



# Analyse Dimensionnelle du Jeu de Données MNIST par Méthode des Composantes Principales

Rapport de Recherche par :

- RAHOLDINA FIARA Anjara Mihavana,
- RANDRIAMIARAMANANA Narivelo Yvan
- ANDRIAMANAKOTO Anjara Tafita
- RALALASON Rodéo Victorieux
- GESY Louis Xavier

# Introduction et Fondements Théoriques

## Défi des Données

MNIST : 70 000 images, 784 dimensions.

Données de grande dimension.

## Solution PCA

Réduit la dimensionnalité.

Maximise la variance conservée.

Facilite la visualisation.

## Principes de la PCA

Calcul de la matrice de covariance.

$$\Sigma = \frac{1}{n} X^T X$$

Décomposition spectrale.

$$\Sigma = W \Lambda W^T$$

Projection pour composantes.

$$Z = XW$$

# Méthodologie : Prétraitement et Implémentation

## Prétraitement

Normalisation des pixels  $[0, 255]$ .

Transformation pour équité.

## Implémentation

NumPy pour calculs matriciels.

Scikit-learn pour optimisation.

## Analyse de Variance

Détermine composantes optimales.

Calcul des ratios expliqués.



# Analyse de la Variance Expliquée

Le graphique de variance cumulée aide à choisir le nombre de composantes.

Environ 200 composantes expliquent 95% de la variance totale.

Ceci souligne la complexité du jeu de données MNIST.

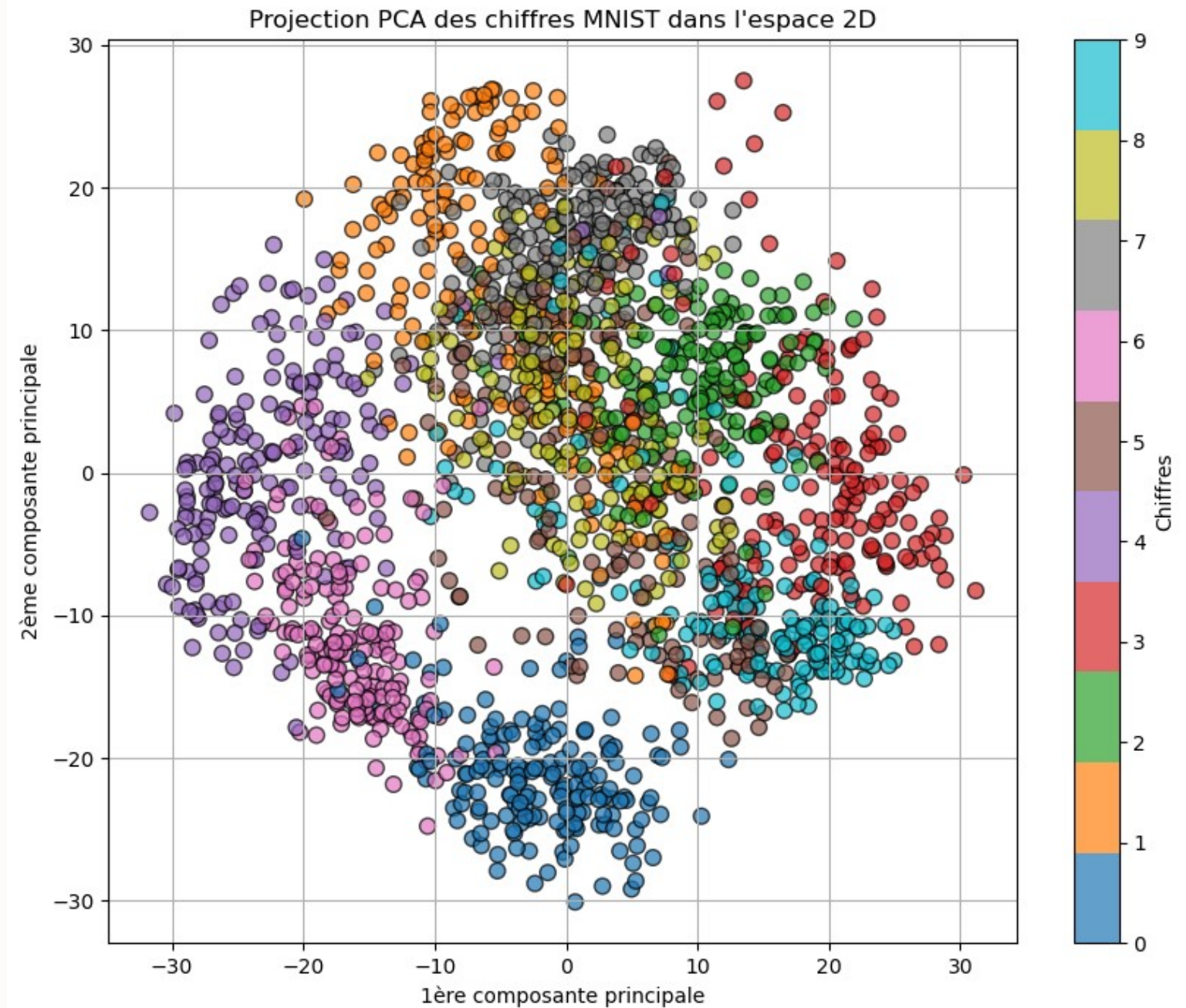
# Résultats : Visualisation des Composantes

## Projection 2D

Séparation partielle des classes.

