

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, SURATHKAL

ರಾಷ್ಟ್ರೀಯ ತಂತ್ರಜ್ಞಾನ ಸಂಸ್ಥೆ ಕರ್ನಾಟಕ, ಸುರತ್ಕಲ್
ಸಾರ್ವಜನಿಕ ಪ್ರಾಣಿಗೋಷ್ಠಿ ಸಂಸ್ಥಾನ ಕರ್ನಾಟಕ, ಸುರತ್ಕಲ್

P.O. SRINIVASNAGAR, MANGALORE-575 025

User Sentiment Analysis - IRIS Labs Proposal

13/08/2019

Rishi Sharma

17C0135

Name and Contact Information

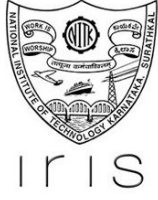
Name: Rishi Sharma

Email: rishi94075@gmail.com

Linkedin: <https://www.linkedin.com/in/rishi-sharma-278999152/>

Blog: <https://medium.com/@rishi94075/the-karate-kid-5c4d0ddfe67>

GitHub: <https://github.com/kampaitees>



NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, SURATHKAL

ರಾಷ್ಟ್ರೀಯ ತಂತ್ರಜ್ಞಾನ ಸಂಸ್ಥೆ ಕರ್ನಾಟಕ, ಸುರತ್ಕಲ್
राष्ट्रीय प्रौद्योगिकी संस्थान कर्नाटक, सुरत्कल

P.O. SRINIVASNAGAR, MANGALORE-575 025

About Me

Education

Department: Computer Science

Degree level: Btech

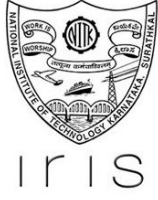
Roll Number: 17CO135

Graduation year: 2021

My previous experience related to the project:

I have been working in AI/ML/DL for more than a year. I have done many projects in the field of deep learning like Automatic Text Generator using NLP, Music Generation using RNNs & LSTMs, Facial Recognition, Neural Style Transfer, Machine Translation, Emoji prediction on the basis of review. I had completed deeplearning.ai specialisation by Andrew Ng. I had done an internship last summer in KSP Computer wing in collaboration with IISC Bangalore which is also based on Deep Learning and data analysis.

Link to the certificates: <https://github.com/kampaitees/Certificates>



NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, SURATHKAL

ರಾಷ್ಟ್ರೀಯ ತಂತ್ರಜ್ಞಾನ ಸಂಸ್ಥೆ ಕರ್ನಾಟಕ, ಸುರತ್ಕಲ್
राष्ट्रीय प्रौद्योगिकी संस्थान कर्नाटक, सुरत्कल

P.O. SRINIVASNAGAR, MANGALORE-575 025

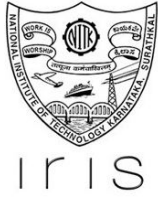
Why this project is important to us?

As we have been exploring Deep Learning/Computer Vision/NLP and would love to have a career in the same, this project would be a great opportunity to solve a real world challenge using our skills and knowledge while being an amazing learning experience. As this project (If successful) would automate the process of analysing feedback from users and will give the insights from their feedback by just one click, it would be a great achievement for us and hopefully a valuable addition to NITK IRIS. It would also serve as an amazing project to showcase our skills in future career endeavours.

