**HUMAN EMOTION IDENTIFIER BASED ON TEXT CLASSIFICATION**

**Team members:**

**Group 2**

1. Varsha Reddy Umannagari -11645255

2. Benerji Vigna Sai Rama Venkat Sumith Thota - 11645256

3. Sumanth Ethamukkala - 11651186

4. Manideep Renikindi – 11651184

**Goals and Objectives:**

* **Motivation:**

The growing demand for automated systems that can analyze and understand human language is the driving force behind the development of an NLP-based emotion detector. Feelings assume a significant part in human correspondence and understanding the feelings communicated through text can give a vague knowledge of how individuals feel about a specific point or circumstance. Businesses can use this information to improve their products or services, psychiatrists can use it to track the emotional state of their patients, and educators can use it to give students specific feedback.

By initializing new methods and models for emotion detection and increasing the accuracy of existing models, this project's development can also contribute to NLP and machine learning.

* **Significance:**

A better customer experience can be achieved by businesses using this approach to assess consumer input and improve their goods and services.

Mental health: Using this method, mental health providers can keep tabs on their patients' emotional health, spot indications of sadness or anxiety, and take immediate action if necessary.

Improvements in Education: Educational institutions can utilize this system to keep track of their students' emotional states and give them personalized feedback to help themstudy more quickly, which will lead to better exam achievement.

In almost every sector, emotion detection improves the user experience. Let us use the social science research: With this method, researchers can examine significant amounts of text-based data and learn about the emotions that various groups of individuals express, improving their comprehension of social interactions and human behavior.

Advanced NLP and Machine Learning Techniques: By studying novel methods and models for emotion recognition, the growth of this project can help advance NLP and machine learning, leading to increased accuracy and generalization.

Ultimately, by offering insightful information about human emotions and behavior, an emotion detector based on NLP has the chance to improve both the lives of its users and society overall. Retail and healthcare sectors as examples to further understand this. With the use of this technology, companies may examine the browsing and purchasing habits of their customers to develop better offers for them. Furthermore, face detection allows physicians to create better treatment plans and deliver services much more efficiently.

With emotion-detection technologies, the product industry can gain a deeper comprehension of the actual emotions of consumers. For identifying and evaluating the changing facial expressions that occur during discussion, businesses can set up a product testing period, record it, and then analyze it.

Using context, emotion detection may estimate and measure a person's feelings. It helps interviewers by providing insight into a person's mindset and personality traits.

The product industry can better understand the actual emotions of customers by using emotion detecting technology. Businesses can set up a product testing period, record it, and evaluate it to identify and assess that change all throughout conversation.

In almost every sector, emotion detection improves the user experience. Let us use the retail and healthcare sectors as examples to further understand this. With the use of this technology, companies may examine the browsing and purchasing habits of their customers to develop better offers for them. Furthermore, face detection allows physicians to create better treatment plans and deliver services much more efficiently.

**Objectives:**

Emotion prediction can be used to forecast client turnover, understand consumer behavior, and create specialized marketing strategies. Companies may create more potent methods to boost customer engagement and loyalty by understanding the emotions that influence consumer behavior. Mental health illnesses like depression and anxiety can be identified and predicted via emotion prediction. This can assist in the creation of more efficient interventions and treatment.

Emotion prediction can be used to improve human-robot interaction by giving robots the ability to recognize and react to human emotions. This could enhance human-robot interaction and result in more effective and natural communication. It might also make robotic technology more usable and popular.

In a range of fields, including web design, mobile apps, and gaming, emotion prediction can be utilized to enhance user experience. Companies can design goods and services that are more interesting to use and engaging by studying customer emotions, which increases customer happiness and loyalty.

* **Features:**

1. A method for collecting and evaluating massive amounts of annotated text data from multiple sources.
2. Preprocessing: A pipeline for filtering and processing unstructured text data into a form that is appropriate for models used in machine learning.
3. Extraction of informative features from preprocessed text data using an efficient feature extraction pipeline, such as Bag-of-Words, Word Embeddings, or other sophisticated methods.
4. Model selection: A range of machine learning models, including Deep Neural Networks, Naive Bayes, Support Vector Machines (SVMs), and Random Forests, are available.
5. Model Training: A procedure for creating and educating machine learning models using the retrieved features and preprocessed text data.
6. Model Deployment: A method for integrating the trained model into an API or web service that can take text input and output the predicted emotion label.
7. User Interface: A simple user interface that allows end users to engage with the system.
8. Scalability: The system must be scalable to accommodate high text data throughput as well as numerous concurrent user requests.
9. Security: Sensitive user data should be protected by the system's proper security measures, which should also guard against unauthorized access.

