

Explicación del Código - Clasificación con Árboles y Random Forest

Explicación General del Código

1. Carga del Dataset Iris:

Se importa el dataset Iris desde sklearn. Contiene 150 flores clasificadas en tres especies (setosa, versicolor, virginica), con 4 características: sepal length, sepal width, petal length y petal width.

2. Creación del DataFrame:

Se crea un DataFrame con los datos y los nombres de columnas, y se agrega la columna 'target', que indica la clase de cada flor (0, 1, 2).

3. División de los datos:

Se separan las variables predictoras (X) del target (y), y se divide en conjunto de entrenamiento y prueba.

4. Árbol de Decisión:

Se entrena un modelo DecisionTreeClassifier con ccp_alpha=0.01 (para poda). Luego se hace el entrenamiento con fit.

5. Predicción y Accuracy:

Se predice con el conjunto de prueba y se calcula la precisión con accuracy_score. El modelo alcanza una precisión de 96.6%.

6. Evaluación del modelo:

Se usa confusion_matrix y classification_report para evaluar la calidad del modelo por clase (precision, recall, f1-score).

7. Importancia de características:

Se muestran los features más importantes para la predicción. Petal width y petal length son los más relevantes.

8. Visualización del Árbol:

Se usa plot_tree para mostrar gráficamente el árbol entrenado y cómo clasifica las flores.

9. Random Forest:

Explicación del Código - Clasificación con Árboles y Random Forest

Se entrena un modelo RandomForestClassifier (con múltiples árboles). Se repiten las predicciones y evaluación, logrando también una precisión del 96.6%, más estable que el árbol único.