The RecSys Pipeline:

A case-study approach

Bereket A. Yilma

The typical RecSys Pipeline

Data Pre-processing →  → Model Training → Post Processing → Evaluation

- Sort
- Filter
- Recommend

The RecSys pipeline: A case-study approach

Task: Design a **Personalised** **Visual Art Recommendation** engine for the National Gallery, London

The RecSys pipeline: A case-study approach

RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

DISCLAIMER

CROSS CULT

LIST

European Commission

HORIZON 2020

The RecSys pipeline: A case-study approach

Personalised Visual Art Recommendation

Context: National Gallery, London

- >= 2,300 paintings dating from the mid-13th century to 1900.
- Total floor area of 46,369 square meters, 3 floors.
- 6.2 million visitors/year (2019)

The RecSys pipeline: A case-study approach

RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

Task — Personalised Visual Art Recommendation

Data — Visual Art (Paintings)

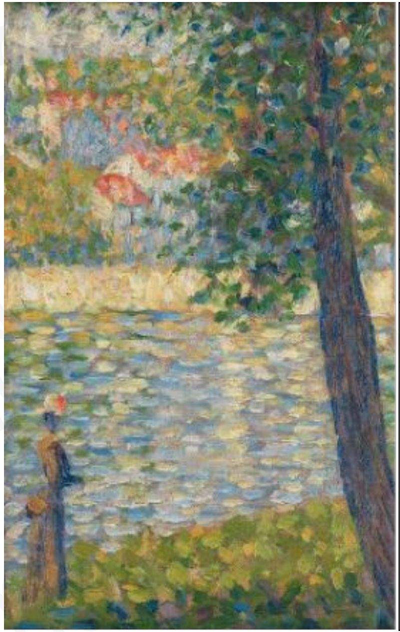
Target User — Visitors (New)

*assumptions Visitor profile

The RecSys pipeline: A case-study approach

1. Data: Visual Art (Paintings)

- 2,368 painting



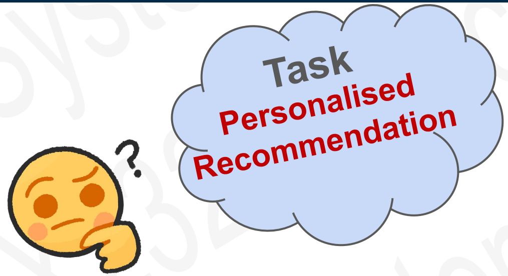
painting_id	000-018P-0000
title	The Morning Walk
artist	Georges Seurat
publication_date	19th_century
size_format	Portrait
size	Very Small
technique	oil painting
description	A woman, silhouetted against the shimmering water, strolls along a riverbank. The red roofs of houses can be made out along the opposite bank. Between 1882 and 1886 Seurat painted numerous such landscape studies on small wooden panels, some as independent works and others in preparation for his large-scale compositions. This sketch provided the starting point for a painting of 1885, 'The Seine at Courbevoie' (private collection).

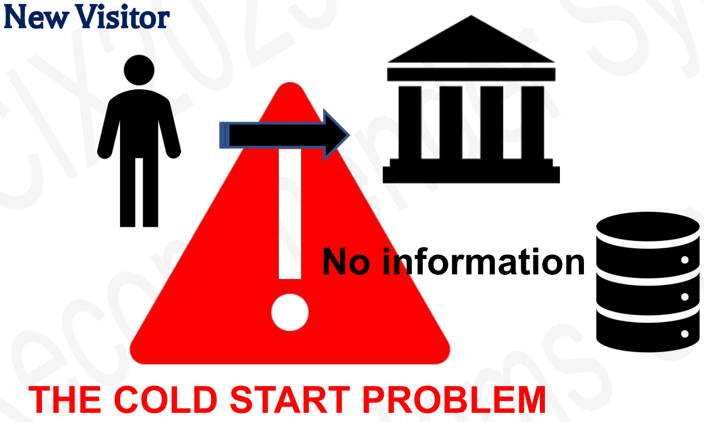
The RecSys pipeline: A case-study approach

RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

Data
Pre-processing





Query User (Profiling)

1. Rate few paintings
2. Popular paintings
3. Visiting style
4. Available time ...

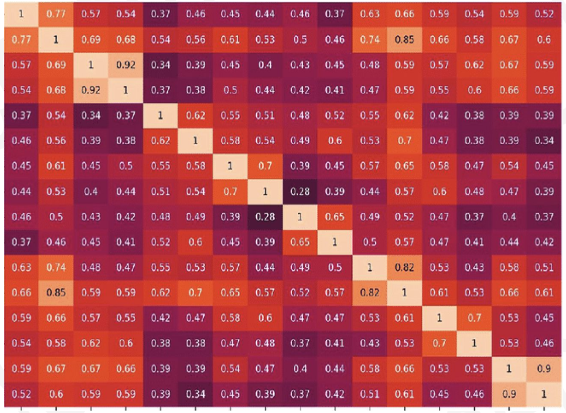
The RecSys pipeline: A case-study approach

Data Pre-processing

Task
Personalised Recommendation

Model Training

$$R^{m \times m}$$



- If a user likes painting **A** find paintings **B, C, D** that are similar to A.

The RecSys pipeline: A case-study approach

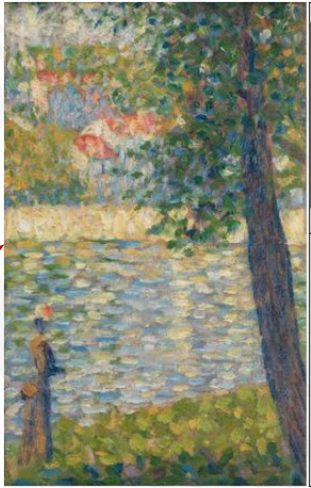
RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

Data Pre-processing

Task
Personalised Recommendation

Image



Metadata

Textual description

painting_id	000-018P-0000
title	The Morning Walk
artist	Georges Seurat
publication_date	19th_century
size_format	Portrait
size	Very Small
technique	oil painting
description	A woman, silhouetted against the shimmering water, strolls along a riverbank. The red roofs of houses can be made out along the opposite bank. Between 1882 and 1886 Seurat painted numerous such landscape studies on small wooden panels, some as independent works and others in preparation for his large-scale compositions. This sketch provided the starting point for a painting of 1885, 'The Seine at Courbevoie' (private collection).

The RecSys pipeline: A case-study approach

RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

Data Pre-processing

Task
Personalised Recommendation

Model Training

Good representation of the data!

$R^{m \times m}$

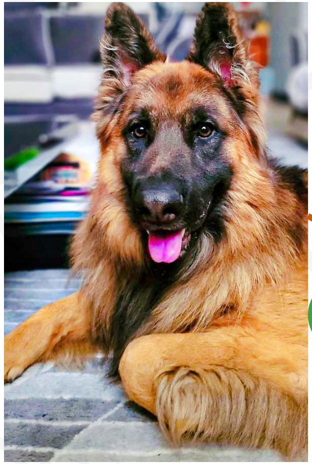
1	0.77	0.57	0.54	0.37	0.46	0.45	0.44	0.46	0.37	0.63	0.66	0.59	0.54	0.59	0.52
0.77	1	0.69	0.68	0.54	0.56	0.61	0.53	0.5	0.46	0.74	0.85	0.66	0.58	0.67	0.6
0.57	0.69	1	0.92	0.34	0.39	0.45	0.4	0.43	0.45	0.48	0.59	0.57	0.62	0.67	0.59
0.54	0.68	0.92	1	0.37	0.38	0.5	0.44	0.42	0.41	0.47	0.59	0.55	0.6	0.66	0.59
0.37	0.54	0.34	0.37	1	0.62	0.55	0.51	0.48	0.52	0.55	0.62	0.42	0.38	0.39	0.39
0.46	0.56	0.39	0.38	0.62	1	0.58	0.54	0.49	0.6	0.53	0.7	0.47	0.38	0.39	0.34
0.45	0.61	0.45	0.5	0.55	0.58	1	0.7	0.39	0.45	0.57	0.65	0.58	0.47	0.54	0.45
0.44	0.53	0.4	0.44	0.51	0.54	0.7	1	0.28	0.39	0.44	0.57	0.6	0.48	0.47	0.39
0.46	0.5	0.43	0.42	0.48	0.49	0.39	0.28	1	0.65	0.49	0.52	0.47	0.37	0.4	0.37
0.37	0.46	0.45	0.41	0.52	0.6	0.45	0.39	0.65	1	0.5	0.57	0.47	0.41	0.44	0.42
0.63	0.74	0.48	0.47	0.55	0.53	0.57	0.44	0.49	0.5	1	0.82	0.53	0.43	0.58	0.51
0.66	0.85	0.59	0.59	0.62	0.7	0.65	0.57	0.52	0.57	0.82	1	0.61	0.53	0.66	0.61
0.59	0.66	0.57	0.55	0.42	0.47	0.58	0.6	0.47	0.47	0.53	0.61	1	0.7	0.53	0.45
0.54	0.58	0.62	0.6	0.38	0.38	0.47	0.48	0.37	0.41	0.43	0.53	0.7	1	0.53	0.46
0.59	0.67	0.67	0.66	0.39	0.39	0.54	0.47	0.4	0.44	0.58	0.66	0.53	0.53	1	0.9
0.52	0.6	0.59	0.59	0.39	0.34	0.45	0.39	0.37	0.42	0.51	0.61	0.45	0.46	0.9	1

- If a user likes painting **A** find paintings **B, C, D** that are similar to A.

Representation Learning

What is a good representation of this image?

Buddy :)



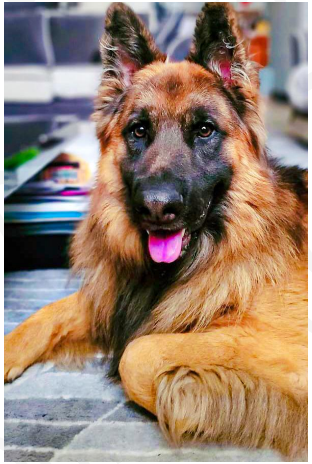
Representation

Summarises → **Dog!** ← Too much information loss!

- Is this dog mine?
- What is the dog doing?
- Is the dog safe?

