A Multilingual Approach for Global Brands

by **Digital:MR** market research evolved

In 2013, a global blue chip organisation commissioned DigitalMR to harvest and analyse social media data related to one of their Personal Care product categories. The objective was to explore the web landscape around the category in six different countries and languages around the world, in order to inform their future marketing campaigns.

What are people saying about the product category? **How do they feel about using the product? Where are the conversations** taking place?

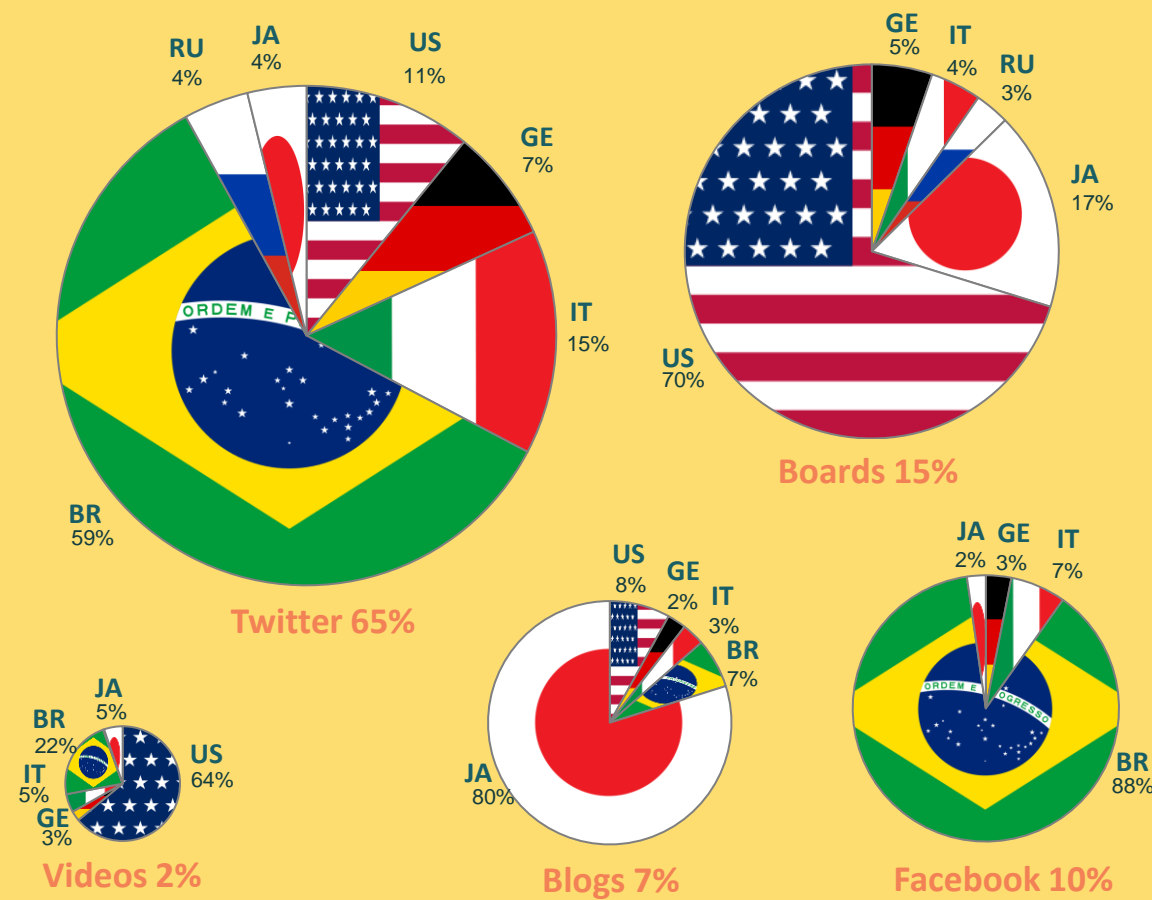Read on to discover our multilingual approach for global brands and 3 main challenges we had to overcome.

///////////////////// Six Countries and Languages

German ~71m
Italian ~35m
Russian ~102m
English (US) ~287m
Japanese ~115m
Brazilian Portuguese ~120m

*Size of Bubbles = Number of internet users per country in 2013, expressed in millions

