COURSE REVIEW ON COURSERA

Asmita Deshpande, Mostafa Chum, Parthiv Gandhi
IS Department University of Maryland Baltimore County

ABSTRACT

Educational videos have emerged as a dominant medium for educational purposes in many Massive Online Course (MOOC) platform., one such online platform is coursera.
Before enrolling in an online course, students always look at the ratings and reviews.
However, reading all the information, particularly the course reviews, might take a lot of time.
Our goal in this research is to provide a text analysis that combines methodologies like text cleaning, text preprocessing, and visualization to give course searchers a quick understanding of the courses and allow them to compare different courses side by side before choosing one.
We have worked with the help of different algorithms and methods such as classification, regression tasks, and clustering (K-means and Agglomerative Hierarchical).

METHODOLOGY

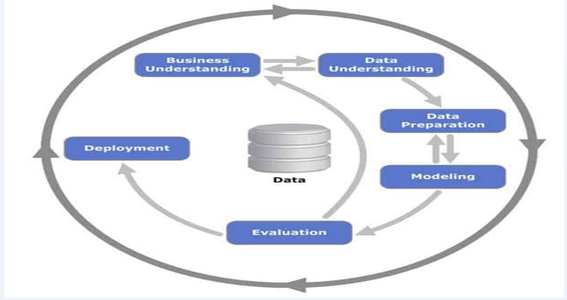

Figure 1: Standard Process for Data Mining

BUSINESS UNDERSTANDING

From the business standpoint understanding the project objectives and needs we selected the dataset.

DATA UNDERSTANDING AND PREPARATION

We have analyzed a large amount of dataset and converted it into attributes for further process. We have preprocessed the data. We did this to make it more valuable and executable. Removed all the random and excess columns for the process of clustering.
Moving forward we also removed numbers, special characters and stops words. Moreover, in the preprocessing task we have done the task of tokenizing and stemming. After saving the preprocessed data we started the implementation of the clustering algorithm.

ALGORITHMS

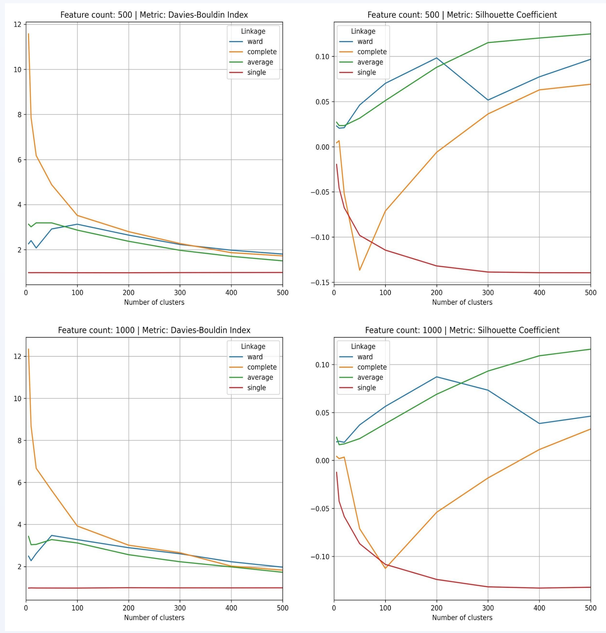

Figure 2: Plot of different linkages

CLUSTERING ALGORITHMS

K-means and feature agglomeration are two techniques that we have utilized for clustering. Additionally, we have employed the Davies-Bouldin Index and the Silhouette Coefficient as two measures to assess and contrast the algorithm. A hierarchical clustering technique called the feature agglomeration algorithm combines features using a bottom-up approach. The K-means method operates by minimizing the squares within a cluster. The Davies-Bouldin Index metric takes in to account the distance between each data point and the centroid of the cluster to which it is assigned as well as the centroid of the closest cluster; a lower score is preferable. The Silhouette Coefficient measure accounts for the average distance between each data point and the nearest cluster samples; 1 is the greatest score, -1 is the worst score, and 0 denotes overlapping clusters.

TF-IDF VECTORIZATION

We used TF-IDF vectorizing method, which considers words frequency in all texts. The purpose of using this method is to deal with the most frequent words by penalizing them.

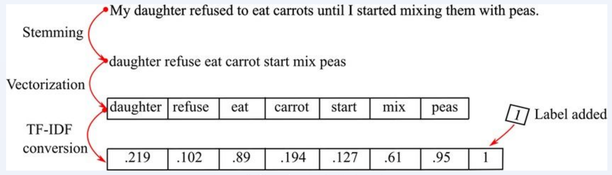

Figure 3: Steps for TF-IDF Vectorization

EVALUATION

Based on the results form applying algorithms for different top feature counts and number of clusters, the best number for top feature counts is 50 and best number for clusters is 500.

	Silhouette Coefficient	Davies-Bouldin Index
KMeans	0.4226	1.1765
Feature Agglomeration	0.3468	1.0578

Table 1 : Evaluation Result

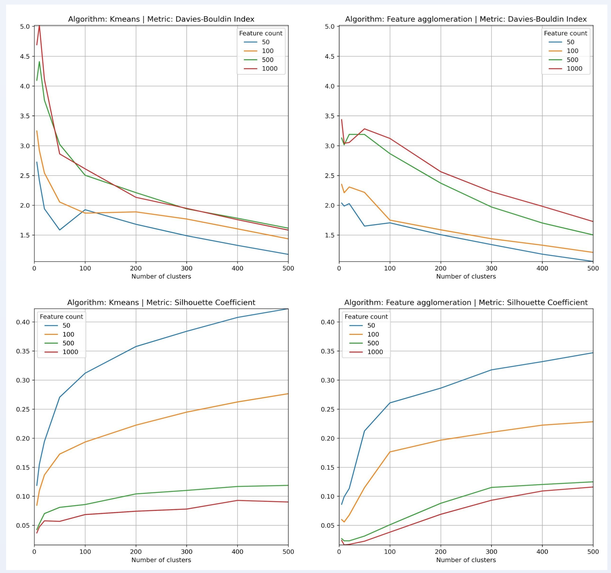

Figure 4: Result for top feature count

CONCLUSION

Coursera Review helps the course taking seekers to gain quick insights including the positive and negative reviews and also enables to make a quick comparison between multiple courses.
This process can help save a lot of time of the course seeker in the course selection process.
The Reviews which are put up below every course can also help the instructors or the course provider to understand the strengths and weakness and modify the course in near future.

FUTURE WORK

We can use all the positive reviews and keywords as an opportunity to promote and encourage more students to join the course.
We can use the information on which users are taking which courses, as an opportunity to recommend them more similar or advanced courses, recommending them books.
We can take this opportunity to also recommend jobs related to those specific courses.
Inform and give a report to Coursera/Instructors about how they can improve a specific course or modify it to make it more attractive and informative.
We can promote and market the courses which have great reviews.

REFERENCES

1. H. Y. Chan, R. Rajamohan, K. H. Gan and N. -H. Samsudin, "Text Analytics on Course Reviews from Coursera Platform," 2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), 2021.
2. B. Wu and Y. Zhou, "The Impact of MOOC Instructor Group Diversity on Review Volume and Rating— Coursera Specialization as an Example," in IEEE Access, vol. 8.
3. Donelson, Curtis & Sutter, Carolyn & Pham, Giang & Narang, Kanika & Wang, Chen & Yun, Joseph. (2021). Using a Machine Learning Methodology to Analyze Reddit Posts regarding Child Feeding Information. Journal of Child and Family Studies.

ACKNOWLEDGEMENT

https://www.kaggle.com/datasets/imuhammad/course-reviews-on-coursera