### Research with solid fill10.3 Hands-on Case Study:

Sentimental Analysis – Movie Review

*Download Source Data Set from GitHub link* : <https://github.com/Smartbrain2024/Mastering_AI_2.git>

**Chapters/Chp\_10/10.3/Hands\_on/IMDB Dataset.csv**

**Background**

Sentiment analysis (or opinion mining) is a natural language processing (NLP) technique used to determine whether data is positive, negative, or neutral. Sentiment analysis is often performed on textual data to help businesses monitor brand and product sentiment in customer feedback and understand customer needs.

**Problem Statement**

Build an ML model on movie reviews data to determine whether the sentiment of a movie review is positive or negative.

**Objectives**

* Preprocess the text and make it ready for model building.
* Build a Decision Tree model to predict review sentiment and observe its performance.

**Data Dictionary**

The description of the different variables in the dataset is provided below.

|  |  |
| --- | --- |
| Column Name | Description |
| Index | Index corresponding to the review |
| URL | URL to the movie review |
| Text | Review text of the movie |
| Sentiment | Sentiment associated with review – Positive (POS) or Negative (NEG) |

Step1 : Source Data set : imdb-reviews.csv

A screenshot of a computer

Description automatically generated

Step 2: Convert Strings to Document

A screenshot of a computer

Description automatically generated

Step 3: Data sorting and Cleansing

A diagram of a document

Description automatically generated

Step 4: Stop Word filter and Case converter and Bag of words.

A diagram of a document

Description automatically generated

Step 5 : Document Vector

A diagram of a document

Description automatically generated

A screenshot of a computer

Description automatically generated

Step 6: Partitioning and Decision Tree AlgoA diagram of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

*Conclusion: The NLP Model is Successfully executed with 91.5 % accuracy with sentiment analysis of a movie review is positive or negative and more tuning opportunity with DT Configuration.*