

# **Systeme de recommandation de musiques**

Présenté par Loïc BOUCHERY, Loïc LE CONTEL,  
Nathan RIHET

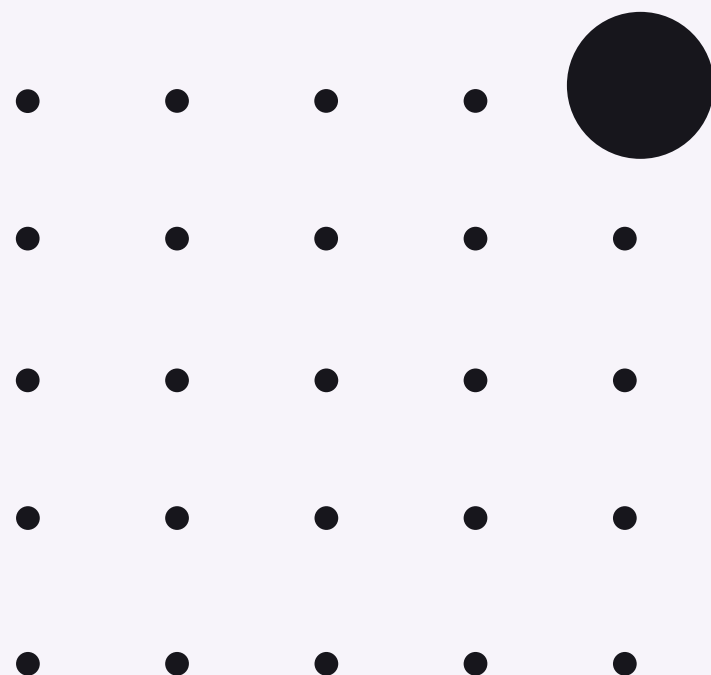
# SOMMAIRE

**I. Dataset et Préparation**

**II. Méthodes de Recommandation**

**III. Performances et Comparaisons**

**IV. Conclusion et Perspectives**



# I. Dataset et Préparation

## Présentation du Million Song Dataset



Description du dataset

Sous ensemble du dataset  
directement fourni sur le site :  
<http://millionsongdataset.com/>

# I. Dataset et Préparation

## Prétraitement des Données

Lecture des fichiers .h5

