Twitter Sentiment Analysis Using Hadoop Eco-System

SYNOPSIS

Institute of Engineering & Technology

Team Members

Sahil Srivastava
171500281
Samarth Ahuja
171500286
Kanishk
171500153
Shivank Garg
171500325

Supervised By:-

Vinay Agrawal

Asst. Professor

Department of Computer Engineering & Applications

About the Project:

In this project, we will showcase how to perform Sentiment Analysis on Twitter data using Pig using Hadoop Eco-System Framework. To begin with, we will be collecting real-time tweets from Twitter using Flume. With the help of AFINN dictionary we can find positive sentiment from the data which we have dumped into our HDFS/Flume. We can prioritize our own customized words in AFINN dictionary as per the need of the client.

As this project is most advance use case of Hadoop in MNC and has wide impact in sentiment analysis. We need most advance concepts of PIG and MAP-REDUCE for getting into this project.

The data from Twitter is in 'JSON' format, so a Pig JsonLoader is required to load the data into Pig.

Name of Pig JsonLoader are :-

- Elephant-bird-hadoop-compat-4.1.jar

- Elephant-bird-pig-4.1.jar

- Json-simple-1.1.1.jar

Execution Module :-

o Invoke PIG grunt shell in map-reduce mode.

o Register all the PIG JSON LOADER with PIG.

o Load AFINN dictionary from HDFS to PIG bag.

Motivation:

The motivation for doing this project was primarily an interest in undertaking a challenging project in an interesting area of Data Science. Our inspiration behind this project was to showcase an efficient and secure form of using data of the users of Twitter to provide them better recommendation, better options to select and cleanliness of vulgarity from their profiles.

Future Prospects:

- Opinion mining can majorly help in discovering hot search keywords. This feature can help the brand in their SEO (Search Engine Optimization). This means that opinion mining will help them make strategies about, how their brand will come up among the top results, when a trending or hot keyword is searched in a search engine.

- It can be used to give your business valuable insights into how people feel about your product brand or service.

- We can predict pre elections polls and exit polls. This will help in checking out the popularity of the leaders among the peoples of a country before the elections.

- This project will impose a barrier on putting a vulgar or abusive comment on any social media platform. It will help the developers to keep an eye on it.

- The benefits of sentiment analysis extend into your bottom line. With sentiment analysis, you can identify a dissatisfied customer as and when they're chatting with your team. This enables your agents to offer a smooth service and quick resolution to appease, and ultimately retain, the customer.

- Sentiment analysis also means you'll be able to detect changes in the overall opinion towards your brand. Because it provides insight into the way your customers are feeling when they approach you, you can monitor trends and see if overall opinion towards your company drops or rises.

- When a company releases a new product or service, it is released as a pilot or beta version. The monitoring of public feedback at this stage is very crucial. So, text mining from social media platforms and review sections greatly helps accelerate this process.

Requirements:

a) **Hardware:**
   - Minimum 4GB RAM
   - Minimum Core i3
   - Minimum 60GB ROM

b) **Software:**
   - Cent-OS Operating System
   - Hadoop Framework using Map-Reduce
   - Flumes
   - Apache PIG