**UnFoolMe - System for detecting Sarcasm and patronising speech during online post and chats**

**DISSERTATION OUTLINE**

**Submitted in partial fulfilment of the requirements of the**

**MTech Data Science and Engineering Degree programme**

**By**

**Jagdish Prasad Saini**

**2020FC04489**

**Under the supervision of**

**Deep Dwivedi, VP Data Science Research and Dev**

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE**

**Pilani (Rajasthan) INDIA**

**May, 2022**

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**

**CERTIFICATE**

**This is to certify that the Dissertation entitled \_\_\_UnFoolMe - System for detecting Sarcasm and patronising speech during online post and chats and submitted by Mr. \_Jagdish Prasad Saini\_\_\_\_ ID No. \_\_\_\_\_\_\_2020FC04489\_\_\_\_\_\_\_\_\_\_\_\_in partial fulfillment of the requirements of DSECLZG628T Dissertation, embodies the work done by him/her under my supervision.**

**Signature of the Supervisor Name- Deep Dwivedi**

**Place: \_Noida\_\_\_\_ Designation- VP Data Science Research and Dev**

**Date: \_\_30-May-2022**

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

**SECOND SEMESTER 2021-22**

DSECLZG628T **DISSERTATION**

Dissertation Title : UnFoolMe - Systemfor detecting Sarcasm and patronising speech during online post and chats

Name of Supervisor : Deep Dwivedi

Name of Student : Jagdish Prasad Saini

ID No. of Student : 2020FC04489

**Abstract**

**Key Words: sarcasm, detection, classification, emotions, sentiment**

Automatic sarcasm detection is a growing field in computer science. For this study, the goal is to use deep learning to classify sarcasm across several areas.

This research intends to bridge the gap between human and machine intelligence in order to recognize and comprehend sarcastic patterns and behavior. An ensemble model of long short term memory (LSTM) gated recurrent unit (GRU) and baseline convolutional neural networks CNN is used to detect sarcasm on the internet. Transformer based model BERT embedding can be uttilized to predict the sarcasm as it is trained to the large corpus.

The goal of Sarcasm Detection is to analyze the tone of a statement to determine whether or not it is sarcastic. Sarcasm is a form of behavior that, when directed against another person, can have a variety of impacts, including upsetting the recipient's established routine of expectations. As a result, the comprehension of sarcasm frequently calls for an in-depth knowledge of a variety of information sources, including the speech and the dialogue and frequently some real world facts.

In this project, we are building a solution that detects whether a chat or post is sarcastic or not.

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

**SECOND SEMESTER 2021-22**

DSECLZG628T **DISSERTATION**

**Dissertation Outline**

**BITS ID No. 2020FC04489 Name of Student:** Jagdish Prasad Saini

**Name of Supervisor:** Deep Dwivedi

**Designation of Supervisor**: VP Data Science Research and Dev at Extramarks Education india pvt. ltd.

**Qualification and Experience:**7 Years, B.tech, PHD(On Going)

**E- mail ID of Supervisor :** deep.dwivedi@EXTRAMARKS.COM

**Topic of Dissertation**: UnFoolMe- System for detecting Sarcasm and patronising speech during online post and chats

**Name of First Examiner:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **Designation of First Examiner**:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **Qualification and Experience: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ E- mail ID of First Examiner:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Name of Second Examiner: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Designation of Second Examiner:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Qualification and Experience:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**E- mail ID of Second Examiner:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

(Signature of Student) (Signature of Supervisor) Date:30-May-2022 Date:30-May-2022

**The Dissertation Outline should contain the following:**

**Student details:**

|  |  |
| --- | --- |
| **Name** | Jagdish Prasad Saini |
| **Roll** | 2020FC04489 |
| **Email id** | jpjacksaini@gmail.com |
| **Mobile** | 7792839346 |

**Date** 30-May-2022

**Dissertation Title : UnFoolMe- System for detecting Sarcasm and patronising speech during online post and chats**

**Supervisor details:**

|  |  |
| --- | --- |
| **Name of Supervisor** | Deep Dwivedi |
| **Designation of Supervisor** | VP Data Science Research and Dev at Extramarks Education india pvt. ltd. |
| **Qualification and Experience** |  |
| **E- mail ID of Supervisor** |  |
| **Problem statement:** |  |

Sarcasm analysis is a technique or linguistic phenomenon to identify people opinion, sentiments, attitude and emotions towards any product, event, organizations and service etc.

