**A**

**PROJECT REPORT**

**ON**

**“Sentiment Analysis of Social Media”**

**SUBMITTED TO**

**SHIVAJIUNIVERSITY, KOLHAPUR**

**IN THE PARTIAL FULFILLMENT OF REQUIREMENT FOR THE AWARD OF DEGREE BACHELOR OF ENGINEERING IN COMPUTER SCIENCE AND ENGINEERING**

**SUBMITTED BY**

**Mr. Akash Balaso Patil 19UCS094**

**Mr. Abhimanyu Pradeep Patil 19UCS093**

**Mr. Rupesh Suresh Pawar 19UCS103**

**Miss. SrushtiBharatkumar Patil 19UCS100**

**Mr. Siddhesh Bharat Mane 19UCS082**

**UNDER THE GUIDANCE OF**

**Prof. P. M. Gavali**



A picture containing text, font, screenshot, logo

Description automatically generated

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**DKTE SOCIETY’S TEXTILE AND ENGINEERING INSTITUTE, ICHALKARANJI**

**2022-2023**

**D.K.T.E.SOCIETY’S**

**TEXTILE AND ENGINEERING INSTITUTE, ICHALKARANJI**

**(AN AUTONOUMOUS INSTITUTE)**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

A picture containing text, font, screenshot, logo

Description automatically generated

**CERTIFICATE**

**This is to certify that project work entitled,**

**“Sentiment Analysis of social media”**

**is a bonafide record of project work carried out in this college by**

**Mr. Akash Balaso Patil 19UCS094**

**Mr. Abhimanyu Pradeep Patil 19UCS093**

**Mr. Rupesh Suresh Pawar 19UCS103**

**Miss. SrushtiBharatkumar Patil 19UCS100**

**Mr. Siddhesh Bharat Mane 19UCS082**

**is in the partial fulfillment of award of degree bachelor’s in engineering in Computer Science & Engineering prescribed by Shivaji University, Kolhapur for the academic year 2022-2023.**

**P. M. Gavali**

**(PROJECT GUIDE)**

**PROF.( DR.) D.V.KODAVADE PROF. (DR.) MRS. L. S. ADMUTHE**

**(HOD CSE DEPT.) (DIRECTOR)**

**EXAMINER: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**DECLARATION**

Weherebydeclarethat,theprojectworkreportentitled“SentimentAnalysis ofSocialMedia”AnwhichisbeingsubmittedtoD.K.T.E.Society’sTextileandEngineeringInstituteIchalkaranji,affiliatedtoShivajiUniversity, KolhapurisinpartialfulfillmentofdegreeB.Tech.(CSE). It is a bonafide report of the work carried out by us. The material contained inthisreport hasnot beensubmitted toany university or institutionfor theaward ofanydegree. Further, we declare that we have not violated any of the provisions under CopyrightandPiracy/ Cyber /IPRActamendedfrom time totime.

**Name PRN Signature**

AkashBalasoPatil 19UCS094

AbhimanyuPradeepPatil 19UCS093

RupeshSureshPawar 19UCS103

SrushtiBharatkumarPatil 19UCS100

SiddheshBharatMane 19UCS082

**ACKNOWLEDGEMENT**

With great pleasure we wish to express our deep sense of gratitude to Prof. P.M.Gavali forhisvaluableguidance,supportandencouragementin the completionofthisprojectreport.

Also, we would like to take the opportunity to thank our head of department Dr. D. V. Kodavadeforhis co-operation in preparingthis project report.

We feel gratified to record our cordial thanks to other staff members of the ComputerScienceand Engineering Department for their support, help and assistance which they extended asandwhen required.

Thankyou,

|  |  |  |
| --- | --- | --- |
| **Mr.** | **Akash BalasoPatil** | **19UCS094** |
| **Mr.** | **AbhimanyuPradeepPatil** | **19UCS093** |
| **Mr.** | **RupeshSureshPawar** | **19UCS103** |
| **Miss.** | **SrushtiBharatkumarPatil** | **19UCS100** |
| **Mr.** | **Siddhesh BharatMane** | **19UCS082** |

**ABSTRACT**

Sentimentanalysishasbecomeapopularresearchareainrecentyearsdue to its potential in various applications, such as social media monitoring,customer feedback analysis, and market research.

The purpose of this project isto develop a sentiment analysis system that can classify user-inputtedstatementsinto one of five different sentiment classes: very negative, negative, neutral,positive, and very positive. The system will be trained on a large dataset of textdataandwilluseamachinelearningalgorithmtolearnpatternsandfeaturesthat areindicativeofdifferentsentiment classes.

The accuracy of the system will be evaluated using various performancemetrics, such as precision, recall, and F1-score. The project aims to provide areliable and efficient tool for sentiment analysis that can be used by businessesandindividualsaliketogaininsightsintocustomeropinionsandattitudes.

**INDEX**

1. [Introduction](#_bookmark0) 7
   1. Problemdefinition 10
   2. Aimand objective of the project 11
   3. Scopeandlimitationoftheproject 12
   4. Timelineoftheproject 13
   5. Projectmanagementplan
2. [Backgroundstudy andliteratureoverview](#_bookmark1) 15
   1. Literature overview 17
   2. Critical appraisalof other people’swork 17
   3. Investigationof currentproject andrelated work 18
3. [Requirementanalysis 19](#_bookmark2)
   1. Requirement Gathering 20
   2. Requirement Specification 21
   3. UsecaseDiagram 22
4. [Systemdesign 24](#_bookmark3)
   1. ArchitecturalDesign 25
   2. User Interface Design 27
   3. Algorithmicdescription ofeach module 29
   4. SystemModeling 31
      1. Sequence Diagram 31
      2. ActivityDiagram 32
5. [Implementation 34](#_bookmark4)
   1. EnvironmentalSettingforRunningthe Project 35
   2. DetailedDescription ofMethods 35
   3. ImplementationDetails 35
6. [Integration and Testing 37](#_bookmark5)
7. PerformanceAnalysis 39
8. FutureScope 42
9. Applications 45
10. Installation Guide andUserManual 47
11. Plagiarism Report 49
12. References 51

**1. Introduction**

The age of the internet has revolutionized the way people communicate and express their opinions. With the rise of social media platforms like Facebook, Twitter, and Instagram, people have found a new medium to share their views, experiences, and emotions with the world. Social media has also given businesses a platform to connect with their customers, gather feedback, and promote their products and services.

However, with the vast amount of user-generated content available online, it has become increasingly challenging for businesses to extract meaningful insights from this data. This is where sentiment analysis comes into play. Sentiment analysis is the process of identifying and classifying the sentiment expressed in a text source, such as a tweet, blog post, or product review.

Sentiment analysis has become a popular research area due to the abundance of sentiment-rich data available on social media platforms. This data can provide valuable insights into the opinions and attitudes of users towards various topics, products, and services. Sentiment analysis can also help businesses to understand their customers' needs and preferences and improve their products and services accordingly.

There are various approaches to perform sentiment analysis, such as BERT, VEDER, Convolutional Neural Networks (CNN), Naive Bayes, and NLP. However, individual use of models like BERT or CNN may not provide accurate results. To tackle this problem, a new approach is to use a combination of BERT and CNN, which has shown promising results in recent studies.

In this project, we aim to develop a sentiment analysis system that can classify user-input statements into one of five different sentiment classes: very negative, negative, neutral, positive, and very positive. The system will be trained on a large dataset of text data and will use a combination of BERT and CNN to improve the accuracy of the classification. The project will provide a reliable and efficient tool for sentiment analysis that can be used by businesses and individuals alike to gain insights into customer opinions and attitudes.

Although BERT, VEDER, Convolutional Neural Networks (CNN), Naive Bayes, and NLP are widely used for sentiment analysis, they may not be as effective if used separately for this project. BERT and VEDER are pre-trained models that work well on general language tasks but may not performas well

when it comes to sentiment analysis.

CNN is a powerful deep learning model for image recognition, but it is less effective when it comes to processing text data. Naive Bayes is a simple and efficient algorithm for text classification, but it may not capture the complex nuances of language required for sentiment analysis.