**Introduction:**

The growing demand for automated systems that can analyze and understand human language is the driving force behind the development of an NLP-based emotion detector. Feelings assume a significant part in human correspondence and understanding the feelings communicated through text can give a vague knowledge of how individuals feel about a specific point or circumstance. Businesses can use this information to improve their products or services, psychiatrists can use it to track the emotional state of their patients, and educators can use it to give students specific feedback. By initializing new methods and models for emotion detection and increasing the accuracy of existing models, this project's development can also contribute to NLP and machine learning.

A better customer experience can be achieved by businesses using this approach to assess consumer input and improve their goods and services.

Mental health: Using this method, mental health providers can keep tabs on their patients' emotional health, spot indications of sadness or anxiety, and take immediate action if necessary.

Improvements in Education: Educational institutions can utilize this system to keep track of their students' emotional states and give them personalized feedback to help themstudy more quickly, which will lead to better exam achievement.

In almost every sector, emotion detection improves the user experience. Let us use the social science research: With this method, researchers can examine significant amounts of text-based data and learn about the emotions that various groups of individuals express, improving their comprehension of social interactions and human behavior.

Advanced NLP and Machine Learning Techniques: By studying novel methods and models for emotion recognition, the growth of this project can help advance NLP and machine learning, leading to increased accuracy and generalization.

Ultimately, by offering insightful information about human emotions and behavior, an emotion detector based on NLP has the chance to improve both the lives of its users and society overall. Retail and healthcare sectors as examples to further understand this. With the use of this technology, companies may examine the browsing and purchasing habits of their customers to develop better offers for them. Furthermore, face detection allows physicians to create better treatment plans and deliver services much more efficiently.

With emotion-detection technologies, the product industry can gain a deeper comprehension of the actual emotions of consumers. For identifying and evaluating the changing facial expressions that occur during discussion, businesses can set up a product testing period, record it, and then analyze it.

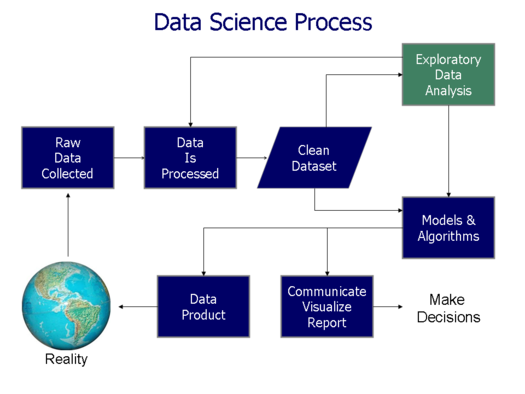
**Background Work:**

The "Human Emotion Identifier based on Text" project aims to develop a machine learning model that can accurately predict human emotions based on text data. Emotions are a fundamental aspect of human behavior and play a crucial role in communication, decision-making, and overall well-being. With the proliferation of text data in various forms, such as social media posts, customer reviews, and online comments, there is a growing interest in leveraging natural language processing (NLP) and machine learning techniques to automatically detect and analyze human emotions from text.

We refer medium website to take the information for our project:

<https://textrics.medium.com/what-are-the-methods-used-for-emotion-detection-in-text-analytics-838d7ca7e435>

**Architecture Diagram:**



**Dataset Description:**

# The dataset comprises of 58k comments from Reddit, meticulously curated and annotated by humans into 27 emotion categories or Neutral. It contains 5426 rows and 3 columns, with the 'text' column containing written statements, expressions, or messages, and the emotion categories including admiration, amusement, anger, smile, laughter, sad and heart.

# Detail design of Features:

# Data Exploration: Analyze the dataset to gain a better understanding of its contents, such as the structure, format, and distribution of the data. This can involve tasks such as data visualization, statistical analysis, and data profiling.

