

CS 574 Machine Learning Final Project

Members: Nitesh Gupta, Tanmay Gore

Topic: Predicting category of trending youtube videos using supervised learning

Summary of Goals: The goal of this project is to categorize the top trending videos on youtube using different supervised learning methods and compare their performance.

This is done using the dataset containing following features for top trending youtube videos:

- Title
- Category_ID
- Channel_title
- Tags
- Description

Basic Approach to be followed:

1. Preprocessing:
 - a. This includes converting strings to corresponding numerical values using **TF-IDF**
 - b. Performing dimensionality reduction using PCA.
2. Training:
 - a. Neural Network
 - b. Logistic Regression
3. Comparing Accuracies:
 - a. Confusion Matrix
 - b. Classification Error
 - c. Training Time

Anticipated Results: Given a training and test dataset of the youtube trending videos, our algorithm will predict the category of videos in test dataset.

Bibliography: www.kaggle.com