Stockage des informations  
utiles

Conversion des durées en  
minutes secondes

Suppression des lignes où la  
durée est nulle

Création des données  
utilisateurs-musiques

# II. Méthodes de Recommandation

## Autoencodeur EASE

### 1. Définition Simple :

- Un modèle linéaire sans couches cachées, spécifiquement conçu pour les recommandations basées sur des données sparses.

### 2. Structure :

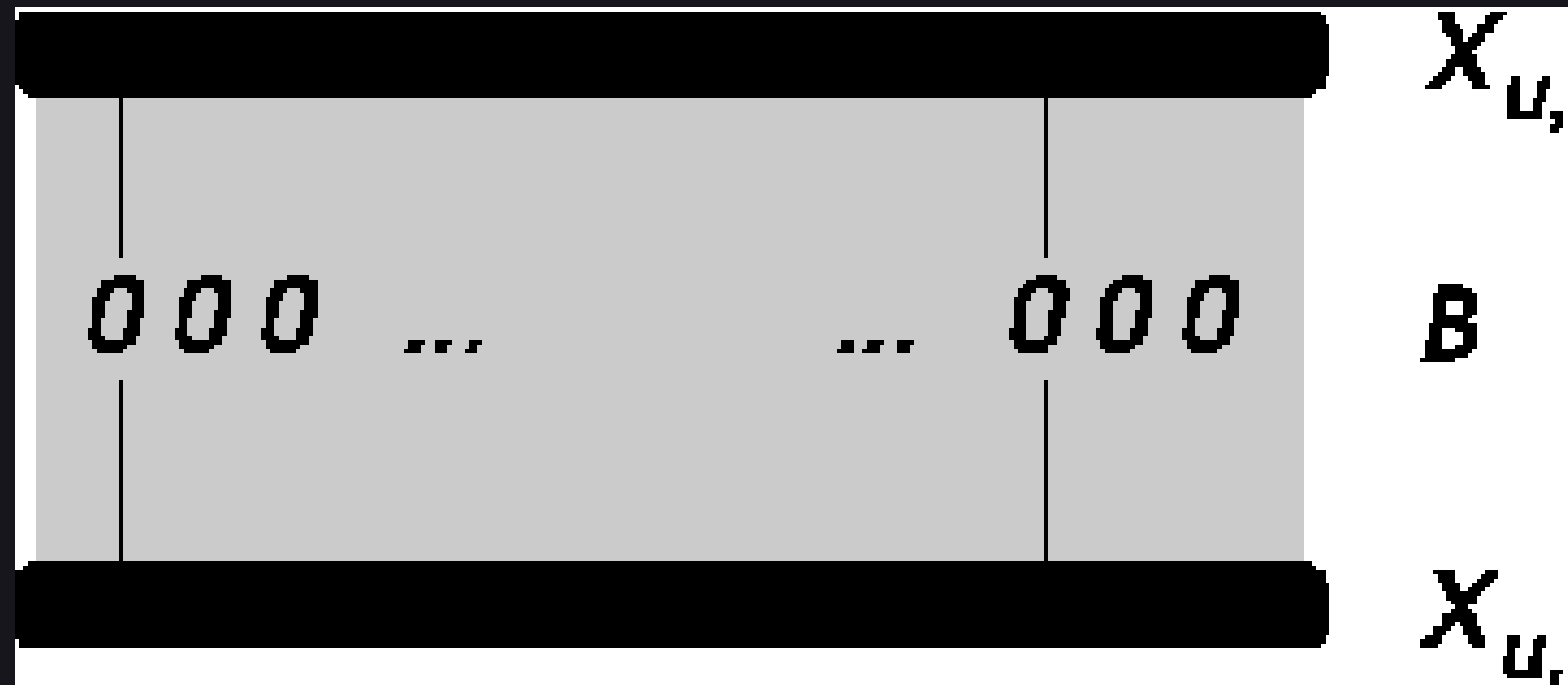
- Une seule matrice de poids  $B$ , sans auto-contribution (diagonale nulle).

