

**fuse** | machines

# PROGRESS PRESENTATION ON RECOMMENDATION UPSKILLING

**Nirajan Bekoju**



# BASIC OVERVIEW

- Week 1, 2, 3 (Completed)
- Week 4 (In progress)
- Week 5 (Not started yet)

# GITHUB REPOSITORY

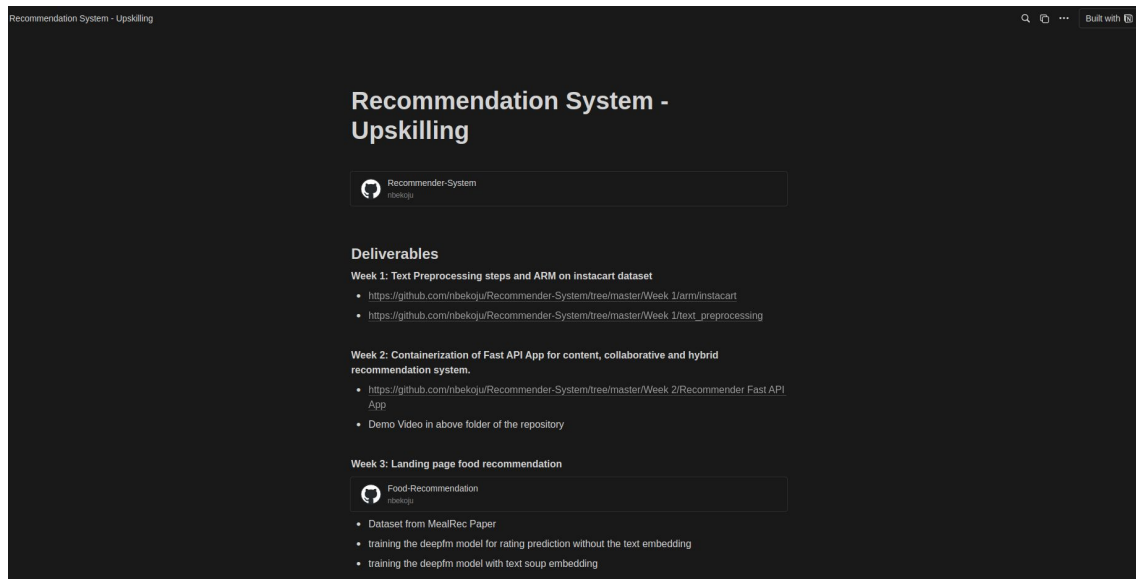
- <https://github.com/nbekoju/Recommender-System>

The screenshot shows the GitHub repository page for 'Recommender-System' by user nbekoju. The repository is public and has 1 branch (master) and 0 tags. The page displays a list of files and folders, including 'Hands-On Recommendation Systems with Py...', 'Week 1', 'Week 2', 'Week 3', 'Week 4', 'instacart\_data', 'paper', '.gitignore', '.pre-commit-config.yaml', 'Readme.md', 'Recommendation Engine - AI Apprenticeship ...', and 'Recommendation Upskilling.pdf'. The commit history shows updates to the README and various files over time.

File/Folder	Commit Message	Commit Hash	Time Ago
Hands-On Recommendation Systems with Py...	add fast api demo video	18c26a7	last week
Week 1	complete collaborative filtering		2 weeks ago
Week 2	complete recommender system from kaggle		last week
Week 3	complete factorization machine implementation from medium		2 days ago
Week 4	work on use modules recbole doc		4 days ago
instacart_data	complete deepctr documentation examples		last week
paper	update readme.md		2 days ago
.gitignore	add fnfm code and download all sequential recommender r...		5 days ago
.pre-commit-config.yaml	rename folder name in week 3		5 days ago
Readme.md	update readme.md		2 days ago
Recommendation Engine - AI Apprenticeship ...	progress day 1 - week 1		3 weeks ago
Recommendation Upskilling.pdf	start sequential recommender quick start using recbole doc		4 days ago

# MY UPSKILLING JOURNEY - RECOMMENDATION SYSTEM

- <https://nirajanbekoju.notion.site/Recommendation-System-Upskilling-487af8811c8d46d2a59049b9a5c9f4f1>
- Resources
- Code Link
- Research Paper Review



