

SENTIMENT ANALYSIS OF TWITTER PRODUCT REVIEWS

A PROJECT BY SREENEHA S

ABSTRACT

- THIS IS MY ATTEMPT AT SOLVING THE PROBLEM STATEMENT GIVEN FOR A COMPETITION HOSTED ON ANALYTICS VIDHYA. "IDENTIFY THE SENTIMENT"
- THE DATASETS USED HERE ARE GIVEN BY THEM. YOU CAN DOWNLOAD THEM HERE:
[HTTPS://DATAHACK.ANALYTICSVIDHYA.COM/CONTEST/LINGUIPEDIA-CODEFEST-NATURAL-LANGUAGE-PROCESSING-1/](https://datahack.analyticsvidhya.com/contest/Linguipedia-Codefest-Natural-Language-Processing-1/)
- I HAVE BUILT TWO MODELS FOR THIS PROBLEM, NAÏVE BAYES CLASSIFIER AND SUPPORT VECTOR CLASSIFIER
- I HAVE ALSO USED WELL KNOWN NLTK TOOLS AND TF/IDF AND OTHER NATURAL LANGUAGE PROCESSING FUNCTIONS

IMPLEMENTATION

- STEP 1: PRE-PROCESSING TEXT

I HAVE TAKEN THE DATA SETS AND PRE-PROCESSED THEM. THE TECHNIQUES I USED FROM THE BASIC NATURAL LANGUAGE TOOL KIT ARE:

- STOP WORDS REMOVAL
- STEMMING
- LEMMATIZATION
- PARTS OF SPEECH FILTERING

- STEP 2: FEATURE ENGINEERING

I HAVE LOGICALLY IMPLEMENTED TECHNIQUES LIKE TF/IDF AND BAG OF WORDS FROM SCRATCH. IN DOING SO, I MADE A DICTIONARY OF WORDS AND WEIGHED THEIR IMPORTANCE BY THEIR OCCURRENCE CHARACTERISTICS. I HAVE MANUALLY COMPUTED THIS TF/IDF SCORE USING THE ROOT FORMULA FOR ALL THE RECORDS AND SENT THIS DATA FOR TRAINING THE MODEL.

- STEP 3: MODEL BUILDING

I HAVE USED TWO CLASSIFICATION MODELS NAMELY, NAÏVE BAYES CLASSIFIER AND SUPPORT VECTOR CLASSIFIER. I IMPLEMENTED THEM USING SKLEARN LIBRARIES

- STEP 4: MODEL PREDICTION:

UPON SUBMISSION OF THE PREDICTIONS FROM EACH OF THE MODELS INDIVIDUALLY, THE NAÏVE BAYES CLASSIFIER MODEL PREDICTED WITH AN ACCURACY OF 0.699 WHEREAS THE SUPPORT VECTOR CLASSIFIER PREDICTED THE VALUES WITH AN ACCURACY OF 0.87