Representation Learning

What is a good representation of this image?



A bit more descriptive yet a summary

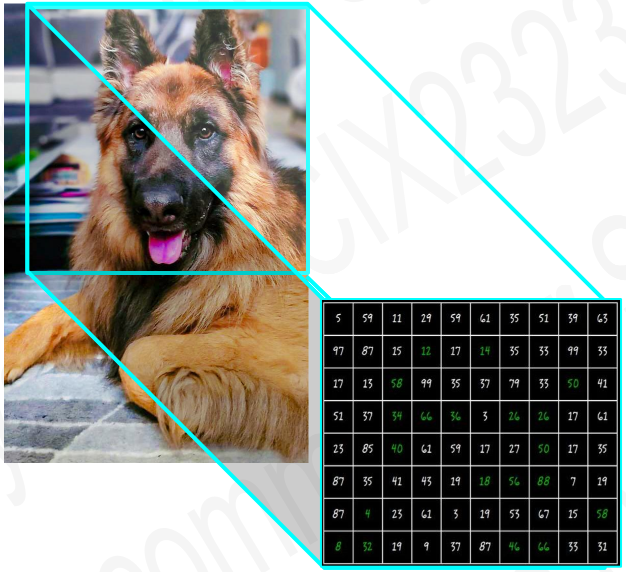
Representation

German shepherd dog resting on a floor in living room.

RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

Representation Learning



Representation is a summary of data which:

➔ **Omits** unnecessary details

➔ **Preserves** important details

Representation

[1, 23, 0.4]

Array of numbers!

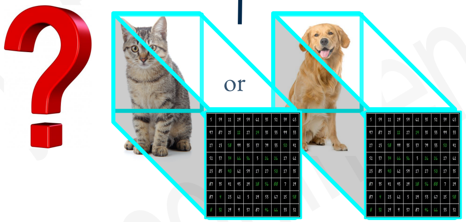
Each dimension would contain some semantic meaning

Evolution of Representation Learning

Classifier on raw data

Task

Classifier



Very high dimensional Array of numbers!

Problem:
Too much redundancy in the input data

Solution
Summarize relevant information
→ Features

Classifier with handcrafted features

Perform reasonably better

Task

Classifier

Features

handcrafted

Dimensionality Reduction

Evolution of Representation Learning

End-to-end training of Neural Networks on task

Good performance

Task

Deep Model

Learn features from data for the task

Whatever features are good to improve the accuracy, the model summarizes those features out

No longer works with predefined set of features

Great! Can we learn features that are useful not only for the specific task but also other tasks?

Learning Transferable features

Task 1

Task 2

Task 3

Features

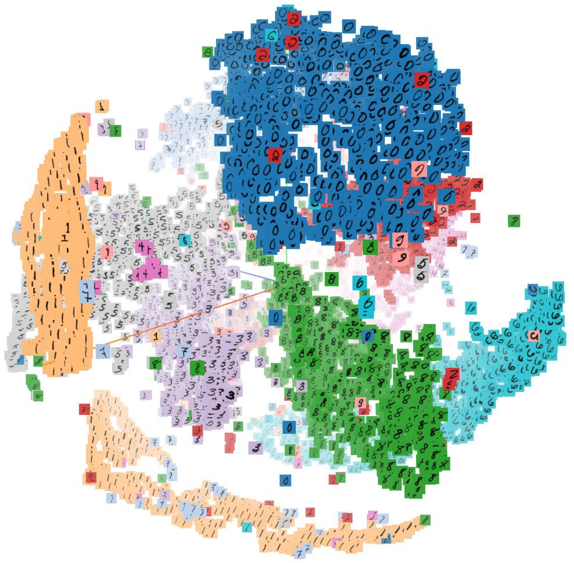
Deep Model

Learned from data

Generally useful for a wide range of tasks

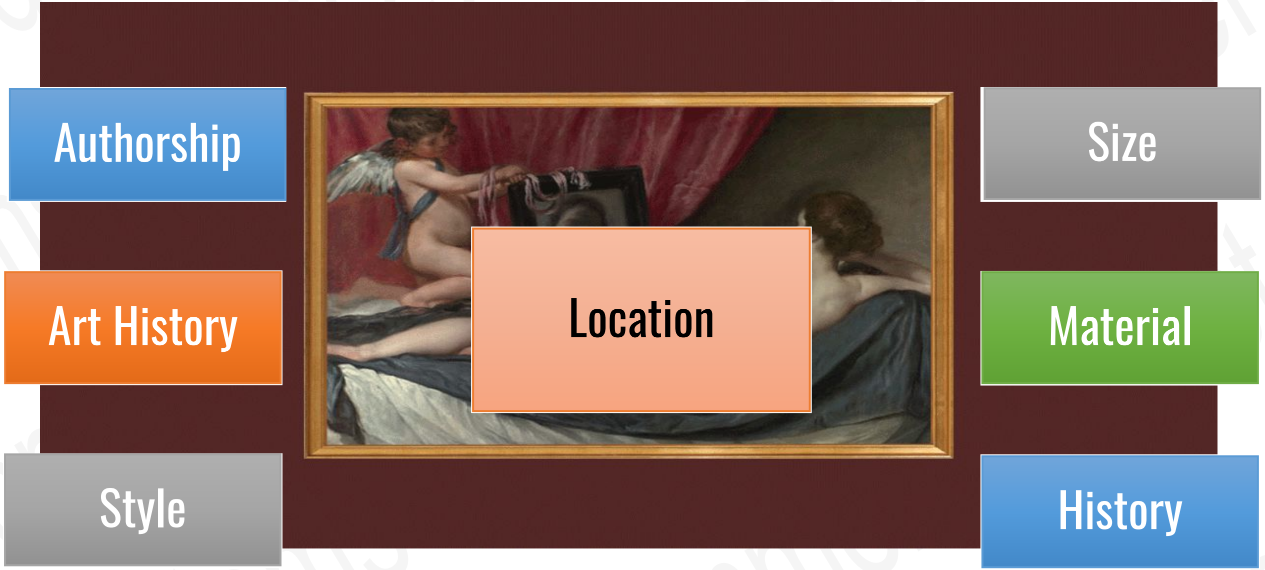
What makes a good Representation?

- ➔ Low dimensional (summarize data)
- ➔ Reusable across tasks
- ➔ Smooth and spatially coherent
- ➔ Disentangled
- ➔ Hierarchical and meaningful



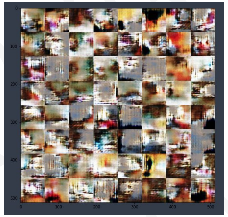
Early approaches in VA RecSys:

Manually curated metadata to drive recommendations.



The RecSys pipeline: A case-study approach

→ Train a model that can learn Visual features from images of paintings



→ Use pre-trained models as feature extractors
- ResNet, AlexNet, GoogLeNet, VGG

How? → Transfer Learning

The RecSys pipeline: A case-study approach

Transfer Learning

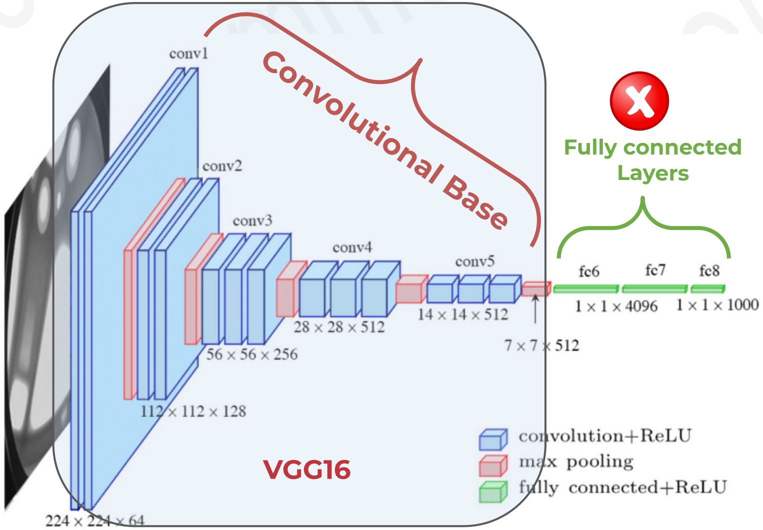
Big Data → **Model 1** → Very Good Results

Knowledge

Small Data → **Model 2** → Very Good Results

New Problem

> 1,000,000 images



RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

Transfer Learning

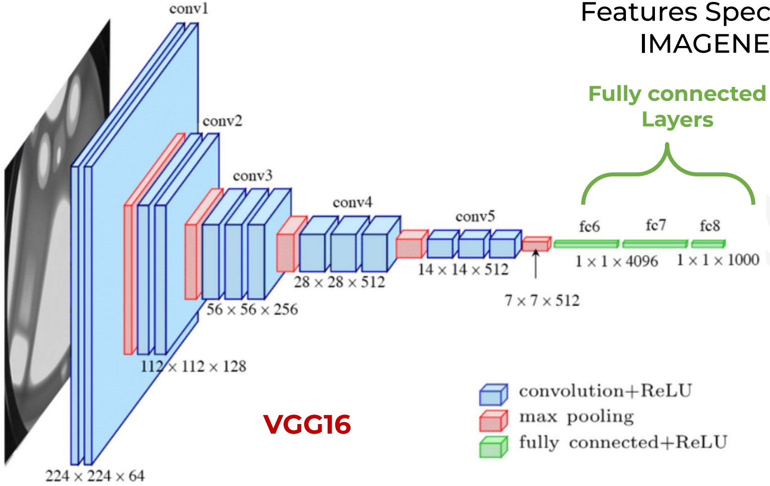
We want Features Very specific to the **Paintings** dataset

