

## CERTIFICATE

This is to certify that Mr. Sunil Premchand Thakur From (institute) Government Polytechnic, Thane having Enrollment No: 1701160154 has completed ***Project Planning Report*** having title Twitter Sentiment Analysis Application in a group consisting of 3 Candidates under the guidance of the Faculty Guide.

Name & Signature of Guide:.....

Name & Signature of HOD:.....

## **Acknowledgments**

If words are considered as a symbol of approval and token of appreciation then let the words play the heralding role expressing our gratitude. The satisfaction that accompanies that the successful completion of any task would be incomplete without the mention of people whose ceaseless cooperation made it possible, whose constant guidance and encouragement crown all efforts with success. We are grateful to Mrs. P. S. Mahajan Ma'am for the guidance, inspiration and constructive suggestions that helpful us in the preparation of this project.

## Abstract

In the past decade, new forms of communication, such as microblogging and text messaging have emerged and become ubiquitous. While there is no limit to the range of information conveyed by tweets and texts, often these short messages are used to share opinions and sentiments that people have about what is going on in the world around them. We propose this task and the development of a twitter sentiment corpus to promote research that will lead to a better understanding of how sentiment is conveyed in tweets and texts. There will be two sub-tasks: an expression-level task and a message-level task participants may choose to participate in either or both tasks.

- **Task A:** Contextual Polarity Disambiguation Given a message containing a marked instance of a word or phrase, determine whether that instance is positive, negative or neutral in that context.

- **Task B:** Message Polarity Classification Given a message, classify whether the message is of positive, negative, or neutral sentiment. For messages conveying both a positive and negative sentiment, whichever is the stronger sentiment should be chosen.

<b>Sr. No.</b>	<b>Table of Contents</b>	<b>Pg. No.</b>
1.	Chapter-1 Introduction and background of the industry or User based Problem. 1.1 Introduction 1.2 Problem Statements	
2.	Chapter-2 Literature Survey for Problem Identification and Specification. 2.1 Literature Survey	
3.	Chapter-3 Proposed Details Methodology of solving the identified problem with action plan. 3.1 Proposed Methodology 3.2 Action Plan	
4.	References and Bibliography	

# Chapter-1 Introduction and background of the industry or User based Problem.

## 1.1 Introduction

In the past decade, new forms of communication, such as microblogging and text messaging have emerged and become ubiquitous. While there is no limit to the range of information conveyed by tweets and texts, often these short messages are used to share opinions and sentiments that people have about what is going on in the world around them. Working with these informal text genres presents challenges for natural language processing beyond those typically encountered when working with more traditional text genres, such as newswire data. Tweets and texts are short: a sentence or a headline rather than a document. The language used is very informal, with creative spelling and punctuation, misspellings, slang, new words, URLs, and genrespecific terminology and abbreviations, such as, RT for “re-tweet” and #hashtags, which are a type of tagging for Twitter messages. How to handle such challenges so as to automatically mine and understand the opinions and sentiments that people are communicating has only very recently been the subject of research

## 1.2 Problem Statements

Another aspect of social media data such as Twitter messages is that it includes rich structured information about the individuals involved in the communication. For example, Twitter maintains information of who follows whom and re-tweets and tags inside of tweets provide discourse information. Modeling such structured information is important because:

- (i) it can lead to more accurate tools for extracting semantic information, and (ii) because it provides means for empirically studying properties of social interactions (e.g., we can study properties of persuasive language or what properties are associated with influential users). We believe that a freely available, annotated corpus that can be used as a common testbed is needed in order to promote research that will lead to a better understanding of how sentiment is conveyed in tweets and texts. Our primary goal in this task is to create such a resource: a corpus of tweets and texts with sentiment expressions marked with their contextual polarity and message-level polarity. The few corpora with detailed opinion and sentiment annotation that have been made freely available, e.g., the MPQA corpus of newswire data, have proved to be valuable resources for learning about the language of sentiment. While a few twitter sentiment datasets have been created, they are either small and proprietary, such as the i-sieve corpus, or they rely on noisy labels obtained from emoticons or hashtags. Furthermore, no twitter or text corpus with expression-level sentiment annotations has been made available so far.

