

# Natural Language Processing Toolkit for Sentiment Analysis, Abuse Analysis and Question Answering System

<sup>1</sup>Prof. Poonam Narkhede, <sup>2</sup>Sayali Bamne, <sup>3</sup>Tanisha Jadhav, <sup>4</sup>Rani Mourya,

<sup>1</sup>Associate Professor, <sup>1</sup>Student, <sup>2</sup>Student, <sup>4</sup>Student,

<sup>1</sup>Department of Computer Engineering,

<sup>1</sup>Shivajirao S. Jondhale College of Engineering, Dombivli, India.

**Abstract :** With reference to this document; it specifically mentions the detail of the project how it was developed, the primary requirement as well as various feature and functionalities of the project and the procedure followed in achieving this functionality. As inter data is growing tremendously, we need to focus of getting maximum out of it. this can achieve if we extract information from it. in this paper Natural Language Processing(NLP) for sentiment analysis ,abuse analysis and question answering. Also, sentiment analysis use for getting masses option on some idea or company's product. This paper has achieved an automatic question answering system support asking question in natural language and has some reference value in other area. Due to huge number of users and data it is not possible to check all comments in social network. We propose a pluggable architecture with reusable components ,able to detect harmful comments.

**Index Terms – Features, Toolkit , Sentiment Analysis , Abuse Analysis, Natural Language Processing.**

## I. INTRODUCTION

Today we are on the verge of data explosion. In 2014, there were 2.4 billion internet users. There are over 2 billion monthly active face book user ,compared to 1.44 billion daily start of 2015 and 1.65 at the start of 2016. With automatic question answering system the system returns answer to user automatically .the system can improve the student learning efficiency as well reducing the teachers' workload .In this system we are going to create 3 modules out of which we would creating 2 modules for handling such big data and then we would be obtaining insights on these data .insights like the sentiment of the masses on particular topic ,masses opinion on any social issue.

### 1.1 Machine Learning:-

Machine learning is that branch of computer studies that give the potentiality to the computer to grasp without being characteristically programmed . It is the method of data analysis that automates analytical model building .it is the branch of intelligence based on idea that system can learn from data , identify pattern and make decisions with minimal human intervention .here the machine learn itself and the divide the data provided into the level of prediction and in a very short period of time gives the accurate result.

### 1.2 Supervised Learning: -

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training.

### 1.3 Need & Application: -

- Our question answering module will be very helpful in building chat bot which intakes the information from same database.
- Moving on the sentiment analysis ,keeping track of customer comments allows you to engage with individual customer in real time.
- With sentiment analysis tools , you will be notified of the negative product report immediately.
- Alternatively , you can keep track of your photos and reputation over time or at any time so you can monitor your progress.

## II. LITERATURE SURVEY

Consider worldwide system for natural language processing toolkit we studied some unique parameter of the system which will help us to define the strong objective about proposed system .Below listed gives us fair idea about the survey we had done .

**[1] Close Domain Question Answering System Using NLP Techniques:-**

Data cleaning and then forming queries from the question and retrieving the answer on the basis of keyword search and then returning the top ranked matching output. Working good for closed domains questions and the question having jargons.

**[2] Computing Machinery and intelligence:-**

Keyword matching .was the first research in this system ,we were able to get to how much the machine is able to understand the context and is able to replicate the human intelligence .was not capable for answering the question which were not in the database.

**[3] Experiment With Open-Domain Textual Question Answering:-**

Knowledge –intensive NLP techniques . Got good result on open domain with this technique 4.Baseball :An Automatic question-answer. Dictionary based matching .was able to answer all the answer to a good extent in related to the football location and dates. Was not to understand the other domain query

**[4] Natural Language Interface To Database : Question Handling And Unit Conversion:-**

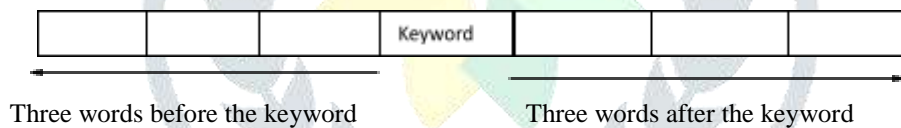
Algorithm was able to identify whether the query is a question type query. Stanford dependency parser technique for getting more knowledge from the context

