

Anno Accademico 2024/2025

Progetto di Machine Learning

## **Machine Learning per l'impronta di Carbonio nei Dispositivi Elettronici**

*Membri:*

*Sinicario Gennaro - 0512116134*

*Pisano Antonio - 0512118258*

### **Identificazione del Problema**

L'impatto ambientale dei dispositivi elettronici è una questione sempre più rilevante. La possibilità di stimare l'impronta di carbonio in base a caratteristiche come peso, metodo di riciclo e componenti tossici può aiutare aziende e consumatori a fare scelte più sostenibili. L'obiettivo del progetto è sviluppare un modello di Machine Learning per prevedere l'impronta di carbonio di un dispositivo elettronico, basandosi su un dataset opportunamente preprocessato.

### **Importanza del Problema**

- Aiuta i centri di riciclo a ottimizzare le operazioni.
- Supporta le aziende nel migliorare la progettazione sostenibile.
- Permette ai consumatori di valutare l'impatto ambientale dei dispositivi.

## Descrizione del Dataset

Abbiamo utilizzato il dataset “*E-Waste Data*” disponibile su Kaggle al seguente link:

<https://www.kaggle.com/datasets/arifmia/e-waste-data>

Il dataset contiene informazioni sui dispositivi elettronici, tra cui:

- Marca e modello
- Anno di acquisto
- Categoria (TV, Mobile, PC, ecc.)
- Peso
- Metodo di riciclo
- Componenti tossici (Piombo, Mercurio, Cadmio, ecc.)
- Impronta di carbonio (**variabile target**)

I dati sono stati puliti ed elaborati per eliminare valori mancanti e convertire variabili categoriche in numeriche.

## Preprocessing dei Dati

- **Gestione valori mancanti:** Rimozione delle righe con target mancante, sostituzione dei valori mancanti con la media.
- **Suddivisione del dataset:** 80% training, 20% test.

- **Encoding delle variabili categoriche:** Trasformazione delle marche e dei metodi di riciclo in valori numerici con il one-hot encoding.

	<b>Brand_LG</b>	<b>Brand_Panasonic</b>	<b>Brand_Sony</b>	<b>Brand_Xiaomi</b>
1	0	1	0	0
2	0	0	0	1
3	1	0	0	0
4	0	0	1	0

## **Modelli Utilizzati**

Abbiamo testato due algoritmi per la regressione dell'impronta di carbonio:

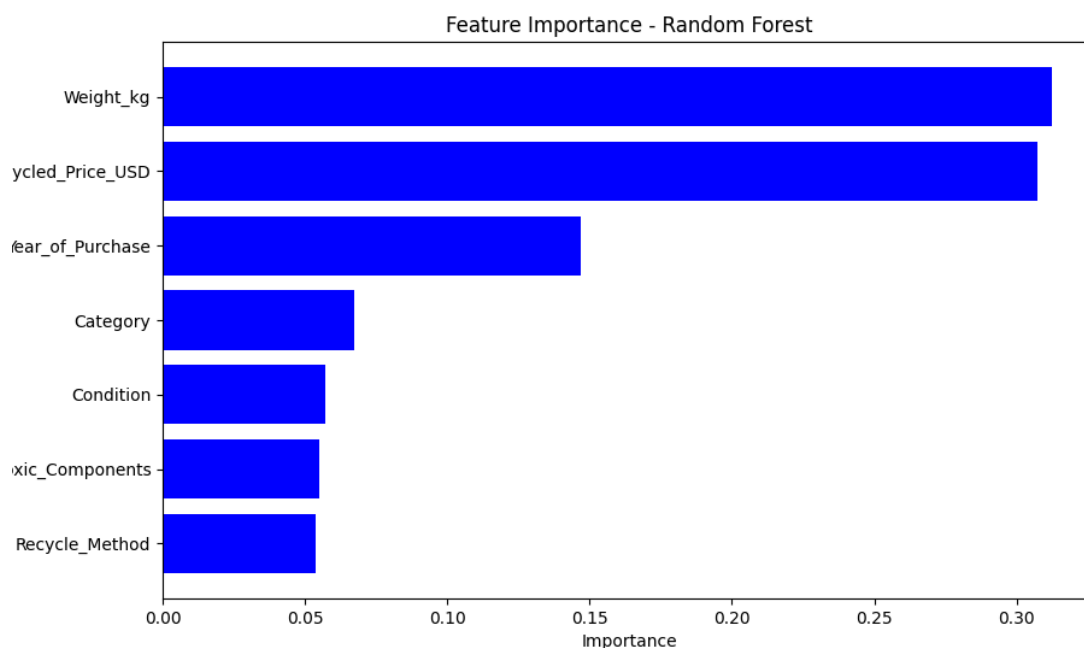
- **Decision Tree**
- **Random Forest**

Per ogni modello sono state calcolate le seguenti metriche di valutazione:

