

Desaster Prediction mit Twitter Daten: Eine kompakte Einführung in die volle Breite der SAS Text Analytics Funktionalitäten

Twitter ist zu einem wichtigen Kommunikationskanal in Krisen- und Notfallsituationen geworden. Eine Twitter Textnachricht besteht aus maximal 240 Textzeichen ist also vom textlichen Inhalt her sehr kurz. Die Allgegenwärtigkeit von Smartphones ermöglicht es Menschen, von einem Notfall, den sie beobachten, in Echtzeit zu berichten. Aus diesem Grund sind immer mehr Behörden daran interessiert, auch den Informationskanal Twitter kontinuierlich im Auge zu behalten. Aber es ist nicht immer klar, ob die geschriebenen Worte eines Tweets tatsächlich auf eine Katastrophe hinweisen. Daher wäre eine Klassifikation hilfreich, die neue Veröffentlichungen automatisch vorsortiert nach absteigender inhaltlich-thematischer Katastrophen-Relevanz.

Wir nutzen einen Datensatz von ca. 7500 Tweets, die zu verschiedenen Schlagworten (z.B. Feuer, Flut, Unwetter usw.) von Hand in Hinblick auf Desaster Ereignisse klassifiziert wurden. Auf dieser Stichprobe soll nachträglich ein maschinelles Text Analytics Model trainiert werden, das möglichst genau vorhersagt, welche Tweets von echten Katastrophen handeln und welche nicht. Ziel soll es sein dieses Modell später als Vorfilter zu nutzen, damit die diensthabenden Mitarbeiter in den Leitstellen nicht jeden einzelnen Tweed selber lesen müssen, sondern sich auf die x% Tweets mit der höchsten Relevanz konzentrieren können. Um ihre Effizienz zu steigern, sollte x möglichst klein sein. Im Vortrag folgen wir dem analytischen Lebenszyklus wie er in der SAS Viya Software abgebildet ist:

- wir schauen wir uns zunächst die Daten an,
- lesen sie ein,
- entdecken und beheben Probleme der Datenqualität,
- führen erste Visualisierungen durch,
- entwickeln Ideen für unkonventionelle Features
- leiten uns neuartige Prädiktoren ab
- entwickeln Maschine Learning Modell für die Klassifikation und validieren die Ergebnisse

Die Teilnehmer lernen

1. Umfang der Text Analytics Komponenten von SAS
2. Die Lösung eines Text Analytics Projektes entlang des analytischen Lebenszyklus in der gesamten SAS Viya Plattform
2. Die vielen Möglichkeiten des Feature Engineering speziell für Text Analytics
3. Die Erkenntnis, dass gute hilfreiche Text Analytics Prognose Features manchmal gar nicht auf Text basieren

Weiterführende Links:

Name: SAS® Visual Text Analytics 8.5: User's Guide (Viya)

URL: <https://go.documentation.sas.com/doc/en/ctxtcdc/8.5/ctxtug/titlepage.htm>

Name: SAS® Text Miner 14.1 Reference Help (V9)

URL: <https://support.sas.com/documentation/onlinedoc/txtminer/14.1/tmref.pdf>

Name: SAS® Text Analytics for Business Applications: Concept Rules for Information Extraction Models

URL: <https://support.sas.com/en/books/authors/teresa-jade.html>

Name: Git Repository für den Text Analytics Flow (Nutzung auf eigene Gefahr!!!)

URL: <https://github.com/Criptic/sas-text-extraction-and-analytics>