Data Science Capstone Project: SwiftKey

Author: Evgeniy Date: April 2017

What is SwiftKey App?

SwiftKey shiny application which help to predict text.

It takes one or few words as inputs then guesses what is the next word to follow as output.

It is very portable and light. It can improve the typing speed of mobile phone users

How it looks like?

Here is a screenshot of the application:

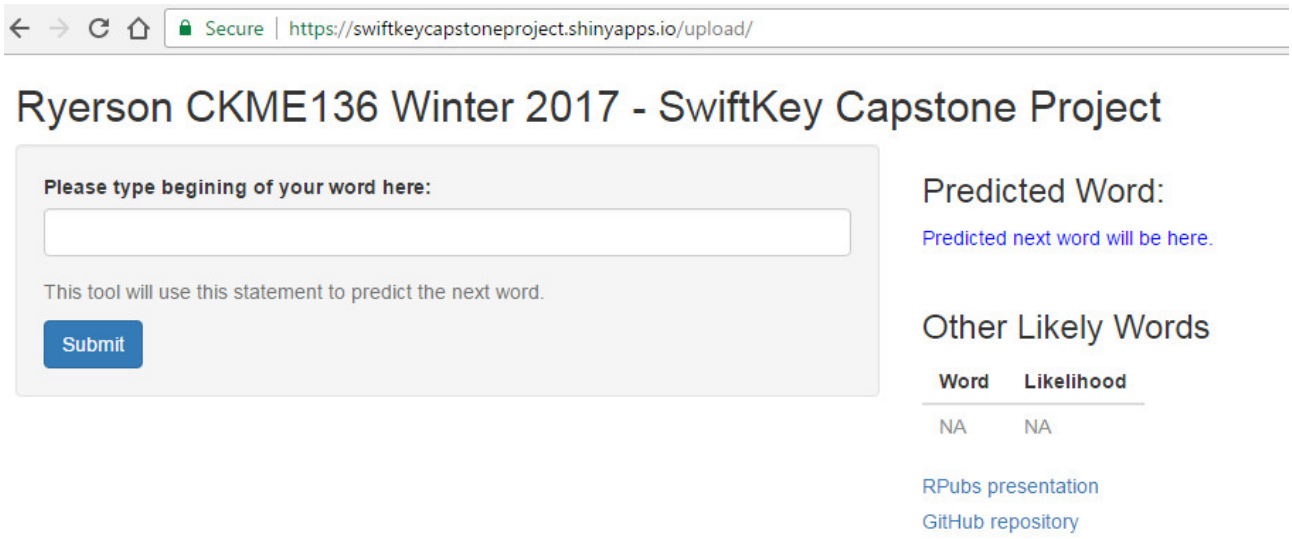

Image:

It can be acceed from: https://swiftkeycapstoneproject.shinyapps.io/upload/ (https://swiftkeycapstoneproject.shinyapps.io/upload/)

How is it working?

Data was cleaned, exploratory analysis were applied for the creation of a predictive infomration.

Natural language processing was done with the usage of a differernt R packages

This applicaiton uses N-Gram model and Tokenization: N-Gram models (http://en.wikipedia.org/wiki/N-gram#n-gram_models) It is widely used for natural language processing and quantitative textual analysis. https://en.wikipedia.org/wiki/Tokenization_%28lexical_analysis%29 (https://en.wikipedia.org/wiki/Tokenization_%28lexical_analysis%29)

The data come from the HC Corpora Dataset it uses US_blogs, US_twitter & US_news datasets.

What Methods & Models were used?

Corpora dataset was cleaned: - Removing punctuation, links, white space, numbers and all kinds of special characters. - Conversion to lowercase

Datta sample then was tokenized into n-grams.

Aggregated BI-gram and TRI-gram frequency matrices were transferred into frequency dictionaries

Summary

Tools used: R github.com shinyapps.io rpubs.com

Repository: https://github.com/ebabiche/SwiftKeyCapstoneProject (https://github.com/ebabiche/SwiftKeyCapstoneProject)