- **MAE** (Mean Absolute Error)
- **MSE** (Mean Squared Error)
- **R<sup>2</sup> Score**

## Random Forest

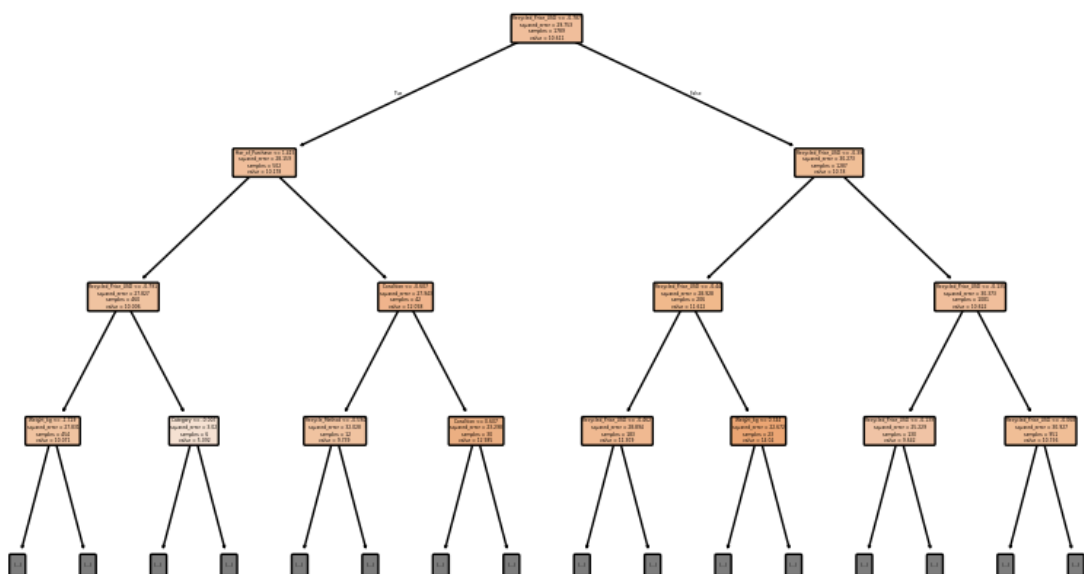
Il Random Forest è un algoritmo di apprendimento automatico ampiamente utilizzato, che combina i risultati di più alberi di decisione per produrre una previsione finale. Si basa su un'estensione della tecnica del bagging, integrando anche la selezione casuale delle caratteristiche per generare una foresta di alberi di decisione tra loro indipendenti. Questo metodo è applicabile sia a problemi di classificazione che di regressione. La sua robustezza deriva dalla capacità di aggregare più alberi, migliorando l'accuratezza complessiva del modello. Con queste caratteristiche, il modello raggiunge le seguenti prestazioni:



# Decision Tree

Un albero di decisione (Decision Tree) è un algoritmo di apprendimento supervisionato, impiegato sia per la classificazione che per la regressione. La sua struttura ad albero è composta da un nodo radice, rami, nodi interni e nodi foglia. Nei nodi interni vengono effettuati test sulle caratteristiche dei dati, mentre i nodi foglia rappresentano le decisioni finali basate sull'analisi degli attributi. L'obiettivo dell'algoritmo è individuare la suddivisione ottimale dei dati per migliorare la previsione. Tuttavia, può soffrire di overfitting, ovvero adattarsi eccessivamente ai dati di addestramento. Il modello, con queste caratteristiche, raggiunge le seguenti prestazioni:

### Decision Tree Visualization



## Risultati e Analisi

Modello	MAE	MSE	R <sup>2</sup> Score	Tempo di addestramento
Decision Tree	6.75	68.28	-1.29	0.03s
Random Forest	4.83	32.27	-0.08	1.62s

Il modello **Random Forest** ha ottenuto le **migliori prestazioni**, con un errore più basso e una maggiore capacità di generalizzazione.

## Conclusioni

I risultati dimostrano come il Machine Learning possa essere applicato con successo per stimare l'impatto ambientale dei dispositivi elettronici.

Miglioramenti futuri potrebbero includere:

- L'uso di modelli più avanzati come Gradient Boosting.
- L'ottimizzazione degli iperparametri per migliorare l'accuratezza.
- L'integrazione di dati aggiuntivi, come il consumo energetico durante l'uso.