**Rapport de Projet : Analyse et Modélisation des Données de Jeux Steam**

**Introduction**

Dans ce projet, nous avons analysé un ensemble de données de jeux Steam en utilisant des techniques d'apprentissage automatique et de deep learning. L'objectif principal était de développer des modèles prédictifs pour estimer la notation d'âge des jeux en fonction de diverses caractéristiques.

**Importation des Bibliothèques**

Les bibliothèques Python suivantes ont été importées pour l'analyse et la modélisation des données :

* pandas pour la manipulation des données
* sklearn pour les modèles de régression linéaire et les métriques d'évaluation
* tensorflow.keras pour la construction et l'entraînement de modèles de réseaux de neurones
* pickle pour la sauvegarde des modèles entraînés

**Exploration et Prétraitement des Données**

Les données ont été chargées à partir du fichier CSV et explorées pour comprendre la structure et les types de données. Des prétraitements ont été effectués, notamment :

* Conversion des prix originaux en format numérique
* Création de variables indicatives pour les genres de jeux
* Détection de thèmes spécifiques à l'aide de mots-clés comme violence, sexe, drogues, etc.

**Modélisation**

**1. Régression Linéaire**

Un modèle de régression linéaire a été ajusté pour prédire la notation d'âge des jeux en fonction des caractéristiques extraites. Les performances du modèle sur l'ensemble de test ont été évaluées avec les métriques suivantes :

* Mean Squared Error (MSE) : 0.0123
* R-squared (R²) : 0.78

Le modèle linéaire a été choisi pour sa simplicité et son interprétabilité par rapport aux modèles de réseaux de neurones.

**2. Modèle Séquentiel (Réseau de Neurones)**

Un modèle séquentiel à deux couches a été construit avec Keras/TensorFlow, utilisant une couche dense avec activation ReLU et une sortie linéaire. Le modèle a été entraîné avec les données d'entraînement et évalué sur l'ensemble de test avec les résultats suivants :

* Mean Squared Error (MSE) : 0.0122

Bien que légèrement meilleur en termes de MSE, ce modèle est plus complexe et considéré comme une boîte noire par rapport à la régression linéaire.

**Comparaison des Modèles**

En termes de performances sur l'ensemble de test, les deux modèles ont montré des résultats similaires en MSE. Cependant, la régression linéaire a été choisie pour sa simplicité et sa facilité d'interprétation.

**Conclusion**

Ce projet a démontré l'application de différentes techniques d'analyse et de modélisation sur un jeu de données de jeux Steam. La régression linéaire a été retenue comme le modèle final en raison de sa capacité à fournir des insights facilement interprétables tout en maintenant des performances compétitives. Ce rapport illustre l'importance de choisir le bon modèle en fonction des objectifs spécifiques du projet.