Chiffres 0 et 1 isolés.

Chevauchement pour 3-5-8.



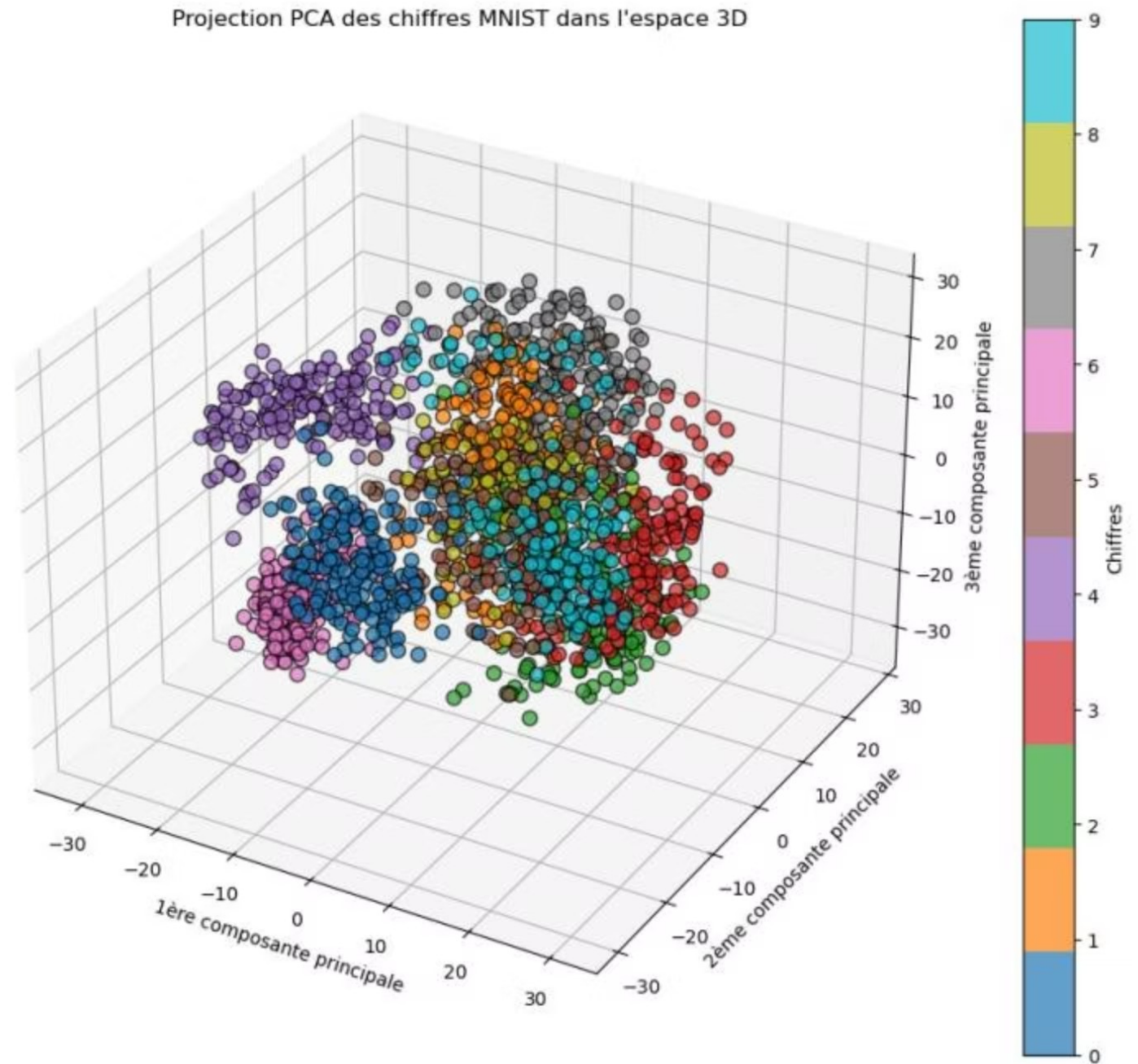


# Résultats : Visualisation des Composantes

## Projection 3D

L'analyse 3D révèle des structures cachées.

La 3ème composante est sensible aux courbures (2 et 6).



# Discussion Critique et Limitations

## Efficacité PCA

Bonne pour visualisation exploratoire.

## Limitations Linéaires

Inadaptée aux relations non linéaires.

Interprétation difficile au-delà de 3 composantes.

## Alternatives

t-SNE ou UMAP pour non-linéarité.

Plus complexes, moins interprétables.

# Conclusion et Perspectives



## Apports et Limites

PCA efficace mais avec limites.



## Pistes de Recherche

Extension à MNIST coloré.

Intégration Deep Learning.



## Évaluation Impact

Combiner PCA et classification.

Évaluer performances prédictives.

Code complet et données disponibles sur GitHub et MNIST.