**III. PROBLEM STATEMENT**

Question Answering(QA) System is an information retrieval system in which a direct answer is expected in response to a submitted query , rather than a set of reference that may contain the answer . It is machine communication device .A basic task in sentiment analysis is classifying the polarity of given document , sentence whether the expressed opinion in the sentence is positive ,negative. All above the approach were using just based the context and the keyword of the keyword of the question but were not considering the semantic meaning of the queried question . No proper language model was used to comprehend the semantic of the text .

**IV. EXISTING SYSTEM**

This approach tries to identify the keyword and the searching for that keywording the entire context given. Once if it gets the keyword matching, some system uses window technique where number of words before the keyword and n number of words after the keywords are taken.



Such approaches are used for finding answers to the question from the context.

**V. PROPOSED SYSTEM**

Design of question answering system with automated question generation. Basically , our proposed QA system gives user a set of candidate query question for user information need , and the candidate question are automatically generated from significant sentence that are expected to contain meaningful facts or event .

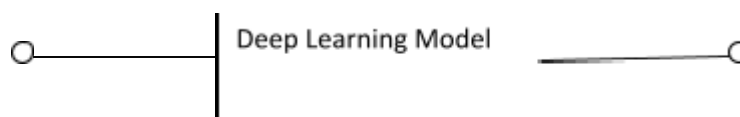


Figure 5.1. Proposed System Architecture

**VI. REQUIREMENT ANALYSIS****6.1 Hardware Requirements:-**

- 16 GB RAM
- 6 GB Graphic Card
- 80 GB HDD
- VGA Resolution Monitor

## 6.2 Software Requirements:-

- Python
- TensorFlow
- Flask

## VII. PLANNING & ESTIMATION

### System Information:-

Our system consisting of three main modules

- Sentiment Analyzer
- Abusive Sentence Detector
- Question Answering Module

In this we are using Transformer, which is a deep learning architecture which is mainly used for sequential information like text data, audio data etc. Also, we used SQUAD dataset, which consisting of questions asked by masses on some Wikipedia articles and passages.

