

**Tech Saksham**

**“Sentiment Analysis AI system”**

**“Rvs College Of Engineering”**

|  |  |
| --- | --- |
| **NM ID** | **NAME** |
| aut21letx11 | P. MANOJ VARMAN |

|  |  |
| --- | --- |
|  | Trainer Name : P.RAJA |
|  | Master Trainer : Edunet Foundation |

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

**ABSTRACT**

Sentiment analysis, also known as opinion mining, is a burgeoning field within artificial intelligence (AI) and natural language processing (NLP). This paper presents a comprehensive overview of a state-of-the-art sentiment analysis AI system designed to analyze and understand human sentiment expressed in textual data. The system employs a combination of machine learning algorithms, deep learning models, and linguistic analysis techniques to accurately classify text into positive, negative, or neutral sentiment categories. The sentiment analysis AI system utilizes a variety of machine learning algorithms, such as support vector machines (SVM), logistic regression, and random forests, to classify text sentiment based on extracted features. Additionally, deep learning models, including recurrent neural networks (RNNs) and convolutional neural networks (CNNs), are employed to capture complex patterns and dependencies in textual data for more accurate sentiment classification. Furthermore, the system incorporates linguistic analysis techniques to enhance sentiment understanding by considering contextual clues, sarcasm, and figurative language in text. This includes the use of sentiment lexicons, syntactic analysis, and semantic parsing to improve the accuracy of sentiment classification. Evaluation of the sentiment analysis AI system is conducted using standard performance metrics such as accuracy, precision, recall, and F1-score. The system is tested on various datasets across different domains to assess its robustness and generalization capabilities. Overall, the proposed sentiment analysis AI system demonstrates promising results in accurately identifying sentiment from textual data, with potential applications in social media monitoring, market research, customer feedback analysis, and sentiment-aware recommendation systems. However, further research is needed to address challenges such as handling noisy data, detecting sentiment in multilingual texts, and improving the interpretability of the model outputs.

**INDEX**

|  |  |  |
| --- | --- | --- |
| **S. No** | **Table of Contents** | **Page No** |
| 1 | Chapter 1: Introduction | 4 |
| 2 | Chapter 2: Services and Tools Required | 6 |
| 3 | Chapter 3: Project Architecture | 7 |
| 4 | Chapter 4: Modeling and Project Outcome | 9 |
| 5 | Conclusion | 18 |
| 6 | Future Scope | 19 |
| 7 | References | 20 |
| 8 | Links | 21 |

**CHAPTER 1**

INTRODUCTION

* 1. **Problem Statement**

In today's digital age, the vast amount of textual data generated on social media platforms, product reviews, and customer feedback channels presents both opportunities and challenges for businesses. Understanding the sentiment conveyed within this data is crucial for making informed decisions, managing brand reputation, and enhancing customer experiences. However, manually analyzing this data is time-consuming, error-prone, and often impractical due to its sheer volume.

To address this challenge, we propose the development of a Sentiment Analysis AI system. The goal of this project is to design and implement an automated solution capable of accurately analyzing textual data and classifying the sentiment expressed within it as positive, negative, or neutral.

1. Data Variety and Volume
2. Accuracy and Generalization
3. Real-time Processing
4. Multilingual Support
5. Contextual Understanding
6. Model Interpretability
   1. **Proposed Solution**

To address the challenges outlined in the problem statement, we propose the development of a comprehensive Sentiment Analysis AI system. The system will leverage state-of-the-art natural language processing (NLP) techniques and machine learning algorithms to accurately analyze and classify sentiments expressed in textual data. The proposed solution consists of the following components:

1. **Data Variety and Volume**:
   * Implement robust data preprocessing techniques to handle diverse sources of textual data, including social media posts, product reviews, news articles, and surveys. This involves cleaning, tokenization, and normalization of text.
   * Utilize scalable data storage and processing frameworks such as Apache Hadoop or Apache Spark to handle large volumes of data efficiently.
   * Employ data augmentation techniques to generate additional training examples and improve model robustness.
2. **Accuracy and Generalization**:
   * Employ transfer learning techniques, such as fine-tuning pre-trained language models (e.g., BERT, GPT) on sentiment analysis tasks, to leverage knowledge from large-scale datasets and improve generalization.
   * Regularize the sentiment analysis model using techniques like dropout or L2 regularization to prevent overfitting and enhance performance on unseen data.
   * Perform extensive hyperparameter tuning and cross-validation to optimize model performance across various topics and domains.
3. **Real-time Processing**:
   * Develop lightweight sentiment analysis models optimized for fast inference, such as shallow neural networks or linear classifiers, to enable real-time or near real-time processing.
   * Implement stream processing frameworks like Apache Kafka or Apache Flink to ingest and analyze streaming textual data in real-time.
   * Utilize cloud-based infrastructure with auto-scaling capabilities to dynamically allocate resources based on processing demands and ensure timely analysis.
4. **Multilingual Support**:
   * Train language-agnostic sentiment analysis models using multilingual embeddings or cross-lingual transfer learning techniques to handle sentiment analysis across multiple languages.
   * Incorporate language detection mechanisms at the preprocessing stage to identify the language of the text and route it to the appropriate sentiment analysis model.
   * Develop language-specific sentiment analysis models for languages with distinct linguistic characteristics and sentiment expressions.
5. **Contextual Understanding**:
   * Incorporate contextual information, such as user demographics, temporal trends, or social context, as additional features in the sentiment analysis model to improve accuracy and relevance.
   * Utilize domain-specific lexicons or embeddings tailored to the specific domain of analysis (e.g., healthcare, finance) to capture domain-specific sentiment nuances.
   * Implement sentiment analysis pipelines that combine multiple NLP tasks (e.g., named entity recognition, topic modeling) to extract relevant contextual information and enhance sentiment analysis results.
6. **Model Interpretability**:
   * Employ techniques such as attention mechanisms or feature importance analysis to provide insights into the sentiment classification decisions made by the AI system.
   * Generate explanation reports or visualizations highlighting key factors influencing sentiment predictions, such as important words or phrases in the text and their corresponding sentiment weights.
   * Incorporate interactive tools or dashboards that allow users to explore and understand the rationale behind the AI system's sentiment analysis predictions.
   1. **Feature**

* **Text Classification**: The system categorizes text into different sentiment categories such as positive, negative, or neutral.
* **Accuracy and Precision**: It aims to provide accurate and precise sentiment analysis results by considering various linguistic nuances and contextual factors.
* **Scalability**: The system should be able to handle large volumes of text data efficiently, making it suitable for processing data from diverse sources like social media, customer reviews, or news articles.
* **Multilingual Support**: Capability to analyze text written in different languages to cater to a global audience.
* **Real-time Analysis**: Ability to provide sentiment analysis results in real-time, enabling timely decision-making and response.
* **Customization**: Some systems offer customization options to adapt to specific domains or industries, allowing users to train the model with domain-specific data for improved accuracy.
* **Sentiment Intensity**: Beyond just classifying sentiment, some systems can also gauge the intensity or strength of the expressed sentiment, providing more nuanced insights.
* **Entity and Aspect-based Analysis**: Ability to identify sentiments associated with specific entities or aspects within the text, providing granular insights, especially in product reviews or feedback analysis.
* **API Integration**: Many sentiment analysis systems offer APIs (Application Programming Interfaces) for seamless integration into existing applications or workflows.
* **Visualization**: Some systems provide visualization tools to present sentiment analysis results in an easy-to-understand format, such as charts or dashboards.
  1. **Advantages**
* **Automated Insights**: Sentiment analysis AI systems automate the process of analyzing vast amounts of textual data, saving time and resources compared to manual analysis.
* **Real-time Analysis**: By providing near real-time insights into customer sentiment, businesses can respond promptly to emerging trends, issues, or opportunities, enhancing their agility and responsiveness.
* **Customer Feedback Analysis**: These systems help businesses analyze customer feedback from various sources such as social media, surveys, reviews, and support tickets, enabling them to understand customer satisfaction levels, identify pain points, and prioritize improvements.
* **Brand Monitoring**: Sentiment analysis AI systems enable organizations to monitor mentions of their brand, products, or services across the web, social media platforms, and news outlets, helping them gauge brand perception and manage reputation effectively.
* **Competitive Analysis**: By analyzing sentiment towards competitors' products or services, businesses can gain insights into market trends, identify competitive strengths and weaknesses, and inform their strategic decision-making.
* **Product Development**: Sentiment analysis helps businesses understand customer preferences, feature requests, and pain points, guiding product development efforts to meet customer needs effectively.
* **Customer Service Optimization**: By analyzing customer interactions with customer service channels, businesses can identify recurring issues, improve response times, and enhance overall customer satisfaction.
* **Marketing Campaign Optimization**: Sentiment analysis aids marketers in evaluating the effectiveness of marketing campaigns by analyzing customer reactions and sentiment towards specific advertisements, promotions, or messaging.
* **Risk Management**: In industries such as finance and insurance, sentiment analysis can help assess market sentiment, identify potential risks or opportunities, and inform investment decisions.
* **Policy and Public Opinion Analysis**: Governments and policymakers can use sentiment analysis to gauge public opinion on various issues, monitor social trends, and assess the impact of policies or initiatives.
  1. **Scope**