NLP techniques such as word embedding’s and named entity recognition can provide useful insights into text data, but they may not be sufficient for sentiment analysis on their own. Hence, a combination of different models and techniques is required to improve the accuracy of sentiment analysis for this project.

Naive Bayes is the simplest and fastest classification algorithm for large data. The Naive Bayes algorithm is a supervised machine learning algorithm based on the Bayes' theorem.

* 1. **Problem Definition**

Sentiment Analysis is a classification problem that ranges on a positive to negative spectrum. However, it can also be viewed as a regression problem when precision is important. While it is a useful tool for businesses to engage with consumers and track brand perception on social media platforms like Twitter, sentiment analysis can be costly and prone to human error due to the challenge of interpreting sentiment polarity and intensity.

Social media platforms like Twitter generate a vast amount of data that can be difficult for businesses to prioritize and respond to. This is why sentiment analysis has become an important instrument in social media marketing strategies. By carefully listening to the voice of the customer using sentiment analysis, companies can keep on top of what’s being said about their brand, competitors, and industry trends.

However, sentiment or emotion analysis can be challenging for machines as they must be trained to analyze and understand emotions like the human brain does. This is further complicated by the nuances of different languages. As data science continues to evolve, sentiment analysis software is becoming better equipped to tackle these issues.

To improve the accuracy of sentiment analysis, businesses may need to combine different approaches, such as using a combination of pre-trained models like BERT and Convolutional Neural Networks (CNN), and techniques like NLP and named entity recognition. By using a combination of different models and techniques, businesses can improve the accuracy of sentiment analysis, thereby providing better insights into customer sentiment and brand perception.

* 1. **Aim**

TodesignanddeveloptoolsforSentimentAnalysisonTwitterdata.

* 1. **objectives**
     + Topre-processthetweetdatainordertoachieveplaintextbyremovingnon-English words,emojiandspellingmistakes.
     + Tobuildthesentimentanalysismodelforidentifyingsentimentoftweettext.
     + Toimprovetheaccuracy ofsentimentanalysis.
  2. **Scope and limitation**

This sentiment analysis project involves several steps, starting with pre- processing the data. This includes cleaning and formatting the data to remove any noise or irrelevant information. The next step is to use Natural Language Processing techniques to transform the text data into numerical representations that can be processed by the machine learning models.

After pre-processing, the data is fed into the machine learning models, which have been trained to classify the text into 5 different sentiment classes - positive, slightly positive, neutral, slightly negative, and negative. Various machine learning techniques such as BERT, VEDER, Convolutional Neural Networks (CNN), Naive Bayes, and NLP can be used to classify the text into these classes.

Once the sentiment classification is complete, the final step is to present the results. This can be done through visualizations such as graphs, charts, and heat maps, which can provide a better understanding of the sentiment trends and patterns. Businesses can use these insights to improve their marketing strategies, product development, and customer service.

Overall, this project involves several steps, including pre-processing the data, transforming it into numerical representations, classifying the text into 5 different sentiment classes using machine learning models, and presenting the results through visualizations.

However, there are limitations to the sentiment analysis project. One major limitation is the accuracy of the sentiment analysis tools. While there are differentapproaches like BERT, VEDER, Convolutional Neural Networks (CNN), Naive Bayes, and NLP, if these tools are used separately, they may not provide an accurate analysis of sentiment polarity and intensity. Moreover, the sentiment analysis tools may not be able to capture the nuances of language and cultural differences that can impact the sentiment of the text.

Another limitation is the ethical and privacy concerns associated with sentiment analysis. As sentiment analysis involves analyzing user-generated content it raises questions about data privacy and security. It is essential to ensure that personal data is not misused or used for unintended purposes.

In conclusion, while sentiment analysis has a broad scope, it is important to consider the limitations of the tools and the ethical concerns. Therefore, it is vital to use appropriate techniques and methods while ensuring data privacy and security.

* 1. **Timeline of the project**

Figure 1 Timeline of project

A picture containing text, screenshot, diagram, line

Description automatically generated

* 1. **Project Management Plan**

Theprojectmanagementplanforthisprojectinvolvesdividingtheprojectinto phases, each with specific tasks and deadlines. The first phase will involvedata collection and pre-processing, while the second phase will focus on modeldevelopmentandtesting.Theteamwillconsistoffivemembers,eachresponsiblefor specific tasks, with a project manager overseeing the entire project.

