# Sentiment analysis of text

Introduction:

With all of the tweets, reviews and feedbacks circulating every second it is hard to tell whether the sentiment behind a specific tweet will **impact a company, or a person's, brand for being viral or devastate profit** because it strikes a negative tone. Capturing sentiment in language is important in these times where decisions and reactions are created and updated in seconds. But, which words actually lead to the sentiment description.

Importance of Sentiment Analysis:

In today's digital age, where information flows ceaselessly through social media, online reviews, and customer feedback, Sentiment analysis empowers us to harness the wealth of unstructured text data to gain insights into public opinion, customer satisfaction, and emotional responses ,**Impact on Decision-Making.**

"Today, I'll be discussing a machine learning task I have done on DEPI internship . We'll go from uploading data to visualizing it, and finally building a full pipeline using Python libraries . the model selected which is SVM(support vector machine). The key tools I'll use include **pandas** for data handling, **NumPy** for numerical operations, and **matplotlib** for visualizations. Everything will be implemented in a **Jupyter Notebook** to ensure a clear and interactive workflow.

The key steps we'll cover are:

* **Exploratory Data Analysis (EDA)**: Understanding the data we have.
* **Data Clean and Preprocessing**: Preparing the data for analysis.
* **Feature Engineering**: Creating features that improve model performance.
* **Model Architecture**: Defining the structure of our machine learning model.
* **Training the Model**: Fitting our model to the training data.
* **Model Evaluation**: Assessing the model's performance.
* **Generating the Full Pipeline**: Integrating all steps into a seamless workflow.



First we load the(train,test) data:-

A white background with black and red dots

Description automatically generated

A screenshot of a computer

Description automatically generatedthen we visualize it using a simple bar chart:-

**Data Clean and Preprocessing**

Define a function textClean to preprocess text :-

A screenshot of a computer code

Description automatically generated

This function do:

* Remove **punctuation**
* **Lowercasing characters**
* **Tokenization**
* **Removing URLs**
* **Removing Stopwords**

A blue and purple gear and a white sheet of paper with a blue dot and a blue square with a green square and a blue dot with a blue dot and a blue dot with a blue dot

Description automatically generated with medium confidence

**Feature Engineering**

**Vectorize the cleaned text using CountVectorizer to transform it into a numerical format for machine learning**

Analyzer purpose is to use “textClean” on all text that will be converted to vector to clean them to improve accuracy

This what vectorizer do:

A white background with black text

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer screen

Description automatically generated

Model  
**Architecture**

A diagram of different colors

Description automatically generated

**Support Vector Machine (SVM):**

SVM is a supervised learning algorithm used for classification and regression tasks.It works by finding the hyperplane that best separates data points belonging to different classes while maximizing the margin between classes.

**Key Characteristics:**

Non-probabilistic binary linear classifier

Effective in high-dimensional spaces

**Use Cases:**

Highlight specific use cases where SVMs excel, such as text classification, image classification, and bioinformatics.

**SVM Architecture:**

* SVM constructs a hyperplane in a multidimensional space that separates data points into different classes.
* The choice of kernel function (linear, radial basis function, polynomial) influences the decision boundary Advantages.

**Model Training:**

* During training, the SVM learns the optimal hyperplane that maximizes the margin between classes.
* The choice of kernel and hyperparameter values significantly influences the model's performance.

**Training the  
 Model**

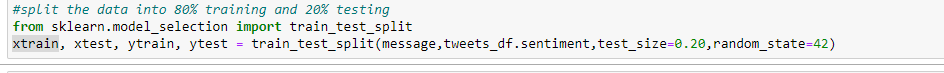
A machine learning logo with blue lines

Description automatically generated with medium confidence

**Spilting data:**

**To data without target value and target value data to make model**

**Train on it**



**Kernel Functions:**

**Linear Kernel:**

**Description: A linear kernel assumes a linear decision boundary.**

**Use Case: Suitable for linearly separable data where classes can be cleanly divided by a straight line or hyperplane.**

**RBF (Radial Basis Function) Kernel:**

**Description: Radial Basis Functions create a non-linear decision boundary by mapping data into a higher-dimensional space.**

**Use Case: Effective for capturing complex relationships and handling non-linear separations. Widely used for its versatility.**

**Polynomial Kernel:**

**Description: Polynomial kernels create decision boundaries with polynomial shapes.**

**Use Case: Useful for data that exhibits polynomial relationships.**

A screenshot of a computer program

Description automatically generated

We used linear because our data is in a straight line (each line don’t depend on other lines)

A clipboard with check marks and a pencil

Description automatically generated**Model Evaluation**

Evaluate the SVM model on the training data:

A screenshot of a computer

Description automatically generated

**Evaluate the SVM model on the test data:**

A screenshot of a computer

Description automatically generated

**Visualize the confusion matrix for the model's predictions:**

A screenshot of a computer

Description automatically generated

Make predictions on new entries text samples:

A screenshot of a computer

Description automatically generated

Making a full pipeline by loading preprocessing and analyzing the data(top 500 rows of TeePublic\_review.csv) then saving the new file with the predicted sentiment as a new column.

A screenshot of a computer

Description automatically generated

Future work:-

* use it in ecommerce website .
* building it in form of webpage where you write text and return sentiment maybe.
* build an API to allow other websites to use it.
* improve model and train it on more data.
* improve vectorizer.

Thanks

For reading

Feel free to send me any question?

Abdelrahmanmoataz888@gmail.com