The scope of sentiment analysis AI systems is expansive, encompassing a wide range of applications across industries and domains. These systems are designed to analyze textual data, such as customer reviews, social media posts, surveys, and customer support interactions, to extract sentiments and opinions expressed within them. Businesses utilize sentiment analysis to gain valuable insights into customer preferences, satisfaction levels, and brand perception. From brand monitoring and reputation management to market research and customer service optimization, sentiment analysis aids organizations in making data-driven decisions and enhancing customer experiences. Additionally, sentiment analysis supports social media analysis, product development, risk management, compliance monitoring, and public opinion analysis, among other applications. By providing actionable insights into consumer sentiment and market trends, sentiment analysis AI systems empower businesses to stay competitive, innovate effectively, and adapt to evolving customer needs and market dynamic.

**CHAPTER 2**

**SERVICES AND TOOLS REQUIRED**

**2.1 Services Used**

* **Natural Language Processing (NLP):** NLP is a foundational service that enables sentiment analysis systems to understand and process human language. It involves tasks such as tokenization, part-of-speech tagging, named entity recognition, and syntactic parsing.
* **Machine Learning Models:** Sentiment analysis often relies on machine learning models, including supervised learning algorithms such as Support Vector Machines (SVM), Naive Bayes, and neural network architectures like Recurrent Neural Networks (RNNs) and Transformers. These models are trained on labeled data to classify text into sentiment categories.
* **Text Preprocessing:** Before sentiment analysis, text data is preprocessed to clean and normalize it. This includes tasks like removing stopwords, punctuation, and special characters, as well as stemming or lemmatization to reduce words to their base forms.
* **Sentiment Lexicons:** Sentiment lexicons or dictionaries contain lists of words or phrases categorized by their sentiment polarity (positive, negative, or neutral). Sentiment analysis systems often use these lexicons to match words in text data and assign sentiment scores.
* **API Integration:** Many sentiment analysis services provide APIs that allow developers to integrate sentiment analysis capabilities into their applications or systems seamlessly. These APIs accept text inputs and return sentiment analysis results, simplifying the integration process.

**2.2 Tools and Software used**

**Tools**:

1. **Containerization Tools**: Containerization tools like Docker and Kubernetes are used to package sentiment analysis models and their dependencies into portable containers. This enables easier deployment, scaling, and management of AI systems across different computing environments.
2. **Data Visualization Tools**: Data visualization tools like Matplotlib, Seaborn, Plotly (Python), ggplot2 (R), and Tableau are used to visualize sentiment analysis results and insights. They help in understanding patterns, trends, and relationships within textual data.
3. **Google Cloud Natural Language API**: Google's NLP API includes sentiment analysis among its offerings. It provides accurate sentiment analysis for multiple languages and scales seamlessly with Google Cloud infrastructure.

**Software Requirements**:

1. **Programming Language**:
   * Python, R, Java, or another language suitable for implementing machine learning algorithms and natural language processing tasks.
2. **Development Environment**:
   * Integrated Development Environment (IDE) such as PyCharm, Jupyter Notebook, RStudio, or Visual Studio Code for writing, testing, and debugging code.
3. **Text Processing and NLP Libraries**:
   * NLTK (Natural Language Toolkit), spaCy, TextBlob, Gensim, Stanford CoreNLP, or other libraries for text preprocessing, tokenization, part-of-speech tagging, and syntactic parsing.
4. **Machine Learning Frameworks**:
   * TensorFlow, PyTorch, Scikit-learn, Keras, or other frameworks for building and training machine learning models, including neural networks for sentiment analysis.

