

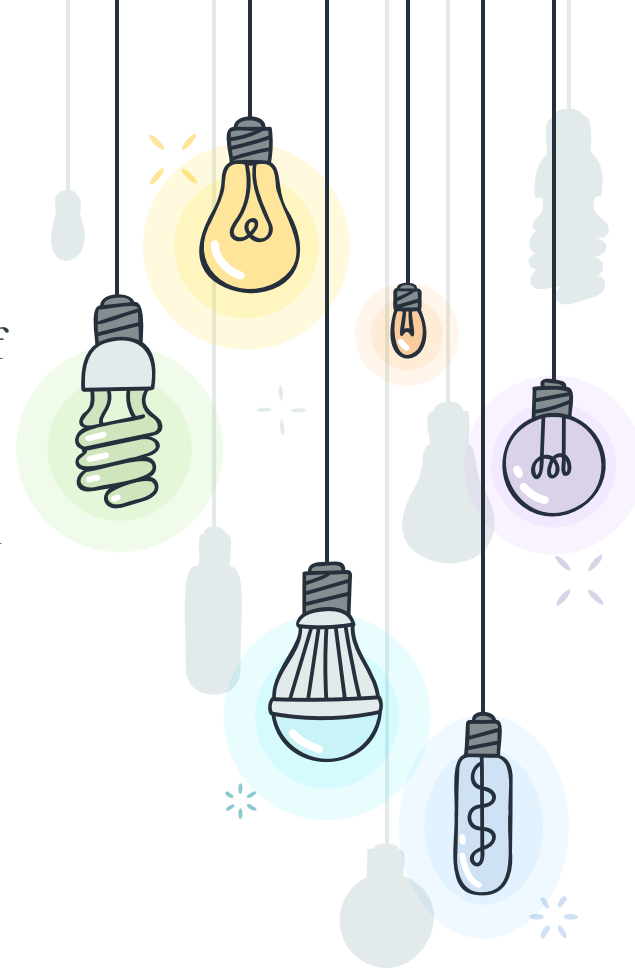


# MOVIE REVIEW & RECOMMENDATION

**Romauli Butarbutar, Saori Sembommatsu, Theo Pratama Silitonga, Alia Al-Sheikh**

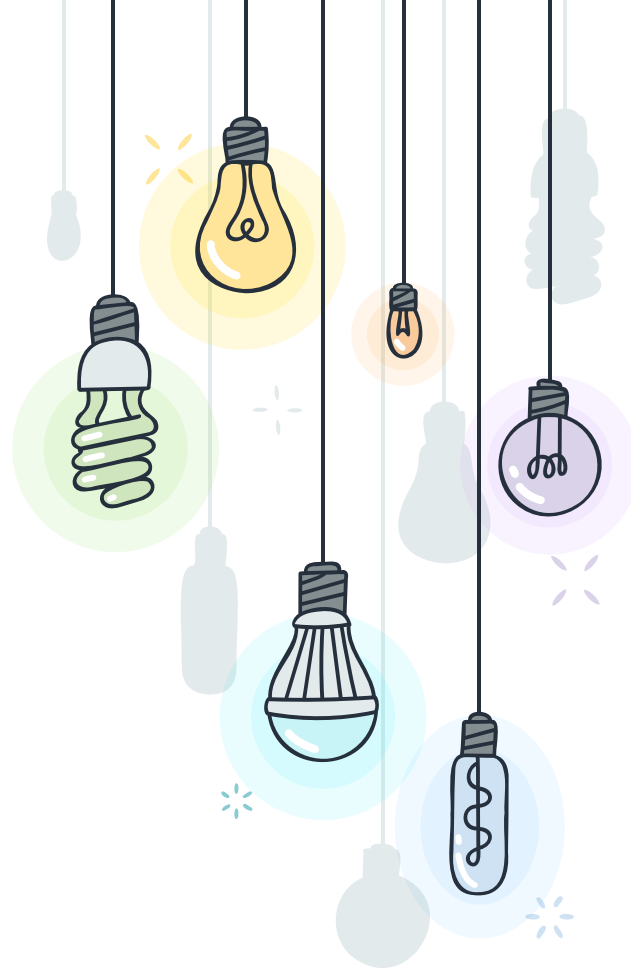
# 1. BACKGROUND + PROBLEM

- ✦ Although more people started watching movies at home after the Covid-19 started, there were not many platforms that provide comprehensive reviews of a movie and information of other recommended movies based on what you searched or watched.
- ✦ From multiple data sources, we have obtained all the scattered reviews, conducted sentiment analysis and created a single platform that houses reviews, sentiment ratings, recommendation lists.



