## **Relatório sobre AutoML com Auto-Sklearn**

### **Introdução**

O AutoML (Automated Machine Learning) é uma abordagem que automatiza o processo de seleção de modelos, otimização de hiperparâmetros, pré-processamento de dados e engenharia de características. O Auto-Sklearn é uma ferramenta de AutoML baseada na biblioteca scikit-learn, projetada para facilitar o uso de machine learning, mesmo para aqueles que não têm experiência profunda na área.

### **Funcionamento do Auto-Sklearn**

O Auto-Sklearn realiza uma série de etapas para encontrar o melhor modelo e as melhores configurações para o problema em questão. Essas etapas incluem:

1. **Pré-processamento de dados**: Inclui técnicas como normalização, padronização, imputation (preenchimento de valores ausentes) e codificação de variáveis categóricas.
2. **Seleção de modelo**: Testa uma variedade de algoritmos de machine learning (como árvores de decisão, florestas aleatórias, SVMs, redes neurais, etc.) para encontrar o mais adequado.
3. **Otimização de hiperparâmetros**: Ajusta automaticamente os hiperparâmetros dos modelos para melhorar o desempenho.
4. **Engenharia de características**: Gera novas características ou transforma as existentes para melhorar a qualidade do modelo.
5. **Validação cruzada**: Avalia o desempenho dos modelos usando técnicas de validação cruzada para garantir a robustez dos resultados.

### **Processo de Funcionamento**

O Auto-Sklearn segue os seguintes passos para realizar suas tarefas:

1. **Exploração do espaço de busca**:
   * **Espaço de modelos**: Inclui diversos algoritmos de machine learning disponíveis no scikit-learn.
   * **Espaço de pré-processamento**: Diversas técnicas para preparar os dados, como escalonamento e transformação.
   * **Espaço de hiperparâmetros**: Conjunto de parâmetros ajustáveis para cada modelo.
2. **Otimização Bayesiana**: Utiliza otimização Bayesiana para explorar o espaço de busca eficientemente. Isso permite a seleção dos melhores modelos e hiperparâmetros com menos avaliações em comparação a métodos de busca exaustiva.
3. **Meta-learning**: O Auto-Sklearn usa meta-learning para iniciar a busca a partir de configurações que foram eficazes em problemas semelhantes no passado. Isso acelera o processo de encontrar boas soluções.
4. **Construção de ensemble**: Após identificar os melhores modelos, o Auto-Sklearn constrói um ensemble (comitê de modelos) para combinar os melhores modelos e melhorar o desempenho geral.

### **Etapas de Pré-Processamento**

O Auto-Sklearn fornece várias técnicas de pré-processamento automaticamente, incluindo:

* **Imputação de valores ausentes**: Substitui valores ausentes por médias, medianas ou métodos mais avançados.
* **Codificação de variáveis categóricas**: Converte variáveis categóricas em forma numérica (one-hot encoding, label encoding).
* **Escalonamento de características**: Normaliza ou padroniza os dados para melhorar o desempenho dos modelos que são sensíveis à escala das características.

**Aplicando o Auto-sklearn**

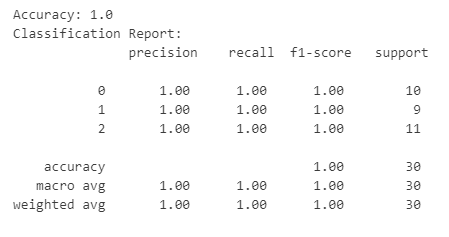
Para aplicar o Auto-sklearn, seguimos os seguintes passos:

1. **Importação das bibliotecas:** Importamos as bibliotecas necessárias, incluindo autosklearn.classification, sklearn.datasets, sklearn.model\_selection, sklearn.metrics, matplotlib.pyplot e seaborn.
2. **Carregamento e divisão do dataset:** Carregamos o dataset desejado (no exemplo, o dataset Iris) e o dividimos em conjuntos de treino e teste.
3. **Criação do classificador Auto-Sklearn:** Criamos uma instância da classe AutoSklearnClassifier, definindo o tempo máximo de execução (time\_left\_for\_this\_task) e o tempo limite para cada execução de modelo (per\_run\_time\_limit).
4. **Treinamento do classificador:** Treinamos o classificador com os dados de treino. O Auto-sklearn automaticamente realiza todas as etapas de pré-processamento, seleção de modelo, otimização de hiperparâmetros e validação cruzada.
5. **Previsões no conjunto de teste:** Fazemos previsões no conjunto de teste usando o classificador treinado.

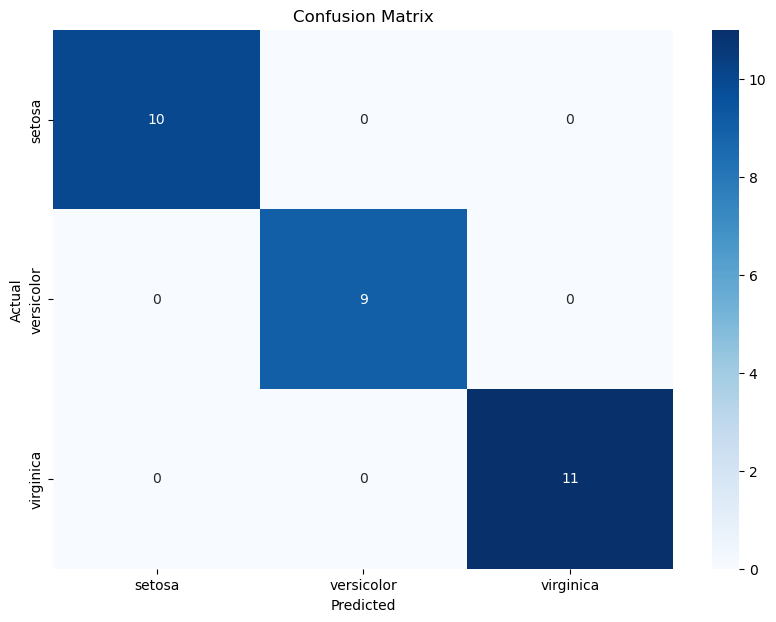
**Resultados Obtidos**

Após o treinamento do classificador Auto-sklearn, obtivemos os seguintes resultados:

* **Precisão (Accuracy):** 1.0
* **Relatório de Classificação:** O relatório mostra que a precisão, revocação e F1-score são perfeitos (1.0) para todas as classes do dataset Iris.



* **Matriz de Confusão:** A matriz de confusão mostra que o modelo classificou corretamente todos os exemplos do dataset de teste.

**A****nalisando o Melhor Modelo:**

Para identificar o modelo final escolhido pelo Auto-sklearn, utilizamos o atributo cv\_results\_. A tabela de classificação do Auto-sklearn (acessível com o comando cls.leaderboard()) mostra que o modelo de **floresta aleatória (random\_forest)** com uma pontuação média de 0.325 foi o modelo que obteve o melhor desempenho durante a validação cruzada.

**Conclusão**

Os resultados obtidos com o Auto-sklearn demonstraram que a ferramenta é capaz de construir um modelo de classificação de alto desempenho com pouca intervenção manual. A precisão perfeita (1.0) e o desempenho consistente em todas as classes do dataset Iris indicam que o Auto-sklearn conseguiu encontrar a melhor combinação de pré-processamento, modelo e hiperparâmetros para este problema.

O Auto-sklearn se apresenta como uma ferramenta valiosa para simplificar o processo de aprendizado de máquina e possibilitar a criação de modelos eficazes, mesmo para usuários com pouca experiência na área