General features for any image classification task

Convolutional Base

Features Specific to IMAGENET

Fully connected Layers



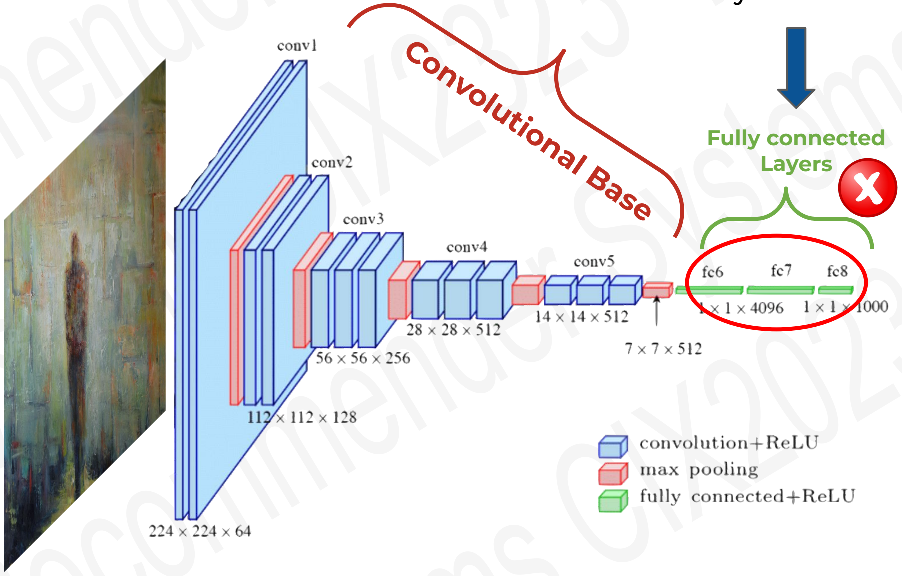
COMPUTER SCIENCE AND ENGINEERING UNIVERSITY OF MICHIGAN

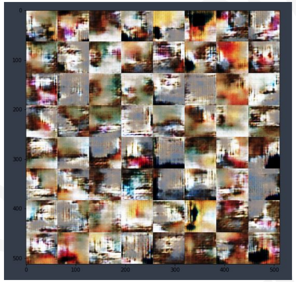
The RecSys pipeline: A case-study approach

Transfer Learning

Replace with Custom Fully connected layer for your task

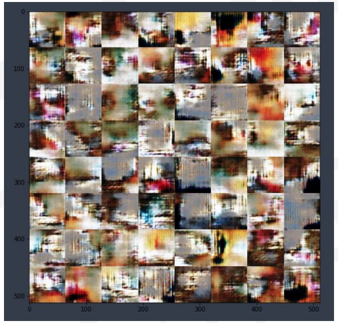
Features Specific to Painting dataset





Visually Similar paintings will be represented close to each other in the representation space

The RecSys pipeline: A case-study approach



Some issues:

- Recommendations don’t have direct interpretation.
- Often fail to capture complex semantics

The RecSys pipeline: A case-study approach

The Challenge in VA Recsys:

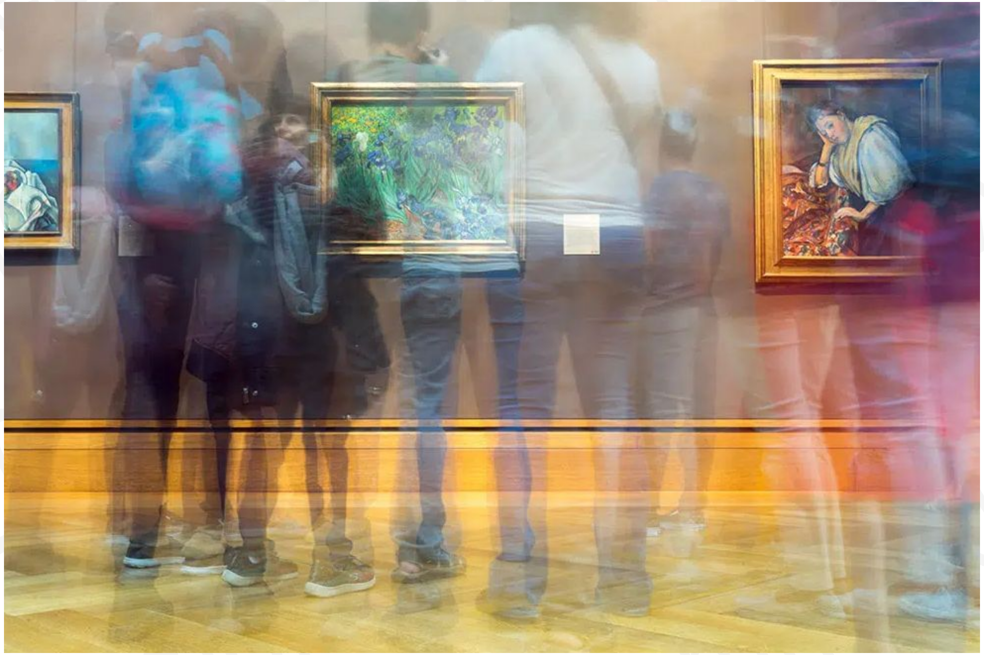
➔ Capturing the complexity of the concepts embedded within the artwork,

RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

The Challenge in VA Recsys:

➔ The emotional and cognitive reflections VA may trigger in users.



The Challenge in VA Recsys:

→ Understanding how users interact with highly subjective content.

The RecSys pipeline: A case-study approach

RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

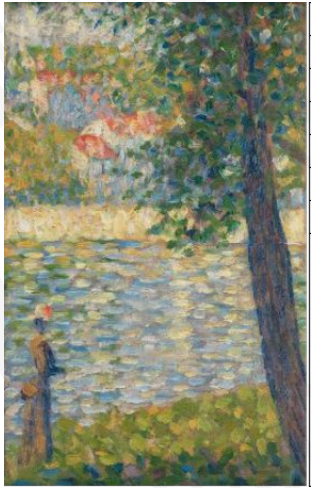
Data Pre-processing

Task
Personalised Recommendation

Model Training

Textual description

- ➔ Public reviews
- ➔ Books
- ➔ Articles, . . .

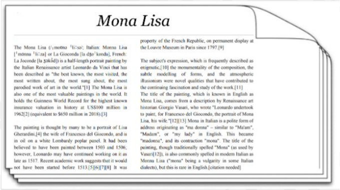


painting_id	000-018P-0000
title	The Morning Walk
artist	Georges Seurat
publication_date	19th_century
size_format	Portrait
size	Very Small
technique	oil painting
description	A woman, silhouetted against the shimmering water, strolls along a riverbank. The red roofs of houses can be made out along the opposite bank. Between 1882 and 1886 Seurat painted numerous such landscape studies on small wooden panels, some as independent works and others in preparation for his large-scale compositions. This sketch provided the starting point for a painting of 1885, 'The Seine at Courbevoie' (private collection).

The RecSys pipeline: A case-study approach

Latent Semantic Representation Learning

- We need representations that capture **hidden semantic relationships** between the paintings

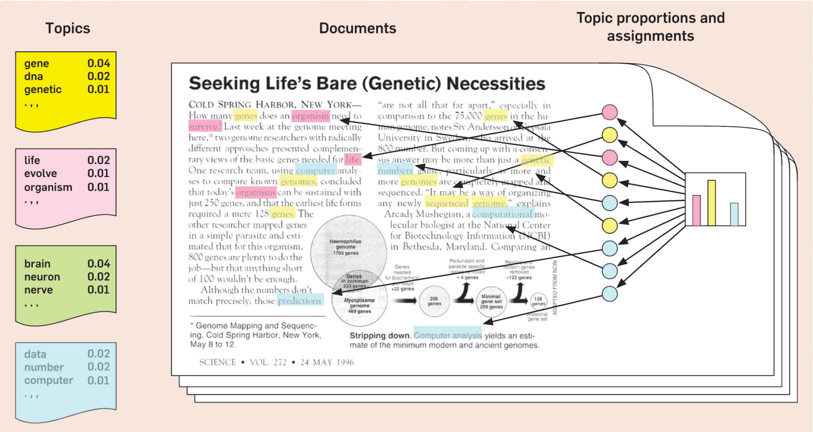


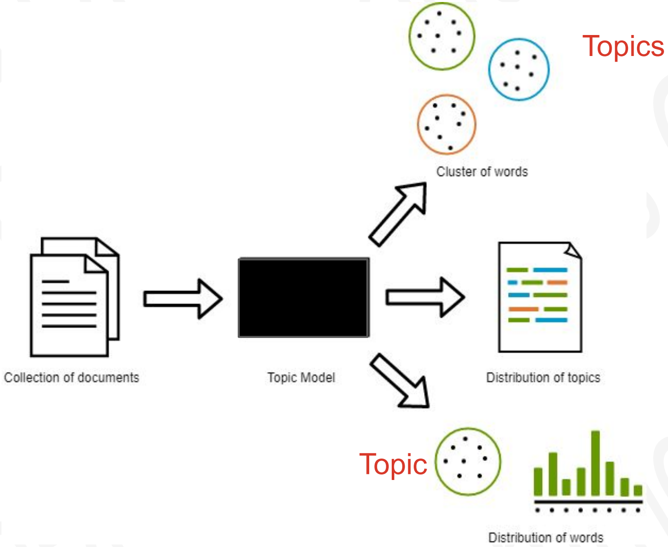
Topic Modelling

Topic Modelling

➔ A way or suit of techniques to identify **latent themes** in a corpus/ collection of documents.

