

Are Positive or Negative Tweets More “Retweetable” in Brazilian Politics?

April 6, 2017

Spotlight on our *Data Scientist Challenge* Winner from the Rosette API Academic Program

This winter, Basis Technology held a *Data Scientist Challenge* to encourage students in the Rosette API [Academic Program](#) to use both [Rosette API](#) and [Rapidminer Studio](#) in the data analytics project of their choice. The aim was to showcase how easy it can be to solve real-world analytics problems with powerful, easy-to-use predictive and text analytics platforms.

We are proud to announce that Delano Lima from Brazil, won first place for his excellent political tweet analysis project, investigating the correlation between the sentiment of a tweet and the number of times it is shared. Previous studies have shown that retweets are more frequent if the sentiment of the initial tweet is negative. Delano sought to determine if this held true in a political context, using the Brazilian political climate as a case study.

Find correlations and interpret results easily

Given its popularity and unstructured, noisy nature, Delano chose to use Twitter as the datasource.

Delano extracted tweets containing the terms “Brazil” and “Michel Temer,” both of which are intrinsically linked to Brazil’s political and economic environment. He used eight operators from Rapidminer and Rosette to aggregate and analyze the tweets collected:

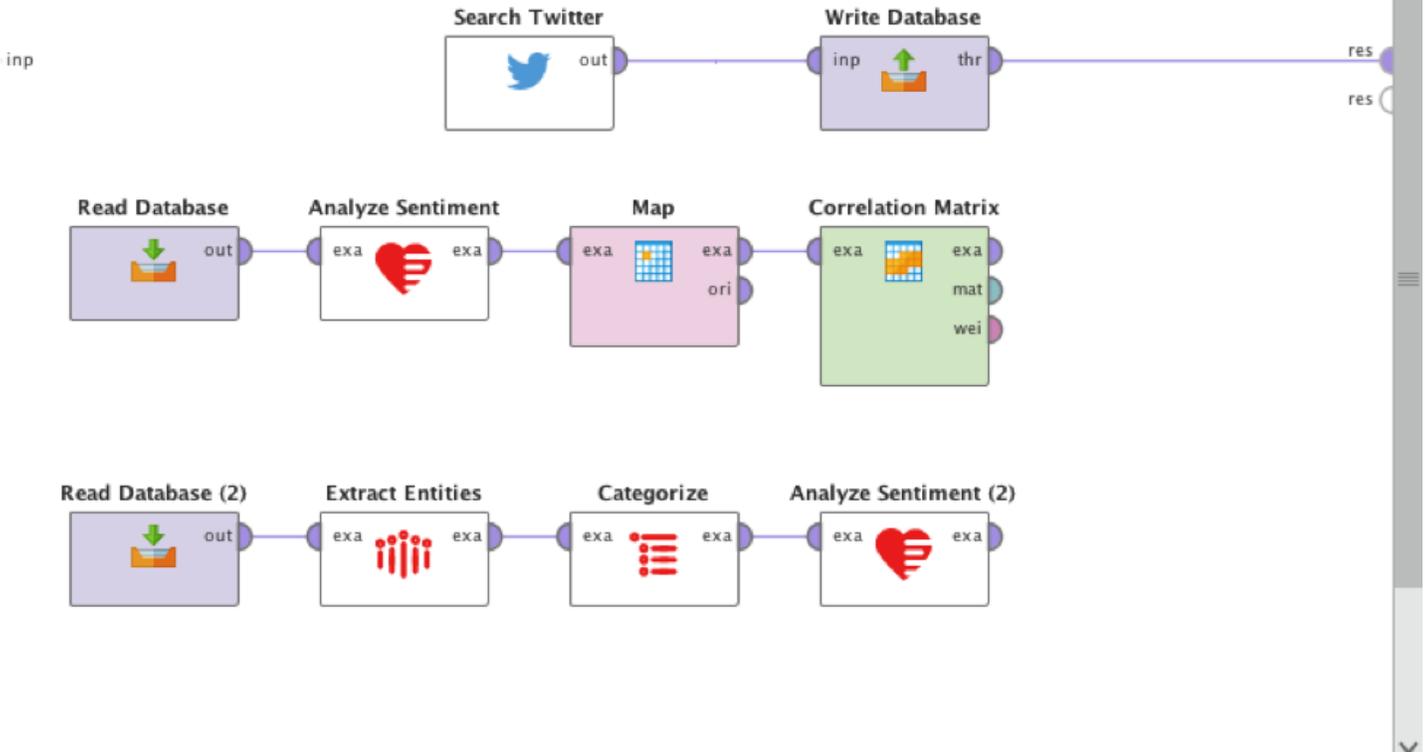
Rapidminer Operators

1. Search on Twitter
2. Write database
3. Read database
4. Mapping
5. Correlation matrix

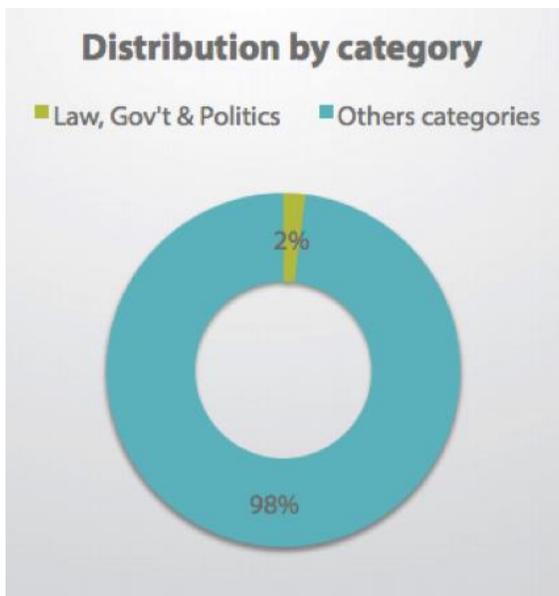
Rosette Operators

1. Entities Extraction
2. Categorization
3. Sentiment Analysis

Process



Delano collected his data over the course of seventeen days using our “entities extraction” operator. A necessary step during this study was to convert the feeling classifications into numerical variables. He also used our “categorization” operator to filter his data down from all tweets containing the words “Brazil” and “Michel Temer,” to only those related to the Laws and Politics category. This ensured he would not include irrelevant noise in his study, like tweets about tourism that also contain the target word “Brazil.”





So *are* people really retweeting news about President Temer more frequently when it relates to a problem or scandal? Rather than spoil the surprise, we encourage you to read Delano's complete study yourself!

Q&A with the data scientist

Did you know the tools before using them for the challenge?

I've been using RapidMiner for some time now. I started using the Rosette API for the challenge. The interesting thing is that if you have some intimacy with unstructured data analysis, using the API and its various functionalities becomes easy and exciting, since the focus is on thinking about the extraction of knowledge. The Rosette API does the "hard work" for you.

If no, how long did it take you to feel comfortable with them?