## 2. DATA SOURCE SPECIFICATION & PROCUREMENT

- ✦ Our data warehouse stores movie reviews from **IMDb** and **New York Times**.
- ✦ Main attributes that have been populated are:
  - × Movie Title
  - × Review Comment
  - × Data Source
  - × Sentiment Categories
  - × Sentiment Scores
  - × A list of recommended movies
- ✦ From two different data sources, we compiled and created a data warehouse in order to run a *Sentiment Analysis* and get sentiment scores & similarity matrix for *Movie Recommendation*.
- ✦ We firstly scraped movie titles and reviews from IMDB and then retrieve reviews for the same movie titles from NY Times via API.



### 3. DATA GOVERNANCE POLICIES AND COST IMPLICATIONS



- ✦ We ensured to uniform column names and formats of our data, even from different data sources
- ✦ We excluded unnecessary info and avoid to share user privacy information
- ✦ We created sentiment scores for each movies to not only create a recommendation list, but also monitor and eliminate overrated reviews which are considered as non-trustworthy data sources.
- ✦ We will only store the files created for the past 3 years to maintain data quality. Others should be automatically deleted from our repositories.
- ✦ Although all the datasets we used are open source data, it may cost more to maintain security (i.e. Having a security system which can protect our database from potential hackers --Security from the 5 pillars of a good architected system).
- ✦ To ensure data quality, below is the data type of our main metrics that have be generated through ETL:

	movie	user_review	source	sentiment	neg
0	The Dark Knight Rises	Christopher Nolan's "Dark Knight Rises" wraps ...	NYT	positive	
1	The Dark Knight	Pitched at the divide between art and industry...	NYT	positive	0.003493
2	Inception	In Christopher Nolan's "Inception" a specialis...	NYT	positive	-1.000000
3	Parasite	In Bong Joon Ho's new film, a destitute family...	NYT	negative	0.560443

**Sentiment  
Scores**

## 4. DESIGN CHOICES & RATIONALE



### **Front-end API service: Python Flask**

**For developing Web Forms  
(movie searching web interface)**

Flask's framework is easier for users to use, less complicated to implement a simple web-Application, the collection of modules and libraries support applications developing without writing the low-level codes such as protocols, thread management, etc.

### **NoSQL data stores : MongoDB**

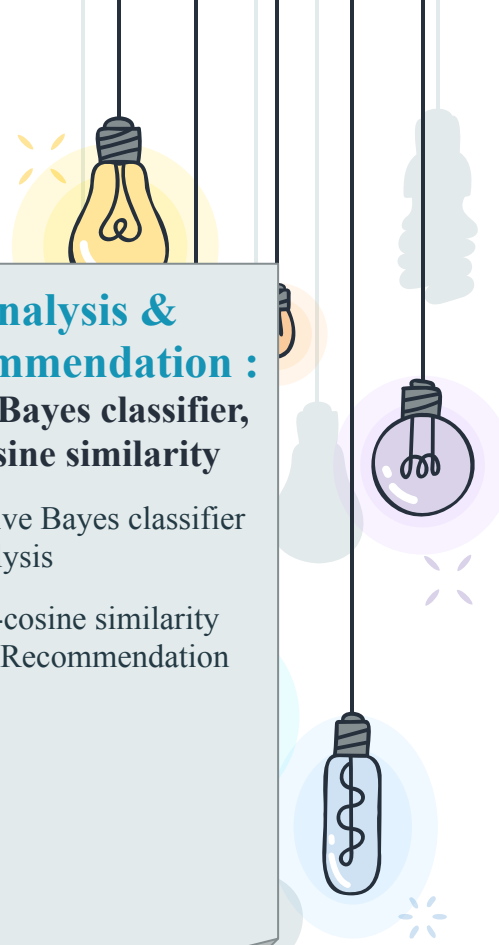
**To save unstructured data from  
news API - Document Databases,**

Schema less (schema optional), Map-Reduce Support, it is like an RDBMS, but it is schema less, has sharding and replication to make it faster than RDBMS

### **Sentiment Analysis & Movie Recommendation :** **NLTK's Naive Bayes classifier, Scikit-learn-cosine similarity**

Using NLTK's Naive Bayes classifier for Sentiment Analysis

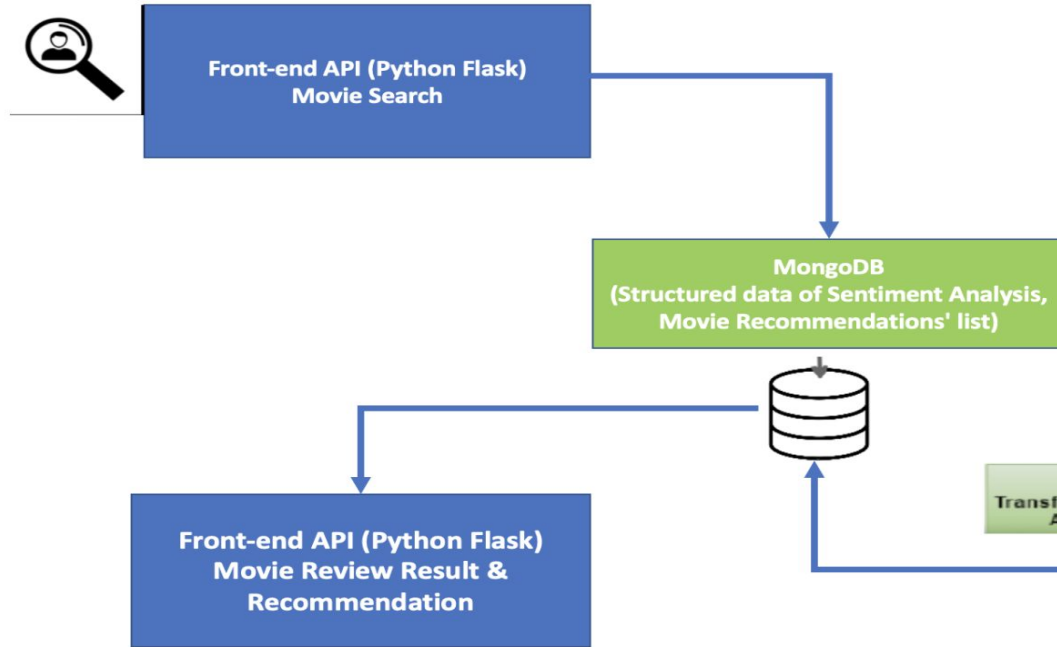
Using Scikit-learn-cosine similarity method for Movie Recommendation analysis.



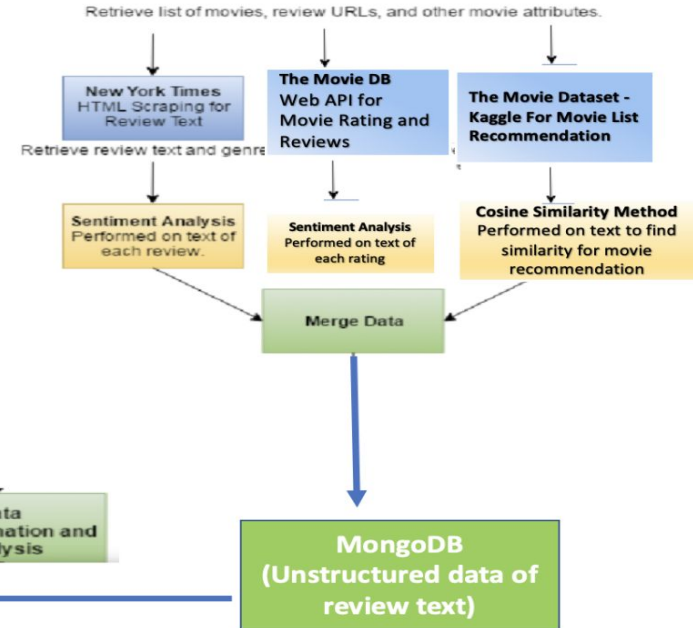
# 5. IMPLEMENTED DESIGN & TECHNICAL PROCESS



## Front End



## Back End



# 6. THE IMPLEMENTATION



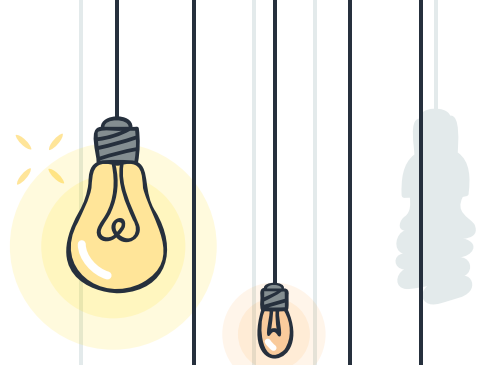
localhost:5001

Enter keyword for a movie title to get the movie reviews:

The Silence of the Lambs

To get appropriate Movies Recommendation,  
Please input complete title as the keyword

SUBMIT



localhost:5001

## Movie Reviews For Keyword: The Silence of the

Review your  
Movies here!

Sentiment  
Scores

Movies  
recommended  
for you

Movie Title	Movie Review	Sentiment	Source	Sentiment Score	Movie Recommendation
The Silence of the Lambs	Hannibal the Cannibal. Fast, witty and jolting.	positive	NYT	0.1214388906955719	[Jurassic Park', 'Misery', 'Mission: Impossible - Fallout', 'The Departed', 'Reservoir Dogs', 'The Handmaiden', 'In Bruges', 'Coraline', 'Repulsion', 'Hell or High Water']
The Silence of the Lambs	<p>The Silence of the Lambs (1991) is a classic psychological suspense brilliant masterpiece well acted and the best thriller ever made! This was my first film I saw on VHS as a teenager and I absolutely loved it. Right after seeing this flick, I heard about Hannibal (2001) Ridley Scot film a sequel to this psychological thriller flick. I rushed buying it right away. Sadly I was disappointed with Ridley Scott Hannibal. "Silence of the Lambs" is a classic and I think people would agree. It is my number 1 favorite thriller ever made, mostly I have enjoyed Jodie Foster as FBI recruit Clarice Starling's. Jodie did a brilliant acting performance I have seen. I love this film to death I absolutely love everything about this film. Since the first time I saw her in Silence I went to seeing her in a movie theater 2002 Panic Room, it was her first film I saw in movie theater since than I love Jodie Foster but Silence will be her best film she ever made. I was disappointed with the sequel Hannibal, no Jodie Foster in it, no brilliant masterful Clarice. Julianne Moore was incompetent for me, she couldn't catch Hannibal. Clarice Starling character was based on Gillian Anderson's FBI Agent Dana Scully and I think The X-Files inspired this film. I am even convinced that The Lone Gunman three geeks were also based on this film. You have a nice shootout on the end with Clarice and Buffalo Bill and Clarice get's her man, spoils her plans saves the day. The Silence of the Lambs is the 1991 thriller that was the feature film adaptation of the novel of the same name by Thomas Harris and marked the first appearance of Anthony Hopkins as Dr. Hannibal Lecter. (Oscar-winner Anthony Hopkins in one of the greatest performances ever on the screen) Dr. Hannibal "The Cannibal" Lecter, is a brutal killer with revolting methods and habits, but he's also very intelligent, charismatic and with good taste (you can interpret that as you like). He is scary and his the same time awesome including brilliant, you can't out smart him. Director Jonathan Demme is masterful directing this flick you have to be brilliant you have to keep thinking to solve the murder mystery. Scott Glenn acts beautiful as Jack Crawford. This flick got five major Academy Awards ("Oscars") for Best Movie, Director, Actor, Actress, Screenplay) it is no wonder why. Plot: FBI recruit Clarice Starling (Jodie Foster) is asked by the Bureau to elicit Lecter's help in tracking down a notorious serial killer who has kidnapped a US Senator's daughter. Directed by Jonathan Demme (The Manchurian Candidate), the film would receive five Academy Awards, including Best Picture, Best Actor for Hopkins and Best Actress for Foster and Best Screenplay. Hopkins would reprise his role as Lecter in two more films: Hannibal (2001) and Red Dragon (2002). Real life killer Ed Gein was the inspiration for Buffalo Bill as well as the inspiration for Norman Bates from Psycho and Leatherface from the Texas Chainsaw Massacre movies. And the director Jonathan Demme was a film student of Roger Corman who makes a cameo in the movie. Roger Corman taught not only him, but I think other filmmakers such as Martin Scorsese, James Cameron, Joe Dante, Francis Ford Coppola, and many others. Ted Levine did a pretty solid job as a serial killer Buffalo Bill. This film is fun, intense, very well fast paced goes quickly around for those two hours. The Silence of the Lambs is a 1991 American horror-thriller film directed by Jonathan Demme and starring Jodie Foster, Anthony Hopkins, and Scott Glenn. The film is based on Thomas Harris' 1988 novel of the same name, his second to feature Hannibal Lecter, a brilliant psychiatrist and cannibalistic serial killer. In the film, Clarice Starling, a young U.S. FBI trainee, seeks the advice of the imprisoned Dr. Lecter to apprehend another serial killer, known only as "Buffalo Bill". This movie is a perfect 10 out of 10, it takes the vision of one of the most imaginative directors on Earth (Jonathan Demme) with a brilliant cast including Thomas Harris novel that was adapted from it. I even saw the Documentary how they were shooting this film on YouTube. I think I do have that documentary film in my Blu-ray special features. Watch this film you want be disappointed at all. The movie is incredibly suspenseful and an absolute must see for fans.</p>	positive	IMDB	0.1214388906955719	[Jurassic Park', 'Misery', 'Mission: Impossible - Fallout', 'The Departed', 'Reservoir Dogs', 'The Handmaiden', 'In Bruges', 'Coraline', 'Repulsion', 'Hell or High Water']

Sentiment Scores:  
calculate positive or  
negative words in the  
reviews.  
Scoring:  
1) positive 0.8466  
2) neutral 0.1458  
3) negative 0.0076  
I.e. if the review contains  
many positive words, the  
sentimental scores for the  
review would be close to  
the full point score,  
0.8466.

# 7. ASSESSMENT OF SUCCESS

**“A search application for movie reviews and recommendations”**

Primary success criteria : retrieval time and relevance of search results of the movie

Additional success criteria : relevance of movie recommendations based on user preferences

## Quantitative metrics:

1. Our website will display complete reviews of related movies up less than 1 minute.
2. Display sentiment scores from reviewers (as a reference) and a list of recommended movies

## Qualitative metrics:

1. More complete and sophisticated movie reviews/sentiment ratings on a single platform
2. Improving user experience on choosing their desired movies



## 8. LIMITATIONS

- ✦ Limited knowledge to connect all into one platform (a recommendation list and sentiment scores).
- ✦ Lack of relevant algorithms to use the sentiment scores for a recommendation list.
- ✦ Lack of Technical Knowledge to display a list of recommended movies.
- ✦ Minimal design of web user interface.
- ✦ The algorithms used for sentiment categories and sentiment scoring are different. Even both used the same review, one may score negatively than another.
- ✦ Lack of relevant data for recommendation analysis using cosine similarity method.

# 9. CONCLUSION AND FUTURE RECOMMENDATIONS

## + **Conclusion**

- + We have been successful on delivering movie reviews platform which gathered data from different sources and inserted into one single source data warehouse through ETL process.

## + **Recommendations:**

- + Find more relevant algorithms for sentiment analysis and creating a recommendation list.
- + Create a solid platform where we provide a recommendation list created with our unique algorithm.
- + Focus more on a decoration part so that users would be more attracted to our platform.

## 10. REFERENCE

- ✦ Flask Web Form
  - ✦ [https://python-adv-web-apps.readthedocs.io/en/latest/flask\\_forms.html](https://python-adv-web-apps.readthedocs.io/en/latest/flask_forms.html)
- ✦ Using New York Times API and jq to collect news data
  - ✦ <https://medium.com/@danalindquist/using-new-york-times-api-and-jq-to-collect-news-data-a5f386c7237b>
- ✦ Kaggle Data
  - ✦ <https://www.kaggle.com/rounakbanik/the-movies-dataset>
- ✦ Sentiment Analysis (Positive/ Negative Reviews)
  - ✦ <https://hub.packtpub.com/how-to-perform-sentiment-analysis-using-python-tutorial/>
- ✦ Cosine Similarity – Understanding the math and how it works (with python codes)
  - ✦ <https://www.machinelearningplus.com/nlp/cosine-similarity/>