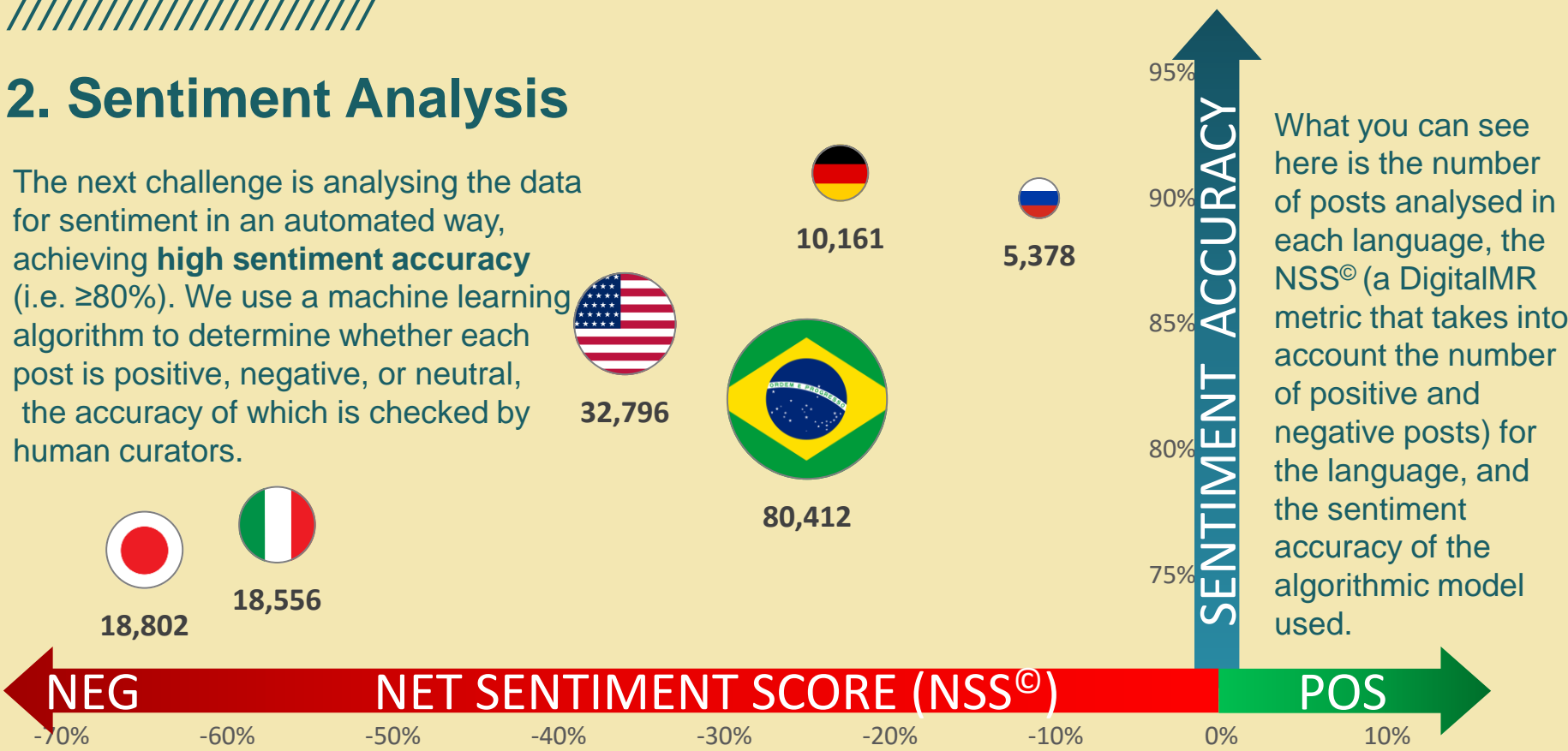## /////////////////////

1. Noise Elimination

Posts harvested from the web often come with a lot of "noise". To be able to clean the dataset and analyse only the truly relevant posts is the first challenge. In this case, we removed noise from tens of thousands of posts from the various countries, to obtain **166,105 actually relevant posts**. Here is the breakdown of posts by Share of Source Type and Share of Language per Source.

*Size of Bubbles = Share of Source Type

RU 4% · JA 4% · US 11% · GE 7% · IT 15% · BR 59%
Twitter 65%

GE 5% · IT 4% · RU 3% · JA 17% · US 70%
Boards 15%

US 8% · GE 2% · IT 3% · BR 7% · JA 80%
Blogs 7%

JA 2% · GE 3% · IT 7% · BR 88%
Facebook 10%

BR 22% · JA 5% · US 64% · IT 5% · GE 3%
Videos 2%

/////////////////////

2. Sentiment Analysis

The next challenge is analysing the data for sentiment in an automated way, achieving **high sentiment accuracy** (i.e. ≥80%). We use a machine learning algorithm to determine whether each post is positive, negative, or neutral, the accuracy of which is checked by human curators.

10,161
5,378
32,796
80,412
18,802
18,556

SENTIMENT ACCURACY — 95% · 90% · 85% · 80% · 75%

What you can see here is the number of posts analysed in each language, the NSS© (a DigitalMR metric that takes into account the number of positive and negative posts) for the language, and the sentiment accuracy of the algorithmic model used.

NEG — NET SENTIMENT SCORE (NSS©) — **POS**
-70% · -60% · -50% · -40% · -30% · -20% · -10% · 0% · 10%

/////////////////////

3. Topic Annotation

In addition to the sentiment, it is important to understand the topic(s) of the posts, the conversation drivers. Manually reading the posts is not the way to go. DigitalMR uses complex hierarchical topic taxonomies tailored to the client's needs for semantic analysis – assigning topics, sub-topics, and attributes to each post. Taxonomies can be adapted to multiple languages, countries, cultures, and of course product categories.

Topics	Maintenance		Emotional Issues		Product	
Sub-topics	Cleaning	Repairing	Feelings	Self-Image	Usage	Types of Product
Attributes	Soaking / Rinsing	Fixing / Repair Glue	Awkward / Nervous / Relieved	Appearance / Age / Attractive	Amount / How to Use	Gel / Strips / Cushion

Want to learn more about accurate social listening?

Get Started with listening247

Digital:MR market research evolved

listening247

www.digital-mr.com

For more exciting content, follow us on social media: