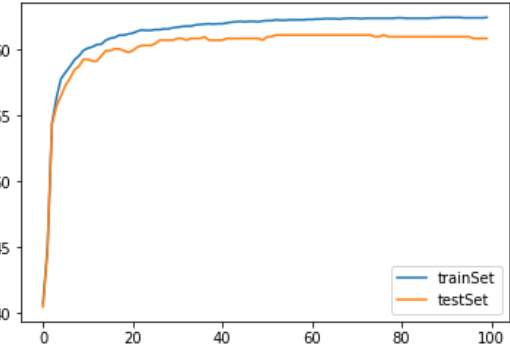


**Problématique 01: La prédiction de note (regression)**

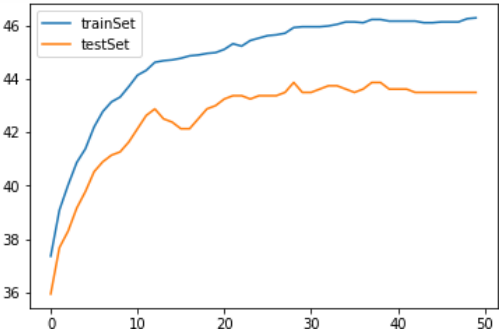
**Déscriptions:** Prédiction de la note moyenne d'un film à travers deux features (budget & gross).

**Méthodes appliquées:** modèle de perceptron gradient batch

**2 Features:**



**Résultat:**  
taux d'accuracy 60%, Nous pouvons justifier cela en se disant que la feature 'budget' a des fois un impact négatif dans la prédiction de la note  
**En utilisant 3 features**

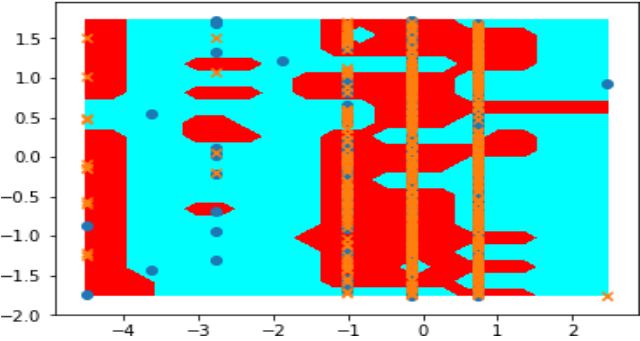


**Resultat :**  
des résultats moins satisfaisants que celle avec 2 features, l'ajoute de l'attribut 'star' n'a pas apporté une meilleure prédiction au contraire, des fois notre modèle n'atteint pas les 35%

**Problématique 02: La prédiction de la catégorie d'un film(classification)**

**1-Drama vs Non Drama:**  
**Description:** prédire si notre film appartient à une catégorie ou pas en utilisant deux features qui sont la 'star' et la restriction d'âge du film  
**Méthodes:** K-nn, Arbre de décision

**Taux de précision :**  
**K-NN:** 60%  
**Arbre de décision:** 60.32%



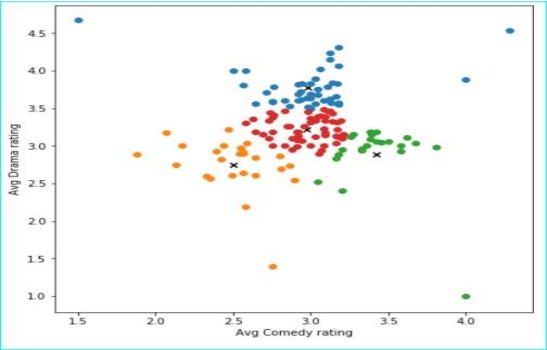
**Résultat:** Les résultats prédis par nos algorithmes produisent tous à peu près un même taux d'accuracy et qui est de 60%

**Problématique 03: catégorisation des groupe d'utilisateur**

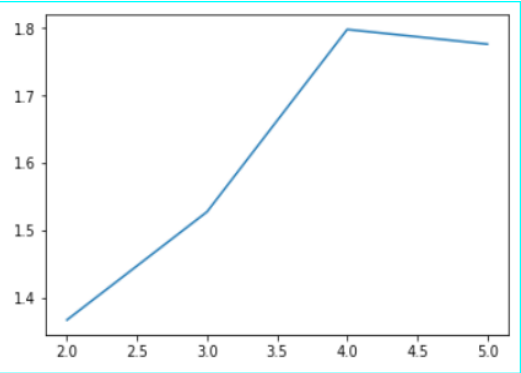
**Déscription:** Déterminer les préférences des internautes selon les notes données aux films

**Méthode appliqué:** modèle non supervisé, algorithme de k-moyennes

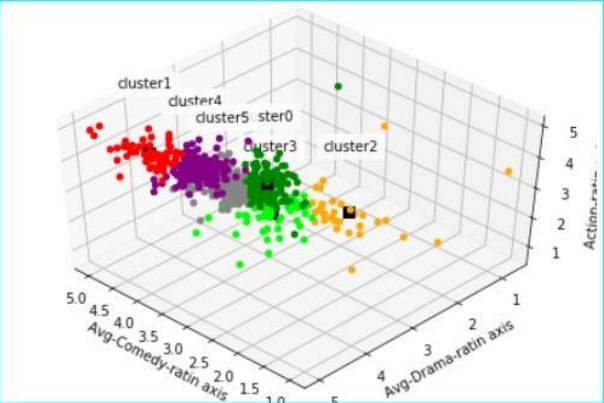
**Resultat-1: (2D)**



**Evaluation: indice de Dunn**  
on remarque que k=4 avec un indice de Dunn maximal(objectif) est le nombre de cluster optimal, ainsi k>4 nous fournit aucune connaissance de plus

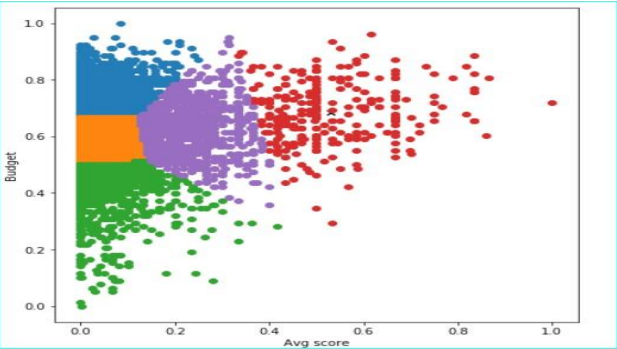


**Resultat-2:(3D)**



**Problématique 04: degré de réussite des films**

**Déscription:** catégoriser des groupes des qui ont fait une réussite selon le budget dépensé et les scores obtenue



**Analyse: On peut distinguer 5 groupe:**  
**Groupe vert:** des films avec un faible score et un faible budget donc on ne peut pas dire que c'est une perte pour l'équipe de production  
**Groupe orange:** un score faible avec un budget un peu plus supérieur à la moyenne  
**Groupe bleu :** un score faible et un budget très élevé on peut classifié ces films dans la catégorie des films qui ont échoué.  
**Groupe violet:** un score qui se rapproche de la moyenne avec un budget moyen  
**Groupe rouge:** un score fort avec un budget moyen, On peut les classifier dans la catégorie des films réussis