

قاعدة البيانات التي سنستخدمها:

عبارة عن 5513 تغريدة مكتوبة باللغة الانكليزية جمعت حول أربع مواضيع مختلفة وهي Apple, Google, Microsoft, Twitter حيث يتم تحديد الموضوع عن طريق البحث عنه في ال Hashtags أو ال Handles الموضوع في التغريدة عن تم جمع هذه التغريدات وتصنيفها يدوياً من قبل شركة sanders للتحليل والهندسة كل حقل في جسم ال Dataset يحتوي المعلومات التالية ترقيم للتغريدة الموضوع نوع الشعور نص التغريدة تاريخ الانشاء اسم المنشئ إلخ.... تم تصنيف كل تغريدة يدوياً من قبل ذكر أمريكي إلى أربع فئات وهي:

Positive (إيجابية) إذا كان النص يبين مشاعر إيجابية تجاه الموضوع.

Negative (سلبية) إذا أظهر النص مشاعر سلبية تجاه الموضوع.

Neutral (محايدة) إذا أظهر النص مشاعر مختلطة (متعارضة) أو ضعيفة أو لم يظهر مشاعر أبداً.

Irrelevant (غير مناسبة أو غير مرتبطة بالموضوع) إذا لم يكن النص باللغة الإنكليزية أو خارج الموضوع الذي يتحدث عنه.