Beykoz University

Department of "Computer Engineering"

"Big Data & Data Analytics - 60613MEEOS-CME0297"

Project II - Final Report

- Twitter Sentiment Analysis/Orange -

Lecturer: FARHAD PANAHIFAR

Leyla Abdullayeva - 1904010038

Aim of the project:

The aim of my project is to perform sentiment analysis on Twitter data using the Orange software, focusing on Big Data and Data Analytics. I will utilize the power of Natural Language Processing (NLP) to analyze tweets and determine whether they convey positive, negative, or neutral sentiment. To achieve this, I will employ VADER, a specialized lexicon and rule-based sentiment analysis tool tailored for social media. By leveraging the capabilities of Orange software, I will gain valuable insights into sentiment trends and patterns within Twitter data, contributing to the field of data analytics. This project will provide a deeper understanding of public opinion on various topics in the realm of social media, showcasing the potential of sentiment analysis in uncovering valuable insights from vast amounts of textual data.

What is “Sentiment Analysis” in the Big Data & Analytics field?

"Twitter Sentiment Analysis" in the Big Data and Analytics field refers to the process of analyzing the sentiments expressed in tweets posted on the social media platform Twitter. It involves using advanced techniques from Natural Language Processing (NLP) and Machine Learning to classify tweets as positive, negative, or neutral based on the emotions, opinions, or attitudes conveyed in the text. The goal is to extract meaningful insights from the massive volume of Twitter data and understand the overall sentiment trends and patterns associated with specific topics, brands, events, or public figures. By analyzing sentiments on Twitter, organizations can gain valuable information about public opinion, customer feedback, brand perception, and emerging trends, allowing them to make data-driven decisions, improve marketing strategies, enhance customer experience, and monitor their online reputation. It is an essential application of Big Data and Analytics, enabling organizations to harness the power of social media data for actionable insights.

About:

Natural Language Processing (NLP) is at the core of research in data science these days and one of the most common applications of NLP is sentiment analysis. Also known as “Opinion Mining” or “Emotion AI”, Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral.

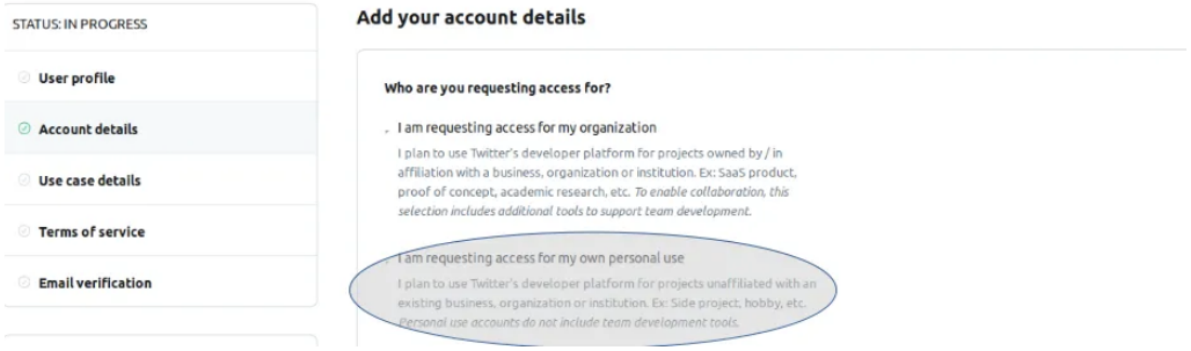
There can be two approaches to sentiment analysis.

1. Lexicon-based methods
2. Machine Learning-based methods.

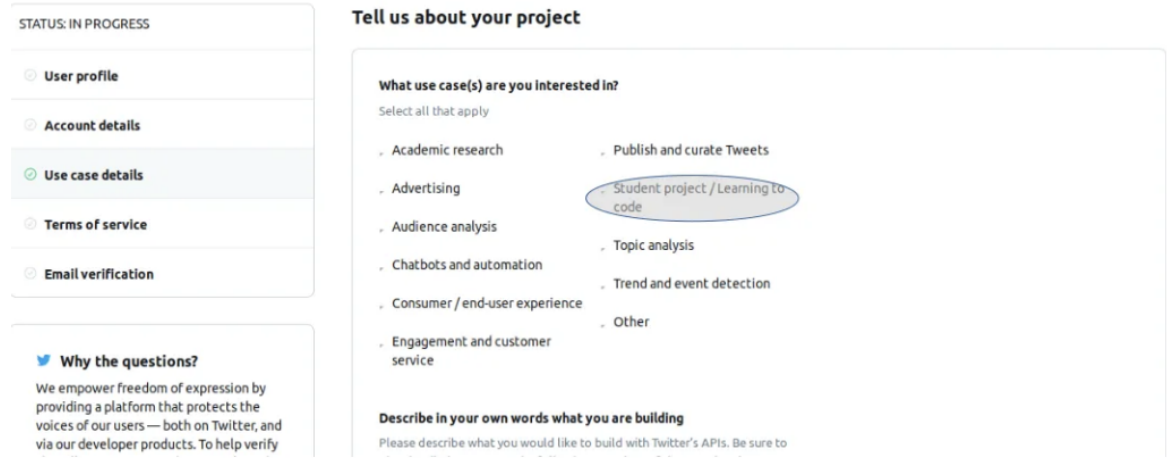
We will be using VADER (Valence Aware Dictionary and sEntiment Reasoner) a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.