## **Chapter-2 Literature Survey for Problem Identification and Specification.**

### **2.1 Literature Survey**

The information needed to make the is collected mainly from some research papers.. some information is given to us by our guide. Also we are making dataset

By collecting data from tweeter this data will be analyzed by algorithms.

To collect the app information we have been guided & helped a lot by our project guide 'Mrs. Mahajan Ma'am' who have been from beginning with us, from selecting a different topic for final year project to providing different path/ideas to build app more productively & still helping us to make project successful.

## Chapter-3 Proposed Details Methodology of solving the identified problem with action plan.

### 3.1 Proposed Methodology

We plan to employ several machine learning techniques to extract users' sentiment (emotion) from the content of their twitter messages ('tweets'). Our general approach will consist of the following steps:

Pre-processing:

- Manually label a set of training and testing instances
- Represent tweets in an appropriate format, such as Bag-of-Words
- Identify any additional features specific to tweets to include in the feature vector, in order to leverage additional information which may increase classification accuracy. Currently, we are considering the addition of **temporal data**, such as the timestamp.
- Investigate attribute selection and transformation possibilities

Modeling:

- Build a machine-learning model from the training data
- Evaluate the model on the testing data

We plan to experiment with a number of machine-learning algorithms and compare their effectiveness. These may include Bayesian classifiers, Maximum Entropy, Support Vector Machines, as well as clustering algorithms.

### 3.2 Action Plan

Sr. No .	Main Activities	Activity Performed (Workwise)	Planned start date	Planned end Date	Name of Responsible Team Members
1.	Information Gathering	Briefly studied about Capstone Project Planning	26-06-2019	02-07-2019	Annu Poddar  Sunil Thakur
		Research started for searching 3 different topic's for selection	03-07-2019	10-07-2019	
		3 topics submitted to our faculty	11-07-2019	18-07-2019	
		1 topic got selected by our faculty	19-07-2019	26-07-2019	
		Information gathered on the selected topic	27-07-2019	03-08-2019	
		The requirements to build the project are specified in SRS	04-08-2019	11-08-2019	
2.	Planning	Guide allocated on	12-08-	19-08-	

	and Designing	project	2019	2019	Usama Tahseen Ul Haque
		In brief Discussion on project ideas with guide	20-08-2019	27-08-2019	
		Presentation Given to guide on app prototype & features	28-08-2019	04-09-2019	
		Planning started on making workflow of project	05-09-2019	12-09-2019	
		Designed the system workflow (Data Flow)	13-09-2019	20-09-2019	
		Submitted System workflow to guide	21-09-2019	28-09-2019	
		System workflow got approved by guide	29-09-2019	05-10-2019	
3.	Implementatio n	Started extracting basic functionality of app required from SRS document	06-10-2019	13-10-2019	
		Made an Algorithm	14-10-2019	21-10-2019	
		Started making a prototype in app development	22-10-2019	29-10-2019	



## References and Bibliography

- [1] G. Mishne, "Experiments with Mood Classification in Blog Posts," in Proceedings of the 1st Workshop on Stylistic Analysis of Text For Information Access, 2005.
- [2] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," Dec 2009. [Online]. Available: <http://www.stanford.edu/~alecmgo/papers/TwitterDistantSupervision09.pdf>
- [3] L. Terveen, W. Hill, B. Amento, D. McDonald, and J. Creter, "PHOAKS: A system for sharing recommendations," Communications of the Association for Computing Machinery (CACM), vol. 40, pp. 59-62, 1997.
- [4] MY Chen. etc. Classifying Mood in Plurks. The 22nd Conference on Computational Linguistics and Speech Processing. Chi-Nan University, Taiwan. .
- [5] Y. Jung, Y. Choi, and S.H. Myaeng, "Determining mood for a blog by combining multiple sources of evidence," in Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence, pp. 271-274, 2007.
- [6] J. B. Walther, and K.P. D'addario, "The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication," Social Science Review, vol. 19, no. 3, pp. 324-347, 2001.
- [7] Confusion matrix. [Online][http://en.wikipedia.org/wiki/Confusion\\_matrix](http://en.wikipedia.org/wiki/Confusion_matrix)