Regularmeetings will be held to discuss progress, identify issues, and make necessaryadjustments to the plan. Additionally, a communication plan will be establishedto ensure effective collaboration and information sharing among team members.

A risk management plan will also be developed to identify potential risks andplan for contingencies. Finally, quality assurance and testing will be conductedtoensurethefinalproductmeetstheproject'sobjectivesandrequirements.

**2. Background study and literature Overview**

The field of sentiment analysis has gained significant attention in recentyears, with the increasing use of social media and other online platforms forsharing opinions and emotions. Sentiment analysis involves the use of naturallanguageprocessingandmachinelearningtechniquestoidentifyandclassifythesentiment expressedintextualdata.

This has numerous applications in various fields, including marketing,customerservice,andpolitics.However,sentimentanalysisisachallengingtask due to the complexity of human emotions and the variability in languageuse.

Therefore, this project aims to develop a sentiment analysis system that canaccuratelyclassifythesentimentinuser-generatedcontentinto5differentclasses.

### 2.1LiteratureOverview

Sentiment analysis has become increasingly important in understandingcustomer opinions and feedback about products and services. To gain insightsabout this topic, we conducted a literature review of research papers publishedbetween 2010 and 2022 using the IEEE Xplore library. Our search termsincluded "Twitter sentiment" and "sentiment analysis."

We found that sentimentanalysishasevolved significantlyovertheyears,withtheemergenceofmachine learning techniques and deep learning models such as BERT, VADER,and Convolutional Neural Networks (CNNs). Many studies focused on applyingthese techniques to twitter data, as it is a rich source of user-generated content.

Other research explored the challenges of sentiment analysis, such as the needfor language-specific models and the difficulty of handling sarcasm and irony intext. Overall, our literature review highlighted the importance of sentimentanalysis in understandingcustomeropinionsandprovidedinsightsintothelatest techniquesandchallengesin this field.

### 2.2 Criticalappraisalofotherpeople’swork

### Critical appraisal of existing research is a crucial step in any project to build on existing knowledge and ensure the validity and reliability of findings. In this project, we conducted a thorough critical appraisal of relevant research papers to understand the latest trends and techniques in sentiment analysis.

### We evaluated the quality of research based on criteria such as study design, data collection and analysis methods, sample size, and validity of results. Through this process, we identified gaps in existing research and opportunities for further exploration.

### We also learned from the strengths and limitations of previous studies and incorporated best practices into our own research design.

### 2.3 Investigationofcurrentprojectandrelatedwork

From the reference papers we study various approaches and techniqueshave been proposed, including machine learning algorithms such as NaiveBayes, Support Vector Machines (SVM), Convolutional Neural Networks(CNN), and Pre-trained Language Models such as BERT but these approachesdot does not provide the accuracy we want thus we are trying to achieve moreaccuracythananycurrent availablemodelpresent forsentimentanalysis.

**3. Requirement Analysis**

**3.1 Requirement Gathering**

The following are the essential requirements that are expected from the system.

•The system should be able collect data from social media platforms like Twitter, Facebook etc. for analysis.

•The collected data needs to be preprocessed to remove any irrelevant information like stop words, special characters, etc.

•The system should be able to classify the sentiment of the input text into five different classes: Positive, Negative, Neutral, Strongly Positive, and Strongly Negative.

•The system should provide accurate results for sentiment classification.

•The system should have an easy-to-use interface for the user to input the text and view the results.

•The system should be able to handle a large amount of data and users simultaneously.

•The system should provide fast and efficient sentiment analysis.

•The system should be easy to maintain and update with new features.

•The system should ensure the privacy and security of user data.

•The system should be compatible with different devices and operating systems.