Project Steps:

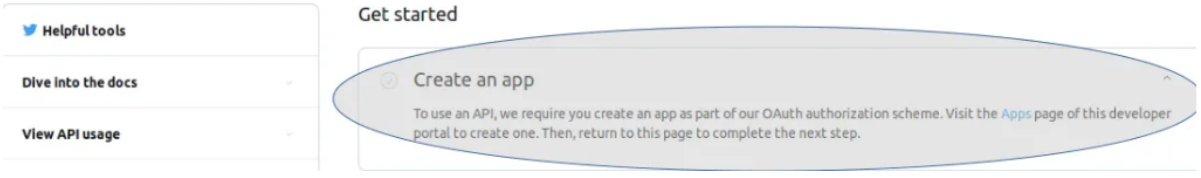
Step 1: Getting the Twitter API Credential
To access the developer account, you need to have a twitter account. To access the Twitter API, you need to register an application at http://apps.twitter.com. On the top-right corner, click on the Apps button, Create an App, Apply and then Continue. Next, we will choose the "I am requesting access for my own personal use" option:



On the same web page, scroll down a bit and input your Account name and Country of operation then click Continue, and you will be redirected to the next web page. Here, you can choose any Use Cases you're interested in. For our case, I chose the following:

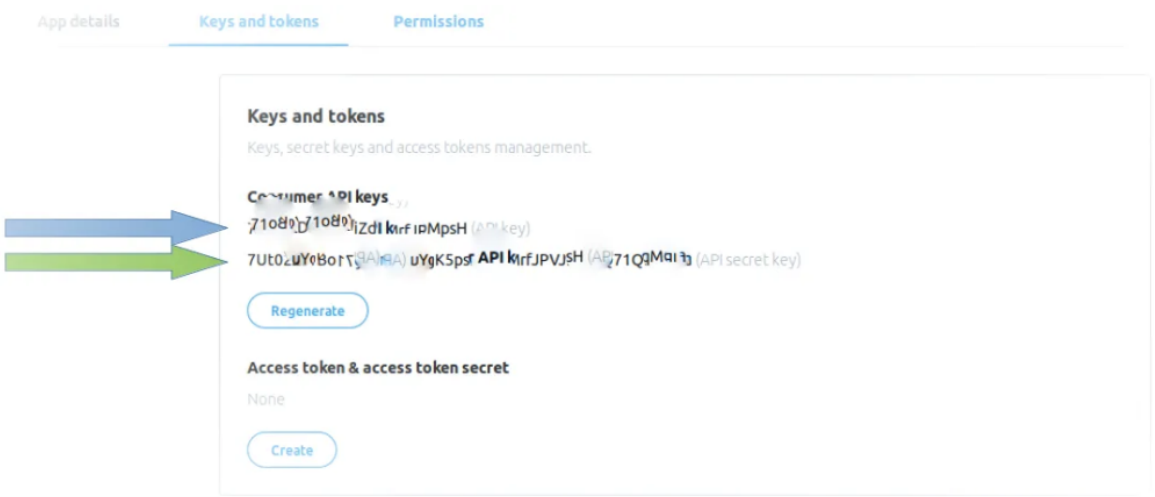


After you make your choice, scroll down and fill out the use case interest paragraph required. This tutorial is for learning, so make sure you emphasize on the application being a self-learning/academic-related project. Choose "No" for the government involvement question, and press "Continue". On the next web page, read the Terms and Conditions list, agree to them then Submit Application. Now, you have to wait for twitter to verify your developer account. When you get the approval email, click on the login link it contains. You will be redirected to the following web page, where you should choose "Create an app".



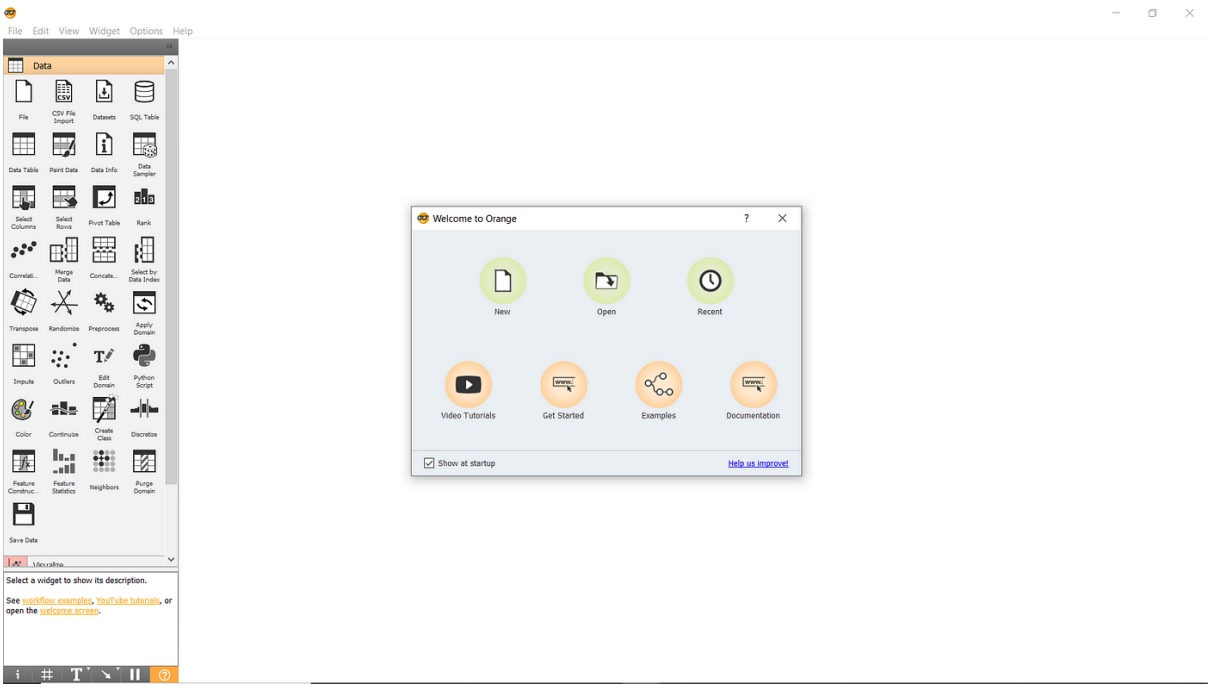
On the next web page, click “Create an app” from the top-right corner. After you are redirected, fill out the required app details, including — if you’d like — that it is for self-learning purposes. Click “Create”.

The next web page will include the app details that you just input, access tokens and permissions. Proceed to the “Keys and tokens” tab. Copy the API key as well as the API secret key into a safe place (a text file, if you’d like), as we will be using them in a bit. We’re done with the credential acquisition part!



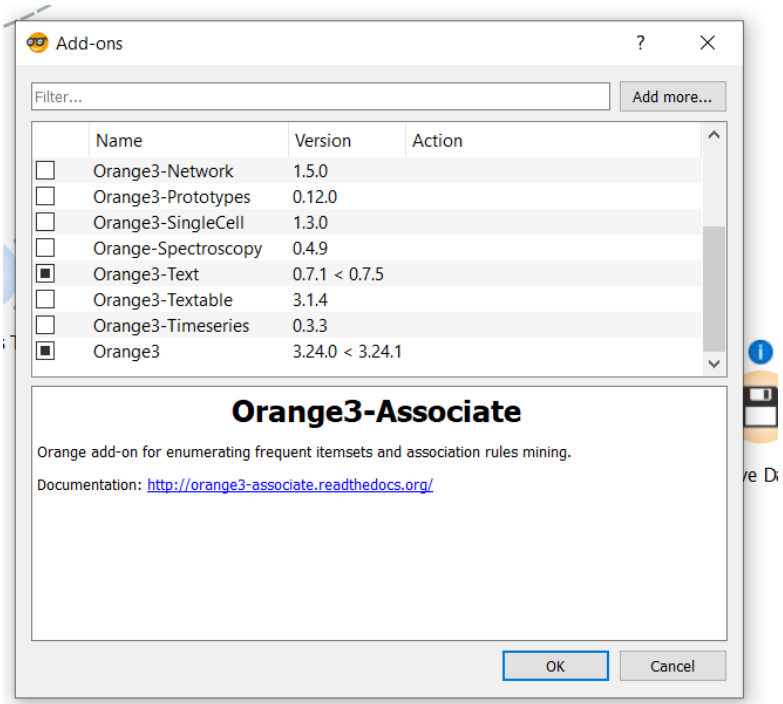
Step 2: Orange3 + Vader for Twitter Streaming and Sentiment Analysis

Now that you have installed Orange3, open the application:

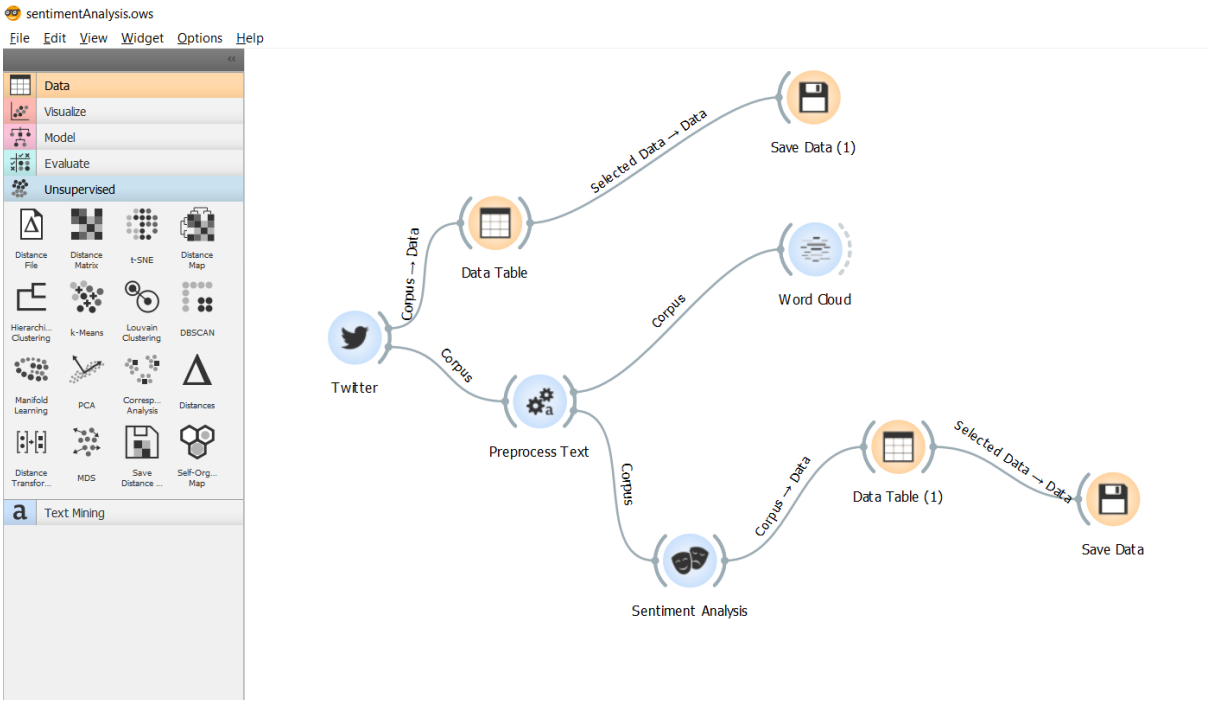


This is the welcome page for Orange3

Click “New” to launch a blank canvas. Orange3 offers a lot of analytics capabilities for data preprocessing, visualization, statistical modeling, and machine learning. Text Analysis doesn’t come with Orange3 by default, so, I need to install the Orange-Text addon. To install, click “Options” on the home ribbon and select “Add-ons…”. Check the Orange3-Text and click okay; wait for the add-on to install.



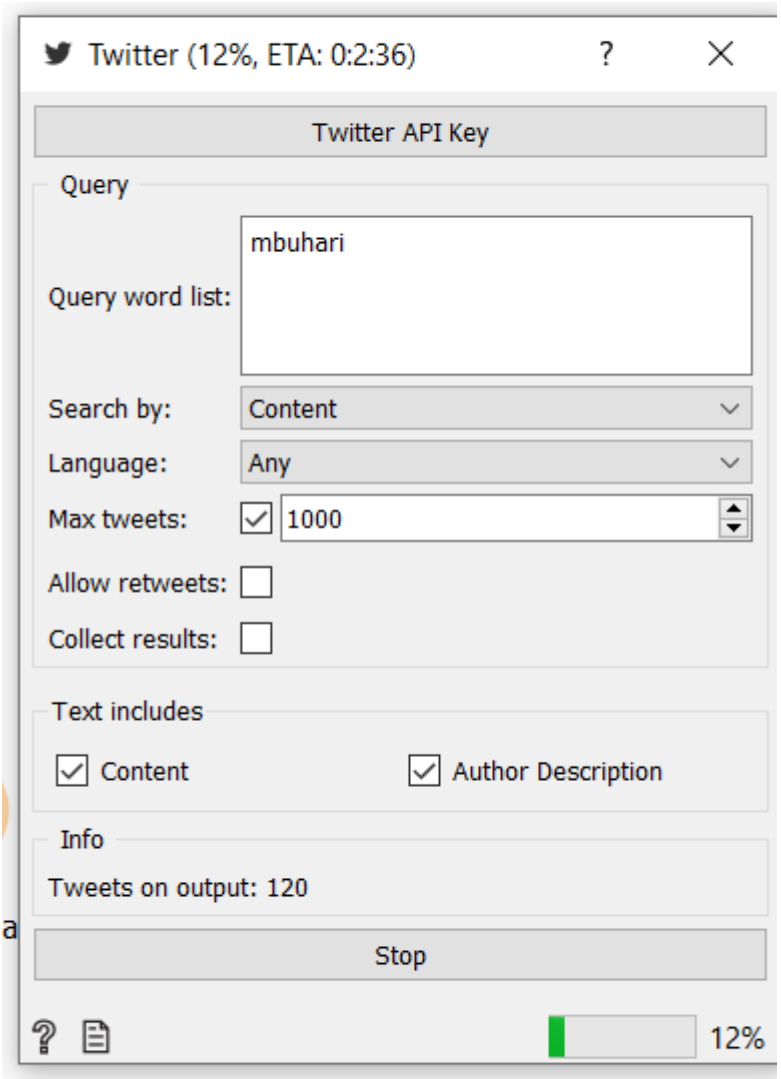
Now that you have the Text add-on installed, let’s build the flow!



Follow these steps to build the Orange3 sentiment analysis flow:

1. The Twitter Widget: Expand the *Text Mining* drop down on the left panel; drag and drop the 'Twitter' widget to the canvas.
2. The Data Table: Expand the *Data* drop-down on the left panel; drag and drop the 'Data Table' widget to the canvas. Connect the 'Twitter' widget to the 'Data Table' by dragging any part of the dotted arc of the 'Twitter' widget to the 'Data Table' widget. NB: This is how you create connections between widgets.
3. The Save Data: Expand the *Data* drop-down on the left panel; drag and drop the 'Save Data' widget to the canvas. Connect the 'Data Table' widget to the 'Save Data'.
4. The Preprocess Text: Expand the *Text Mining* drop down on the left panel; drag and drop the 'Preprocess Text' widget to the canvas. Connect the 'Twitter' widget to it.
5. The Word Cloud: Expand the *Text Mining* drop down on the left panel; drag and drop the 'Word Cloud' widget to the canvas. Connect the 'Preprocess Text' widget to it.
6. The Sentiment Analysis: Expand the *Text Mining* drop down on the left panel; drag and drop the 'Sentiment Analysis' widget to the canvas. Connect the 'Preprocess Text' widget to it.

Double click the 'Twitter' widget and the configuration window opens up:



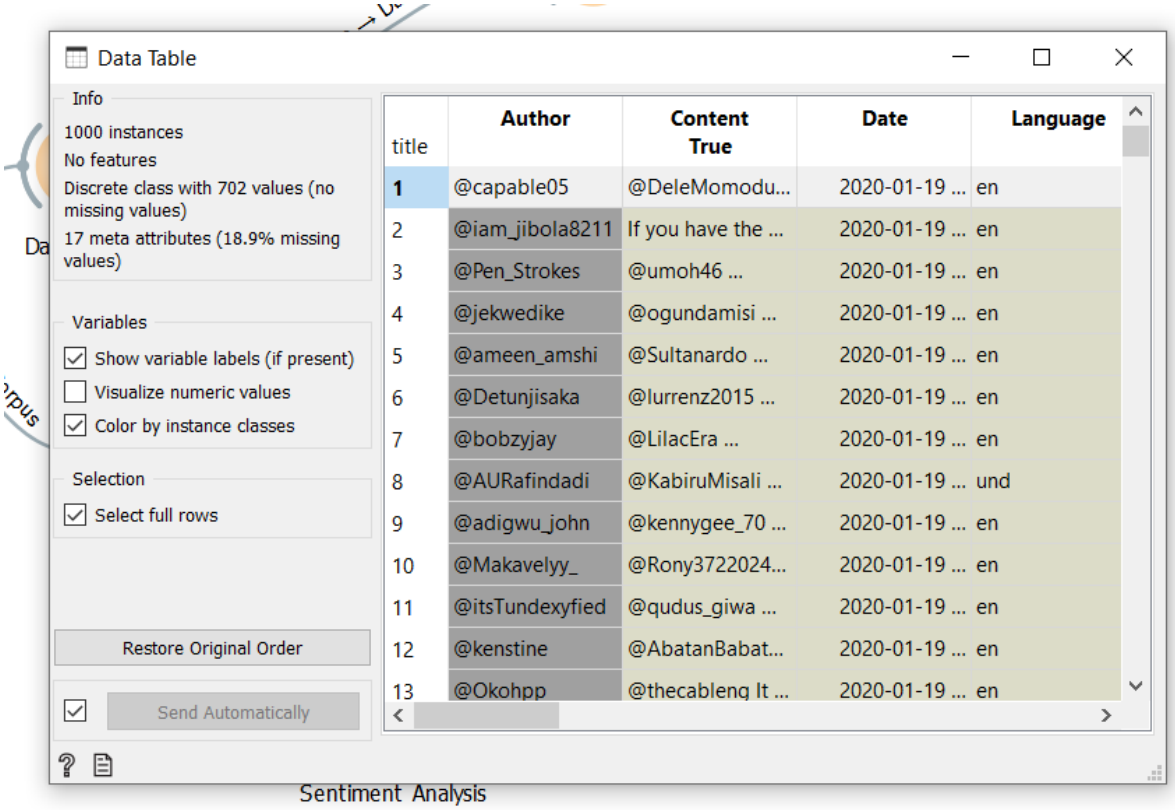
Click the ‘Twitter API Key’ button and input you ‘Consumer API Key’ and ‘Secret Key’ that was generated in

Step 1

For this tutorial, we will stream 1000 tweets where the word mbuhari(the official twitter handle of the Nigerian President, President Muhammadu Buhari GCFR) was mentioned. So, if you would like to follow this article religiously, input ‘mbuhari’ in the query word list box and set the max tweets to 1000. Click ‘Start’ to start streaming data from twitter. (tweets streamed on January 19th, 2020)

The Data Table Widget

In this workflow, we made use of two ‘Data Table’ widgets. This widget allows us to view the data in a table format. The first data table is connected to the ‘Twitter’ widget. We can view this data by double-clicking the ‘Data Table’ widget.

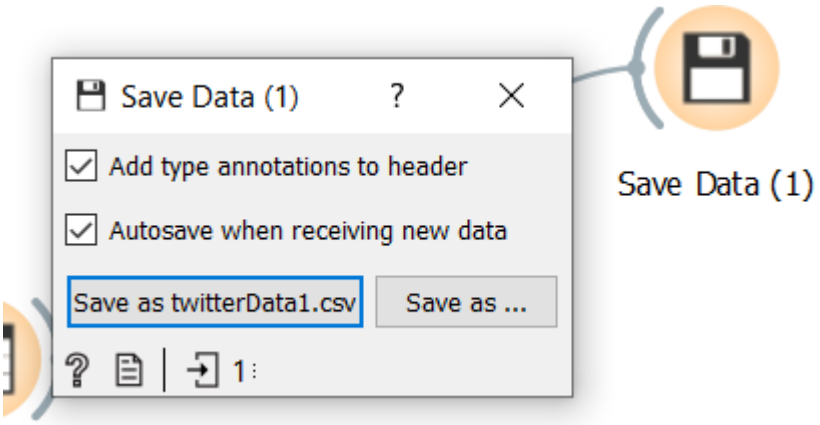


title	Author	Content True	Date	Language
1	@capable05	@DeleMomodu...	2020-01-19 ...	en
2	@iam_jibola8211	If you have the ...	2020-01-19 ...	en
3	@Pen_Strokes	@umoh46 ...	2020-01-19 ...	en
4	@jekwedike	@ogundamisi ...	2020-01-19 ...	en
5	@ameen_amshi	@Sultanardo ...	2020-01-19 ...	en
6	@Detunjisaka	@lurrenz2015 ...	2020-01-19 ...	en
7	@bobzyjay	@LilacEra ...	2020-01-19 ...	en
8	@AURafindadi	@KabiruMisali ...	2020-01-19 ...	und
9	@adigwu_john	@kennygee_70 ...	2020-01-19 ...	en
10	@Makavelyy_	@Rony3722024...	2020-01-19 ...	en
11	@itsTundexyfied	@qudus_giwa ...	2020-01-19 ...	en
12	@kenstine	@AbatanBabat...	2020-01-19 ...	en
13	@Okohpp	@thecableng It ...	2020-01-19 ...	en

