# **E08** – Sentiment Analysis (Python)

Business Intelligence

Exercise

Winter Term 2025/2026

# Introduction - Usage of Sentiment Analysis

- Social networks represent a fundamentally new instrument to conduct social measurements. Millions of people voluntarily express opinions on any topic one could imagine. And such data source is extremely valuable for both research and business. Researchers from behavioral sciences could potentially try confirm or reject a particular hypothesis about people just based on a data collected (anonymized) from such sources. For instance, companies could try to understand opinions about their products and initiatives on much granular level. For example, researchers have shown that the "mood" of communication on twitter via sentiment analysis could be potentially estimated with a very high precision.

- In this assignment, you will
  - estimate a perception (a sentiment) of a particular phrase
  - estimate a public's perception (a sentiment) of a particular term or phrase

# Notes on Software (I/III) - Python

- The current exercise uses Python / Jupyter Notebook, which is an open platform-as-a-service that provides a very popular opportunity to work with Python directly in your browser. Python as a programming language is quite widespread nowadays and got an amazing community. Despite availability of a *Natural Language Processing (NLP)* capabilities in in form of various packages in Python ecosystem, your task will be to implement sentiment analysis on your own and then try to perform sentiment analysis by using provided inside the database methods.

- **NOTE:** If you are familiar with any similar tool/software, which you could potentially use to implement the current assignment you are free to use it

# Notes on Software (I/III) - Jupyter Notebook

Use one of the options below to setup working enviroment

- Option 1. Locally installed Python
  - Python
    - https://www.python.org/downloads/release/python-3128/
  - Information on tools (IDE, packages, etc.)
    - https://github.com/vdmitriyev/uol-data-analytics?tab=readme-ov-file#gear-tools
- Option 2. Locally installed Jupyter Notebook
  - You can install it on your PC and then use it, here you are going to find more information about the installation of this tool (as part of the Anaconda Distribution) - https://www.anaconda.com/products/individual
- Options 3. Remote SaaS
  - You can also use online service that offer Python environment (e.g., Google Colab)

# Notes on Software (III/III) - Infrastructure

If you are new to Python, here you are going to find all required for this assignment materials to master Python programming

https://github.com/vdmitriyev/uol-data-analytics

# Sentiment Analysis

According to the Wikipedia a **sentiment analysis** determines the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation, affective state (emotional state of the author when writing), or the intended emotional communication (the emotional effect the author wishes to have on the reader).

Sentiment analysis is widely applied to review social media in a variety of applications (e.g., marketing, customer service, etc.).

**Examples of the sentiments in the given text**

| Text | Sentiment |
|---|---|
| I really **like** new book of that author. | Positive |
| I **hate** new regulations about importing policies. | Negative |

# Task 1 – Understanding Sentiment of an Input Text

## Task Description

You task is to implement a small system that will provide sentiment analysis of a single sentences that will be given to it as an input. Response of the system should be **positive**, **negative** or neutral based on the "estimated" sentiment of input text. The estimation should be done according to the sentiment algorithms provided. Algorithm is very simple, but yet can be useful to derive interesting insights from an unstructured texts of social media.

## Sentiment Algorithms Description

*The sentiment of each input should be compute based on the sentiment scores of the terms in given sentences.* The sentiment of a input text is equivalent to **the sum of the sentiment scores** for each term **found inside input text**.

The file AFINN-111.txt contains a list of pre-computed sentiment scores. Each line in the file contains a word or phrase followed by a sentiment score. Each word or phrase found in a inside input text, but not in AFINN-111.txt should be given a sentiment score of 0. See the file AFINN-README.txt for more information. Check supplementary materials archive for the mentioned files.

## Sample Input/Output Texts

| Text | Sentiment |
|---|---|
| I really **like** new book of that author. | Positive |
| I **hate** new regulations about importing policies. | Negative |
| Look at that door, it's still open. | Neutral |

# Task 2 – Understanding Sentiment of New Terms

### Task Description

In this part you need to create a functionality within your existing solution that computes a new sentiment for the terms that **do not** appear in the file AFINN-111.txt and save them into another data frame or CSV file called ‘sentiments_new’.

### Algorithms Description

Certain words can be used to “estimate” the sentiment of a input text. Once you know the sentiment of the input texts that contain some new term, you can assign a sentiment to the new term itself. However, there are a number of words called “stop-words” that are not actually not bringing much sentiment to the text directly. You should ignore such words, list of them can be found in ‘stop-words-english.zip’ file.

### Sample Input/Output Texts

| Text | New Terms Sentiment |
|---|---|
| I really **like** new book of that author. | author 2.0<br>book 2.0 |

# Submission

- Use StudIP to upload your solution (PDF report)

- You should upload the following document
  - Report as a PDF file with your solutions and explanations
    - You could also use Jupyter Notebook to directly export your solutions and explanations as HTML and convert it to PDF
  - Each answer (e.g., screenshot, source code, etc.) should be shortly annotated by you
    - Original question  (including the serial number of the task)
    - Provided answer

- Name convention for your submission file (without extension)
  - **E08**_FIRSTNAME LASTNAME

- Submission deadline (it is a “soft” deadline)
  - 10 days after this exercise starts
  - Some exercises could take a bit more time and could be submitted later
  - NOTE: to receive feedback, you should first submit your progress

# Credits and Materials

- https://github.com/vdmitriyev/uol-data-analytics
- https://www.python.org/
- https://www.anaconda.com/products/individual
- https://jupyter.org/