About The Project

1. User Sentiment Analysis

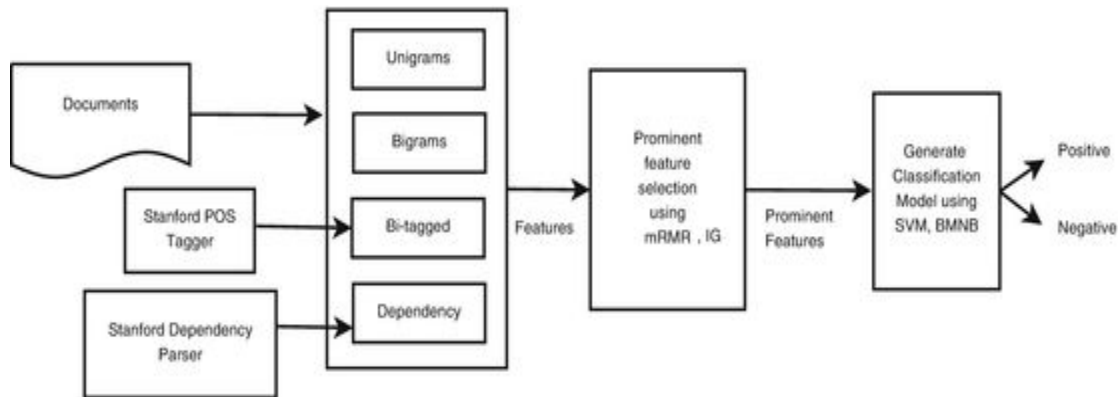
Our day-to-day life has been always influenced by what people think. Ideas and opinions of others have always affected our own opinions. The explosion of Web has led to increased activity in Podcasting, Blogging, Tagging, Social Marketing, and Social Networking. As a result there has been an eruption of interest in people to mine these vast resources of data for opinions. Sentiment Analysis or Opinion mining is the computational treatment of opinions, sentiments and subjectivity of text. In this proposal I will discuss various algorithms I will be using to perform a computational treatment of sentiments and opinions. Various supervised or data-driven techniques to Sentiment Analysis like Naive Bayes, Support Vector Machine or using various Natural Language Processing models like XLNet, BERT, SentiWordNet, Bi-directional LSTM or GRU approach to Sentiment Analysis.



2. Approach

For doing the Analysis, there can be two approaches:

Approach 1

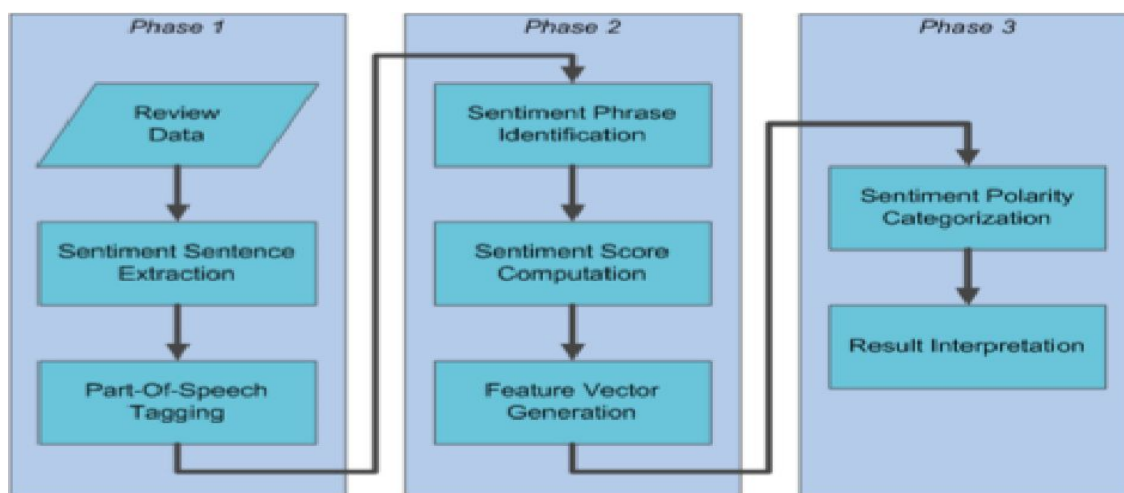


The machine learning based text classifiers which are a kind of supervised machine learning paradigm, where the classifier needs to be trained on some labeled training data before it can be applied to actual classification task. The algorithms which can be used in this paradigm are Naive Bayes(statistical classifier), Support Vector Machine(vector space classifier).

