**Sentiment Analysis of Amazon Fine Food Reviews Using SpaCy and NLTK**

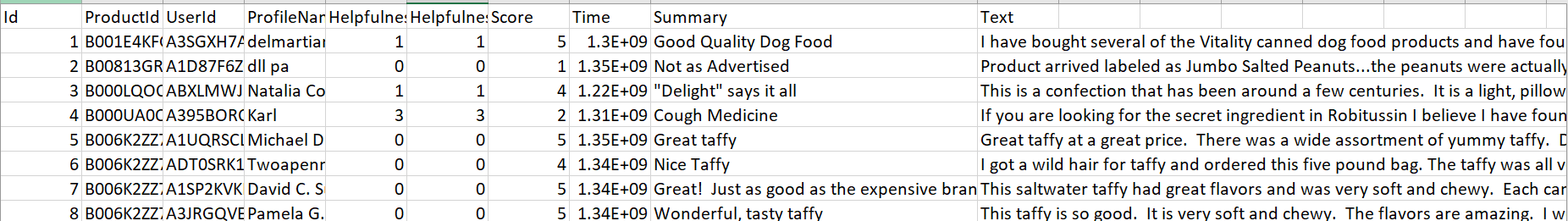
**Goal**

To establish that the trendier spaCy is better than the conventional NLTK in terms of speed for natural language processing.

**Dataset**

Amazon Fine Food Reviews sourced through <https://www.kaggle.com/snap/amazon-fine-food-reviews>

**Data Snapshot**



**Data Dictionary**

|  |  |
| --- | --- |
| **Dimension** | **Description** |
| Id | Unique identifier for each row of entry |
| ProductId | Unique identifier for the product |
| UserId | Unique identifier for the user |
| ProfileName | User’s profile name |
| HelpfulnessNumerator | Number of users who found the review helpful |
| HelpfulnessDenominator | Number of users who indicated whether they found the review helpful |
| Score | User rating between 1 and 5 (1 being worst and 5 being best) |
| Time | Timestamp for the review |
| Summary | Brief details of the feedback on product |
| Text | Detailed feedback on the product |

**Process Summary**

**Results and Comparisons**

**--Speed Comparison**

SpaCy takes 22% less time for natural language processing in comparison to NLTK when performing sentiment analysis for the Amazon reviews data.

**--Model Results**

As visible from the following graphs, NLTK and SpaCy render nearly identical results.

|  |  |
| --- | --- |
| **NLTK** | **SpaCy** |
| C:\Users\John Stephen\AppData\Local\Microsoft\Windows\INetCache\Content.Word\figure_1-NLTK.PNG | C:\Users\John Stephen\AppData\Local\Microsoft\Windows\INetCache\Content.Word\figure_1_spacey.png |
| C:\Users\John Stephen\AppData\Local\Microsoft\Windows\INetCache\Content.Word\figure_2-NLTK.PNG | C:\Users\John Stephen\AppData\Local\Microsoft\Windows\INetCache\Content.Word\figure_2_spacey.png |

Click on the link to view our prezi presentation of the project <https://prezi.com/p/8zw5oitol3rj/>