# WEEK 1 - RESOURCE EXPLORATION

- Text Preprocessing
  - Tokenization, stemming, lemmatization, stop words and punctuation removal
  - Removal of URLs, HTML Tags
  - Chat Word Conversion (AFK → Away From Keyboard)
- Text Encoding
  - BOW, TF-IDF, Word2Vec, sBERT
- Association Rule Mining (ARM)
  - Apriori Method
  - FP-Growth Method

# WEEK 1 - DELIVERABLES 1 : Text Preprocessing

Recommender-System / Week 1 / text\_preprocessing / text\_preprocessing.ipynb

nbekoju complete text preprocessing and ecommerce product embedding c918b25 · 3 weeks ago History

Preview Code Blame 2248 lines (2248 loc) · 68.2 KB Code 55% faster with GitHub Copilot Raw Copy Download Edit

## Text Preprocessing

For an unstructured data like text, preprocessing is one of the most important phase. Some of the common text preprocessing / cleaning steps are

- lower casing
- removal of punctuations
- removal of stopwords
- removal of frequent words
- Removal of Rare words
- Stemming
- Lemmatization
- Removal of emojis
- Removal of emoticons
- Conversion of emoticons to words
- Conversion of emojis to words
- Removal of URLs
- Removal of HTML tags
- Chat words conversion
- Spelling correction

We need to carefully choose our preprocessing steps based on our use case. For example: in sentimental analysis, we need not remove the emojis or emoticons as it will convey information about the sentiment.

# WEEK 1 - DELIVERABLE 2

## **ARM ON INSTACART DATASET**

# ARM ON INSTACART DATASET : Data Preparation

- Instacart Data shows order id and its corresponding product metadata

	order_id	product_id	add_to_cart_order	reordered	product_name	aisle_id	department_id
0	1	49302	1	1	Bulgarian Yogurt	120	16
1	1	11109	2	1	Organic 4% Milk Fat Whole Milk Cottage Cheese	108	16
2	1	10246	3	0	Organic Celery Hearts	83	4
3	1	49683	4	0	Cucumber Kirby	83	4
4	1	43633	5	1	Lightly Smoked Sardines in Olive Oil	95	15



# ARM ON INSTACART DATASET : Data Preparation

- For an order, get the list of product from previous data

	order_id	products
0	1	[Bulgarian Yogurt, Organic 4% Milk Fat Whole M...
1	36	[Grated Pecorino Romano Cheese, Spring Water, ...
2	38	[Shelled Pistachios, Organic Biologique Limes,...
3	96	[Roasted Turkey, Organic Cucumber, Organic Gra...
4	98	[Natural Spring Water, Organic Orange Juice Wi...

