

과제- 프로세스

2020년 8월 26일 수요일 오후 12:01



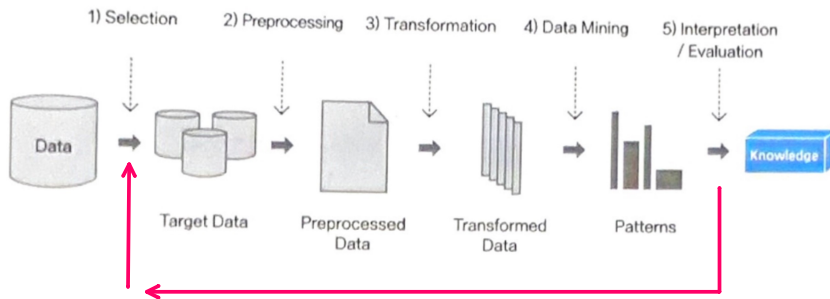
유튜브 영상/댓글 데이터를
통해 새로운 영상/댓글이
좋아요를 많이 받을 확률 계산

프로젝트 계획 : 좋아요 수 예측

프로젝트 모델 : 나선형 모델로써,

시제품 만들고 -> 고치고 -> 두번째 시제품 만들고 -> 고치 고를 반복

프로젝트 목표 : Kaggle 에 제출하기



프로젝트 모델 운영 ->



프로젝트 툴 : notion, 깃헙, 주피터, vscode, 카톡

notion

로드맵

<https://www.notion.so/invite/cdfc26a683695f14d158a02bc720c85715be1732>

로드맵 템플릿을 사용하면 모든 프로젝트의 작업을 기록할 수 있습니다.

- ▲ 에픽은 하위 프로젝트와 업무를 포괄하는 상위 이니셔티브입니다.
- 📌 스프린트는 일련의 작업을 완료할 수 있도록 하는 푸시입니다.
- 🔍 작업은 에픽을 구성하는 작업입니다.
- 🔴 버그는 버그를 수정하는 작업입니다.

1. 에픽, 스프린트, 작업, 버그로 작업을 구분하려면 **상태별**을 클릭하세요. 상태, 엔지니어, 제품 관리자별로 작업을 정렬할 수도 있습니다. 작업의 완료 예정일을 보려면 캘린더 보기로 전환하세요.

상태별 ▾

속성 그룹화 상태 필터 정렬 🔍 검색 ... 새로 만들기 ▾

시작 전 1	진행 중 2	완료 0	숨긴 열
<div>쿼리 캐싱 로직 재작성</div> <div>Zeno Wu</div> <div>Zeno Wu</div> <div>작업 <</div> <div>스프린트 22 스프린트 23</div> <div>스프린트 24</div>	<div>새 이미지 렌더링 실패</div> <div>Zeno Wu</div> <div>Shirley Miao</div> <div>버그</div> <div>스프린트 21</div>		상태 없음 1

+ 새로 만들기

타사 통합 개선

Shawn Sanchez

Cory Etzkorn Zeno Wu

Chet Corcos

에픽 ▲

깃헙: 각자 깃헙 생성

데이터 준비 : <https://www.kaggle.com/datasnaek/youtube-new?select=CAvideos.csv>

Dataset

Trending YouTube Video Statistics

Daily statistics for trending YouTube videos

Mitchell J • updated a year ago (Version 115)

Data Tasks (5) Notebooks (1,085) Discussion (40) Activity Metadata

Download (514 MB) New Notebook

Back to task list

Predict number of likes


rajan1002 · 1 Submission

Task Details

No. likes depends on several factor like comment disabled, number of views etc.

< CAvideos.csv (61.1 MB) 📄 🗃️ 🔍

Detail Compact Column 10 of 16 columns

video_id	trending_date	title	channel_title	category_id	
24427 unique values	205 unique values	24573 unique values	5076 unique values		
n18pP7lowLc	17.14.11	Estimote - Walk on Water (Audio) ft. Beyoncé	EstimoteVEVO	10	24 14
0d8Bx04Hz1M	17.14.11	PLUSH - Bad Unboxing Fan Mail	1DubbbzTV	23	24 14
5qpJK50gCt4	17.14.11	Racist Superman I Rudy Mancuso, King Bach & Lele Pons	Rudy Mancuso	23	24 14

참고자료 :

<https://www.kaggle.com/rahulanand0070/youtube-view-like-comment-prediction>



Youtube-View,Like&Comment prediction

Python notebook using data from [Trending YouTube Video Statistics](#) · 826 views · 71 regression

<https://www.kaggle.com/datasnaek/youtube-new/notebooks?sortBy=relevance&group=everyone&search=predic&page=1&pageSize=20&datasetId=4549>

Youtube Likes,views Prediction

Table of content

- 1.Machine Learning Formulation
 - 1.1 Data Overview
 - 1.2 Attribute-information
- 2.EDA
- 3.Feature Engineering
 - 3.1 publish_weekday-Day at which video is publish
 - 3.2 No of Tags-No of tag video contain
 - 3.3 Length of description-Length of video description
 - 3.4 Ratio's
 - 3.4.1 Ratio of View and likes
 - 3.4.2 Ratio of view and dislikes
 - 3.4.3 Ratio of view and comment_count
 - 3.4.4 Ratio of likes and dislikes

1651		🔥 Creating, Reading and Writing	📅 2mo ago
1132		🔥 Indexing, Selecting & Assigning	📅 18d ago
726		🔥 Summary Functions and Maps	📅 2mo ago
722		🔥 Grouping and Sorting	📅 2mo ago
413		🔥 Data Types and Missing Values	📅 2mo ago
385		🔥 Renaming and Combining	📅 2mo ago
215		🔥 Extensive USA Youtube [EDA]	1y ago 🌐 internet, beginner, exploratory data analysis, feature engineering

- 4. Correlation matrix
- 5. Machine Learning (metric = r^2 score)
 - 5.1 - View Prediction
 - 5.1.1 - Splitting the data into train and Test (80:20)
 - 5.1.2 - Linear Regression
 - 5.1.3 - Random Forest
 - 5.2 - Like Prediction
 - 5.2.1 - Splitting the data into train and Test (80:20)
 - 5.2.2 - Linear Regression
 - 5.2.3 - Random Forest
 - 5.3 - comment Count Prediction
 - 5.3.1 - Splitting the data into train and Test (80:20)
 - 5.3.2 - Linear Regression
 - 5.3.3 - Random Forest
- 6. Conclusion

1. Machine Learning Formulation

1.1 Data Overviews

Contain one file

1.2 Attribute-information

- video_id - Unique video id
- trending_date - the date at which video start trending
- title - Title of video
- channel_title - video posted by channel
- category_id - there are 15 Category value
- publish_time - at what time video is uploaded
- tags - tag given to video
- views - no of views
- likes - no of likes
- dislikes - no of dislikes
- comment_count - no of comment

오늘 필수적으로 해야하는 일 :

pdf 읽어보기, 데이터 다운로드 받기, 참고자료 살펴보기, 깃허브 프로젝트 생성해서 카톡에 공유하기