# The below code shows the shape of the dataset and also display top 5 values of the dataset.

# Graphical user interface, text, application Description automatically generated

# Analysis of data:

# Below is the code to check null values of the dataset.

Graphical user interface, text, application, email

Description automatically generated

Applying all the preprocessing steps stop words removal, stemming and lemmatization.

Graphical user interface, text, application

Description automatically generated

The bar chart displays the count of occurrences for each class number in the "Length of classes" column of data.

Chart, box and whisker chart

Description automatically generated

The below graph displays the count of occurrences for each emotion in the dataset.

Graphical user interface, application, Word

Description automatically generated

Below code is the function word cloud dictionary. This function useful to create the word cloud based each emotion.

Graphical user interface, text, application, Word

Description automatically generated

The below wordcloud shows each emotion in the dataset.

Smile laugh

 Text

Description automatically generated with medium confidence

Disappointing Joy

Text

Description automatically generated A picture containing text

Description automatically generated

**Implementation:**

**#Load the dataset into a pandas DataFrame**

**function pd.read\_csv()**

**#Preprocess the text data by removing stop words, stemming and lemmatization the words**

**//importing necessary libraries**

**import re**

**from nltk.corpus import stopwords**

**from nltk.tokenize import word\_tokenize**

**nltk.download('stopwords') //downloading stopwords**

**nltk.download('wordnet')**

**function preprocess\_text():**

**text = text.translate(str.maketrans('', '', string.punctuation))**

**text = text.lower()**

**text = re.sub(r'\d+', '', text)**

**text = re.sub(r'\s+', ' ', text).strip()**

**stop\_words = set(stopwords.words('english'))**

**ps = PorterStemmer()**

**tokens = [ps.stem(word) for word in tokens]**

**lemmatizer = WordNetLemmatizer()**

**tokens = [lemmatizer.lemmatize(word) for word in tokens]**

**text = ' '.join(tokens)**

**return text**

**#Extract features from the preprocessed text data using wordcloud**

**wordcloud = {} //declaring empty wordcloud dictionary object**

**for i in data.unique():**

**wordcloud [i] = {}**

**odds\_list = data [['emotion'] == i]['odds'].tolist()**

**word\_list = data [['emotion'] == i][ 'word'].tolist()**

**wordcloud[i] = dict(dict(zip(word\_list,odds\_list)))**

**return wordcloud**

**#Split the data into training and testing sets using train\_test\_split**

**from sklearn.model\_selection import train\_test\_split**

**X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2)**

**#Select a machine learning model, such as Naïve Bayes, Logistic Regression, SVM, Random forest**

**//Naïve Bayes model**

**from sklearn.naive\_bayes import MultinomialNB**

**model = MultinomialNB(alpha=1.0)**

**model.fit(X\_train, y\_train)**

**//logistic regression model**

**Function LogisticRegression()**

**model.fit(X\_train, y\_train)**

**//Random forest model**

**from sklearn.ensemble import RandomForestClassifier**

**function RandomForestClassifier(n\_estimators=100, random\_state=42)**

**model.fit(X\_train\_vec, y\_train)**

**//Support vector machine model**

**from sklearn.svm import SVC**

**function SVC()**

**model.fit(X\_train\_vec, y\_train)**

**#Evaluate the model performance on the testing data using metrics like accuracy, precision, recall, and F1-score**

**model.predict(X\_test)**

**function accuracy\_score(y\_test, y\_predict)**

**print(f"Accuracy: {accuracy:.2f}")**

**//Print classification report and confusion matrix**

**print("Classification Report:")**

**print(classification\_report(y\_test, y\_predict))**

**#Using the emoji library in Python to convert class labels into corresponding emojis based on a predefined mapping.**

**import emoji**

**y\_new\_pred\_emojis = []**

**for label in y\_predict\_log:**

**str\_label = str(label) if str(label) in emoji\_mapping else '1'**

**emoji\_alias = emoji\_mapping[str\_label]**

**emoji\_emoji=emoji.emojize(emoji\_alias,language='alias') y\_new\_pred\_emojis.append(emoji\_emoji)**

**print("Predicted Emojis for Test Data:")**

**for i in range(len(X\_test)):**

**print("Sentence: " + X\_test.iloc[i])**

**random\_emoji=random.choice(y\_new\_pred\_emojis)**

**random\_emoji = emoji.emojize(random\_emoji, language='alias')**

**print("Predicted Emoji is: " + random\_emoji)**

Data Preprocessing: The first component involves collecting, cleaning, and preprocessing the text data. This can include steps such as removing stopwords, stemming, and lemmatizing. In the project, the dataset is read from a dev file and split into training and test sets using the train\_test\_split function from scikit-learn.