**3.2 RequirementSpecification**

|  |  |  |
| --- | --- | --- |
| **Requirement** | **Essential/Desirable** | **Description** |
| UserInput | Essential | The system should be able toreceive user input in the form oftextdata. |
| Pre-processing | Essential | The input text data shouldundergo pre-processing stepssuch as tokenization, stop-wordremoval, and stemming beforeanalysis. |
| SentimentClassification | Essential | The system should be able toclassify the sentiment of theinput text data into one of thefive predefined classes:positive, negative, neutral,strongly positive, stronglynegative. |
| Accuracy | Essential | The system should have a highaccuracy in sentimentclassification. |
| Speed | Desirable | The system should process theinput data quickly and provideresultswithinareasonabletimeframe. |
| UserInterface | Desirable | The system should have a user-friendly interface to allow usersto input data and view theresultsofsentimentanalysis. |
| Language Support | Desirable | The system should be able tohandle input data in multiplelanguages. |
| Scalability | Desirable | Thesystemshouldbescalableto handle large volumes ofinputdata. |

### 3.3 UseCaseDiagram

As shown in diagram the user provides twitter data to the system. They collect this data from the user, and it is stored in a labeled manner.

The module is trained with the tweets provided within the dataset. Then the trained model is tested with test data to calculate its accuracy. This trained model is used to perform sentiment analysis. Result is displayed in output window.

A picture containing text, diagram, screenshot, line

Description automatically generated

Figure 2 Use case diagram.

The user is provided with a login id and password to enter the system. If users don’t have the login credentials, then there is a register option available.

After Authentication user enters on the home page which contains the search box for the input tweet text. When user clicks on view sentiment button trained module runs in the background and result is generated and displayed on the home page.

**4. System Design**

**4.1 Architectural Design**

A picture containing text, diagram, parallel, plan

Description automatically generated

Preprocessing the data is a crucial step in any machine learning project, and this project is no exception. In the first step of the project, we will preprocess the Twitter data by removing any irrelevant information such as URLs, usernames, and special characters.

We will also remove any stop words and perform stemming or lemmatization to reduce the dimensionality of the data. This will make it easier for the machine learning algorithm to identify the important features of the text. Once the data has been preprocessed, we will divide it into training and testing sets.

The training set will be used to train the machine learning model, while the testing set will be used to evaluate its performance. This will allow us to assess how well the model generalizes to new, unseen data and make any necessary adjustments before deploying it in a real-world scenario.

The architectural design of this project involves the use of a Convolutional Neural Network (CNN) with Global Max Pooling. The CNN is fine-tuned on the BERT model's pre-trained weights, which has been trained on a large corpus of text data. The BERT model is used to extract rich contextual information from text.

The CNN layers are followed by an Average Pooling layer and then concatenated with the BERT output. The concatenation is passed through a Dropout layer to reduce over fitting and then fed into a Classification layer.

The output layer is designed to classify the sentiment into five different classes: strongly positive, weakly positive, neutral, weakly negative, and strongly negative. The architecture is designed to extract features from the input text and then use them to predict the sentiment of the text.

**4.2User Interface Desig**A screenshot of a computer

Description automatically generated with medium confidence**n**

Figure 3 Login Page UI.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 4 Positive sentiment Result.

A screenshot of a computer

Description automatically generated

Figure 5 Positive sentiment Result.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 6Negative sentiment Result.

**4.3 Algorithmic Description of each model**

1. START: The algorithm begins.

2. Collect Twitter Data: Download the Sentiment 140 dataset, which consists of

one million tweets. This dataset is commonly used for sentiment analysis tasks.

3. Preprocess the Tweets: Apply preprocessing steps to the tweets to clean and prepare them for analysis. The specific steps include:

3.1: Remove punctuation from the texts

3.2: Remove hashtags.

3.3: Remove mentions.

3.4: Convert the text into tokens.

4. Train the Preprocessed Data on Word2Vec Model: Use the Gensim library to train a Word2Vec model on the preprocessed tweet data. This step involves converting the tokens into word-level embeddings using Word2Vec.

5. Create a Convolutional Neural Network (CNN): Build a CNN model to process the sentence-level embeddings obtained from the previous step. The CNN layer extracts features from the embeddings.

6. Use Pre-trained BERT Model:

6.1: Fine-tune a pre-trained BERT model using the Hugging Face Transformers library on the Twitter data. Fine-tuning allows the model to learn from the specific sentiment analysis task.

