Ying Liu: yl1206

Chaoying Luo: cl1419

Spam Content Detection and Classification using Naive Bayes Classifier

Objective:

The spam content increases as people extensively use social media and the time spent by people using social media is also overgrowing, especially in the time of the pandemic. While users get a lot of text messages through social media they cannot recognize the spam content in these messages. To improve social media security, the detection and control of spam text are essential. Therefore, we want to choose spam content detection as the main topic in this project.

Dataset:

For the dataset, "MS Spam Collection Data Set" will be used. It is a public set of SMS labeled messages that have been collected for mobile phone spam research. To be more specific, it consists of 5574 instances. A collection of 425 SMS spam messages was manually extracted from the Grumbletext Web site. A subset of 3375 SMS randomly chosen ham messages of the NUS SMS Corpus (NSC). A list of 450 SMS ham messages collected from Caroline Tag's PhD Thesis. And finally, 1,002 SMS ham messages and 322 spam messages were collected from the SMS Spam Corpus v.0.1 Big. For attribute information, the dataset is composed of just one text file, where each line has the correct class followed by the raw message.

Methods & Metrics:

Naïve Bayes algorithm will be used for learning and classification of messages as spam and ham. Bayes theorem has strong independence property and it gives the probability of an event based on the prior knowledge of a related event. Finally, accuracy, precision, recall, f1-score, and support cases(how many cases supported that classification) will be used to evaluate the model performance.

Citation 1:

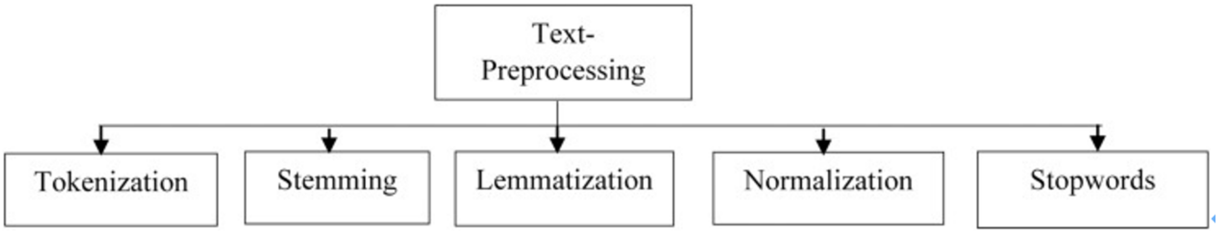
Maram, Sai Charan Reddy's *SMS Spam and Ham Detection Using Naïve Bayes Algorithm* (Available at SSRN: https://ssrn.com/abstract=3908998) is cited in the project. This paper introduces how to use traditional machine learning classification algorithms such as Naive Bayes to differentiate between spam and ham messages, which is exactly the same as the topic here. In the data-preprocessing, it cleaned the text by removing unwanted or redundant data by using stop words and noisy data was removed by stemming and lemmatization. Vectorization was also used to convert text data into integer format. In the training, the Multinomial Naive Bayes model was built. Multinomial NB calculation is a probabilistic learning technique which is for the most part utilized in Natural Language Processing. The paper achieved a high accuracy of 98.13% at last so its steps including preprocessing and training will be followed in this project. We will also use stemming, lemmatization, and vectorization in the preprocessing section, and Multinomial NB will be trained.

Citation 2:

Almeida, T. A., Hidalgo, J. M. G. & Yamakami, A. 's Contributions *to the study of SMS spam filtering: new collection and results* is quoted in the project. The paper presented a new mobile phone spam collection and compared the performance achieved by several established machine learning methods. This paper also used the MS Spam Collection Data Set, which we plan to use. The dataset is a new real, public and non-encoded SMS spam corpus that is the largest one as far as we know. Besides, it detailed the tokenization in the first stage in the classification pipeline and trained a multinomial term frequency NB model. Therefore, we can compare the final accuracy between our model and the paper model at last.

Citation 3:

Kaddoura, S., Chandrasekaran, G., Elena Popescu, D., & Duraisamy, J. H.'s *A systematic literature review on spam content detection and classification* is cited in the project. This paper presents various techniques involved in spam detection and classification involving Machine Learning, Deep Learning, and text-based approaches. This paper described numerous strategies for spam text identification in-depth in our systematic literature review on spam content detection and categorization. Our research also investigated the same techniques for pre-processing and feature extraction. The paper stated before extracting features from the text, it is necessary to eliminate any undesired data from the dataset. Unwanted data in the text dataset include punctuation, HTTP links, special characters, and stop words. As illustrated in the flowchart, there are numerous text-preprocessing techniques available that can be used to remove superfluous information from incoming text input. In our final project, we will flow the flowchart to pre-process the text.



Citation 4:

Parmar, Nandan & Sharma, Ankita & Jain, Harshita & Kadam, Amol's *Email Spam Detection using Naïve Bayes and Particle Swarm Optimization* is quoted in our project. In this paper, an integrated approach using the Naïve Bayes algorithm along with Particle Swarm Optimization is used for email spam detection. The evaluation of the experiment in the paper is done based on F-measure, precision, accuracy, and recall. By evaluating the results, the paper can explain that the integrated concept results in increased accuracy and precision than the individual Naïve Bayes approach. In our final project, we also use F-measure, precision, accuracy, and recall to determine how well our model will do overall on the entire dataset. These parameters are calculated with the assistance of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

Reference:

1. Kaddoura, S., Chandrasekaran, G., Elena Popescu, D., & Duraisamy, J. H. (2022). A systematic literature review on spam content detection and classification. PeerJ. Computer science, 8, e830.
2. Parmar, Nandan & Sharma, Ankita & Jain, Harshita & Kadam, Amol. (2020). Email Spam Detection using Naïve Bayes and Particle Swarm Optimization. Volume 6. 367 - 373.
3. Sai Charan Reddy, M. (2021) SMS Spam and Ham Detection Using Naïve Bayes Algorithm. SSRN Electronic Journal, 251 - 256.
4. Almeida, T. A., Hidalgo, J. M. G. & Yamakami, A. (2011). Contributions to the study of SMS spam filtering: new collection and results.. In M. R. B. Hardy & F. W. Tompa (eds.), ACM Symposium on Document Engineering (p./pp. 259-262), : ACM. ISBN: 978-1-4503-0863-2