# II. Méthodes de Recommandation

Autoencodeur EASE

## 3. Fonctionnement :

prédit directement les préférences des utilisateurs en se basant sur une matrice de poids sans couches cachées, avec une régularisation qui empêche l'auto-influence des items.



# II. Méthodes de Recommandation

Autoencodeur EASE

## 4. Génération des Recommandations :

Classement des items par score pour recommander.

## 5. Avantages :

Rapide, efficace et facile à comprendre.

Performant sur les données sparses (Peu denses).

# II. Méthodes de Recommandation

## Implémentation de EASE

- Transformation des données d'interactions en une matrice utilisateur-item
- Construction de la Matrice de Cooccurrence des Items
- Résolution de la Régression Ridge
- Génération des Recommandations



# II. Méthodes de Recommandation

Métrique d'évaluation des résultats : Normalized Discounted Cumulative Gain

- Mesure la qualité des recommandations en tenant compte de la pertinence des éléments recommandés et de leur position dans la liste.
- Combinaison de différentes informations sur les chansons et attribution d'une valeur de pertinence à chacune.
- Evaluation de la qualité des 100 premières recommandations en considérant à la fois leur pertinence et leur position dans la liste.

# II. Méthodes de Recommandation

## Modèle CDAE

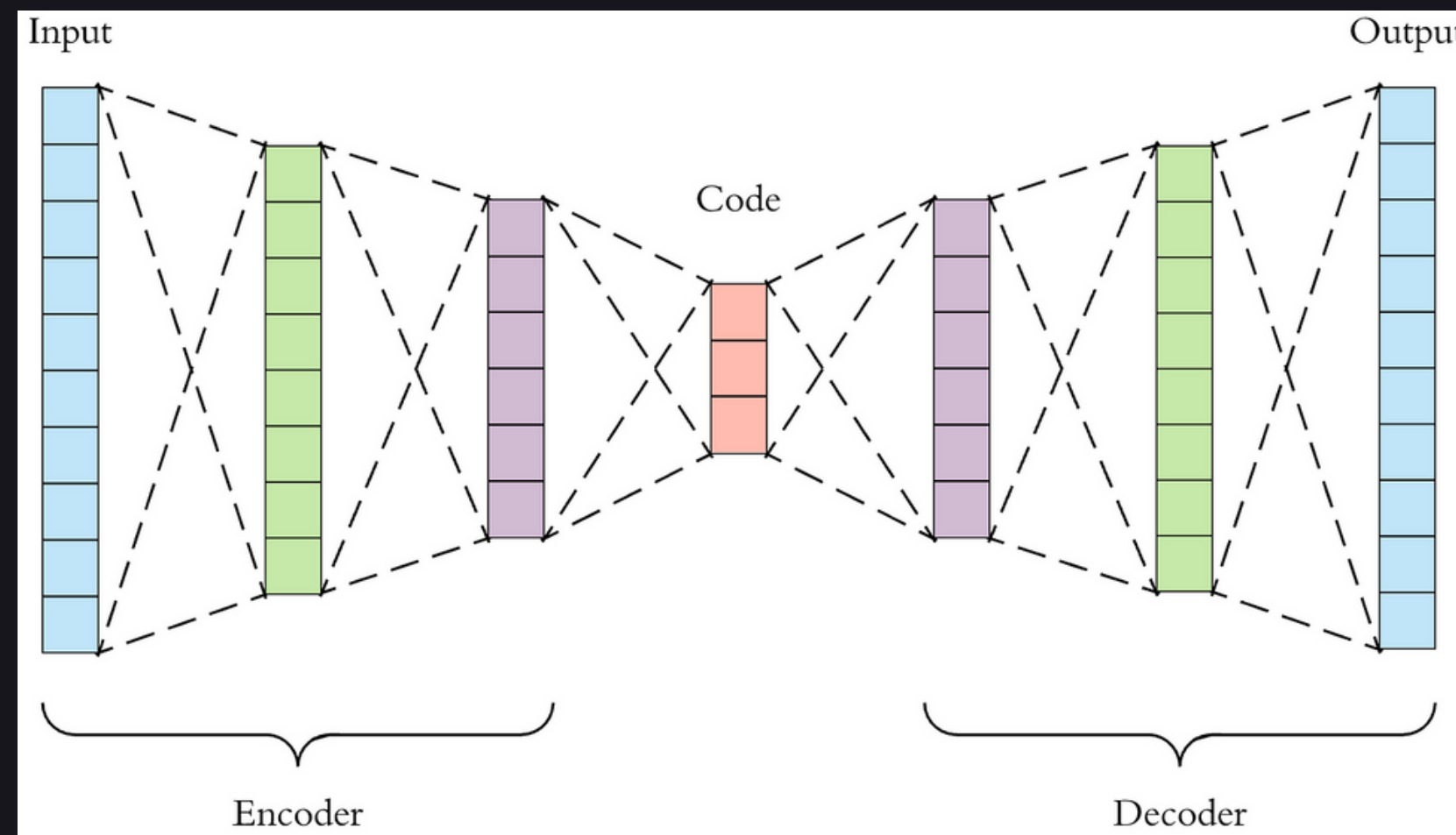
- **Définition Simple :**
  - Modèle non-linéaire conçu pour améliorer la précision des recommandations en utilisant des techniques de "denoising" et d'apprentissage collaboratif.
- **Structure :**
  - Modèle neuronal avec  $X$  couches cachées

# II. Méthodes de Recommandation

## Modèle CDAE

- **Fonctionnement :**

- Utilisation d'un mécanisme de denoising où du bruit est ajouté aux données d'entrée pour renforcer la capacité du modèle à généraliser.



# II. Méthodes de Recommandation

Modèle CDAE : Collaborative Denoising Auto-Encoders

- **Génération des Recommandations :**
  - Un score pour chaque item pour classer et recommander les items les plus pertinents
- **Avantages & inconvénients :**
  - Plus flexible et précis que les modèles linéaires, mais moins adapté aux données sparses

