

1. Data Mining Task: What is your data mining task? This task could be a series of exploratory questions that you want to investigate or analyze. What is your motivation behind choosing this task for your project?

For my data mining task, I have decided to perform some analysis on YouTube video statistics. The main question that I want to answer is to what is the relationship between YouTube views, likes and dislikes are and can that relationship be used to predict the number of likes and dislikes that a video will receive. My main motivation behind choosing this task for my project is my interest in how online media becomes popular. Additionally, its interesting to see the difference in levels of engagement that videos from certain categories receive compared to an average or videos from other categories.

2. Dataset: What is the source of your data? Provide a link to your data source if you acquired it online.

The source of my dataset is YouTube trending data for several months throughout the year across multiple regions throughout the world. The dataset includes data from USA, Great Britain, Germany, Canada, and France. Every day the top 200 trending videos were collected for each region and their video ID, Trending Date, Title, channel, category, publish date, tags, views, likes and dislikes are recorded. The data source is found online as a Kaggle Data set. <https://www.kaggle.com/datasnaek/youtube-new>.

3. Methodology: How will you solve the data mining task? You should have some idea of the algorithms or software tools you plan to investigate. Please feel free to use existing data mining and machine learning tool kits (e.g., Weka, Scikit-Learn) as needed for your project.

I will solve this data mining task using the Python programing language to analyze the data found within the above Kaggle Dataset. The main components within Python that will help me apply data mining techniques to this data is the popular Machine learning library Pandas, as well as the data analysis library Numpy. With these libraries I will have the capabilities that I need to solve my exploratory questions and analysis. Additionally, once the data has been successfully analyzed I have the options to use the matplotlib python library to generate graphs and other visual aids that will help myself and others view the results of this project.

4. Final product: What will be the outcome of this project? How will you measure the success of your course project? Will this project help you explore or learn something new

The outcome of this project will be values that shows the engagement on each video and the relationship that has been determined between likes, dislikes, and views on videos. The success of this course project will be datasets that show the engagement levels for videos in different regions, engagement levels for videos during different times throughout the year, and engagement levels broken down by category. This project will help me learn something new by allowing me to apply what I have learned in class to real world data that I have prior experience with. I will also gain a deeper insight into the statistics that trending YouTube videos have and

what is required for a video to become trending on YouTube during different times of year and in different regions.