The Data Table

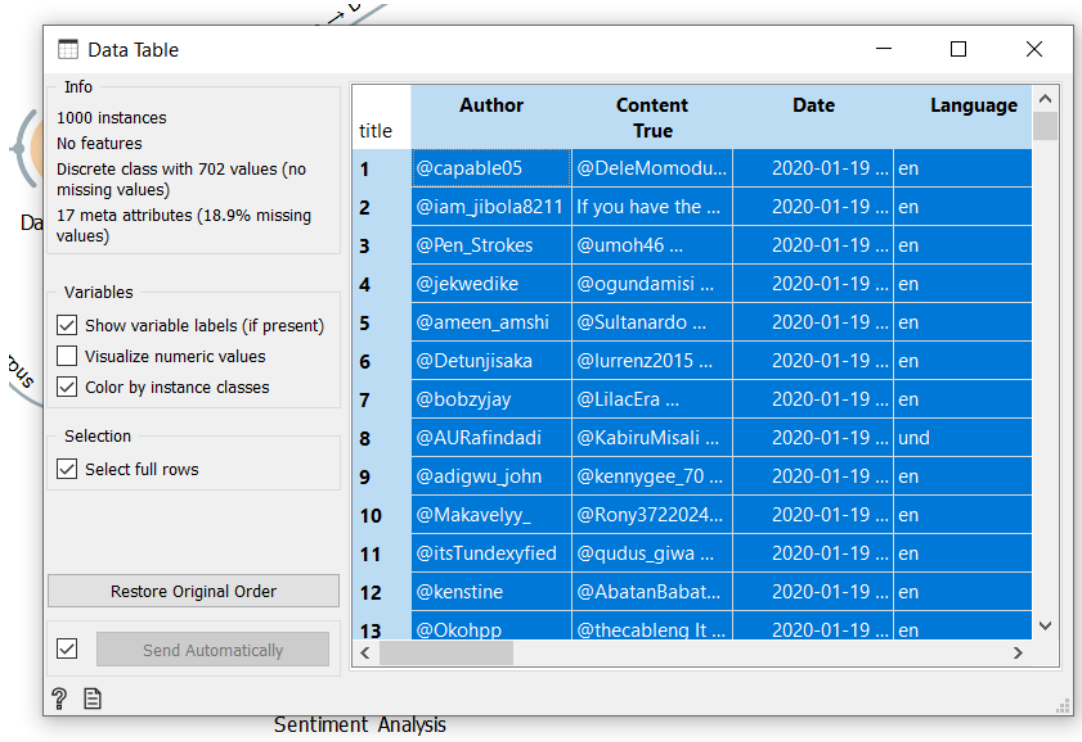
The Save Data Widget

This widget saves data from the data table as .csv. Double click this widget to define the name and path for your file.



The Save Data window

Note: 'Save Data' will only save the highlighted data on the 'Data Table'. So, to highlight data on the data table, double click the 'Data Table' widget to open the 'Data Table' Window. Double click the 'title' on the top left corner of the table, the entire data table is highlighted and automatically saved to .csv



title	Author	Content True	Date	Language
1	@capable05	@DeleMomodu...	2020-01-19 ...	en
2	@iam_jibola8211	If you have the ...	2020-01-19 ...	en
3	@Pen_Strokes	@umoh46 ...	2020-01-19 ...	en
4	@jekwedike	@ogundamisi ...	2020-01-19 ...	en
5	@ameen_amshi	@Sultanardo ...	2020-01-19 ...	en
6	@Detunjisaka	@lurrenz2015 ...	2020-01-19 ...	en
7	@bobzyjay	@LilacEra ...	2020-01-19 ...	en
8	@AURafindadi	@KabiruMisali ...	2020-01-19 ...	und
9	@adigwu_john	@kennygee_70 ...	2020-01-19 ...	en
10	@Makavelyy_	@Rony3722024...	2020-01-19 ...	en
11	@itsTundexyfied	@qudus_giwa ...	2020-01-19 ...	en
12	@kenstine	@AbatanBabat...	2020-01-19 ...	en
13	@Okohpp	@thecableng It ...	2020-01-19 ...	en