# ARM ON INSTACART DATASET : RESULT

- ARM using apriori shows association between items in descending order by lift

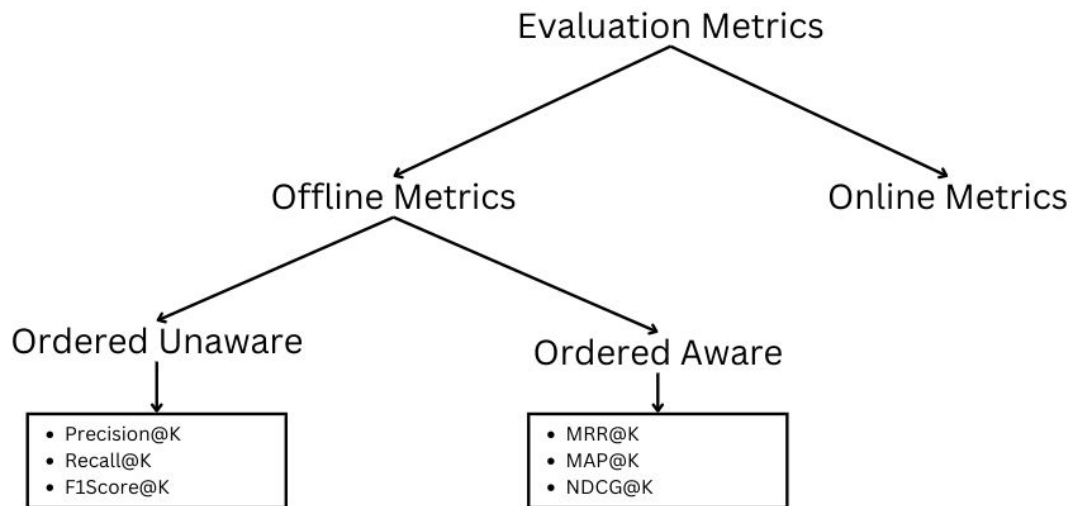
	Item #1	Item #2	Support	Confidence	Lift
2880	Unsweetened Whole Milk Peach Greek Yogurt	Unsweetened Whole Milk Strawberry Yogurt	0.000381	0.531915	606.887142
2878	Unsweetened Whole Milk Blueberry Greek Yogurt	Unsweetened Whole Milk Strawberry Yogurt	0.000328	0.500000	570.473913
2879	Unsweetened Whole Milk Mixed Berry Greek Yogurt	Unsweetened Whole Milk Strawberry Yogurt	0.000312	0.493976	563.600733
2877	Unsweetened Blackberry Water	Unsweetened Watermelon Water	0.000305	0.449438	440.077142
1300	Organic 4 Months Butternut Squash Carrots Appl...	Stage 1 Apples Sweet Potatoes Pumpkin & Bluebe...	0.000335	0.453608	407.654004
1538	Organic Blended Raspberry Whole Milk Greek Yogurt	Organic Greek Whole Milk Blended Strawberry Yo...	0.000305	0.416667	364.469444
57	Almond Milk Blueberry Yogurt	Almond Milk Peach Yogurt	0.000594	0.430939	353.394406
2433	Organic Pears, Peas and Broccoli Puree Stage 1	Stage 1 Apples Sweet Potatoes Pumpkin & Bluebe...	0.000320	0.365217	328.217868

# WEEK 2

## WEEK 2 - RESOURCE EXPLORATION

- Content-based, Collaborative-based and Hybrid Recommendation
  - These topics are covered by **Book: Hands on Recommendation System With Python (Chapter 4, 5, 6, 7)**
- Cold Start and **Long Tail Problems**
- **Evaluation Metrics**
- Deployment using FastAPI
- Containerization using Docker

# WEEK 2 - EVALUATION METRICS



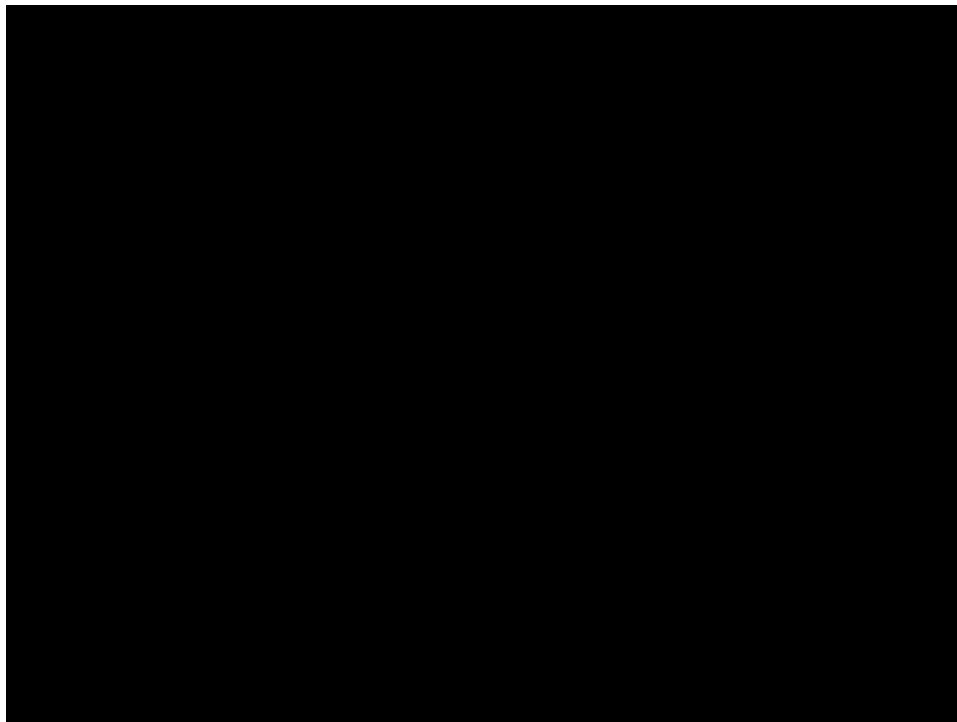
MRR → Mean Reciprocal Rank  
MAP → Mean Average Precision  
NDCG → Normalized Discounted Cumulative Gain

