Quotes Recommender System

Semester Project, Practical Machine Learning and Deep Learning, Fall 2023, Innopolis University

Team

Name	E-mail	Responsible for
Vladimir Makharev	v.makharev@innopolis.university	Data, method, demo, presentation
Artem Batalov	a.batalov@innopolis.university	Method, demo
Georgii Budnik	g.budnik@innopolis.university	Method, organization, presentation

Responsibilities can be combined between us

GitHub repository

https://github.com/kilimanj4r0/quotes-recsys

Description

"If I quote others, it is only to better express my own thought"

— Michel de Montaigne is a French philosopher of the Renaissance

Inspirational quotes for everyday help a person achieve goals and believe in himself, do not give up and move on. The goal of our project is to instantly select a quote based on the answer to the question "How was your day?" to lift the mood and improve the next day.

The goal of our project is to develop a system of recommendations for quotations based on the answer to the question to lift your mood and improve your mood for the next day. The recommendation engine will be based on NLP mechanisms and extraction of valuable features for deep learning.

Tags

RecSys, NLP, PML, DL, Feature Extraction

Data

We will use the Quotes Kaggle dataset. The data was created to build a Content-based Recommender System using Text Data. We are planning to enrich this dataset by adding more quotes from other resources and extracting new features.

Method

We will ask a user a single open question and analyze their answer using NLP methods. This will include preprocessing of the answer by applying different NLP techniques of data preprocessing, such as tokenization, lemmatization, and building embeddings. After preprocessing, we will extract features of the answer applying text classification, sentiment analysis, and we will try to understand the user profile. Next, we will match features of the answer and quote features by using different methods: text similarity, number of features matched, etc. The output of the system will be the fittest quote for the user input.

Status

At the moment, the project is in the status of working out the idea and clarifying the details. We want to make the product high-quality and user-friendly, so the process of analyzing information and selecting a specific question for users takes some time. Now we are focusing on the following points:

- Exploring the data
- Enriching the dataset
- Developing of the question
- Implementing the method

Further work

From the point of view of the technological stack, the project can be supplemented with technologies of federated training and data distribution to improve the security of the service. The collection of quotations will be developed as a local repository, within which users with their own unique preferences will connect to it. Also, generative models for response generation will be considered, and advanced quality assessment metrics will be introduced.

The project can be integrated into psychological support services for people as part of the assistance of existing specialists to understand and analyze the patient. The predicative aspect of our solution can be considered as a labeling of the user's mood in larger projects of a similar nature.

Also, the recommendation system can be integrated into the educational process for students as part of a further analysis of the workload, work with children and the level of quality of teachers from the point of view of the teaching.

-