online communities Over the past decade, sentiment analysis has become increasingly popular as a tool to mine user opinions and emotions from online platforms, such as social media, blogs, and forums (**citePang?**). Numerous studies have focused on the application of sentiment analysis in various domains, such as politics (**citeTumasjan?**), finance (**citeBollen?**), and consumer reviews (**citeDave?**). These studies demonstrate the potential of sentiment analysis as a valuable technique for understanding user-generated content and informing decision-making processes.

1.4 2.2 Sentiment analysis techniques in R

R is a widely used programming language for statistical analysis and data visualization, with several libraries and packages available for natural language processing and sentiment analysis (R Core Team 2020). Some of the most popular sentiment analysis packages in R include tidytext (**citeSilge?**), syuzhet (**citeJockers?**), and sentiment (**citeRinker?**). These packages provide various algorithms and techniques for sentiment analysis, such as lexicon-based approaches, machine learning models, and deep learning methods. Researchers have employed these techniques to analyze sentiment in diverse contexts, ranging from Twitter data (**citeWijeratne?**) to news articles (**citeKearney?**)

1.5 2.3 Applications of sentiment analysis in Machine Learning research

Sentiment analysis has been employed in Machine Learning research to understand user perceptions, opinions, and trends related to the field. For example, Aliza (**citeAliza?**) analyzed the sentiment of tweets related to Machine Learning conferences, revealing insights into the community's responses to specific presentations and events. In another study, Wang (**citeWang?**) investigated the sentiment of user comments on popular Machine Learning platforms, such as GitHub and Stack Overflow, to identify trends and patterns in the community's engagement with specific algorithms and techniques. These studies highlight the potential of sentiment analysis to uncover valuable insights into the Machine Learning community and inform future research directions.

The literature demonstrates the growing interest in sentiment analysis as a valuable tool for understanding user-generated content in various domains. R has emerged as a popular platform sentiment analysis due to its extensive libraries and packages. However, there remains a need for more comprehensive studies that investigate the sentiment of user comments related to Machine Learning, particularly in the context of online communities where users share their knowledge and opinions on the topic. This study aims to address this gap by employing sentiment analysis techniques in R to analyze a dataset of post IDs and comments related to Machine Learning.