With the increase in the use of the smart mobile devices and with the facility of the high speed internet, users are able to use the social media sites like facebook, twitter, whatsapp, instagram etc.

The goal of this project is to identify sarcasm in plain text. The project plans to exploit the property of a general sarcastic statement of possessing contrasting sentiments by using Natural Language processing. The project aims at training a deep learning model to detect if a given statement is a sarcastic or regular.

**Objective of the project:**

* To determine whether the attitude of the mass is positive or negative towards the subject of interest.
* To identify whether the statement is sarcastic or non sarcastic.

**Uniqueness of the project :**

* Use of transformer architecture to detect sarcasm.

**Benefit to the organization :**

* It can be integrated with chatbot.
* It can be deployed on discussion forum.

**Scope of work :**

* Collecting data set from multiple sources.
* Literature survey on the existing work.
* EDA on the dataset to visualize different keywords.
* Building deep learning models(RNN, LSTM, Transformer based model)
* Evaluating performance of multiple model and selecting the best one.
* Deployment of model and web app development.

**Mid Sem** – Solution Analysis,Data collection and EDA, Hypothesis of algorithm,Mid-Semester Report.

,

**End Sem** – Implementation, Demonstration, Testing on realtime data, Final Report.

**Resources Needed for the project, including people, hardware, software, etc.**

* Python
* Jupyter Notebook
* Keras, Hugging face
* Visual Studio IDE
* React.js for frontend development
* Fast API for backend development

**Potential challenges & risks in doing the project**

Sarcasm detection is known as the ‘Achilles heel of sentiment analysis. It is the most difficult area in sentiment analysis. Some researchers are proved that correctly detecting sarcasm in a sentence can increase the sentiment analysis of that particular sentence. The main challenges in sarcasm detection are followed:

a) It is an easier task to detect sarcasm from speech when it is compared to the sarcasm detection from text. Because, the use of a certain tone of speech, body language, and facial expression can be useful while identifying sarcasm from speech.

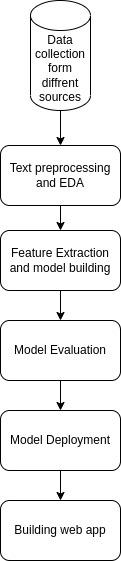
b) The quality of the data set also a crucial factor in sarcasm detection. The general nature of sarcastic sentences are ambiguous and doubtful. The presence of hashtags which indicates the sarcasm solves this ambiguity. But without hashtags, sarcastic sentences are complicated to understand.

c) Feature selection is another important task in sarcasm detection. So, introducing new features and using them with already existing features can increase the accuracy of sarcasm detection. Selecting an appropriate new feature should involve deeper study abut semantic, punctuation-based and hyperbole features, etc.

d) Selection of appropriate classification techniques is also significant. Data sets may be balanced or imbalanced. So the correct classification techniques should be used on the dataset for precise categorization of sentences into non-sarcastic and sarcastic.

e) The sarcasm detection from a noisy text is very challenging. Because generally short and noisy texts would not reveal much about context and provides lesser features.

**Solution architecture :**



1. **Detailed Plan of Work** (as follows) :

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr No.** | **Task** | **Expected Date** | **Names of Deliverables** |
|  |  |  |  |
| 1. | Solution Analysis | 10-June-2022 | Literature survey, Solution approch |
|  |  |  |  |
| 2. | Data collection and EDA | 24-June-2022 | Data insights, Text preprocessing |
|  |  |  |  |
| 3. | Hypothesis of algorithm | 8-July-2022 | Feature engineering and model building |
|  |  |  |  |
| 4. | Mid-Semester Report | 15-July-2022 | Progress report |
|  |  |  |  |
| 5. | Implementation | 5-August-2022 | Model deployment and UI development |
|  |  |  |  |
| 6. | Testing | 12-August-2022 | Application testing |
|  |  |  |  |
| 7. | Final Project Report | 27-August-2022 | Final Project Report |

**Supervisor’s Rating of the Technical Quality of this Dissertation Outline**

EXCELLENT / GOOD / FAIR/ POOR (Please specify): EXCELLENT

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Signature of Supervisor)

Name of the supervisor : Deep Dwivedi

Email Id of Supervisor :

Mob # of supervisor :