**Spotify uses  
MRR@30 as the  
evaluation metric**

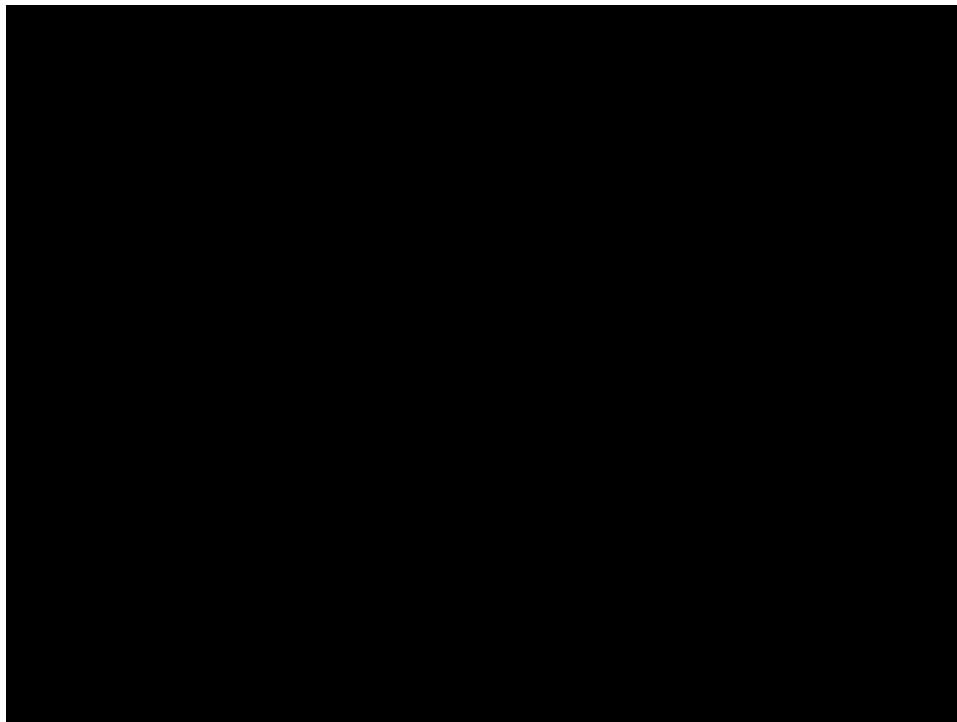
## WEEK 2 - DELIVERABLES

### **FastAPI Application: Content-based, Collaborative-based and Hybrid Movie Recommendation System**

## WEEK 2 - DELIVERABLE : DEMO VIDEO



## WEEK 2 - DELIVERABLE : DEMO VIDEO






# WEEK 2 - DELIVERABLE : DOCKERFILE

Code

Blame

29 lines (20 loc) · 799 Bytes

 Code 55% faster with GitHub Copilot

```
1 FROM continuumio/miniconda3
2
3 WORKDIR /app
4
5 # Update Conda
6 RUN conda update -n base -c defaults conda
7
8 # Install build dependencies
9 RUN apt-get update && apt-get install -y \
10     build-essential \
11     libatlas-base-dev
12
13 # create the conda environment
14 COPY environment.yml .
15
16 # Update Conda and create the Conda environment
17 RUN conda env create -f environment.yml
18
19 # Make RUN commands use the new environment:
20 SHELL ["conda", "run", "-n", "recommender_fast_api", "/bin/bash", "-c"]
21
22 # Copy the rest of the application code to the working directory
23 COPY . .
24
25 # Expose the port that app runs on
26 EXPOSE 8000
27
28 # Activate the environment and run the FastAPI application
29 CMD ["conda", "run", "--no-capture-output", "-n", "recommender_fast_api", "uvicorn", "main:app", "--host", "0.0.0.0", "--port", "8000"]
```