I did not need 30 minutes of use to get comfortable with using Rosette API for RapidMiner. The usage becomes very intuitive if you are already accustomed to working with data analysis in text. It flows in a very interesting way.

"The biggest challenge was to find the right angle for analysis, not to use the technologies"

How long did it take you to perform this analysis?

I needed about 15 hours to complete this work. I did not do it continuously. I must have used it for 5 days. The API has responded well to requests and the variety of API operators makes the process of constructing the metrics faster since they work well together.

How did you define the angle of analysis?

I thought about how a press officer that had to work on the international image of Brazil or the presidency of the republic in particular would act. A clear example of this is to use the entity extraction operator and evaluate which Brazilian institutions appear in greater volume in these tweets. It is possible to create targeted speeches, understand the real size and focus of the country's poor political image. It's like doing real-time opinion polls, finding out what really matters to the public, and generating the right answers.



What were the benefits of using Rapidminer and Rosette?

That the real challenge is to find an interesting subject to deal with, not to learn how to use the platforms. Rosette API and RapidMiner are incredible tools to find accurate answers, according to the problem that needs results, a solution. Rosette Text Toolkit gives you the chance to analyze social media data much more deeply than other tools with standardized reports. It is possible to continuously create new metrics and think about the process of knowledge discovery in the database.

At the enterprise level, why would you recommend it?

The use of both tools allows professionals from different areas, with the need to analyze this type of information, to do so with low levels of technical knowledge. Companies will not need a senior programmer to dig into information, at least at the first levels of analysis. In terms of knowledge in T.I, the required level of knowledge needs not be high. With Rosette you can achieve interesting results with a small team. Low cost and great results.

Champion spotlight: Delano Lima

Delano Lima, graduated in Advertising and is a postgraduate in Marketing Management by [UNIFOR](#), University of Fortaleza located in Northeast Brazil. He is founder and CEO of [DataImobi](#), a Brazilian startup focused on big data and analytics for the real estate market.



Want to conduct a study of your own?

Rapidminer and Rosette API make it easy to collect and analyze data yourself with minimal technical expertise. First, sign up for a [free API key](#) (no credit card required), then check out our [Rapidminer walkthrough](#). Find some interesting results, [let us know](#) and we'll feature you on our blog.