أكاديمية سدايا
SDAIA Academy

TED talks using unsupervised learning

Banan Alhethlool
banan.alhethlool@gmail.com

Ghadah Alharbi
Ghadah.msh@gmail.com

Abstract:

Our aim is to use NLP to understand what words or topics make the most persuasive talks and if any relationships among them. Finally, we would like to build a linear regression model to predict the number of views. based on the TED.com dataset.

TED is devoted to spreading powerful ideas in just about any topic. These datasets contain over 4,000 TED talks including transcripts in many languages.

Question/need:

The goal of this project is to know the most persuasive talks TED users and speakers can benefit from the modeling.

Data Descriptions:

The dataset is a TED Talks dataset found on Kaggle that has over 4000 talks, almost all of them in English. It has a column that has the transcription of each talk. Additional features of the dataset include: "views, speaker, (speaker) occupations, recorded_date, published_date, event, available_languages, duration, (number of) comments, topics" which could be used for aggregating info/ modeling later on.

Tools:

Technologies: Python, Jupyter Notebook.

Libraries: Pandas, NumPy for EDA, Matplot, Seaborn for Visualization, Scikit-learn for modeling.