

Conseils pour le TP Mushrooms_RF



A n'utiliser que si vous êtes bloqué.

Vous apprendrez mieux si vous n'utilisez pas ces conseils, mais cherchez par vous même.

1.1

Utiliser la méthode pandas `read_csv`.

Utiliser la méthode pandas `display` pour afficher le dataframe.

Utiliser la méthode `value_count()` pour compter les champignons toxiques.

1.2

Regarder le paramètre `sparse_output` et la méthode `get_feature_names_out()` du modèle `OneHotEncoder`

[https://scikit-](https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html#sklearn.preprocessing.OneHotEncoder)

[learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html#sklearn.preprocessing.OneHotEncoder](https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html#sklearn.preprocessing.OneHotEncoder)

1.3

Le dataframe peut être créé en utilisant le constructeur `pd.DataFrame()` et les bons arguments

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html>

Appliquer `sample(frac=1)` pour mélanger le dataframe

1.4

Fonction scikit-learn `train_test_split`.

Bien faire attention à ne passe en paramètre que les colonnes pertinentes.

Utiliser un `display` dans un boucle pour afficher les 4 ensembles de données

2.1

Voir exemple avec `DecisionTreeClassifier`.

2.2

Voir exemple. Générer un fichier graphviz dot, puis le convertir en png et l'afficher.

Pour améliorer la lisibilité, renseigner les paramètres `feature_names` et `class_names` dans la fonction `export_graphviz`.

2.3

Utiliser `predict` sur les données dev, et la fonction `accuracy_score`.

3.1

Voir l'exemple avec RandomForestClassifier.

3.2

Tirer 3 nombres au hasard entre 0 et le nombre d'arbres – 1 (fonction python randInt).
Utiliser Graphviz dans une boucle.

3.3

Idem 2.3

4.1

Pour trouver les meilleurs paramètres, vous pouvez utiliser GridSearchCV.

https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

4.2

Voir documentation GridSearch.

La précision est comme pour les modèle précédents.

4.3

Utiliser la propriété `feature_importances_` du meilleur estimateur.

5.1

Voir exemple.