

SDAIA Academy T5C04 Bootcamps: Data Science

Natural Language Processing (NLP) Module Project Proposal

Topic Modeling: What's trending in App store?

Date: November 4, 2021

Prepared for: Nawras Boufaied

Prepared by: Nada Rambu

Objective

For app makers, it is very important to overlook the users' likes, dislikes, and preferences about an app. Thanks to data science, millions of app ratings and reviews can be analyzed to understand where the app is failing to provide a good customer experience.

The goal of this project is to learn more about the most popular app genre with high ratings. With the help of apps data, a machine learning model will be built to analyze app descriptions using NLP, then to find genre patterns. Some common topic modeling methods such as Latent Dirichlet association (LDA), Non-negative matrix factorization (NMF) and Latent Dirichlet Allocation (LDA) will be used to achieve the desired goal.

Data Description

The data for this project will be read into a CSV file (Find the dataset on the following link: <https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps>). The obtained dataset contains more than 7000 Apple iOS mobile application details, including their ratings and description.

Tools

The following tools will be used to carry out the project:

1. **NLTK** toolkit to perform common NLP tasks.
2. **Pandas** library will be used to create data frames for easier data manipulation.
3. **Scikit-learn** library will be to implement the clustering models.
4. **Tableau** tool to visualize and discuss the results of the analysis.