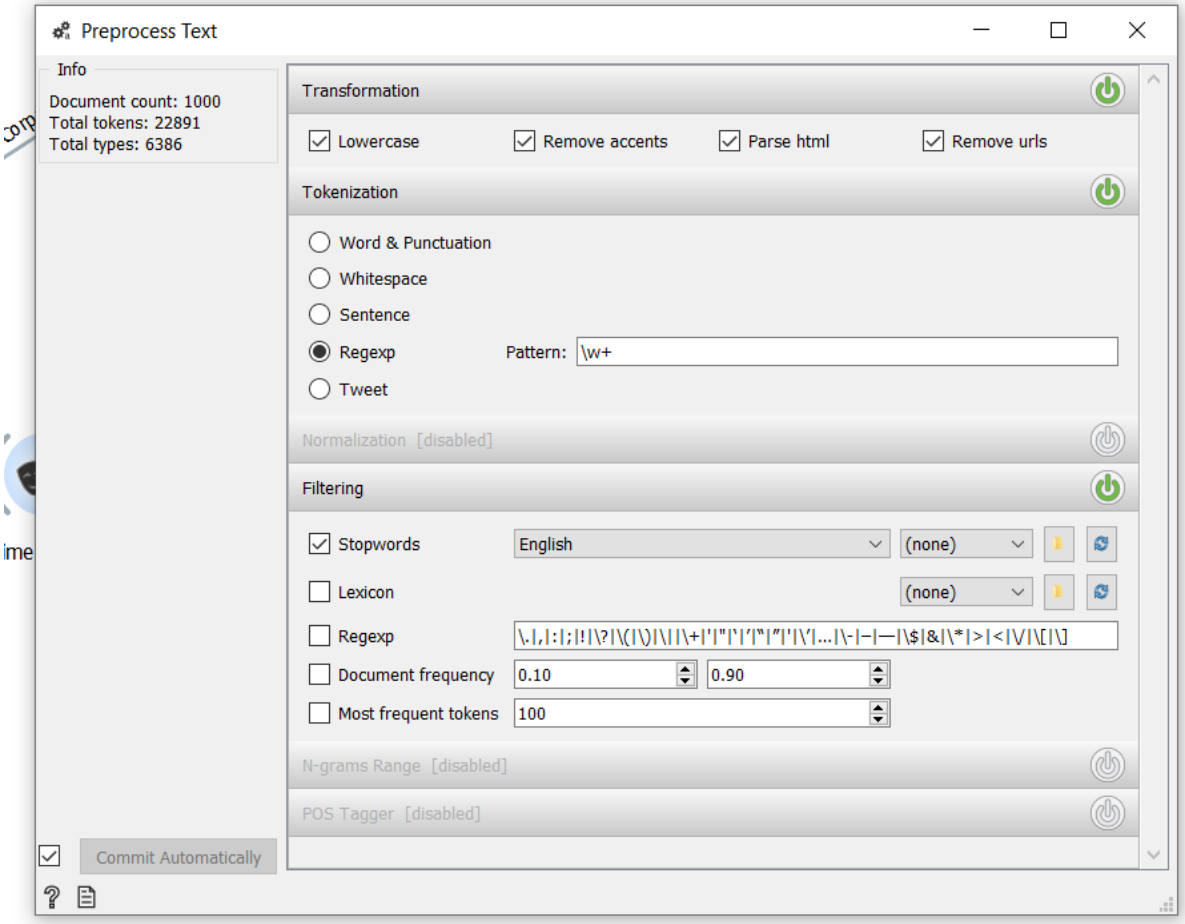
Highlighted data table will be saved

The Pre-process Text

This constructs a text pre-processing pipeline. It allows us to transform, tokenize and filter our data. Double click the widget to open the Preprocess window. We want to **transform** our data by maintaining lower case in all tweets, removing accents, parse HTML and removing URLs; so, please check all the boxes under the **Transformation** section.

Under **Tokenization,** we are only interested in splitting by regular expressions and keeping only words. Select **Regexp** and type ***\w+*** as the pattern.

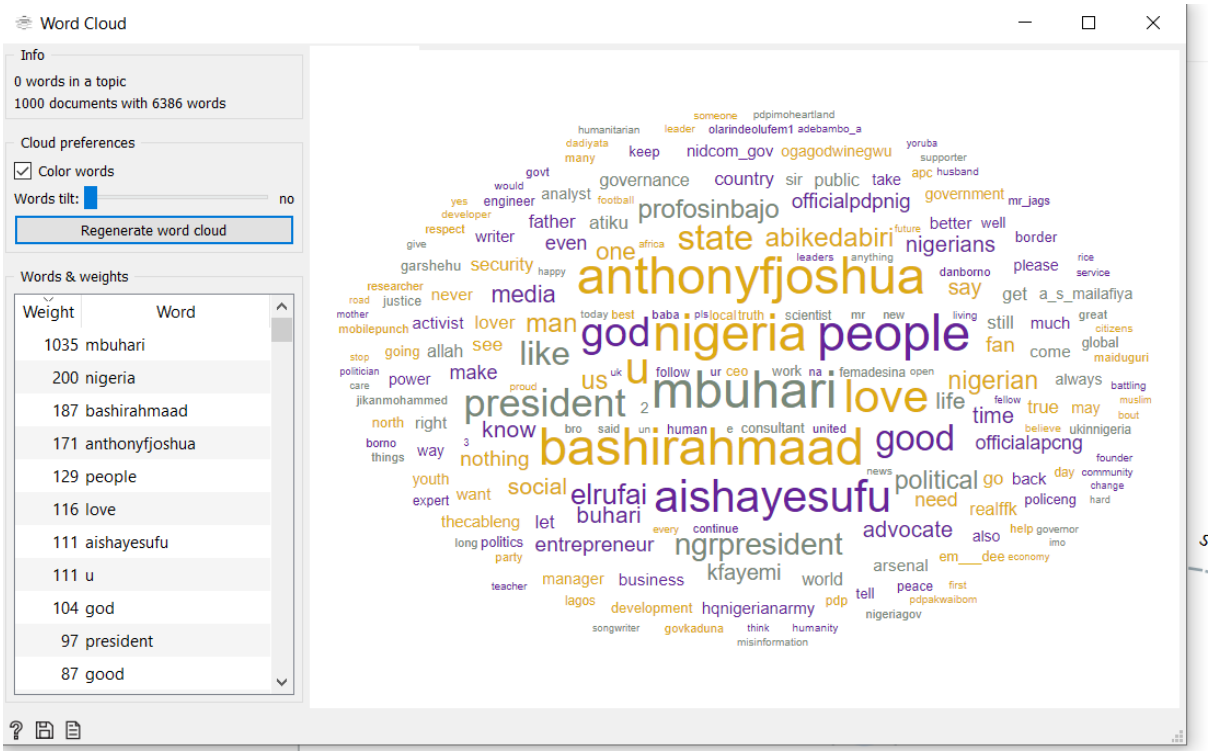
Under Filtering, we will remove stopwords in the English language. So, check 'Stopwords' and set language to 'English'