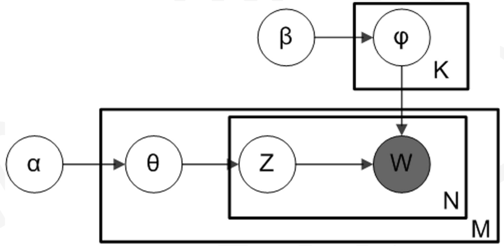
1. **Latent Dirichlet Allocation (LDA)**





Blei, D. M. (2012). Probabilistic Topic Models. Communications Of The Acm, 55(4), 77-84

Latent Dirichlet Allocation (LDA)



M denotes the number of documents

N is number of words in a given document (document *i* has N_i words)

α is the parameter of the Dirichlet prior on the per-document topic distributions

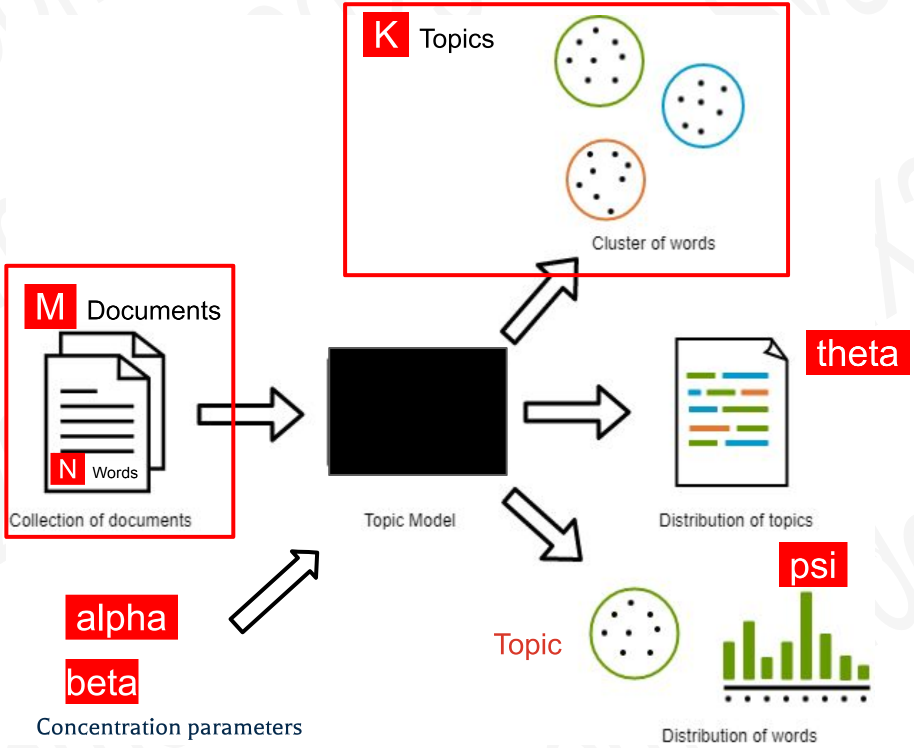
β is the parameter of the Dirichlet prior on the per-topic word distribution

θ_i is the topic distribution for document *i*

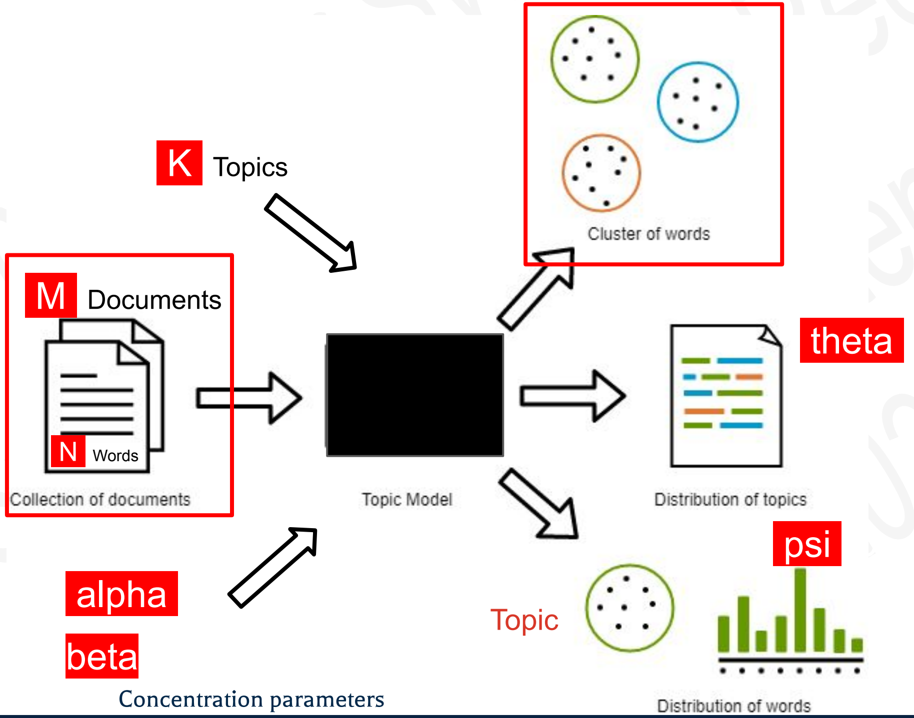
φ_k is the word distribution for topic *k*

z_{ij} is the topic for the *j*-th word in document *i*

w_{ij} is the specific word.



The RecSys pipeline: A case-study approach



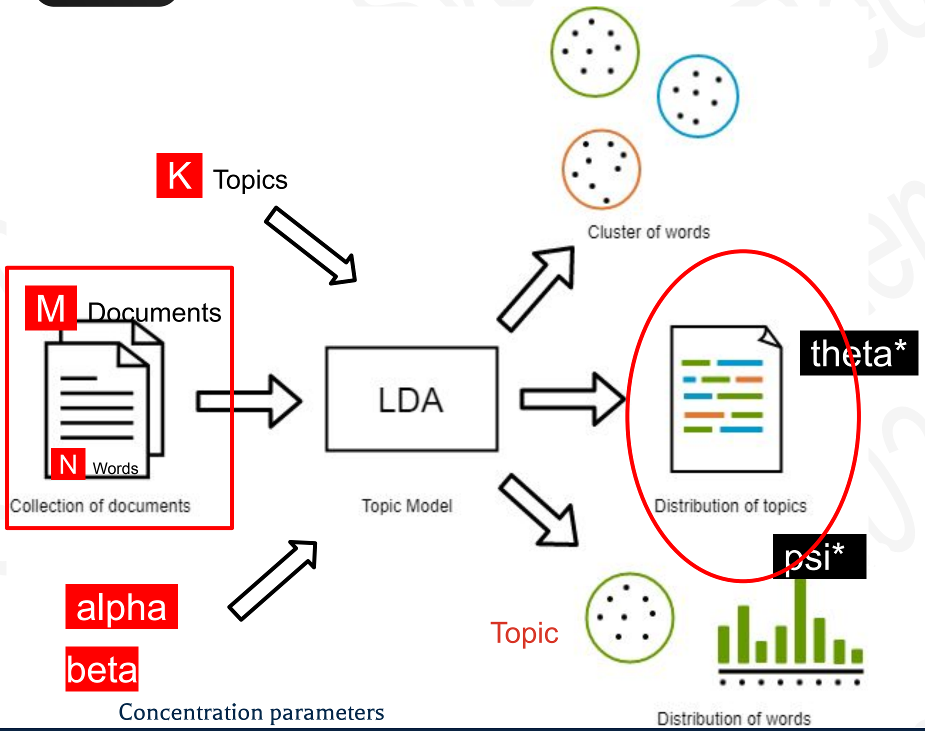
LDA Algorithm

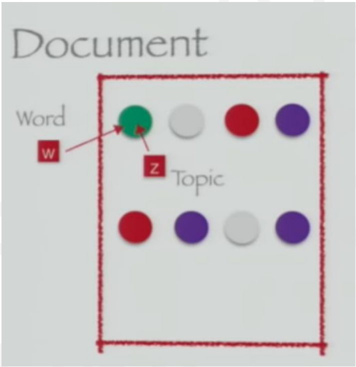
Iterative Algorithm

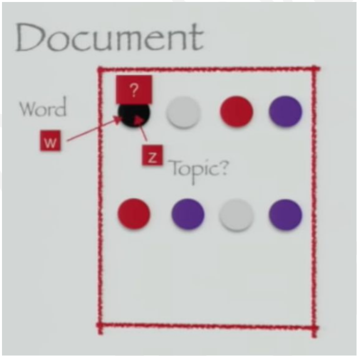
1. Initialize parameters
2. Initialize Topic assignments randomly
3. Iterate : For each word in each document
   - Resample topic for word given all other words and their current topic assignments

5. Get results
6. Evaluate model

The RecSys pipeline: A case-study approach



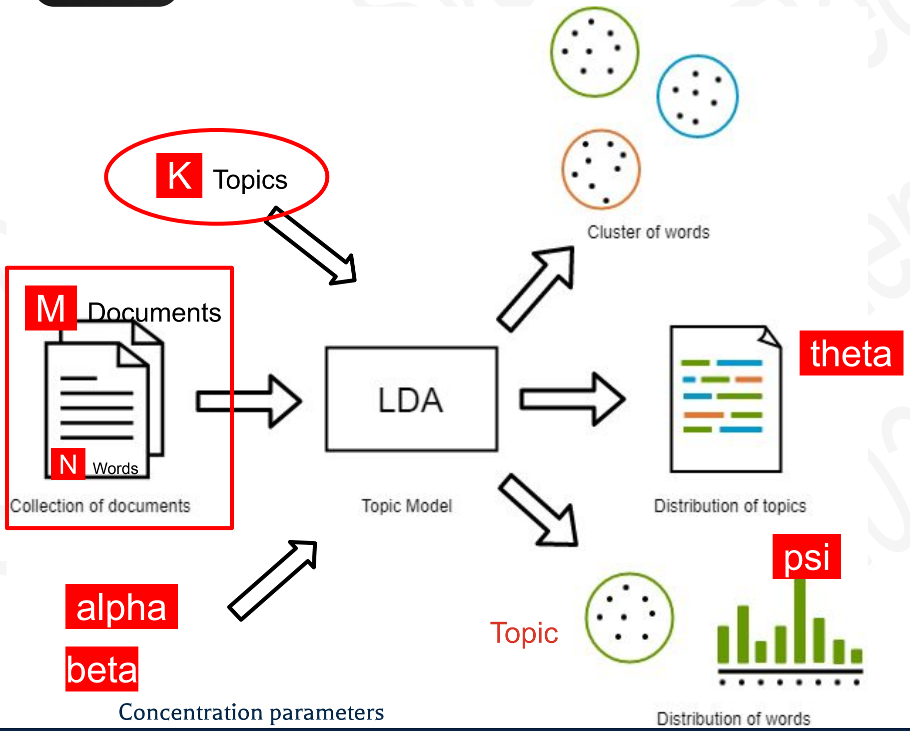




Resample topic for word given all other words and their current topic assignments

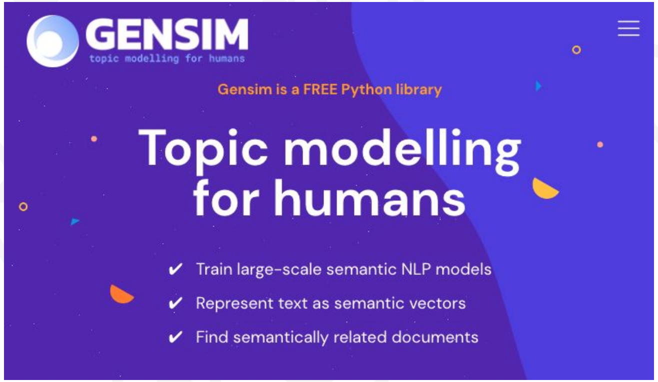
1. Which topics occur in this document?
2. Which topics like the word X

The RecSys pipeline: A case-study approach



➔ What is the optimal number of Topics?

