MESANO BIN KIMBESA

TITANIC

Dans le cours de machine Learning il nous a été demandé de traiter le DATASET qui nous a été donné par l’enseignant et pour moi je pris le DATASET appeler TITANIC qui se trouver sur https://www.kaggle.com/competitions/titanic/data et le DATASET a question contient les informations nécessaire sur le bateau de Titanic enfin de prédire la possibilité de survivre ou de mourir dans un naufrage quelconque. Et on nous demander d’utiliser deux algorithme de classement en machine Learning c’est-à-dire KNN et décision tree. Et enfin nous allons comparer entre l’aglo KNN et décision tree laquelle est le meilleur algorithme de classement pour le genre de DATASET.

KEY WORDS

DATASET=c’est une base de données en IA

KNN=c’est l’algorithme de classement

DECISIONTREE=c’est aussi en algorithme de classement

APPRENTISSAGE supervise=c’est l’apprentissage qui consister en apprendre le modele en étiquetant

INTRODUCTION

Objectif de la recherche

Nous avons décidé de faire cette recherche enfin de s’avoir quel est l’algorithme la plus meilleurs entre KNN et DECISION TREE pour résoudre le problème de classement pour prédire la possibilité de survivre dans un naufrage grâce a un Dataset qui enregistre les information de tout le passager de TITANIC qui était mort et survécu. Et notre recherche a pour objectif de donner lumière au futur chercheur sur les deux algo KNN et DECISION TREE.

Grandes parties

Les parties essentielles pour mon ce travail sont :

LA METHODOLOGIE : ici nous allons montrer nous avons utilisé pour résoudre le problème

L’ANALYSE DE LA PERFOMANCE DE DEUX ALGO : ici nous allons passer a la comparaison de deux algorithme qu’on nous a présenter

LA CONCLUSION : ici nous allons donner notre point de vue ou choix parmi ces deux algorithmes.

METHODOLOGIE

* DATASET

Nous avons prix notre DATASET sur <https://www.kaggle.com/competitions/titanic/data?select=train.csv> et nous avons et on a téléchargé ce DATASET possédant 891 lignes et 7 colonnes après et le DATASET a été filtre par nous même pour supprimer les informations indésirable et puis nous allons appliquer notre algorithme sur le donnée bien nettoyer pour comparer leurs performances.

* DATA PREPOCESSING

Quant a ce qui concerne le procédure que nous avons utilisé est que nous avons eu le plaisir d’avoir un DATASET de donner déjà collecte et donc notre tâche était de voir comment nettoyer le dataset.

* ENTAINEMENT

Pour entrainer notre modele nous alons utiliser :

1. Knn

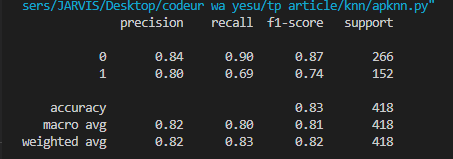
La méthode des K plus proches voisins (KNN) a pour but de classifier des points cibles (classe méconnue) en fonction de leurs distances par rapport à des points constituant un échantillon d'apprentissage (c'est-à-dire dont la classe est connue a priori).

1. Desicion Tree

Les algorithmes de Decision tree, que l’on appelle également arbres de décision font partie de la catégorie des algorithmes supervisés, ils permettent de prédire une valeur (prédiction) ou une catégorie (classement).C’est une méthode très populaire en Data Science et qui a donné naissance à d’autres algorithmes plus puissants tels que Random Forest ou XGBoost par exemple. Comme son nom l’indique, cet algorithme se base sur la construction d’un arbre ce qui rend la méthode assez simple à expliquer et plus facile à interpréter.

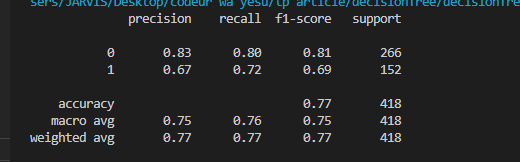
PERFOMANCE ANALYSIS

1. KNN



Le tableau ci haut explique en détaille la précision, le recall, f1-score ainsi que le support pour l’algorithme knn.

1. DECISION TREE



Voici ci-haut la precision, le recall, f1-score, accuracy et support pour decision tree

1. COMPARAISON

Pour une étude comparative nous voyons que l’algorithme de KNN est plus performant que DECISION TREE pour bien faire la prédiction avec TITANIC.csv comme dataset car la performance de KNN leurs résultats tend vers 1 ou 100% que DECISION TREE. Et donc nous déclarons KNN étant comme la meilleure.

CONCLUSION

Pour clore notre travaille nous remercions l’enseignant qui nous a donné ce travail qui nous permis de découvrir en détail tout ce qui concerne KNN et DECISION TREE et donc nous avons conclu que nous trouvons KNN meilleure que DECISION TREE pour le problème que nous avons eu a traiter et cela dépend du problème.

REFERENCE

Voici le lien github ou se trouve tout nos code python et dataset que nous avons traité :

<https://github.com/CODEUR-WA-YESU/tp1>