# WEEK 3

# WEEK 3 - RESOURCE EXPLORATION

- Factorization Machine
- Deep Factorization Machine
- Neural Factorization Machine
- Field Aware Neural Factorization Machine
- DeepCTR
  - SparseFeat vs DenseFeat
  - Embeddings
  - Converting data to model format
  - Training, Evaluation and Inference
- MLFlow

# WEEK 3 - REPOSITORY OVERVIEW

Recommender-System / Week 3 /

Add file



nbekoju complete factorization machine implementation from medium

fb8992b · 2 days ago

History

Name	Last commit message	Last commit date
..		
FNFM: deepctr avaze CTR prediction	start sequential recommender quick start using recbole doc	4 days ago
Factorization Machine - Kaggle	complete factorization machine rating prediction	2 weeks ago
Factorization Machine-Medium	complete factorization machine implementation from medium	2 days ago
MLFlow	complete quick start to mlflow	last week
Notes	add fnfm code and download all sequential recommender resources	5 days ago
SVD	complete SVD image compression	2 weeks ago
SVM	complete SVM classifier	2 weeks ago
deepCTR examples from documentation	rename folder name in week 3	5 days ago
ReadMe.md	add readme file in week 3	2 weeks ago

## WEEK 3 - DELIVERABLES

### LANDING PAGE FOOD RECOMMENDATION USING deepCTR

<https://github.com/nbekoju/Food-Recommendation>

Dataset: <https://github.com/WUT-IDEA/MealRec>

DeepFM: <https://arxiv.org/abs/1703.04247>

# LANDING PAGE FOOD RECOMMENDATION: DATA PREPARATION

	recipe_id	recipe_name	review_nums	category	aver_rate	ingredients	cooking_directions	nutritions	tags
0	6698	Mom's Zucchini Bread	9355	appetizer	4.754207	purpose flour salt baking soda baking powder g...	directions u prep n20 ncook n1 h nready n1 h 4...	u niacin u hascompletedata true u name u niaci...	north american breads easy beginner cook inexp...
1	6709	Zucchini Walnut Bread	58	appetizer	4.478873	chopped walnuts eggs white sugar vegetable oil...	directions u whisk together flour baking soda ...	u niacin u hascompletedata true u name u niaci...	weeknight breads fruit vegetables kid friendly...

## Recipe Data

- Text preprocessing

## User Recipe

### Interaction Data

- Date to year, month...