**CHAPTER 3**

**PROJECT ARCHITECTURE**

**3.1 Architecture**

+---------------------+

| |

| Data |

| Collection |

| |

+---------+-----------+

|

|

v

+---------+-----------+

| |

| Preprocessing |

| and Feature |

| Extraction |

+---------+-----------+

|

|

v

+---------+-----------+

| |

| Model Training |

| |

+---------+-----------+

|

|

v

+---------+-----------+

| |

| Inference and |

| Prediction |

| |

+---------+-----------+

|

|

v

+---------+-----------+

| |

| Post-processing |

| and Analysis |

| |

|  |
| --- |
|  |

Here’s a high-level architecture for the project:

1. **Data Collection**: Textual data is collected from various sources such as social media, customer feedback, or product reviews.
2. **Preprocessing and Feature Extraction**: The collected data is preprocessed to clean, tokenize, and extract features, such as word frequencies or embeddings, for analysis.
3. **Model Training**: Machine learning models, such as neural networks or traditional classifiers, are trained on labeled data to learn patterns and relationships between text and sentiment labels.
4. **Inference and Prediction**: Deployed models process new, unseen text data to predict sentiment labels or scores for each input text.
5. **Post-processing and Analysis**: Post-processing steps involve aggregating sentiment scores, generating reports, or visualizing sentiment analysis results to derive actionable insights.

This architecture diagram provides a high-level overview of the sentiment analysis AI system, illustrating the sequential flow of data and operations from data collection to analysis.

1. Top of Form

**CHAPTER 4**

**MODELING AND PROJECT OUTCOME**

**Manage relationship**

The “disp” file will be used as the main connector as it contains most key identifier (account id, client id and disp id) which can be use to relates the 8 data files together. The “district” file is use to link the client profile geographically with “district id”







**Modelling for Gender and Age data**

Notice that the Gender and age of the client are missing from the data. These can be formulated from the birth number YYMMDD where at months (the 3rd and 4th digits) greater than 50 means that client is a Female. We can create a column for Gender.



For birthday, we need to reduce the birth month of the female by 50 and then change the date format to DD/MM/YYYY adding 1900 to the year.



For Age, we shall assume it is year 1999 as explain previously and use it to minus from the birth year.



**Replacing values**

Set some fields to English for easy understanding, we replace values to English with the Power Query Editor.







Changing the order of Region name at Power Query

Duplicate the “district /region” then split column using space as delimiter.



Then merge column by Region and direction. Refer to applied steps for details.



**Grouping of age by ranges**

As the customers’ age ranges from 12 to 88, we shall group them into different generation age range for easier profiling, we will group the ages into 5 groups.

The Gen Y are youths,

Gen X are young working adults, some starting their families

Baby Boomer are working adults with families.

The silent Generations some are working and retired, living on pensions.

The greatest Generation, retired elderly living on pensions.



**Credit Rating and Loan Status**

As the Loan status uses A, B, C, D which are not reader friendly. We can add a column to represent what it stands for, we also simplify the classification of those with late or default on payment as bad credit, refer to the table below for details on the new columns added.



Values of such as “account Id” have also been set as Text.

And District name have been categorized as place to be use for the map to show the sum of the inhabitants in each region.

**Dashboard**

A screenshot of a credit card

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a credit card

Description automatically generated

**CONCLUSION**

The project “Real-Time Analysis of Bank Customers” using PowerBI has successfully demonstrated the potential of data analytics in the banking sector. The real-time analysis of customer data has provided valuable insights into customer behavior, preferences, and trends, thereby facilitating informed decision-making. The interactive dashboards and reports have offered a comprehensive view of customer data, enabling the identification of patterns and correlations. This has not only improved the efficiency of data analysis but also enhanced the bank’s ability to provide personalized services to its customers. The project has also highlighted the importance of data visualization in making complex data more understandable and accessible. The use of PowerBI has made it possible to present data in a visually appealing and easy-to-understand format, thereby aiding in better decision-making.

**FUTURE SCOPE**

The future scope of this project is vast. With the advent of advanced analytics and machine learning, PowerBI can be leveraged to predict future trends based on historical data. Integrating these predictive analytics into the project could enable the bank to anticipate customer needs and proactively offer solutions. Furthermore, PowerBI’s capability to integrate with various data sources opens up the possibility of incorporating more diverse datasets for a more holistic view of customers. As data privacy and security become increasingly important, future iterations of this project should focus on implementing robust data governance strategies. This would ensure the secure handling of sensitive customer data while complying with data protection regulations. Additionally, the project could explore the integration of real-time data streams to provide even more timely and relevant insights. This could potentially transform the way banks interact with their customers, leading to improved customer satisfaction and loyalty.

**REFERENCES**

<https://medium.com/analytics-vidhya/analysis-of-bank-customers-using-dashboard-in-power-bi-a366f2b3e563>

# **GIT Hub Link of Project Code:**

<https://github.com/githubtraining/hellogitworld.git>