Feature Extraction: The next component is featuring extraction, which involves converting the raw text data into numerical features that can be used as input to machine learning models. In the project, the CountVectorizer class from scikit-learn is used to convert the text data into a matrix of word counts.

Model Training: The third component is model training, which involves using the feature matrix and corresponding labels to train a machine learning model. In the project, three models are trained: Naive Bayes, Logistic Regression, and Random Forest. These models are implemented using scikit-learn.

Prediction: The final component is prediction, which involves using the trained models to predict the emotions associated with new text data. In the project, the models are used to predict the emotions associated with the test data. The predicted emotions are then mapped to corresponding emojis using a dictionary and the emoji module.

**Results:**

# Model Selection and Training: Various machine learning models such as logistic regression, support vector machines and Naïve bayes, can be used for emotion identification. The steps involve selecting an appropriate model or ensemble of models, training them on the preprocessed data, and tuning hyperparameters to optimize their performance.

# The below codes show the accuracy of Logistic regression and Support vector machine models.

# Text Description automatically generated with low confidence

# Text Description automatically generated

# Below is the code for Naïve bayes model and we are getting high accuracy value compare to other models.

# Table Description automatically generated with medium confidence

# We are trying to predict the emojis on logistic regression test data.

# Graphical user interface, text, application, email Description automatically generated

# We are trying to predict the emojis on SVM model test data.

# Graphical user interface, text, application Description automatically generated

# We are trying to predict the emojis on Naïve Bayes model test data.

# Graphical user interface, text, application, email Description automatically generated

# We are trying to predict the emojis on Random forest model test data.

# Graphical user interface, text, application, email Description automatically generated

# Project Management:

**Responsibility:**

**1.Varsha Reddy Umannagari:** I have contributed for the project by using SVM Model which is a supervised learning machine used for classification, here sk-learn module is used to find the accuracy with the training data and metrics. and I have predicted the emojis on SVM Model for the data set.

**2. Benerji Vigna Sai Rama Venkat Sumith Thota**: My contribution is on logistic regression is a linear model in sk-learn module. to find the accuaracy of the given dataset and printed the classification report along with it. and I have predicted the emojis on Logistic Regression Model for the data set.

**3. Sumanth Ethamukkala –** I have done the naive bayes model for the project to find the accuracies and print the Precision, recall and f1 scores, and I have predicted the emojis on Naive Bayes model for the data set.

4. **Manideep Renikindi-** My contribution in project is to find accuracies using random forest model along with that precision, recall and f1 scores, and I have predicted the emojis on Random Forest Model for the data set.

**Contributions:**

1. Varsha Reddy Umannagari -25%

2. Benerji Vigna Sai Rama Venkat Sumith Thota – 25%

3. Sumanth Ethamukkala -25%

4. Manideep Renikindi –25%

# References:

**Dataset-** [**https://www.kaggle.com/datasets/debarshichanda/goemotions**](https://www.kaggle.com/datasets/debarshichanda/goemotions)

**GitHub-** <https://github.com/UVarshaReddy/HUMAN-EMOTION-IDENTIFIER-USING-TEXT-CLASSIFICATION.git>

1.Y. An, S. Sun and S. Wang, "Naive Bayes classifiers for music emotion classification based on lyrics," 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, China, 2017, pp. 635-638, doi: 10.1109/ICIS.2017.7960070.

2. F. Noroozi, D. Kaminska, T. Sapinski, and G. Anbarjafari, "Supervised Vocal-Based Emotion Recognition Using Multiclass Support Vector Machine, Random Forests, and Adaboost," J. Audio Eng. Soc., vol. 65, no. 7/8, pp. 562-572, (2017 July.). doi: https://doi.org/10.17743/jaes.2017.0022