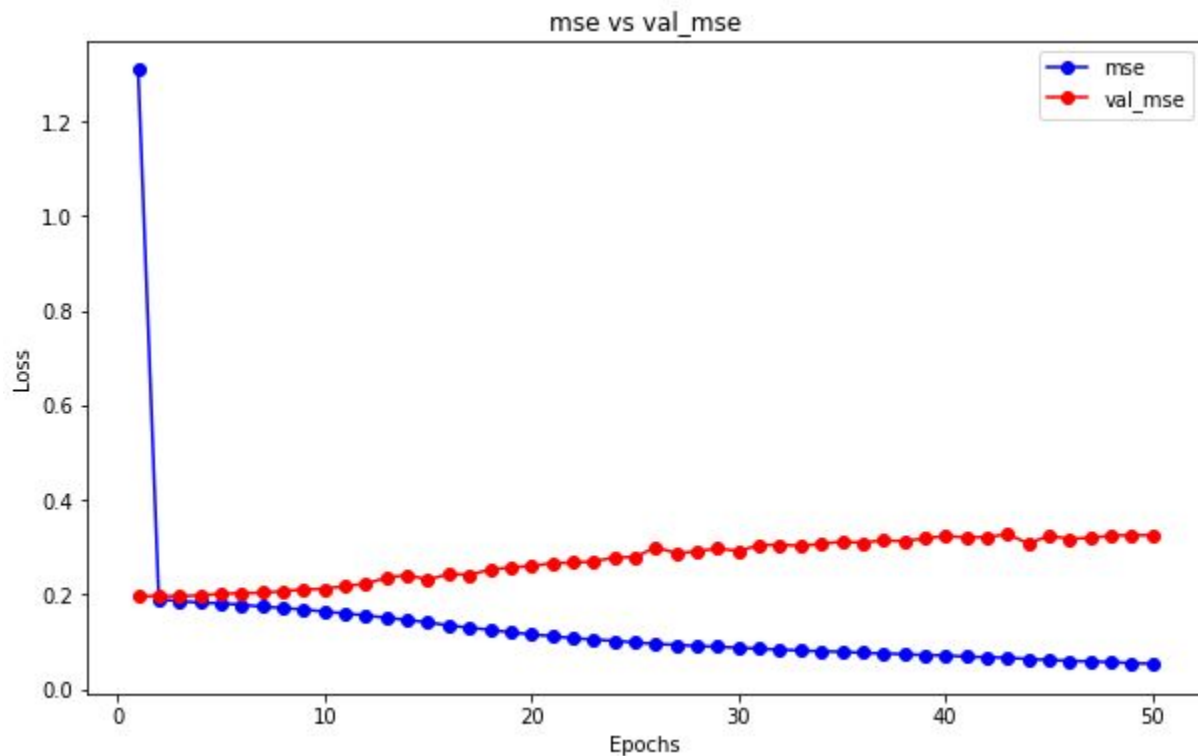
	user_id	recipe_id	rating	year	month	day	hour
0	39	61727	4	2001	3	26	7
1	39	7612	5	2004	8	2	16
2	39	12009	4	2001	2	28	10
3	39	88185	5	2009	5	18	14
4	39	24445	5	2015	5	7	15

# LANDING PAGE FOOD RECOMMENDATION: DATA PREPARATION

- Join the two dataset
- Merge all text data into **soup** column
- Use sBERT for text encoding of soup column

	user_id	recipe_id	rating	year	month	day	hour	review_nums	category	aver_rate	soup
0	39	61727	4	2001	3	26	7	2443	main-dish	4.439774	Ten Minute Enchilada Sauce vegetable oil self ...
1	39	7612	5	2004	8	2	16	848	dessert	4.683846	Chocolate Lovers' Favorite Cake devil food cak...
2	39	12009	4	2001	2	28	10	4573	main-dish	4.676716	Cajun Chicken Pasta linguine pasta boneless sk...
3	39	88185	5	2009	5	18	14	235	dessert	4.793919	Sopapilla Cheesecake cream cheese white sugar ...
4	39	24445	5	2015	5	7	15	21731	dessert	4.554303	Chewy Chocolate Chip Oatmeal Cookies butter pa...