Preprocess Text window

The Word Cloud

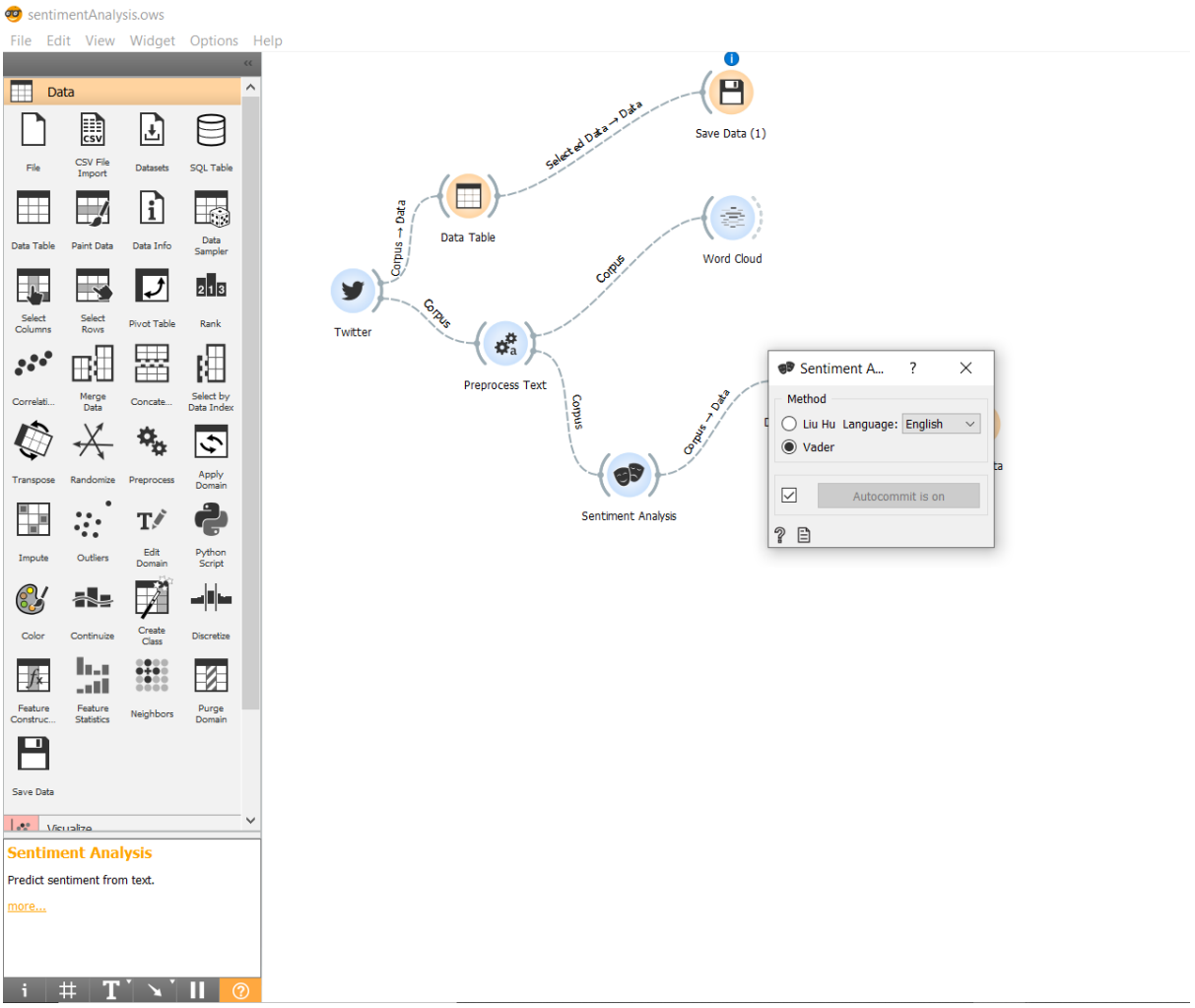
This widget is one of my favorite text analytics visuals. It helps you see the most mentioned words in text data. See what our cloud looks like:



The Word Cloud

The Sentiment Analysis

Double click this widget and select Vader. VADER uses a combination of a list of lexical features (e.g., words) which are generally labeled according to their semantic orientation as either positive or negative. VADER not only tells about the Positivity and Negativity score but also tells us about how positive or negative a sentiment is.



Future use:

The future use of this project holds great potential in various domains. With the continuous growth of social media and the increasing importance of online public sentiment, Twitter Sentiment Analysis using Big Data and Analytics techniques will remain highly relevant. Organizations can leverage the insights gained from this project to refine their marketing strategies, enhance brand reputation management, and improve customer experience. The ability to accurately analyze and understand the sentiments expressed on Twitter can aid in predicting consumer behavior, identifying emerging trends, and conducting market research. Additionally, in the field of social and political analysis, this project can contribute to understanding public opinion, sentiment shifts, and the impact of events or policies. Moreover, the findings from this project can be extended to other social media platforms, providing a broader scope for sentiment analysis and enabling comprehensive monitoring of online sentiment. As technology advances, the integration of advanced machine learning algorithms and real-time data processing capabilities will further enhance the accuracy and timeliness of Twitter sentiment analysis, making it an indispensable tool for decision-making in the era of Big Data and Analytics.

REFERENCES:

1. Orange: Data Mining Toolbox in Python. Retrieved from https://orange.biolab.si/
2. Twitter API Documentation. Retrieved from https://developer.twitter.com/en/docs
3. "Twitter Sentiment Analysis with Python" - Step-by-step guide: https://towardsdatascience.com/twitter-sentiment-analysis-classification-using-nltk-python-fa912578614c
4. "Sentiment Analysis on Twitter Data" - Case study and code implementation: https://www.datacamp.com/tutorial/simplifying-sentiment-analysis-python