6.2: Obtain word-level embeddings from BERT.

6.3: Perform global average pooling to obtain a sentence representation embedding. Concatenate Static Word Embeddings and BERT Embedding: Combine the word embeddings obtained from Word2Vec (static) and BERT (contextual) into a single embedding representation.

7. Pass the Embeddings to Neural Dropout Layer: Apply a neural dropout layer to regularize the model and prevent overfitting.

8. Use Classification Layer with Softmax Activation Function: Add a classification layer with softmax activation function to perform sentiment

9. classification. Softmax assigns probabilities to each sentiment class (e.g., positive, negative, neutral).

10. Evaluate Accuracy of the Model: Assess the performance of the model by evaluating its accuracy.

This step involves measuring how well the model predicts sentiment on a separate test dataset.

11. Deploy the Model to a Web Server using Flask:

11.1 : Create a login and registration page for user authentication.

11.2 : Build a home page where users can enter tweets for sentiment analysis.

11.3 : Display the sentiment analysis results to the user.

12. END: The algorithm concludes.

### 4.4 SystemModeling

#### 4.4.1 SequenceDiagram

As shown in the above figure, it describes the sequence diagram of theproject. Sequence diagram shows all the activities that are performed on textdatafromfirststeptill resultisdisplayedon thescreen.

*A diagram of a server

Description automatically generated with medium confidence*

Figure 7 Sequence Diagram

The above Sequence diagram gives a clear perspective of the whole project and itsworkinginsequential manner.

Themodelistrainedwiththedatasetprovidedthenitsaccuracyischecked.Thenpreprocessingofuserinputdocumentstarts afteritsvalidation. Theinputtextisalsosavedtothedatabaseandtrained modelcanevaluatetheresultandsentbackto theuserThesentimentanalysisofthetweetisdoneby themodelandtheresultsareshown totheuser.

#### 4.4.2Activity Diagram

The activity diagram provided above illustrates the workflow of a sentiment analysis tool. It depicts various steps involved in the process, starting from user registration to obtaining the sentiment analysis results.

When a user interacts with the system, the first check is whether the user is registered or not. If the user is not registered, they are prompted to register on the system by providing necessary details. Once the registration process is completed, the user proceeds to the next step.

For registered users, the system requires authentication to ensure secure access. The user provides their credentials, and upon successful authentication, they gain access to the system's features.

The user can then enter a tweet in the provided textbox. After entering the tweet, they send a request to the server, which triggers the sentiment analysis process. The server utilizes a trained model to analyze the sentiment of the tweet based on various factors.

Once the sentiment analysis is complete, the server sends the predicted sentiment output back to the user. This output can be viewed by the user, allowing them to understand the sentiment conveyed by the tweet. Additionally, the system may provide a visualization of the sentiment analysis results, which can offer a graphical representation or summary of the sentiment.

Finally, the user has the option to log out from the system, terminating their current session and ensuring their privacy and security.

Overall, the activity diagram outlines the sequential steps involved in utilizing the sentiment analysis tool, providing a clear understanding of the user's journey and the system's functionality.

A picture containing diagram, text, sketch, technical drawing

Description automatically generated

Figure 8 Activity Diagram

**5. Implementation**

### 5.1 Environmental Setting for Running theProject

* Torunthisproject,weneedabrowser. (Chrome/Firefox/Edge)
* Gotothespecified URL.
* Loginwithcredentials.
* Inputthestatement andviewtheresultfortheinputstatement.

### 5.2 DetailedDescriptionofMethods

* LoginMethod:Acceptsregistereduserswithvalidmailandpassword.
* RegisterMethod:Registerusersusingvalidmailandpassword.
* Take Input: Take the tweet as input from the user after successful loginforits sentimentanalysis.
* Identify the sentiment: BERT + CNN model will predict the sentiment ofthetweet.
* Display the Sentiment: After successful model execution the sentiment oftweetisdisplayed.