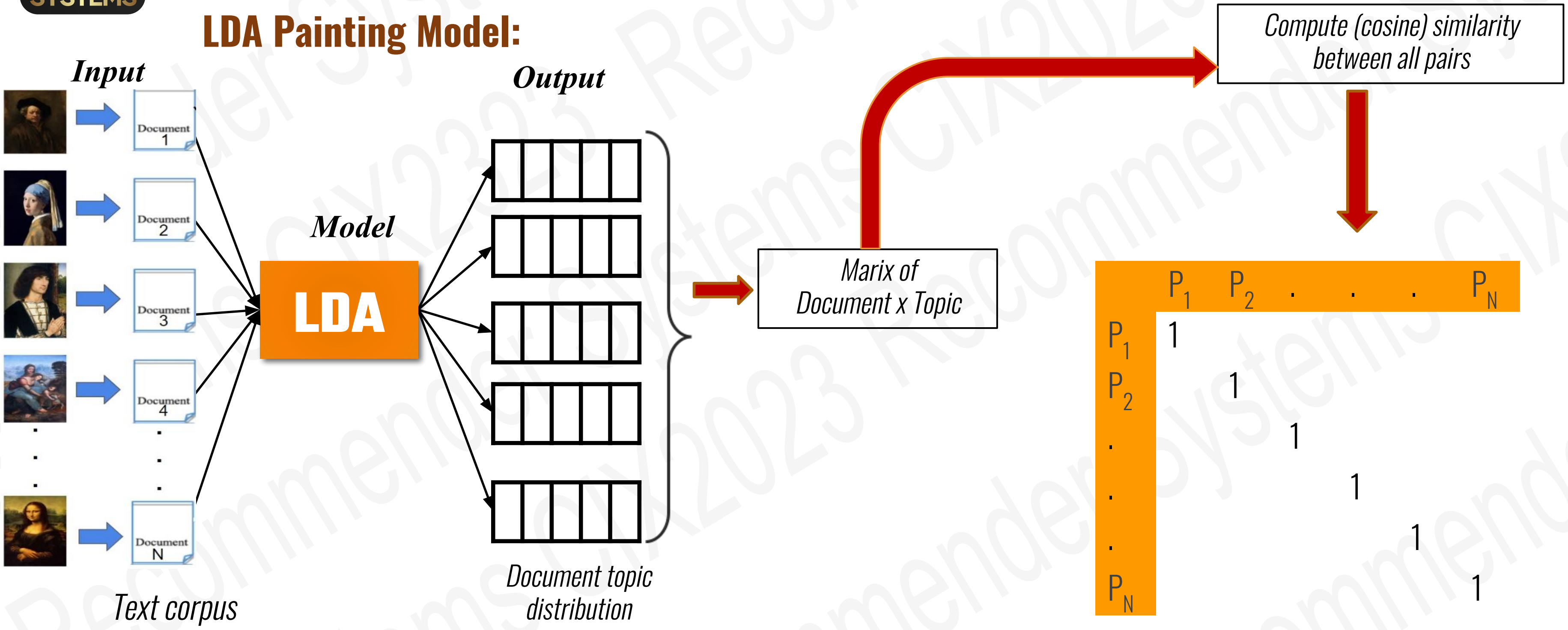
Topic coherence



More on LDA → Notes on Dirichlet Processes by Timothy Hopper: https://dp.tdhopper.com/

The RecSys pipeline: A case-study approach

LDA Painting Model:

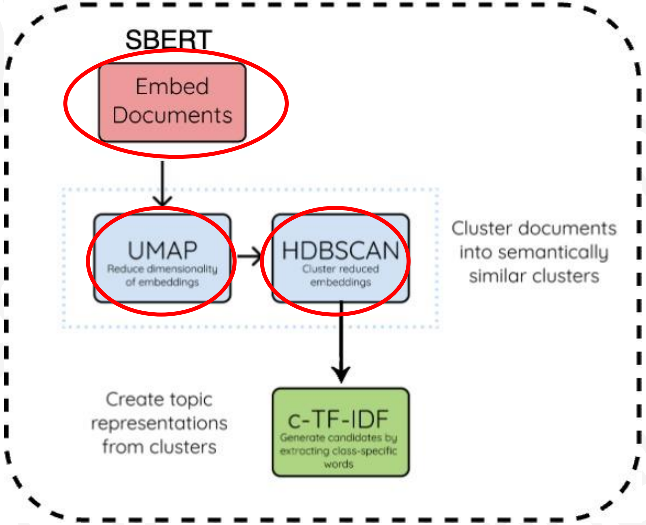


Pretrained Language model

Sentence Transformers

Google

BERT

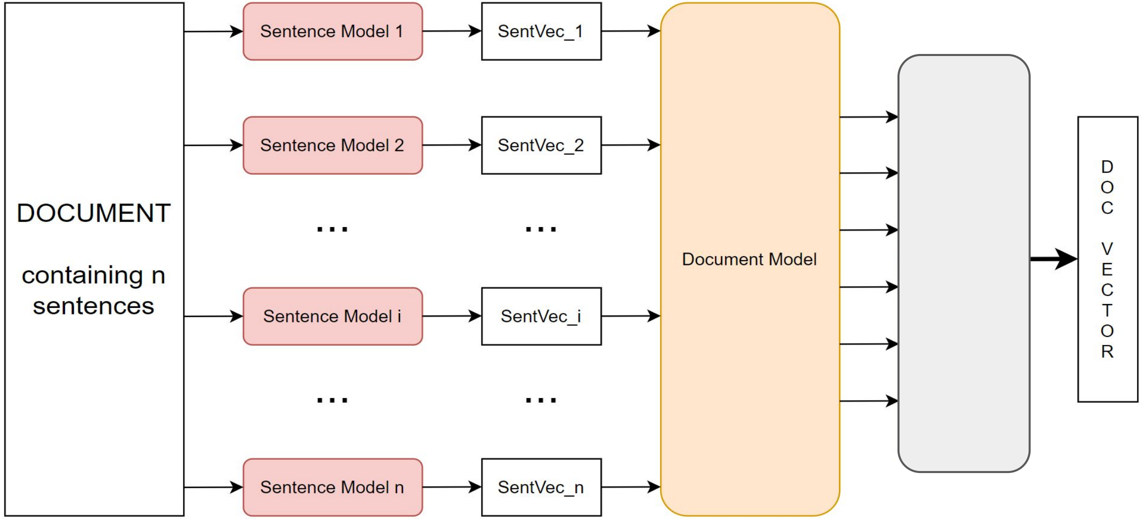


384-dimensional dense vector

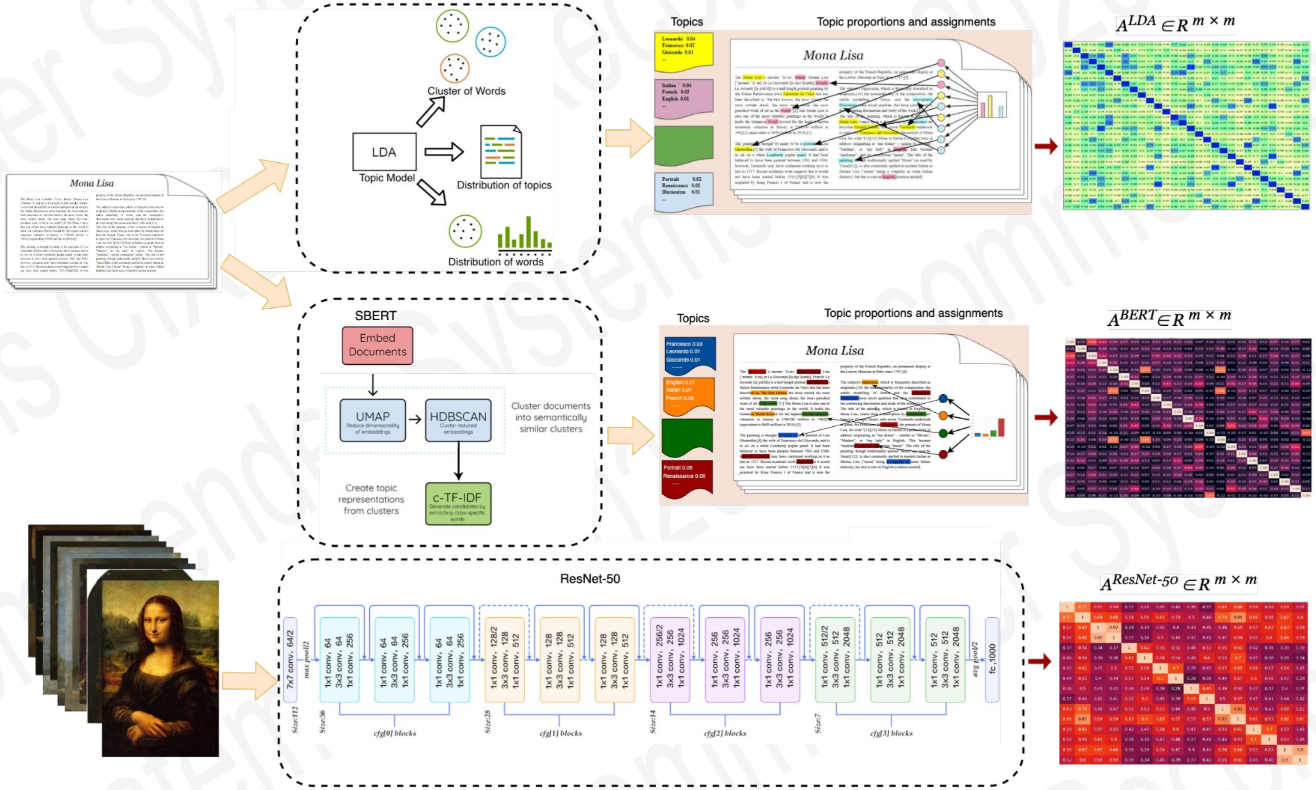
Neural topic modeling with a class-based TF-IDF procedure. (Maartin Grootendorst 2022)