So if I go with approach I will first give a try to Naive Bayes then record its performance and then give a try to SVMs and record its performance too.

Tokenization → Feature Extraction → Classification either by SVM or Naive Bayes

Approach 2



In this approach Semantics are used to understand the language. Major tasks in NLP that helps in extracting sentiment from a sentence

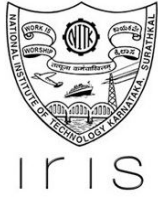
1. Extracting part of the sentence that reflects the sentiment
2. Understanding the structure of the sentence
3. Different tools which help process the textual data

This approach will use SentiWordNet lexicon which consists of positive, negative score for each of the term occurring in WordNet. The implementation done by extracting the adjectives out of the sentence and then searching it in the SentiWordNet to find out its positive, negative score. In this way the total net score of the sentence is calculated and whichever is greater (either positive or negative) becomes the review for the sentence.

So this will be procedure followed for using this approach:

Data → Tokenization → Sentence Splitter → Adjective Extraction → SentiWordNet Interpretation → Aggregating Score

Models which are used here will be Transformer, GPT2, XLnet, Bert, StanfordNLP



NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, SURATHKAL

ರಾಷ್ಟ್ರೀಯ ತಂತ್ರಜ್ಞಾನ ಸಂಸ್ಥೆ ಕರ್ನಾಟಕ, ಸುರತ್ಕಲ್
राष्ट्रीय प्रौद्योगिकी संस्थान कर्नाटक, सुरत्कल

P.O. SRINIVASNAGAR, MANGALORE-575 025

Components and functionality used will be Term frequency , Parts of speech(POS) for finding adjectives, Glove vector for Embedding. These need not be developed but will directly be used from the best sources.

I will use agile SDLC model because

- The small projects can be shown in small period of time say 2 weeks.
- Both team members and Scrum masters are going to work closely.
- Early and frequent changes to the final product.
- Documentation attends less priority than software development.

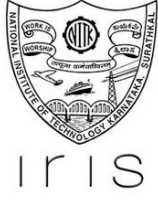
3. Deliverables

Milestones of project:

- Learning Naive Bayes, SVMs, going through Stanford NLP course and brushing up some other concepts required for sentiment analysis
- Brushing up libraries by going through documentation of various libraries used in this project like spacy, nltk, keras, numpy, re, sklearn etc.
- Going through survey papers for the already done comparison of different methods for sentiment analysis.
- Learning ensemble methods for doing ensembles of the models which will be used in Approach 2 like XLNET, BERT, Transformer, GP2

Final outputs of project:

- Naive Bayes, SVM performance report and blog
- N-gram and K-NN performance report and blog
- Different other models performance report and blog
- Best performance models or ensemble of all models will be used as a final product



NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, SURATHKAL

ರಾಷ್ಟ್ರೀಯ ತಂತ್ರಜ್ಞಾನ ಸಂಸ್ಥೆ ಕರ್ನಾಟಕ, ಸುರತ್ಕಲ್
राष्ट्रीय प्रौद्योगिकी संस्थान कर्नाटक, सुरत्कल

P.O. SRINIVASNAGAR, MANGALORE-575 025

4. Plan

22 Aug - 10 Oct : Going through Stanford Lectures where they uses vanilla Recurrent neural network for sentiment analysis, will show the results and write a blog.

10 Oct - 10 Nov : Understanding Naive Bayes, SVM, KNN and use it for prediction, show the results and write a blog.

10 Nov - 25 Nov : Going through survey papers to see the models better than vanilla recurrent neural network, will show the results and write a blog.

25 Nov - 15 Dec : Brushing up knowledge of different libraries used for the project , will show the results and write a blog.

15 Dec - 15 Feb : Use Bi-directional LSTM or GRUs and NLP techniques for prediction, show results and write a blog.

15 Feb - 20 Feb : Final Project and Presentation ready.