### 5.3 ImplementationDetails

* + 1. LoginProcedure–

Loginprocedureisdoneinnodejswithhelpofmongodatabaseanditsconnection.

const url = 'mongodb://localhost:7000 ';const dbName ='myproject';

// Use connect method to connect to the serverMongoClient.connect(url,function(err,client)

* + 1. Pre-processingtheuserinputteddata

Thedataenteredinthetextareaofthehomepageissendto theserverapiin which computation is done for classifying it into different classes. Thepreprocessing of this data is done in python in many ways the most basiclineinthisprojectis

# Remove all non-alphanumeric characterspreprocessed\_input=re.sub(r'\W+','',user\_input)

* + 1. Classificationofsentimentintodifferentclasses

The classification of statement is done by the following way the text iscovered into the sequence of input and the sent to model inside predict()method

# Use the model to predict the sentiment of the input textprediction =model.predict([input\_text])

* + 1. Sendingtheresultstothebrowser

Browser receives the response from the API. and the corresponding classandresultaredisplayedtothebrowserwithhelpofHTML5andbasicJS.

**6. Integration and Testing**

Integration and testing are crucial phases in the software developmentprocess. In this project, the integration phase involves combining the variousmodules of the system, such as the data preprocessing, feature extraction, andclassification modules. The testing phase involves evaluating the performanceof the system using appropriate evaluation metrics such as accuracy, precision,recall,and F1-score.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr.**  **No.** | **TestcaseDescription** | **ExpectedOutput** | **ActualOutput** | **Status** |
| 1 | Login | System needstorecognize  validusernameandpassword. | System able torecognize  validusernameandpassword. | Pass |
| 2 | Registernewuser | System shouldbeinposition  toregisternewuser. | Systemabletoregister newuser | Pass |
| 3 | Takinginput | Systemshouldbeinposition  to take inputfromtheuser. | Systemabletotake inputfromtheuser. | Pass |
| 4 | Pre-process the inputtedtext | Systemshouldbe able to pre-process theinputted textand should beableto  recognize itssentiment. | Systemisableto pre-processthe inputtedtext and alsoable torecognize itssentiment. | Pass |
| 5 | DisplaySentiment | Systemshouldbe able todisplaysentimentof  theinputtedtext. | Systemisable  to displaysentiment ofthe inputtedtext. | Pass |

**7. Performance Analysis**

The fine-tuned BERT + CNN model has been shown to be significantly betterthan the simple BERT model in sentiment analysis tasks. While BERT providespowerfullanguagerepresentation,theCNNlayerhelpstocaptureimportantlocalfeatures andpatternsintheinputtext.

By fine-tuning the BERT + CNN model on the specific sentiment analysistask, the model can learn to extract the most relevant information from the textandclassifyitintooneofthesentimentcategories.Thisresultsinamoreaccurateand precise sentiment analysis, making the model a better fit for real-worldapplications where accuratesentiment classificationis crucial.

The graphical representation of the accuracy of both models is done isdisplayed belowpictures.

A picture containing diagram, line, plot, text

Description automatically generated

Figure 9 on basis of accuracy

A picture containing line, diagram, plot, slope

Description automatically generated

Figure 10. On basis of Loss function

**8.Future scope**

**MarketResearch:**Sentimentanalysiscanprovidevaluableinsightsintoconsumer opinions and preferences. Companies can use sentiment analysis toanalyzesocialmediaconversations,customerreviews,andfeedbacktounderstand market trends, evaluate product perception, and gain a competitiveedge.

**BrandReputationManagement:**Monitoringandanalyzingsentimentaroundabrandinreal-timecanhelpcompaniesidentifypotentialissues,addresscustomerconcerns,andmanagetheironlinereputationmoreeffectively.Sentimentanalysisenablesproactivebrandmanagementbypromptlyidentifyingandaddressing negativesentimentorpublic relations crises.

**CustomerService:**Sentimentanalysiscanplayacrucialroleincustomerserviceinteractions. By analyzing the sentiment of customer messages, companies canprioritize and escalate urgent issues, identify dissatisfied customers, and providetimelysupportandsolutions.

**Political Analysis:** Sentiment analysis can be applied to political discourse onsocial media platforms to gauge public opinion and sentiment towards politicalcandidates, parties, or policies. This information can help political campaigns,policymakers,andresearchersunderstandpublicsentimentandmakedata-drivendecisions.