The RecSys pipeline: A case-study approach

Pretrained Language model   Sentence Transformers



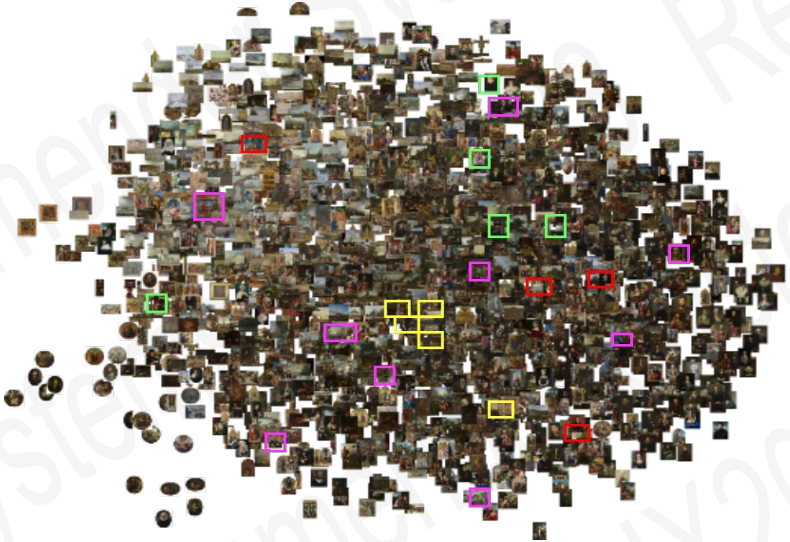
The RecSys pipeline: A case-study approach



The RecSys pipeline: A case-study approach

RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG



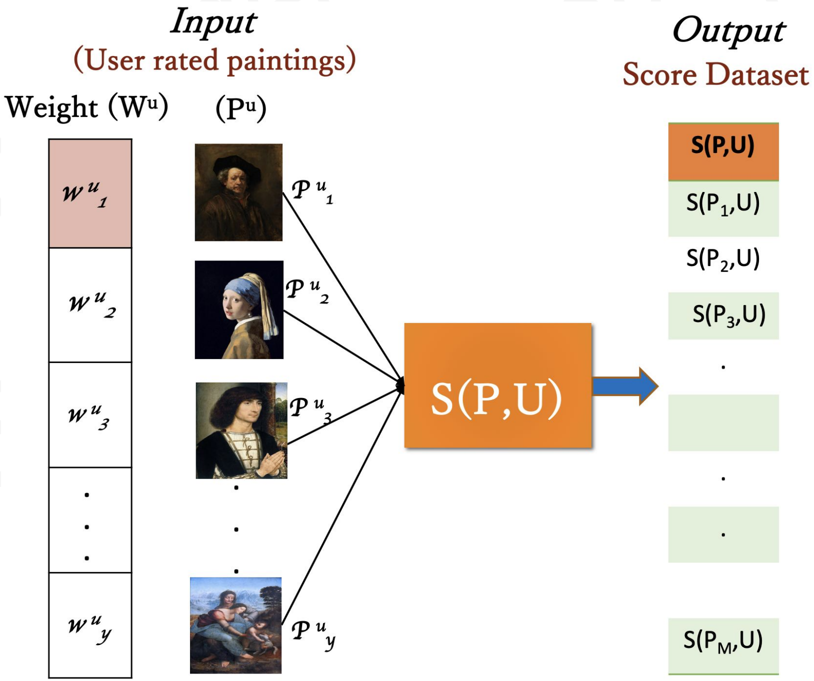
Latent Semantic Representation learning:

1	0.77	0.57	0.54	0.37	0.46	0.45	0.44	0.46	0.37	0.63	0.66	0.59	0.54	0.59	0.52
0.77	1	0.69	0.68	0.54	0.56	0.61	0.53	0.5	0.46	0.74	0.85	0.66	0.58	0.67	0.6
0.57	0.69	1	0.92	0.34	0.39	0.45	0.4	0.43	0.45	0.48	0.59	0.57	0.62	0.67	0.59
0.54	0.68	0.92	1	0.37	0.38	0.5	0.44	0.42	0.41	0.47	0.59	0.55	0.6	0.66	0.59
0.37	0.54	0.34	0.37	1	0.62	0.55	0.51	0.48	0.52	0.55	0.62	0.42	0.38	0.39	0.39
0.46	0.56	0.39	0.38	0.62	1	0.58	0.54	0.49	0.6	0.53	0.7	0.47	0.38	0.39	0.34
0.45	0.61	0.45	0.5	0.55	0.58	1	0.7	0.39	0.45	0.57	0.65	0.58	0.47	0.54	0.45
0.44	0.53	0.4	0.44	0.51	0.54	0.7	1	0.28	0.39	0.44	0.57	0.6	0.48	0.47	0.39
0.46	0.5	0.43	0.42	0.48	0.49	0.39	0.28	1	0.65	0.49	0.52	0.47	0.37	0.4	0.37
0.37	0.46	0.45	0.41	0.52	0.6	0.45	0.39	0.65	1	0.5	0.57	0.47	0.41	0.44	0.42
0.63	0.74	0.48	0.47	0.55	0.53	0.57	0.44	0.49	0.5	1	0.82	0.53	0.43	0.58	0.51
0.66	0.85	0.59	0.59	0.62	0.7	0.65	0.57	0.52	0.57	0.82	1	0.61	0.53	0.66	0.61
0.59	0.66	0.57	0.55	0.42	0.47	0.58	0.6	0.47	0.47	0.53	0.61	1	0.7	0.53	0.45
0.54	0.58	0.62	0.6	0.38	0.38	0.47	0.48	0.37	0.41	0.43	0.53	0.7	1	0.53	0.46
0.59	0.67	0.67	0.66	0.39	0.39	0.54	0.47	0.4	0.44	0.58	0.66	0.53	0.53	1	0.9
0.52	0.6	0.59	0.59	0.39	0.34	0.45	0.39	0.37	0.42	0.51	0.61	0.45	0.46	0.9	1

Similar paintings will be represented close to each other in the representation space

Data Pre-processing → Task: Personalised Recommendation → Model Training → Post Processing

Active user preferences → Normalized weight [0,1]

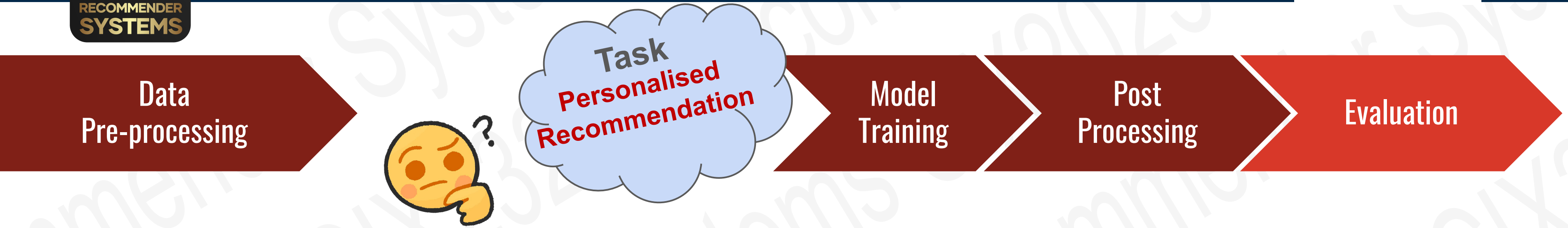


- The predicted Score $S(P_i, U)$ for a novel painting P_i with respect to an active user u is based on the weighted average score from all other paintings that have been rated by the active user.

$$S(P,U) = \frac{1}{N}\sum_{j=1}^{N} w_j * d(p_i, p_j)$$

- $d(p_i, p_j)$ is the similarity between p_i and p_j according to our model (LDA, BERT, ResNet..)
- Top K recommendation

The RecSys pipeline: A case-study approach



- Proxy for how relevant/ good are recommender systems.

1. Offline Experiment.

- Easiest to conduct
- Requires no interaction with users.

2. User Studies

- Small group of Subjects
- Controlled setting
- Qualitative/quantitative

3. Online Experiments

- The most trustworthy
- A pool of real users (unaware of the experiment)

- Common evaluation protocols in ML, IR

Evaluation

A couple of basic guidelines in general experimental studies

1. **Hypothesis:**

What do you want to evaluate? Form a **concise** and **restrictive** Hypothesis.

E.g. Algorithm **A** better predicts rating than Algorithm **B**

→ **Predictive Accuracy** not other factors

Algorithm **A** generates diverse recommendations than Algorithm **B**

→ **Diversity** not other factors, etc.

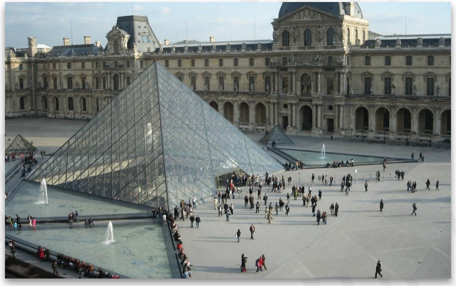
A couple of basic guidelines in general experimental studies

2. **Controlling variables:** Variables that are not being tested should stay fixed.

E.g. Algorithm **A** better predicts rating than Algorithm **B** → **Predictive Accuracy**

Train A on NG dataset

Train B on Louvre dataset



Can not tell why A > B or B < A

- Superior model?
- better input data?

Train both A & B on the same dataset.

RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

Evaluation: Offline Experiments

Performed using a pre-collected dataset of users choosing or rating of items.

- Using this data we can simulate behaviour of users that interact with the a Recommender System: (assuming users will behave the same when the Recommender System is deployed.)
  - ➢ To make reliable decision.

- **Attractive & easy:**
  - ➢ Requires no interaction with real users.
  - ➢ We can compare wide range of candidate algorithms at low cost.

RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

Evaluation: Offline Experiments

- **Downside:** We can not measure the recommender's influence on user behavior.

