Sentiment Analysis On Reddit Data And Analyzing Attacks on Social media

Sahithi NallaniChakravartula (SNC)

# Abstract

In this paper I wanted to do research on a model that classifies user reviews from real-time reddit data into positive, neutral and negative categories using ML models. To do this I need to analyze the Attack and defense methods through some reference papers and give the insights. The growing use of social media has led to the development of several Machine Learning (ML) and Natural Language Processing (NLP) tools to process the unprecedented amount of social media content to make actionable decisions. Sentiment Analysis is the process of computationally identifying and categorizing words expressed in a piece of text, especially in order to determine whether the writer’s attitude towards a particular topic, product, etc. is positive or negative. Any brand’s presence on social networks has a significant impact on emotional reactions of its users to different types of posts on social media. If a company understands the preferred types of posts (photo or video) of its customers, based on their reactions, it could make use of these preferences in designing its future communication strategy. In sentiment analysis, an attacker can launch adversarial attacks by adding small perturbations to text to generate different perceptions than the actual opinions.

## 1 Introduction

Social media posts regarding how people are feeling can be utilized to gain a better knowledge of public health, everyday decision-making, and people's assessments of their quality of life. Evidence reveals that people are more prone to disclose their feelings online, particularly on social media platforms, during times of crisis. Sentiment analysis helps in real-time monitoring of products across public platforms. It helps in understanding customer feelings. Hackers are constantly developing new attacking tools and hacking strategies to gain malicious access to systems and attack social media networks thereby making it difficult for security administrators and organizations to develop and implement the proper policies and procedures necessary to prevent the hackers’ attacks. The increase in cyber-attacks on social media platforms calls for urgent and more intelligent security measures to enhance the effectiveness of social media platforms. A more realistic idea on how well the product is doing ,since we are considering all the possible social media platforms. ( eg: Reddit vs Twitter) The product owners can streamline their promotions / advertisements concentrating majorly on a specific social media platform The limitation is filtering out spam tweets from twitter and reddit because of its ambiguity

## 2 Related Work

Social media adversarial attack refers to a wide range of hostile operations carried out through human interactions on social media. It manipulates users' minds to make them make security mistakes or reveal important information. Attacks on social media can take several forms. To carry out the assault, a perpetrator first examines the intended victim to obtain background information such as possible avenues of entry and weak security mechanisms. The attacker then attempts to acquire the victim's trust and give stimuli for later acts that violate security protocols, such as disclosing sensitive information or granting access to key resources.

### 12 References

1. An Analysis of Security in Social Networks

Weimin Luo1 , Jingbo Liu1 , Jing Liu2

1. Adversarial Attacks and Defenses for Social Network Text Processing Applications: Techniques, Challenges and Future Research Directions Izzat Alsmadia , Kashif Ahmadb , Mahmoud Nazzalc , Firoj Alamd , Ala Al-Fuqahab , Abdallah Khreishahc ,

Abdulelah Algosaibie

1. A Guide to Differential Privacy Theory in Social Network Analysis by Christine Task, Chris

Clifton

1. Evasion attacks against machine learning at test time Battista Biggio1 , Igino Corona1 , Davide

Maiorca1 , Blaine Nelson2 , Nedim Srndi´c ˇ 3