

NAME:

Sentiment Analysis of Bengali languages using Multinomial Naive Bayes classifier.

Author : Anik Chowdhury

DATASET

newds3c.csv X

Edit as csv

Users > anik\_chowdhury > ML\_Projects > Sentiment\_analysis\_bangla\_svm > newds3c.csv

```
1 sad বরাবর, মাননীয় প্রধানমন্ত্রী গণপ্রজাতন্ত্রী বাংলাদেশ সরকার। মাননীয় প্রধানমন্ত্রী, আমি আপনার স্বপ্নের ডিজিটাল বাংলাদেশের স্বপ্ন বুকে লালনকারী একজন সাধারণ
2 sad তদন্ত করে লাভ কী ? কোন দিন কোন তদন্তের পর কিছু হয়েছে বলে কী আপনার মন পাবে ? রাজনৈতিক শক্তির সামনে তদন্ত কিছুই করতে পারবে না ! আর তা
3 angry দেশকি মধ্যম আয়ের দেশে রূপান্তর হচ্ছে নাকি মগের মূল্যের দেশে পরিণত হচ্ছে?
4 happy আমি গর্বিত আমি মুসলিম। আমি সংগ্রামী। আমি যোদ্ধা। আমি চির রণবীর। আমি আল্লাহ হারা কারও কাছে নিচু করি না আমার শির। নারায়ণ তাকবির আল্লা
5 angry রাতের বেলা আবার কিসের সকাল।
6 sad মনে রাখবেন এই বেহাইয়াপনাতো মা ও বোনদের সসন্ধান বাড়ছেন। তোমাদের ইজ্জত লুটে নিচ্ছে।ঐ ছায়াবট।সময় আছে নিজের সন্ধান বাচানো।ভেবে দেখুন আপনার
7 angry নাস্তিকরা কখনো তো
8 sad কেয়ামতের তোমাদের ছায়া নষ্ট।
9 angry বাংলাদেশের নদীর ওপর ভারতীয় আগ্রাসনের বিরুদ্ধে ঐক্যবদ্ধ হোন
10 angry মরালে বুজবে ছাইয়ান কাকে বলে
11 angry ভন্ডামি শালীরা। অমুসলিম।
12 happy ১৯৮৬ সালে চাকপীঠ নামের একটি প্রতিষ্ঠান যশোরে প্রথমবারের মতো নববর্ষ উপলক্ষে আনন্দ শোভাযাত্রার আয়োজন করে। যশোরের সেই শোভাযাত্রায় ছিল পাঁচ
13 angry তারে এইদেশে দনকার নাই পাকিস্তান যা।
14 happy আমি সেই বাঙ্গালি, যে তাদের ইতিহাস, ঐতিহ্য, সংস্কৃতিকে দারন, বহন ও লালন, সন্ধান এবং শ্রদ্ধা করতে ঝানি!!!! পান্তা- ইলিশ খাবো। কারণ এটা আম
15 angry অমঙ্গল শোভাযাত্রা বাপ দাদার নাম নাই গোদের উপর টাম হিন্দুদের হিন্দি কালচার পালন করে আবার বলে বাঙালি ছি?
16 sad মঙ্গল পোতা যাত্রা অ ইসলামিক কাজ
17 sad এটাই আসল বাংলা সংস্কৃতি, ঐ চাকার মঙ্গল শোভাযাত্রা, পান্তা-ইলিশের মিথ্যা চল আর হেলির মত রঙ ছিটানো বৈশাখী সংস্কৃতি নয়।
18 happy এটাই তো আমাদের অহংকার,
19 sad বহু বছরের পুরনো স্মৃতি মনে পারছে, যখন ছোট বেলা দাদা বাড়ি যেতাম,,,,,,
20 happy দারুন।
21 happy মামা খুশ হে না
22 angry তাদের মতো দালাল ৭১ যতদিন থাকবে ততদিন বাংলাদেশের উন্নতি হবে না, যারা ফাশান নামে নগ্নতাকেই প্রাধান্য দেয়.....
23 happy একাত্তর চিভি মুক্তিযুদ্ধের চেতনাকে ধারণ করে তার বস্তুনিষ্ঠ সংবাদ পরিবেশন করছে। এর সাফল্য কামনা করছি।
24 sad আমরা কেউ কাউকে দেখতে পারিনা তাই আমরা পিছিয়ে আছি
25 happy বাংলাদেশ দাঁড়িয়ে আছে -উন্নততর অর্থনৈতিক সেপানে, মাফিয়া অর্থনীতি, রাজনৈতিক দেউলিয়াপনা ও দুর্ভোগপনা, ধর্মান্ধতার পালকবীধায়। *ভাসের ঘর।
26 sad তেমনি আরেকটি ব্যর্থতা হচ্ছে যে মহিয়সী নারীর জন্য রংপুর সিটি তে পরিণত হল তার গ্রামেই পরেনি সিটিতে তাহলে এর কি প্রয়োজন ছিল
27 happy ফেইসবুকের আলস সময়কে কাজে লাগান আর মাএ ২ মিনিটে আয় করুন ৫০০ থেকে ১০০০ টাকা। নাম মাএ কিছু কাজ করেই আপনি পেতে পারেন ১০০০ টা
28 happy লেখা ভাল কিন্তু কাফনের কাপড় পরে এটা কোথায় -----,,,?
29 happy ভাই অসংখ্য ধন্যবাদ।
30 sad আমরা বাঙ্গালী। এটাই আমাদের পরিচিতি। ওই ভু-খাস্তের সকল মুসলমানই বাঙ্গালী কিন্তু সকল বাঙ্গালীই মুসলমান নয়।
31 happy প্রথমে মানুষ ,তারপর বাংলালি, এবং মুসলমান।
32 happy আপনার সাহসি বক্তব্যের জন্য আপনাকে ধন্যবাদ । আপনার মত ২/৪ হলে এ দেশের আমূল পরিবর্তন আসবে । আল্লাহ পাক আপনার হায়াত এর বরকত দান
33 happy আমি প্রথমে একজন মুসলিম তারপরে একজন দেশপ্রেমিক বাংলাদেশী আগে তো মানুষ হন . তারপরে নাহয় অন্য কিছু হাবন.....
```

master\* 0 CSVLint Query Align Rainbow OFF Ln 7, Col 26 Spaces: 4 UTF-8 CRLF CSV

In this project, i scraped the comments and posts available on social sites like Facebook and some famous bengali blog like somewherein blog, priyo.com etc. I modified the raw data using NumPy and Pandas. The dataset contains label(sentiment) and text. I have used three class for sentiment analysis: **Angry, Sad and Happy**. From this dataset I have trained a model that can detect the sentiment of text if it is angry, sad or happy.

Since this is a sentiment analysis project, I only used a **Multinomial Naive Bayes** classifier as is standard for text-based data science projects.

I have attached images of code and output of Confusion matrix and plot of both training and test result, so that the performance of the model could be understood properly.

Accuracy is 70%.

USED LIBRARIES AND IMPLEMENTED KNOWLEDGE:

- 1. Pandas
- 2. Matplotlib
- 3. NLTK

implemented Knowledge from course work 1. Data wranggling and preprocessing 2. Reducing overfitting, underfitting 4. Machine learning Algorithm 5. Model Evaluation, Confusion Matrices

All the images of codes and their outputs are attached. **Images(codes, outputs)** folder contains all the images of code and output of this project.