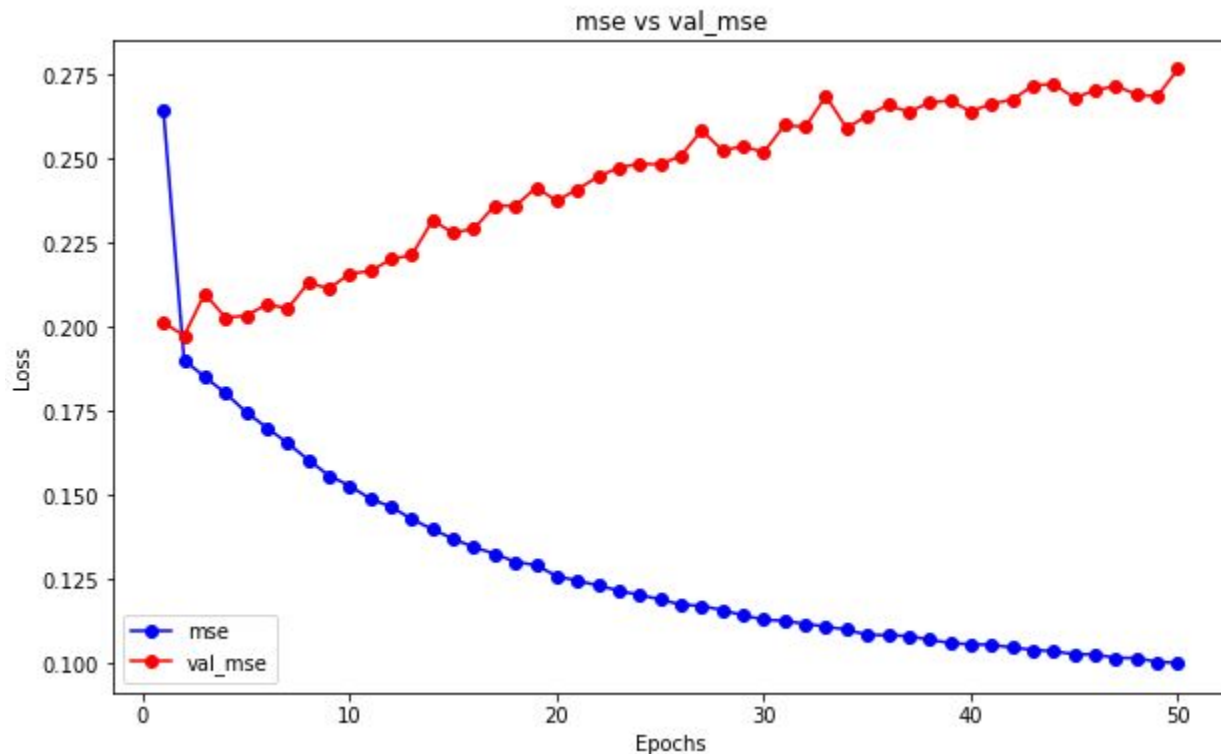
# LANDING PAGE FOOD RECOMMENDATION: TRAINING WITHOUT SOUP



TEST MSE : 0.3246

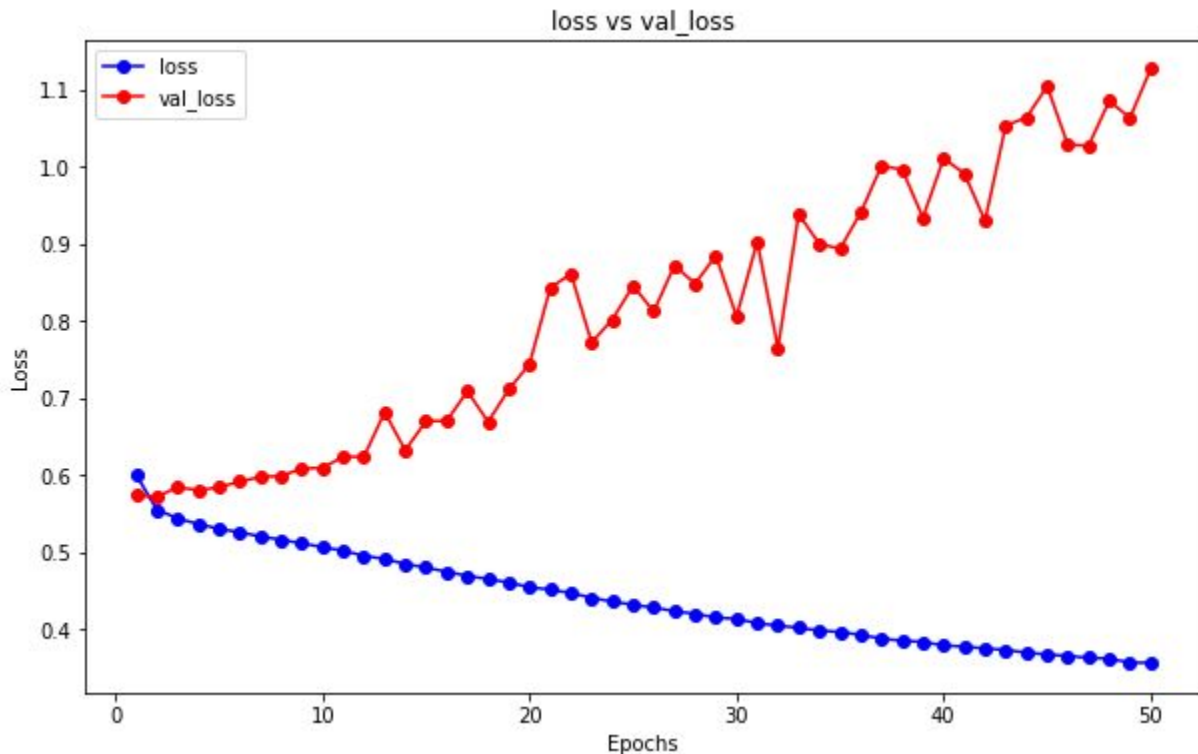


# LANDING PAGE FOOD RECOMMENDATION: TRAINING WITH SOUP



TEST MSE : 0.282

# LANDING PAGE FOOD RECOMMENDATION: TRAINING WITH SOUP AS A BINARY TASK



Loss = Binary  
CrossEntropy

TEST MSE : 0.256

# WEEK 4 : Sequential Recommender System

# WEEK 4 - Sequential Recommender System: Challenges, Progress and Prospects

- <https://nirajanbekoju.notion.site/Sequential-Recommender-System-Challenges-Progress-and-Prospects-11ffcfbce6634de196e68bc3b063f3b4>

**Sequential Recommender System: Challenges, Progress and Prospects**

Introduction

