Johns Hopkins University Data Science Capstone

Zain Naboulsi

2023-03-22

Contents

Abstract

In today's digital landscape, individuals worldwide devote more time to their mobile devices, engaging in various activities such as emailing, social networking, and banking. However, typing on these devices can pose difficulties. SwiftKey, the corporate partner for this capstone project, offers a solution with its smart keyboard that eases typing on mobile devices. A key feature of SwiftKey's keyboard includes predictive text models, which suggest potential words to users based on their previous input. For example, after a user types "I went to the", the keyboard could suggest "gym", "store", or "restaurant". The goal of this capstone project involves both understanding and developing predictive text models similar to those SwiftKey employs.

This project starts with analyzing a large body of text documents to understand data structure and word organization. This process includes cleaning and analyzing text data, followed by the construction and sampling of a predictive text model. The final goal is to create a predictive text product. We will be utilizing all the skills acquired during the Data Science Specialization, with a focus on text data analysis and natural language processing (NLP).

References

Appendix: All Source Code

```
# set a seed in case we use any random items
set.seed(1337)


# Set the names of the packages and libraries you want to install
# Most notably load up all the quanteda packages we will need
required_libraries <- c("quanteda","quanteda.textmodels","quanteda.textstats",
                        "quanteda.textplots")

# Install missing packages and load all required libraries
for (lib in required_libraries) {
  if (!requireNamespace(lib, quietly = TRUE)) {
    install.packages(lib)
  }
  library(lib, character.only = TRUE)
}
```