Useful for :

- ➢ Filtering out inappropriate approaches, select candidate algorithms for more costly user studies/online experiments.
- ➢ Tuning parameters of algorithms.

RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

Evaluation: User Studies

Conducted by recruiting a set of test subject, and asking them to perform several tasks requiring an interaction with the recommendation system.

- During interaction we observe and record their behaviour.

Quantitative measures: % of completed task, accuracy, time, etc.

Qualitative measures: pre/post interaction questions that are not directly observable.

- ➢ weather the subject enjoyed the UI
- ➢ weather the subject perceived the task easy, etc.

RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

Evaluation: User Studies

Downsides:

- Expensive to conduct (large set of subjects, large set of tasks)
- Subjects (volunteers or employed)
  - ➢ Motivations (intrinsic/ extrinsic)--> quality of response
  - ➢ Budget
- Testing all possible scenarios can be challenging.
- Finding subjects that represent the entire population. (Bias)

Pilot user studies: to test the systems for bugs and malfunctions.

RECOMMENDER SYSTEMS

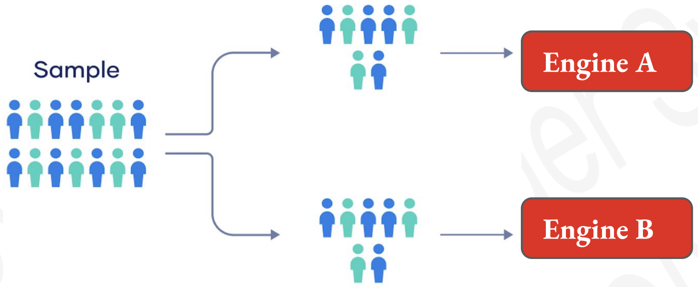
UNIVERSITÉ DU LUXEMBOURG

Evaluation: User Studies

Between vs. Within Subjects

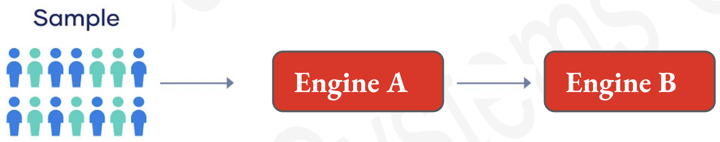
Few candidate approaches: each method must be tested on the same task.

1. Between Subjects (A-B testing)



- Easier to setup and analyse correctly
- No learning across conditions
- Test long term effects

2. Within Subjects



- More informative (superiority of methods)
- Can ask comparative questions about candidates

Evaluation: Online Evaluation

Done with a set of real users unaware of the experiment.

- Many realistic RecSys applications wish to influence the behaviour of users. (engagement, purchase)
- Measure the change in user behavior while interacting with recommendations.

  - ➢ Did the user follow recommendations? for how long?
  - ➢ Weather some utility gathered from users of system A exceeds the one from system B. (which system is superior?)

Evaluation: Online Evaluation

➢ **Most reliable: Real users with real needs in the context.**

- The real effect of Recommendation systems depends on several factors
  - ➢ Users' intent
    - How specific are their needs?
    - How much novelty are they looking for?
    - How much risk are they willing to take?
  - ➢ Users' context
    - What items are they already familiar with ?
    - how much they trust the system?
  - ➢ The Recommendation interface

Evaluation: Online Evaluation

Downsides:

- May discourage users from using the real system ever again.
- Financial risk in commercial applications.

To reduce such risks it is best to run an online evaluation last

- After an extensive Offline study followed by a user study provided evidence that the candidate approaches are reasonable.

RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

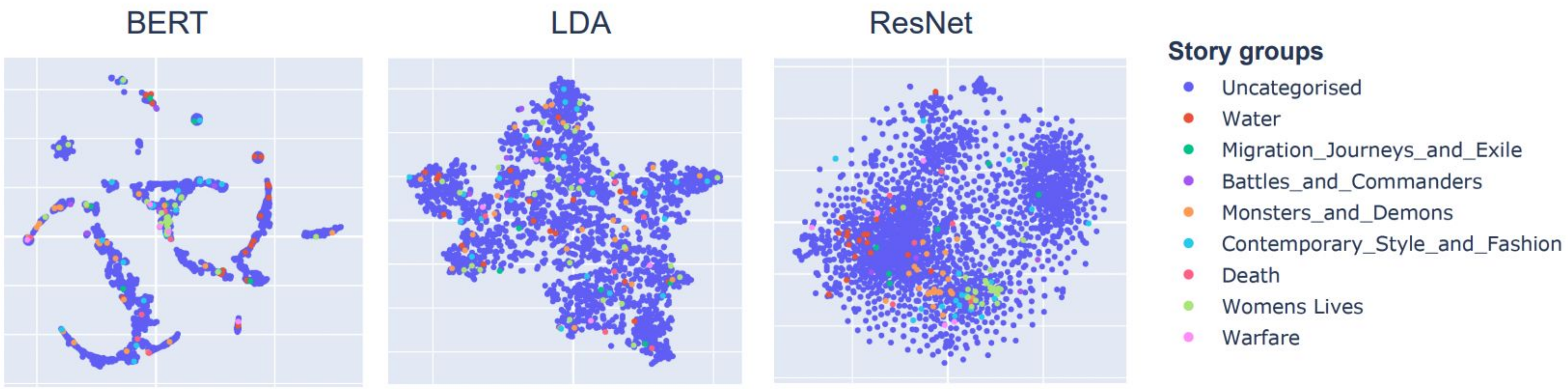
Evaluation: Drawing Reliable Conclusions

Even after a carefully conducted experiment Recsys might fail in unseen scenarios when deployed in real life.

To reduce such risks it is best to perform significance testing on the results.

The Elements of Visual Art Recommendation

Latent space projection (t-SNE) of the curated story groups

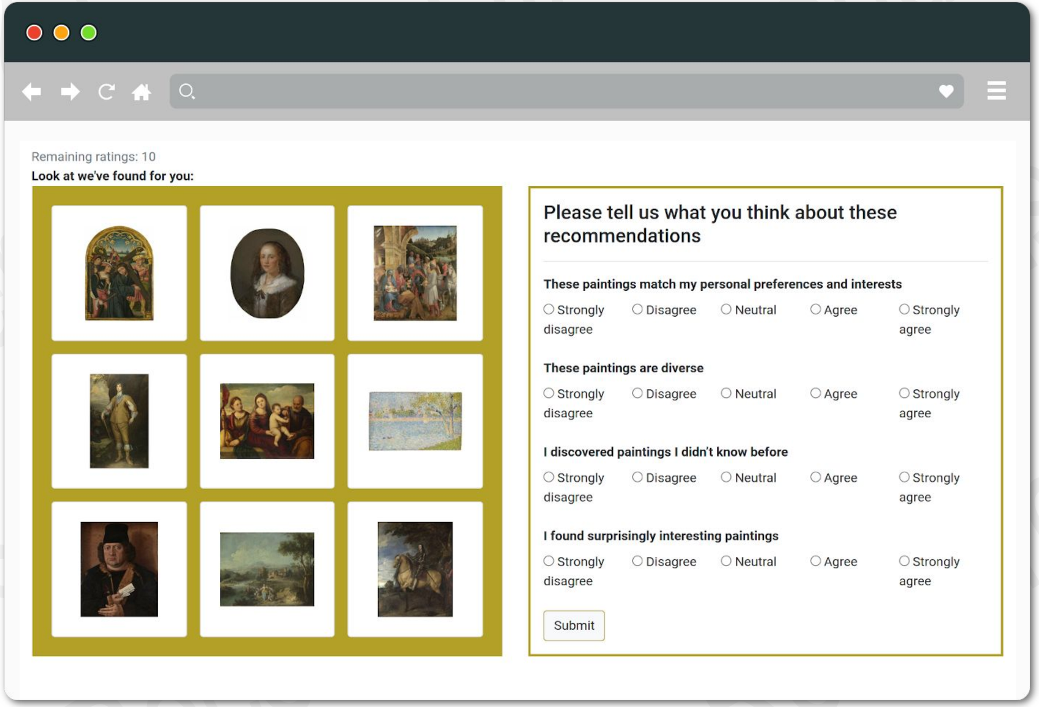


- We developed Visual art recommendation engines

BERT **LDA** **ResNet**

Evaluation: User Study

User-centric evaluation



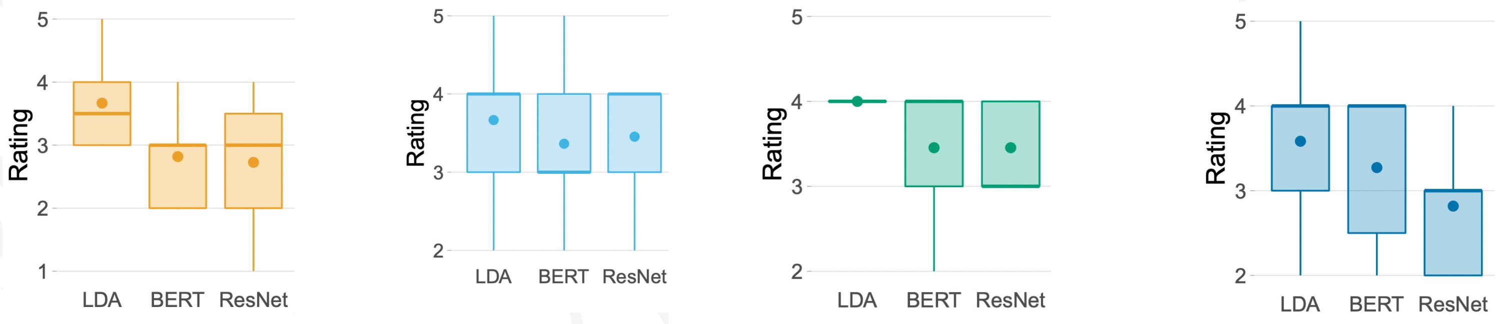
The RecSys pipeline: A case-study approach

RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

Evaluation: User Study

User-centric evaluation



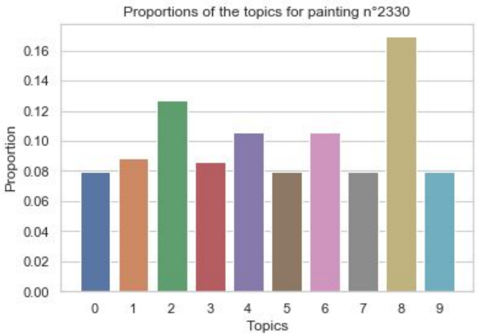
(a) Accuracy (b) Diversity (c) Novelty (d) Serendipity

