

Project Big Data

Sentiments Analysis

Objective:

The objective of this project is to perform **Sentiment Analysis** using the MapReduce programming model. You will analyze a dataset of IMDB reviews and classify the sentiment (positive, negative, neutral) based on the presence of predefined keywords.

Prerequisites:

- Basic understanding of the MapReduce framework.
- Python installed (if using Hadoop streaming) or access to a Hadoop cluster.

Dataset Example

We will use the **IMDB_Review_Dataset.csv**, which contains customer reviews of various movies.

- Each line in the dataset represents a single review.
- Example

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	review													
2	One of the other reviewers has mentioned that after watching just 1 Oz episode you'll be hooked. They are right, as this is exactly what happened with me. The first thing ta													
3	A wonderful little production. The filming technique is very unassuming- very old-time-BBC fashion and gives a comforting, and sometimes discomforting, sense of realism to													
4	I thought this was a wonderful way to spend time on a too hot summer weekend, sitting in the air conditioned theater and watching a light-hearted comedy. The plot is simplistic, but th													
5	Basically there's a family where a little boy (Jake) thinks there's a zombie in his closet & his parents are fighting all the time. This movie is slower than a soap opera... and sudc													
6	Petter Mattei's "Love in the Time of Money" is a visually stunning film to watch. Mr. Mattei offers us a vivid portrait about human relations. This is a movie that seems to be telling us wh													
7	Probably my all-time favorite movie, a story of selflessness, sacrifice and dedication to a noble cause, but it's not preachy or boring. It just never gets old, despite my having seen it some													
8	I sure would like to see a resurrection of a up dated Seahunt series with the tech they have today it would bring back the kid excitement in me.I grew up on black and white TV and Seahu													
9	This show was an amazing, fresh & innovative idea in the 70's when it first aired. The first 7 or 8 years were brilliant, but things dropped off after that. By 1990, the show was not really fu													
10	Encouraged by the positive comments about this film on here I was looking forward to watching this film. Bad mistake. I've seen 950+ films and this is truly one of the worst of them - it's													
11	If you like original gut wrenching laughter you will like this movie. If you are young or old then you will love this movie, hell even my mom liked it. Great Camp!!!													
12	Phil the Alien is one of those quirky films where the humour is based around the oddness of everything rather than actual punchlines. At first it was very odd and pretty funny													
13	I saw this movie when I was about 12 when it came out. I recall the scariest scene was the big bird eating men dangling helplessly from parachutes right out of the air. The horror. The ho													
14	So im not a big fan of Boll's work but then again not many are. I enjoyed his movie Postal (maybe im the only one). Boll apparently bought the rights to use Far Cry long ago even before t													
15	The cast played Shakespeare. Shakespeare lost. I appreciate that this is trying to bring Shakespeare to the masses, but why ruin something so good. Is													
16	This is a fantastic movie of three prisoners who become famous. One of the actors is george clooney and I'm not a fan but this roll is not bad. Another good thing about the movie is the so													
17	Kind of drawn in by the erotic scenes, only to realize this was one of the most amateurish and unbelievable bits of film I've ever seen. Sort of like a high school film project. What was Ros													
18	Some films just simply should not be remade. This is one of them. In and of itself it is not a bad film. But it fails to capture the flavor and the terror of the 1963 film of the same title. Liam													
19	This movie made it into one of my top 10 most awful movies. Horrible. There wasn't a continuous minute where there wasn't a fight with one monster or another. There was													
20	I remember this film,it was the first film i had watched at the cinema the picture was dark in places i was very nervous it was back in 74/75 my Dad took me my brother & sister to Newbu													
21	An awful film! It must have been up against some real stinkers to be nominated for the Golden Globe. They've taken the story of the first famous female Renaissance painter and mangle													
22	After the success of Die Hard and it's sequels it's no surprise really that in the 1990s, a glut of 'Die Hard on a' movies cashed in on the wrong guy, wrong place, wrong time concept. Th													

Tasks

1. Collect the Sentiment Keywords Dictionary

- **Positive Keywords:**
Examples include *love, amazing, happy, good, excellent, wonderful, fantastic, brilliant, satisfied, success*.
- **Negative Keywords:**
Examples include *horrible, frustrating, terrible, bad, worst, disappointing, awful, regret*.

2. Preprocess the Dataset

- Explore preprocessing techniques to clean the data:
 - Remove irrelevant information, such as **stop words** (e.g., *the, is, and, or*).
 - Strip special characters and punctuation to standardize the text.

3. Create the MapReduce Program

- Implement a program to process the dataset and classify reviews based on the sentiment dictionary.

4. Upload Dataset to HDFS

- Load the IMDB_Review_Dataset.csv file into the Hadoop Distributed File System (HDFS).

5. Run MapReduce Job

- Execute the MapReduce program to classify reviews as **Positive, Negative, or Neutral**.

6. Fetch Results

- Collect the sentiment analysis results from the HDFS output.

7. Display Results in a Graph

- Use visualization tools (e.g., Matplotlib, Excel, or Tableau) to create a bar chart or pie chart showing the count of positive, negative, and neutral reviews.

Analysis Questions

1. Which sentiment dominates the dataset?

Analyze the visualized results to determine whether the majority of reviews are positive, negative, or neutral.

2. What does this imply about the overall sentiment in the IMDB reviews?

Reflect on the broader implications of your findings (Consumer Perception, Platform Trustworthiness, Economic Insights).