BDAT1007 - Assignment 1

In this assignment you will analyze the sentiment of some documents you select and perform a text mining analysis on its contents.

Prerequisites:

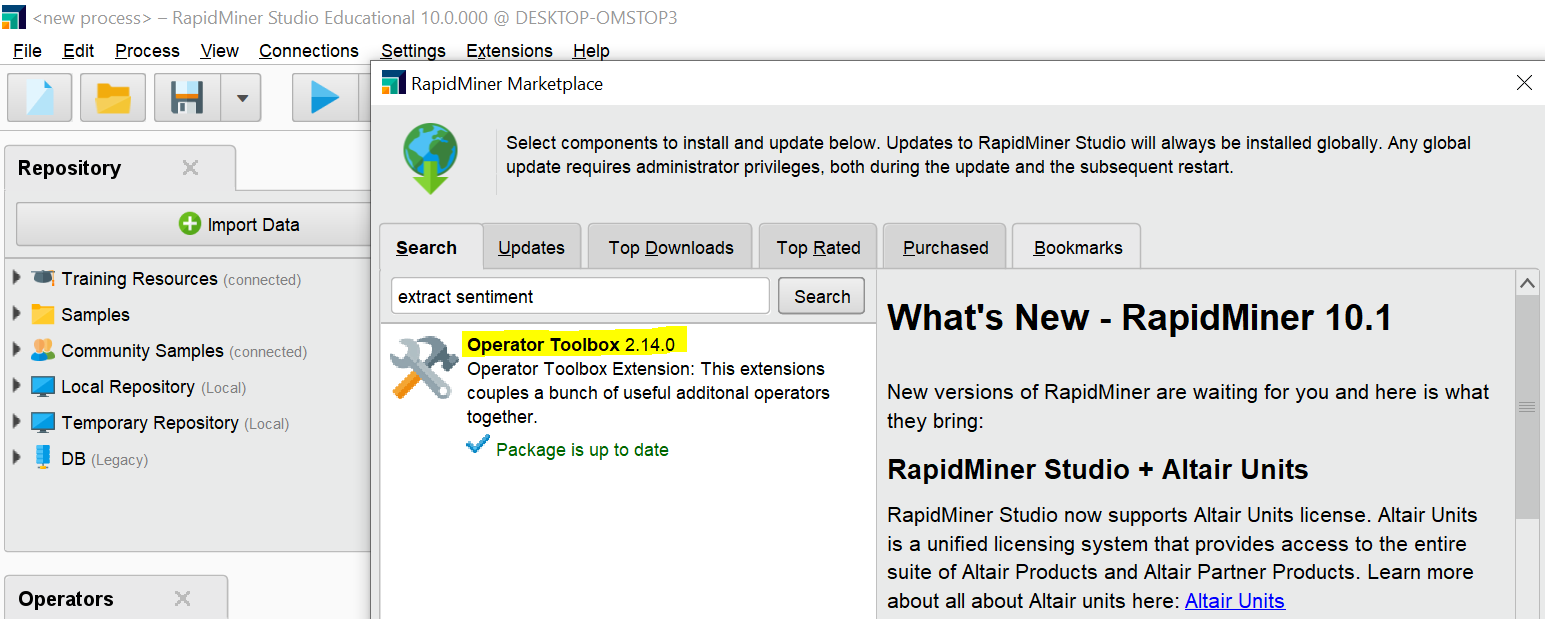
1. RapidMiner software package with text processing extension and text analysis extension added.

Note:

1. If you don’t like RapidMiner, feel free to complete this assignment with other tool / languages (python, pyspark, R, Tableau, Power BI, Excel, etc.)
2. This is individual work – however if you find it challenging, you can pick one partner to work together – just specify it in the submission. You can only have one partner by maximum.
3. Please respect Georgian College’s Academic Integrity policy. Plagiarized work will receive 0 marks.

Procedure:

1. Research for some user generated text dataset (Kaggle, Twitter, New York Times, etc.)
2. Use RapidMiner to import the dataset
3. Apply below steps:
4. Apply Extract Sentiment operator - Apply sentiment analysis on text data from extract sentiment (you would need to install Operator Toolbox). On extract Parameter use model as: VADER and apply the text attribute as Text.



1. Add Generate Attribute operator to create another field as sentiment that shows the sentiment. if(Score > 0, "Positive", if (Score < 0 , "negative" , if (Score == 0, "neutral","" )))
2. Add Write Excel operator to save the result to the excel file on your computer.

Submission – Please provide below:

Objective: 85%

* Full screenshots of your work including the process design

Here I retrieve he data I downloaded in my machine. The name is smileannotationfinal

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The file contains 278 records and 3 attributes ID Text and Emotion

A screenshot of a computer

Description automatically generated

This is file in results tab

A screenshot of a computer

Description automatically generated

Here I use operator extract sentiment.

Model is vader and text attribute is Emotion .

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Here I use generate attribute operator and give it the function expression and name the new column as sentiment

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Then I use write excel operator to store my output in the form of a excel file

A screenshot of a computer

Description automatically generated

Here I name the excel file as sentiment and format is xlsx

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

This I the output the with scoring of all the text in the found in the emotion column. The score of happy is 0.69

A screenshot of a yellow paper with black text

Description automatically generated

Disgust -1.333

Sad -0.538

A screenshot of a yellow and black data

Description automatically generated

Angry -0.59

Surprise 0.2

A screenshot of a computer

Description automatically generated

The new file I created is saved as sentiment.

And my process is saved as Assigment1.

* Explanation on the various operators / steps used within your process.
* Copy of the data that you extracted (.csv/.xls, .xlsx). It can be zipped.
* Your RapidMiner process file (.rmp format). Note: if you are using other tools and languages, please provide relevant files (python, R, Tableau, Power BI, Excel, etc.)
* The Excel file with the sentiment analysis calculated.

Subjective: 15%

* Elaboration on why you choose your topic & dataset, and what’s your findings.

I searched many datasets and I chose this to start my assignment.

I took this dataset from twitter and the link is

[SMILE Twitter Emotion dataset (figshare.com)](https://figshare.com/articles/dataset/smile_annotations_final_csv/3187909?file=4988956)

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Findings

“This dataset is collected and annotated for the SMILE project http://www.culturesmile.org. This collection of tweets mentioning 13 Twitter handles associated with British museums was gathered between May 2013 and June 2015. It was created for the purpose of classifying emotions, expressed on Twitter towards arts and cultural experiences in museums.

It contains 3,085 tweets, with 5 emotions namely anger, disgust, happiness, surprise and sadness. Please see our paper "SMILE: Twitter Emotion Classification using Domain Adaptation" for more details of the dataset.”

Citation “[SMILE Twitter Emotion dataset (figshare.com)](https://figshare.com/articles/dataset/smile_annotations_final_csv/3187909?file=4988956)”

* Elaboration on your feedback of the exercise – is it easy to do? And what have you learned?

This exercise helps us learn sentiment analysis via rapid miner.

This is not very difficult some challenges I came across were where dataset search.

And in the last where my saved excel file is saved.

I learned that how it create score for our text words. And negativity and positivity on certain word we gave it.