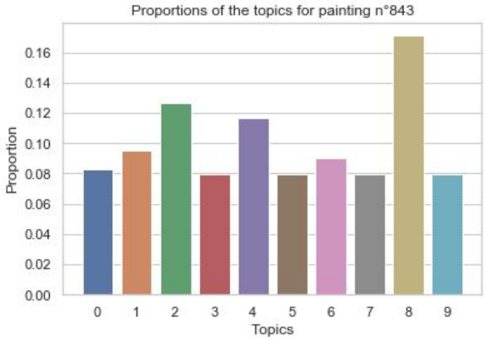
The RecSys pipeline: A case-study approach

Explaining Recommendations

Target painting

Most similar painting



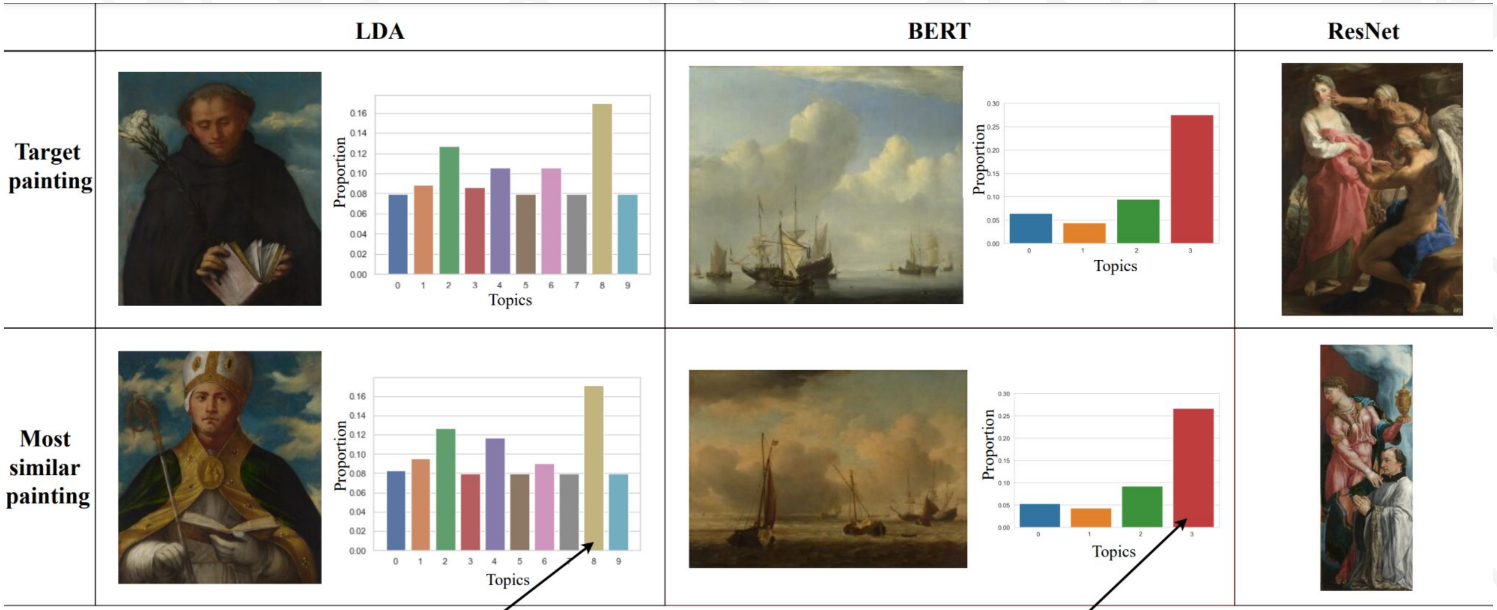


Topic 8
CHRIST
SAINT
JESUS
EVANGELIST
CROSS
CHURCH

Explainable recommendations have a positive impact on user experience.

The RecSys pipeline: A case-study approach

Explaining Recommendations



(christ, saint, altarpiece, panel, Jesus, new testament, evangelist, cross, church crucification)

(landscape, oil, van, anchor, 17th_century, river, view, scene, 17th_century landscape)

The RecSys pipeline: A case-study approach

UNIVERSITÉ DU LUXEMBOURG

Explaining Recommendations

Beauty

Old Age

Time

Batoni intends to encourage considering the brevity of youth and the inevitable passing of time.

Time orders Old Age to destroy Beauty
by Pompeo Girolamo Batoni
18th century

Semantic Gap

Mary Magdalene

precious ointment

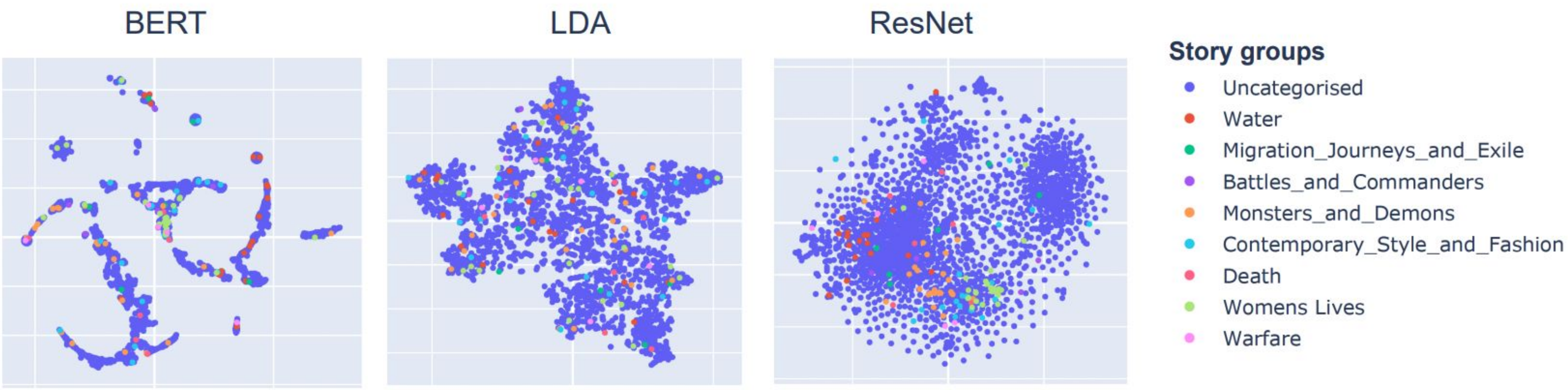
Religious

The Donor and Saint Mary Magdalene
by Marten van Heemskerck.
16th century

The Elements of Visual Art Recommendation

uni.lu
UNIVERSITÉ DU LUXEMBOURG

Latent space projection (t-SNE) of the curated story groups



- We developed Visual art recommendation engines

BERT **LDA** **ResNet**

Reciprocal rank fusion

ResNet + BERT		ResNet + LDA	
25%	75%	25%	75%
50%	50%	50%	50%
75%	25%	75%	25%

Late Fusion Results

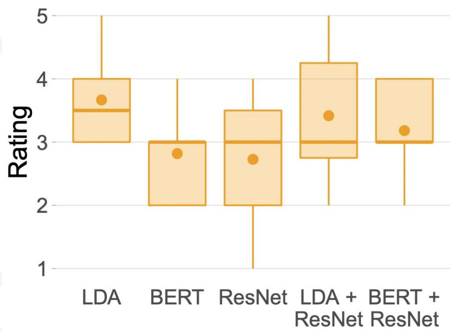
RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

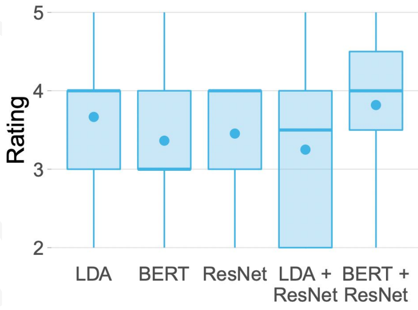
Evaluation: User Study

Distribution of ratings from museum users.

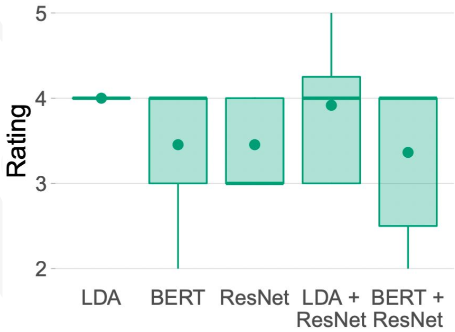
Dots denote mean values.



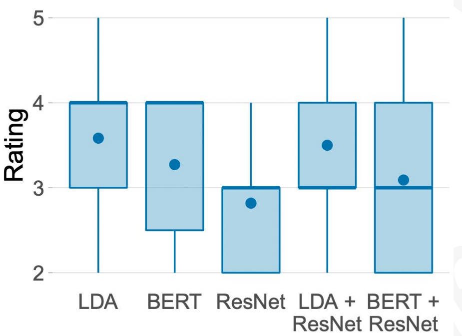
(a) Accuracy



(b) Diversity



(c) Novelty



(d) Serendipity

Yilma et al. (CHI23 The Elements of visual Art recommendation)

Late Fusion Results

RECOMMENDER SYSTEMS

UNIVERSITÉ DU LUXEMBOURG

Evaluation: User Study

➔ Latent semantics inherent in visual arts are effectively captured through the fusion of modalities.

➔ Distinct modalities lead to varying degrees of exposition of latent semantics

The RecSys pipeline: A case-study approach

Data Pre-processing →  → Model Training → Post Processing → Evaluation

- Sort
- Filter
- Recommend

Next session:

- **A Multi-Stakeholder aware RecSys**
- **How to formulate the RecSys Problem? →(A framework)**

Next → Session 3

- → A Multi-Stakeholder aware RecSys
- → Formulating the RecSys Problem? →(A framework)
- → Modern RecSys Paradigms
  - ◆ Large Language Models as RecSys (Zero & Few shot)
- → Common Issues and Challenges in RecSys