## VIII. RESULTS

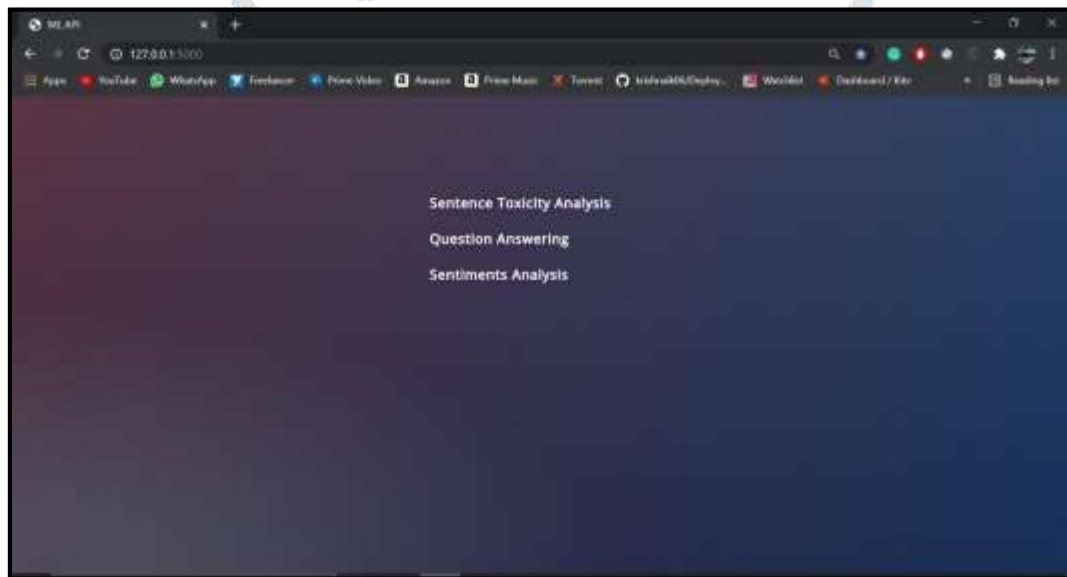


Figure 8.1 User Interface Home Page

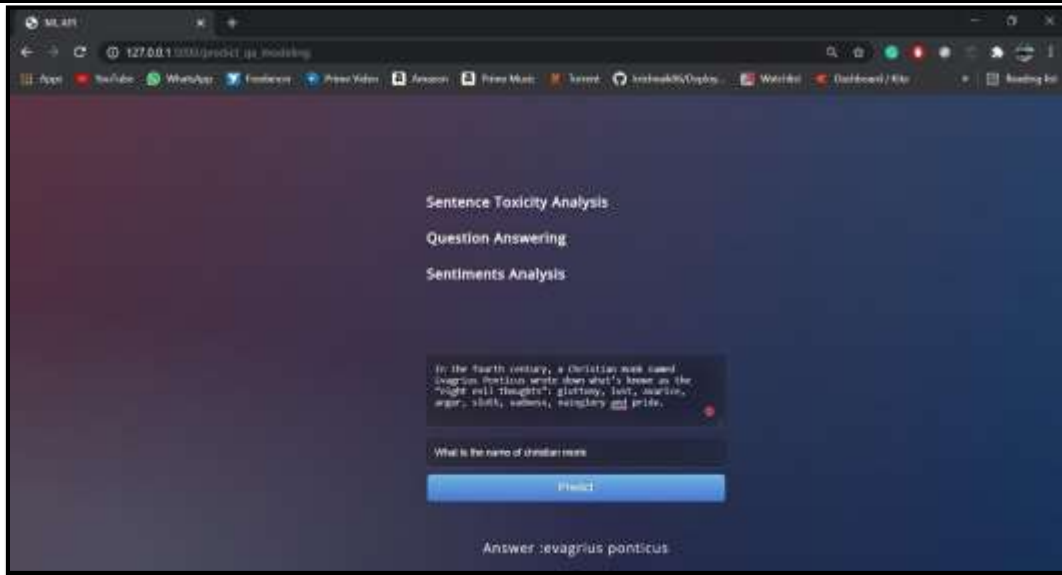


Figure 8.2 Interface For Question Answering Module

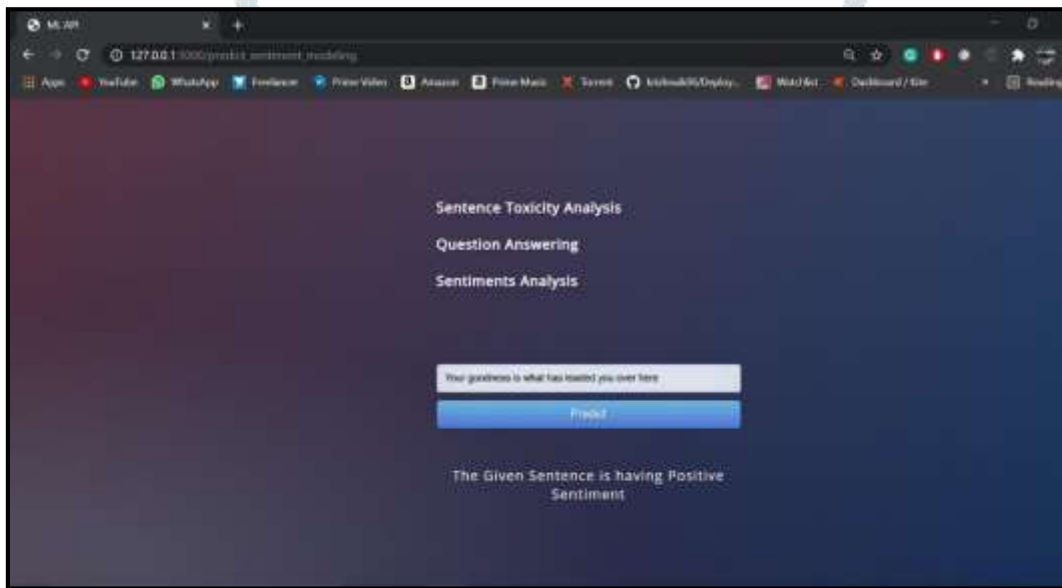


Figure 8.3 Interface For Sentiment Analysis

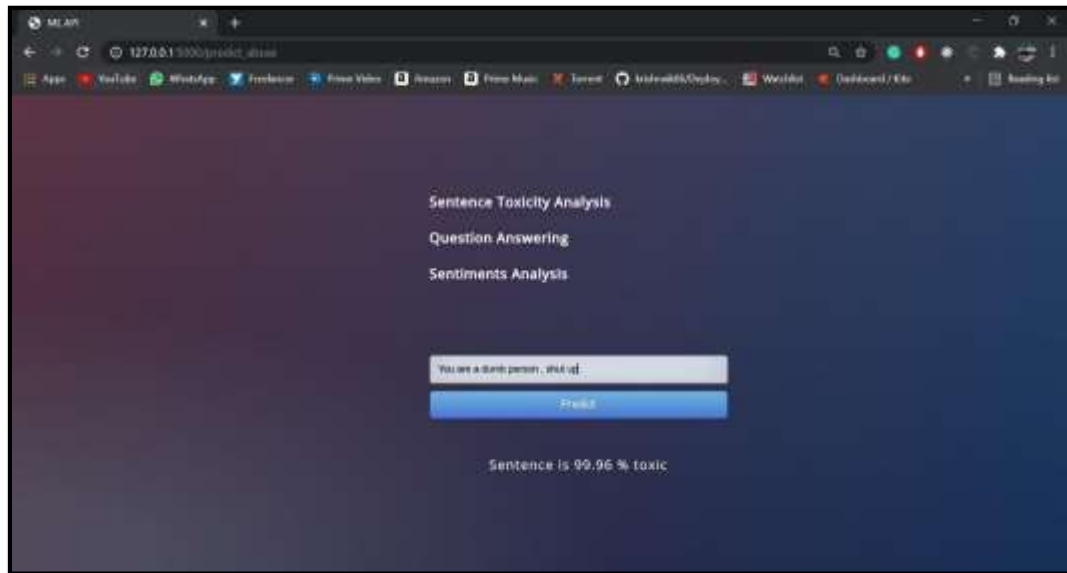


Figure 8.4 Interface For Toxic Analysis

## IX. CONCLUSION

In this paper, we are proposed a natural language toolkit with three modules as Question answering system, Sentiment analysis and Abuse detection. Here we try to propose the best(accuracy) results.

## X. ACKNOWLEDGEMENT

We sincerely wish to thank the project guide **PROF. POONAM NARKHEDE** for her encouraging and inspiring guidance helped us to make our project a success. Our project guide makes us endure with her expert guidance, kind advice and timely motivation which helped to us determine our project.

We would like to thank our project coordinator **DR. UTTARA GOGATE** for all the support needed from her our project.

We also express our deepest thanks to our HOD **DR. PRAMOD RODGE** whose benevolent help us making available the computer facilities to us for our project in our laboratory and making it true success. Without his kind and keen co-operation our project would have been stifled to standstill.

Lastly, we would like to thank our college principal **DR. J. W. BAKAL** for providing lab facilities and permitting to go on with our project. We would also like to thank our colleagues who helped us directly or indirectly during our project.

## REFERENCES

- [1] <https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/>
- [2] Denis Lukovnikov,Asja Fischer,Jens Lehmann PretrainedTrnsformers for Simple question answering over knowledge graph.
- [3] Jacob Dev lin,Ming-Wei Chang,Kenton Lee,Kristina toutanova BERT:Pre-training of Deep Biodirectional Transformers for Language Understanding.
- [4] Rushlene Kaur Bakshi;Navneet Kaur;Ravneet Kaur;Gurpreet Kaur,Opinion mining and sentiment analysis.