Youtube data Analysis and Predicting number of likes, views & comment count

Rahul Anand
rahulanand1103@gmail.com

1.Introduction

Youtube is one of the largest video-sharing website with humongous amount of video on it .The site allows users to upload, view, rate, share, add to favorites, report and comment on videos.

link-https://www.youtube.com

1.1. Data Overview

FileName-INvideos.csv

Size-56.84 MB

link-https://www.kaggle.com/datasnaek/youtube-new#INvideos.csv

Number of rows-35,183

Number of columns-15

Name	**Description**	**dtype**
video_id	Unique video id	string
trending_date	Date at which video start trending	string
title	Title of video	string
channel_title	Channel title which post the video	string
category_id	There are 15 Category value	Number
tags	tag given to video	No of tags
views	Number of views	No of views
likes	Number of Likes	No of likes
dislikes	Number of Dislikes	No dislikes
comment_count	Number of comment Count	boolean
thumbnail_link	Thumbnail image link	string
comments_disabled	Comment is unable or disable	boolean
ratings_disabled	rating is unable or disable	boolean

video_error_or_removed	Video is available or not	boolean
description	Description of video	String

2. Feature Engineering

1.-publish_weekday-Day at which video is publish

2.-No of Tags-No of tag video contain

3.-Length of description-Length of video description

4.-Ratio's

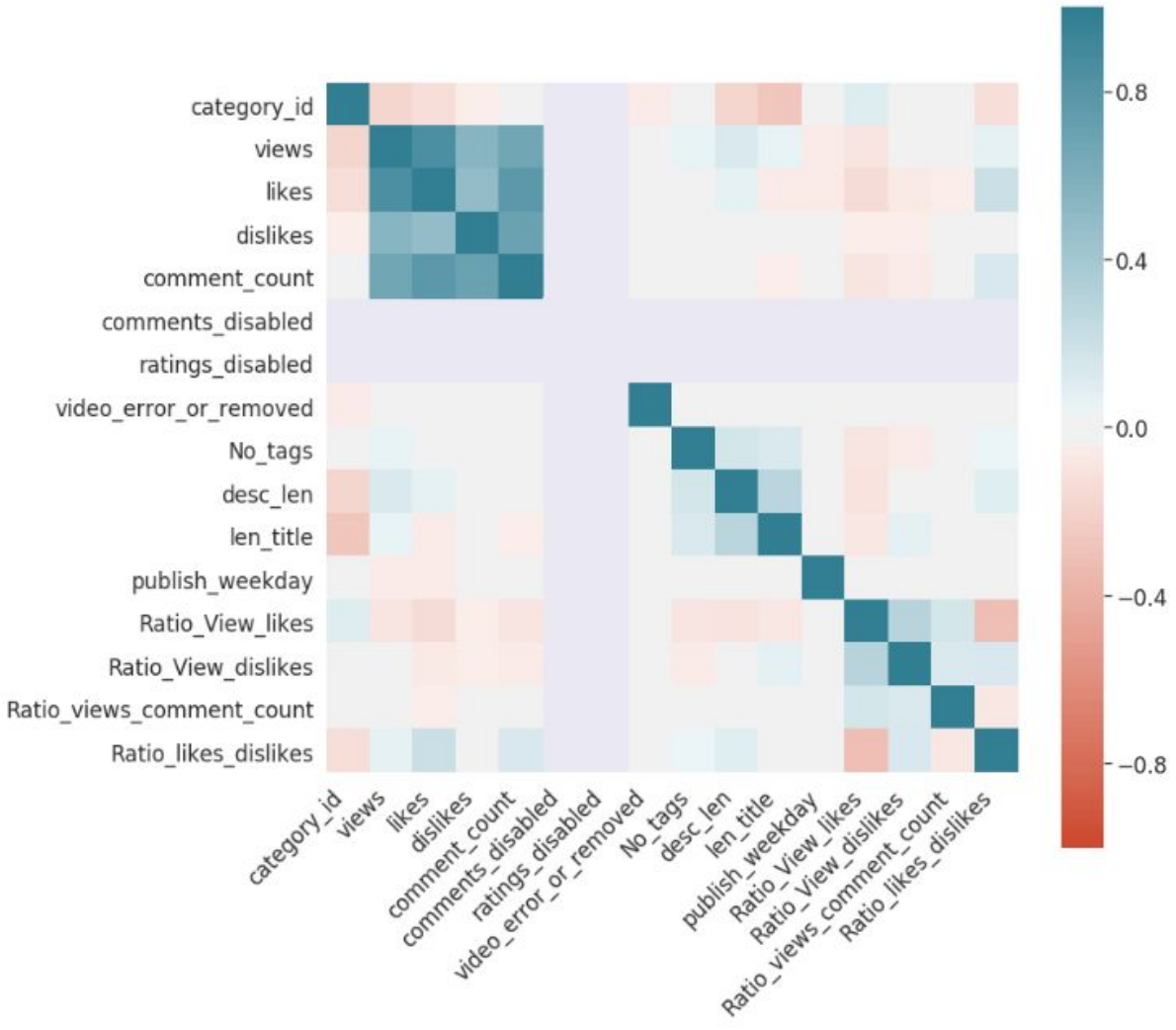
4.1-Ratio_View_likes-Ratio of View and likes

4.2 Ratio_View_dislikes-Ratio of view and dislikes

4.3 Ratio_views_comment_count-Ratio of view and comment_count

4.4 Ratio_likes_dislikes-Ratio of likes and dislikes

3.Correation matrix



4.Machine Learning Models

4.1 For Views

Models	Variance
Linear Regression	0.73
Random Forest	0.984

4.2 For Likes

Models	Variance
Linear Regression	0.41
Random Forest	0.96

4.3 For CommentCount

Models	Variance
Linear Regression	0.35
Random Forest	0.80