**Nigeria Cultural Heritage Topic Modelling using BERTopic & LDA**

**Overview**

This project aims to explore the cultural heritage of Nigeria through topic modelling techniques using BERTopic and Latent Dirichlet Allocation (LDA). The dataset used contains textual information about various aspects of Nigeria's cultural heritage.

**Table of Contents**

1. [Introduction](https://chat.openai.com/c/57d3b499-6c79-4c45-b478-dd1655cfc0e3#introduction)
2. [Dataset](https://chat.openai.com/c/57d3b499-6c79-4c45-b478-dd1655cfc0e3#dataset)
3. [Usage](https://chat.openai.com/c/57d3b499-6c79-4c45-b478-dd1655cfc0e3#usage)
4. [Results](https://chat.openai.com/c/57d3b499-6c79-4c45-b478-dd1655cfc0e3#results)
5. [Evaluation](https://chat.openai.com/c/57d3b499-6c79-4c45-b478-dd1655cfc0e3#evaluation)
6. [Dependencies](https://chat.openai.com/c/57d3b499-6c79-4c45-b478-dd1655cfc0e3#dependencies)
7. [License](https://chat.openai.com/c/57d3b499-6c79-4c45-b478-dd1655cfc0e3#license)

**Introduction**

Nigeria is rich in cultural heritage, with diverse ethnic groups and traditions spread across the country. Understanding and preserving this cultural heritage is crucial for maintaining Nigeria's identity and promoting cultural exchange. This project utilizes topic modelling techniques to uncover latent topics within the textual data related to Nigeria's cultural heritage.

**Dataset**

The dataset consists of text documents containing information about various aspects of Nigeria's cultural heritage. It includes descriptions of traditions, folklore, historical events, rituals, and more. Each document represents a unique aspect of Nigeria's cultural heritage.

**Usage**

1. **Data Preprocessing**: Preprocess the text data by removing noise, special characters, and stopwords. Tokenization and lemmatization may also be applied to enhance data quality.
2. **Topic Modelling with BERTopic**: Utilize BERTopic, a topic modelling technique based on BERT embeddings and hierarchical clustering, to identify latent topics within the dataset.
3. **Topic Modelling with LDA**: Employ Latent Dirichlet Allocation (LDA), a generative probabilistic model, to discover topics within the dataset.
4. **Evaluation**: Evaluate the performance of both BERTopic and LDA models using appropriate metrics such as silhouette score (BERTopic) and coherence score (LDA).
5. **Interpretation**: Interpret the discovered topics and gain insights into Nigeria's cultural heritage based on the model results.

**Results**

The results of the topic modelling process reveal a set of latent topics present in the dataset. Each topic represents a thematic cluster of documents related to specific aspects of Nigeria's cultural heritage. Visualization techniques such as word clouds or topic distribution plots may be employed to enhance result interpretation.

**Evaluation**

Evaluation metrics such as silhouette score (for BERTopic) and coherence score (for LDA) are used to assess the quality of the discovered topics. Higher scores indicate better coherence and separation between topics, indicating the effectiveness of the topic modelling approach.

**Dependencies**

* Python 3.x
* Libraries: scikit-learn, nltk, pandas, numpy, matplotlib, seaborn, gensim, bertopic

**License**

This project is licensed under the University of Salford License.

Feel free to customize this README file according to your specific project details and requirements.

Top of Form