# II. Méthodes de Recommandation

## Modèle MultVAE

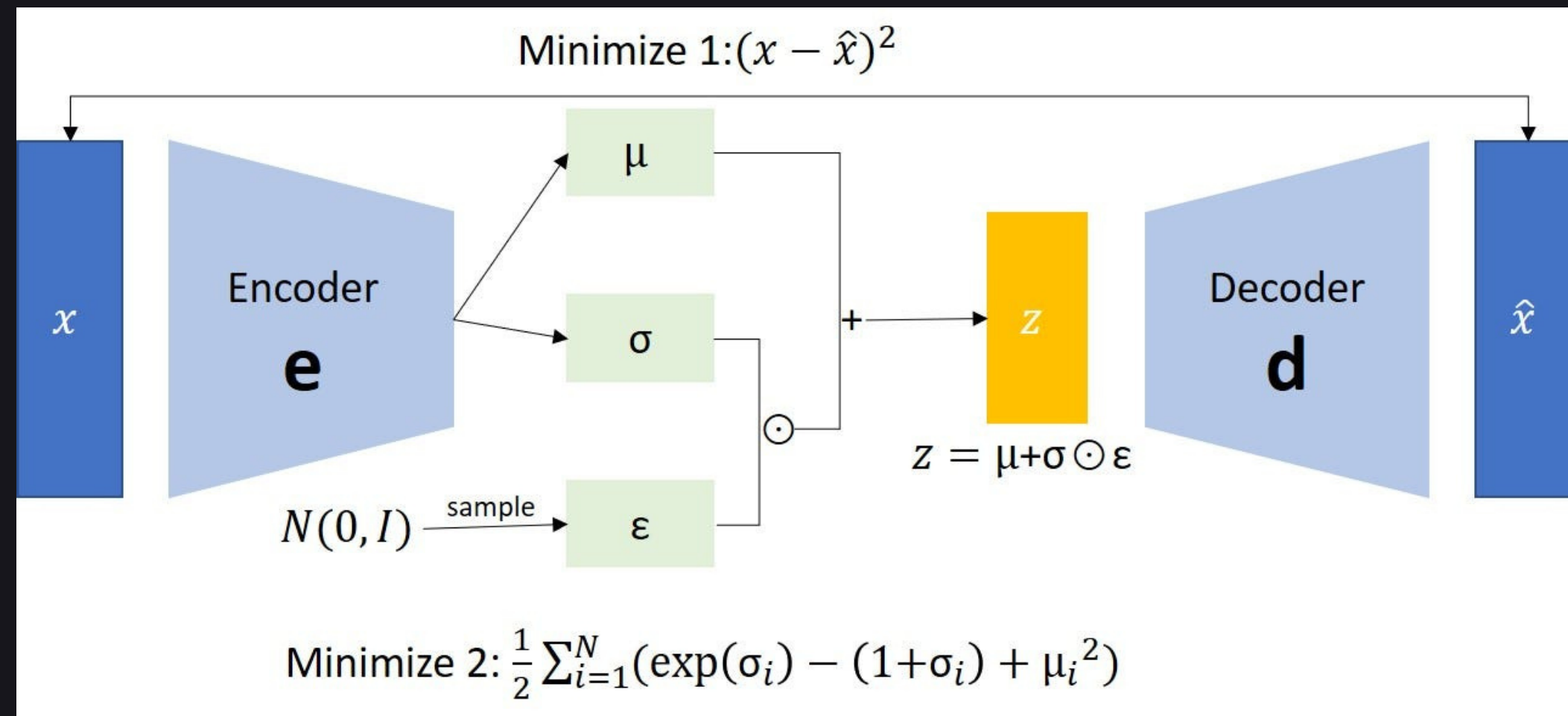
- **Définition Simple :**
  - Modèle probabiliste basé sur un auto-encodeur variationnel pour la modélisation des distributions de données utilisateurs.
- **Structure :**
  - Modèle neuronal avec  $X$  couches cachées

# II. Méthodes de Recommandation

## Modèle MultVAE

### 3. Fonctionnement :

- Mécanisme variationnel créant une distribution de probabilité pour les couches cachées plutôt qu'une représentation déterministe.
- Loss mixte (log-vraisemblance + KL divergence)



# II. Méthodes de Recommandation

Modèle MultVAE : The Multi-Entity Variational Autoencoder

- Génération des Recommandations :
  - Apprend une distribution latente des données d'entrée pour générer des scores de probabilité pour chaque item
- Avantages & inconvénients :
  - Permet une modélisation plus riche des préférences et comportements des utilisateurs grâce à l'approche probabiliste.
  - Approche utilisant des éléments complexes, besoin de plus d'époques et de puissance de calcul

# II. Méthodes de Recommandation

## Méthodes Simples

### SVD (Singular Value Decomposition)

décompose les matrices de données utilisateur-item en composants latents, révélant des schémas cachés pour des prédictions de recommandation efficaces.

### ALS (Alternating Least Squares)

optimise séparément les facteurs utilisateurs et items dans un modèle de factorisation de matrice, traitant efficacement les données manquantes



# III. Performances et Comparaisons

## Évaluation des Modèles

Modèle	Moyenne NDCG	Variance NDCG
SVD	98.709	0.00027
ALS	94.695	0.00271
EASE	95.743	0.00011
CDAE	89.986	0.00028
Mult-VAE	89.776	0.00024

# IV. Conclusion et Perspectives

Apport du projet point de vue personnel

Nous a permis de réaliser bon tour d'horizon des différents modèles et auto encodeur pour des systèmes de recommandation qui nous permettrait de réappliquer ces connaissances dans un autre contexte

# IV. Conclusion et Perspectives

## Limitations et Améliorations Futures

Utilisations de données utilisateurs fictive amène à des résultats légèrement faussé pour les méthodes simple

Nous pourrions envisager de refaire l'analyse avec un autre dataset ou avec le même mais contenant des données utilisateurs réelles