**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |  |
| --- | --- |
| **Team Member’s Name, Email and Contribution:** | |
| |  |  |  | | --- | --- | --- | | **NAMES** | **E-MAIL** | **CONTRIBUTION** | | **Prince Jain** | princejain1195@gmail.com | Worked on coding, analyzing the database and visualization of data.  Drafting observations and conclusions.  Prepared project summary.  Prepared Technical Documentation.  Prepared project presentation. | | **Rishabh Patidar** | rishabhpatidar1999@gmail.com | Worked on data visualization and codes for visualization.  Drafting observation and conclusion. | | **Vikas Shrivas** | vicky.07shrivas@gmail.com | Helped in the colab formation.  Helped in mistakes detections. | | |
| **Please paste the GitHub Repo link.** | |
| **GitHub Link:-**  **https://github.com/Prince11jain/Cluster-of-netflix**  **Drive Link :-**  **https://drive.google.com/drive/folders/1OpqQU7ic44s-qvGJ2cJpydru0wTzeOKr?usp=sharing** | |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** | |
| We hereby present the summary of the project with the title **NETFLIX MOVIES & TV SHOWS CLUSTERING**  Netflix is a company that manages a large collection of TV shows and movies, streaming it anytime via online. This business is profitable because users make a monthly payment to access the platform. However, customers can cancel their subscriptions at any time. Therefore, the company must keep the users hooked on the platform and not lose their interest. This is where recommendation systems start to play an important role, providing valuable suggestions to users is essential.  Here we have used libraries like NumPy and Pandas to analyze the data and Seaborn, Matplotlib,  Plotly, Machine learning models such as Hierarchical Clustering, DBSCAN, k-means clustering.  Firstly, the data is loaded using the panda’s library, then we use shape () and describe () methods for the basic understanding of the database. Before jumping to the exploration stage, we perform some basic data pre-processing methods like null value imputation and removal of unwanted data. In order to create a curated dataset, we rename all the columns post null-value treatment and drop the unwanted columns.  Now our analytics come into picture. The deep study of the dataset helped us in making different operations to derive a set of important observations which encapsulate many crucial facts about the terrorism and its wide impacts.  In this article, we looked at classification. Classifiers represent the intersection of advanced machine theory and practical application. These algorithms are more than just a sorting mechanism for organizing unlabeled data instances into distinct groupings. Classifiers include a unique set of dynamic rules that include an interpretation mechanism for dealing with ambiguous or unknown values, all of which are suited to the kind of inputs being analyzed. Most classifiers also utilize probability estimates, which enable end-users to adjust data categorization using utility functions.   * Netflix has 5372 movies and 2398 TV shows, there are more number movies on Netflix than TV shows. * TV-MA has the highest number of ratings for tv shows i.e., **adult ratings** * Highest number of **movies** released in **2017 and 2018.** * We saw a huge increase in the number of movies and television episodes after 2015. * There is a significant drop in the number of movies and television episodes produced after 2020. * It appears that Netflix has focused more attention on increasing Movie content that TV Shows. Movies have increased much more dramatically than TV shows. * **India is the country having maximum numbers of movie on Netflix.** * **USA having maximum numbers of TV shows followed by India.** * Those movies that have a rating of **NC-17 have the longest average duration.** * When it comes to movies having a TV-Y rating, they have the shortest runtime on average * **October to January, maximum number of movies and TV shows were added**. * The USA, India, the United Kingdom, Canada, and Egypt are the top five producer countries.   **k -mean cluster score:** |