From Raw Data to Informed Decisions: Analyzing Amazon Book Reviews

Alberti A. • Ligari D. •Andreoli C. [1]

[1]*Department of Computer Engineering - Data Science, University of Pavia, Italy*
Course of Data science and big data analytics

Github repository: https://github.com/DavideLigari01/data-science-project

Abstract

Keywords— *DNS reflection and amplification Attacks • Amplification factor • Ping • Wireshark • DIG • Mitigation measures*

Contents

1 Introduction

In the age of digital commerce, customer reviews play a pivotal role in shaping product perception and influencing purchasing decisions. With the proliferation of online bookstores, Amazon has amassed an immense repository of book reviews spanning nearly two decades. These reviews contain valuable insights, sentiments, and trends that can unlock a treasure trove of information for authors, publishers, and book enthusiasts. This project embarks on a journey to harness the power of data, employing a comprehensive workflow to dissect and understand the vast collection of Amazon Books Reviews. Our mission is to develop a scalable solution that allow us to discover patterns, sentiment trends, and hidden correlations within the world of book reviews. We leverage cutting-edge tools and technologies, including Hadoop, Spark, MongoDB, and Python libraries such as Pandas and Scikit-learn. In this report, we embark on a detailed exploration of our project, delving into each stage of our workflow, from initial data discovery and preparation to feature extraction, model building, and rigorous evaluation.