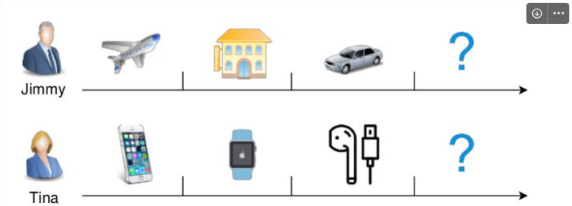


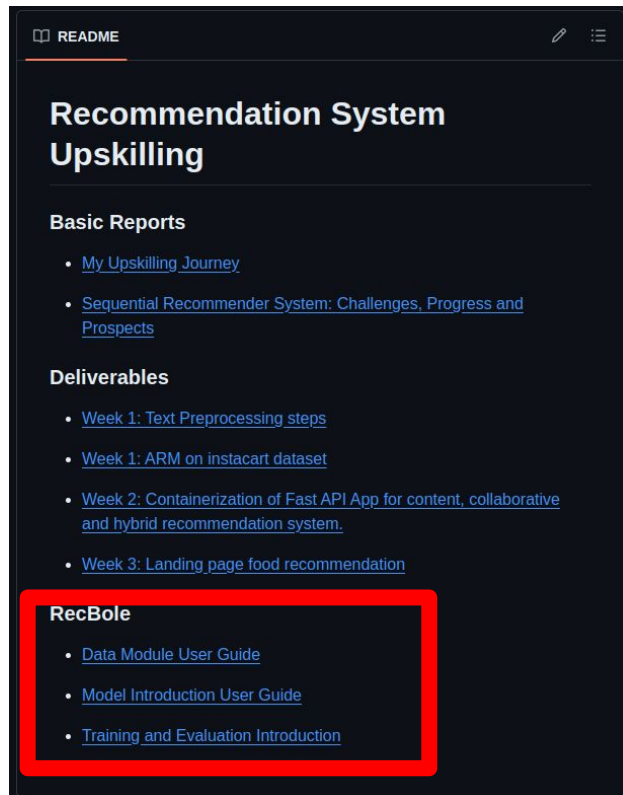
Figure 1: Two examples of SRSs: (1) After Jimmy has booked a flight, a hotel and rented a car, what will be his next action? (2) After Tina has bought an iPhone, an iWatch and a pair of AirPods, what would she buy next?

Sequential Dependency

- The hotel may be close to the destination airport for the flight
- Location for renting car may not be far away from the hotel.

Sequential Recommendation System takes the prior sequential interactions as a context to predict which items would be interacted in the near future.

# WEEK 4 - RecBole Notes



**THANK YOU**