**PublicOpinionAnalysis:**Sentimentanalysiscanbeusedbygovernmentagencies, public figures, and organizations to analyze public opinion on varioustopics,policies,orsocialissues.Thisanalysiscaninformpolicy-making,campaignstrategies,and publicoutreachinitiatives.

**HealthcareandMentalHealth:**Sentimentanalysiscancontributetohealthcarebyanalyzingpatientfeedback,socialmediaposts,andonlinehealthforums.Thiscanhelpidentifypotentialmentalhealthissues,monitorpatientsentimentduringtreatment,andprovide insights forimprovinghealthcareservices.

**FuturescopeAdditionofmultilingualsupport:**Currently,theprojectsupportssentiment analysis for English language only. In the future, the scope of theproject can be extended to support other languages as well, which will make itmoreuseful forbusinessesoperatingin multilingualenvironments.

**Integrationwithmoresocialmediaplatforms:**Currently,theprojectislimitedto analyzing sentiments from Twitter. In the future, the project can be extendedtosupportsentimentanalysisonothersocialmediaplatforms likeFacebook,Instagram,andLinkedIn.

**Additionofemotiondetection:**Sentimentanalysiscanprovideagoodunderstandingofwhetheratweetorpostispositive,negativeorneutral.However, it does not provide any information about the underlying emotionsbehindthesentiment.Inthefuture,theprojectcanbeextendedtoincludeemotiondetection.

**9. Applications**

1. **Customer feedback analysis:** Companies can use sentimentanalysis to analyze customer feedback on social media and adjust theirproductsorservicesaccordingly.
2. **Public Opinion Analysis:** Sentiment analysis can be applied tosocial media data to gauge public opinion on various topics, events, orsocial issues. This can be valuable for government agencies,policymakers, or researchers seeking to understand public sentiment,identify areas of concern, and shape their policies or initiativesaccordingly.
3. **Sentiment-based Recommender Systems**: Your sentimentanalysis model can be integrated into recommender systems to providepersonalizedrecommendationsbasedonusersentiment.Byanalyzingthesentiment associated with user reviews, feedback, or preferences,businesses can offer more relevant and tailored recommendations to theircustomers.

**10. Installation Guide and User**

**Manual**

**Step 1:** User needs a good internet connection and any web browser likeGooglechrome,microsoftedge,etc.

**Step 2**: New user must register before logging into the website otherwise if newusertries tologin directlythenitshowserror ofusernot found.

**Step3**:Theregistereduserneedstologintothewebsiteusinglogincredentials.

* + - * 1. If the user enters the correct username and password, then it logs insuccessfully.
        2. Iftheuserentersan invalidusernameand password,thenitgivesanerror.

**Step5:**Userwillseethe homepagewheretheinputboxisprovided.

**Step6:**Nowenterthetextyouwanttoanalyzeintheinputboxprovidedonthehomepage.

**Step7**:Clickonthe"Predict"buttontosubmit yourtext.

**Step8:**Waitforthemodeltoprocessyourinputandanalyzeitssentiment.

**Step 9:** The result of the analysis will be displayed on the page as either"Positive","Negative" or "Neutral".

**Step 10:** You can then choose to analyze another text by clicking the "Clear"button tocleartheinputbox andstartover.

**10. Plagiarism Report**

A screenshot of a computer

Description automatically generated with medium confidence

**10. References**

#### Referencepaper

1. Shihab Elbagir, Jing Yang, “Twitter Sentiment Analysis based on ordinalRegression”,IEEEaccess,Vol.7,PP.163677–163685,07November2019
2. IsraaAlghanmi, Luis Espinosa-Anke, Steven Schockaert, “CombiningBERTwithStaticWordEmbeddingsforCategorizingsocialmedia”,WNUT,Vol.ProceedingsoftheSixthWorkshoponNoisyUser-generatedText(W-

NUT2020),PP.28-33,November2020

1. Lei Wang, JianweiNiu, “SentiDiff: Combining textual information andsentiment diffusion patterns for twitter sentiment analysis”, IEEE TransactionsonKnowledgeandDataEngineering,Vol.32,PP.2026 -2039,26April2019