

# Raport Aplicație practică 2

## Analiza Setului de Date

Setul de date conține următoarele atribute: **Country**, **Category**, **Visitors**, **Revenue**, **Rating**, **Accommodation Available**. Am realizat o etapă de preprocesare pentru a normaliza datele și a aplica *Label Encoding* pe atributele categorice. În plus, am adăugat o coloană nouă, **Revenue per Visitor (RpV)**, pentru a evalua mai precis venitul generat per vizitator.

## Algoritm Selectat și Justificare

Pentru această problemă, cele mai bune rezultate le are **Random Forest Regressor** datorită capacității sale de a modela relații non-liniare complexe și de a gestiona variabile de intrare eterogene. Algoritmul a fost comparat cu alți candidați, precum *XGBoost*, *AdaBoost*, *Linear Regression* și *ID3*, utilizând diferite rapoarte din împărțire între datele de antrenare și testare. Dintre algoritmi învățați la curs însă, cel mai bun este **ID3**-ul, cu aproape aceeași acuratețe ca Random Forest, iar cel mai bun raport de antrenare-testare este de 75-25 (justificare în figura 1)

## Rezultate Experimentale

Am evaluat performanța algoritmului utilizând *Root Mean Square Error (RMSE)* și am generat următoarele grafice pentru a ilustra:

1. Performanța algoritmilor comparați pe setul de date de testare.
2. Ierarhia categoriilor pentru fiecare țară, bazată pe **Predicted Revenue** și **Predicted RpV**.

## Performanța Algoritmilor

Loc pentru grafic: **Compararea RMSE pentru algoritmi.**

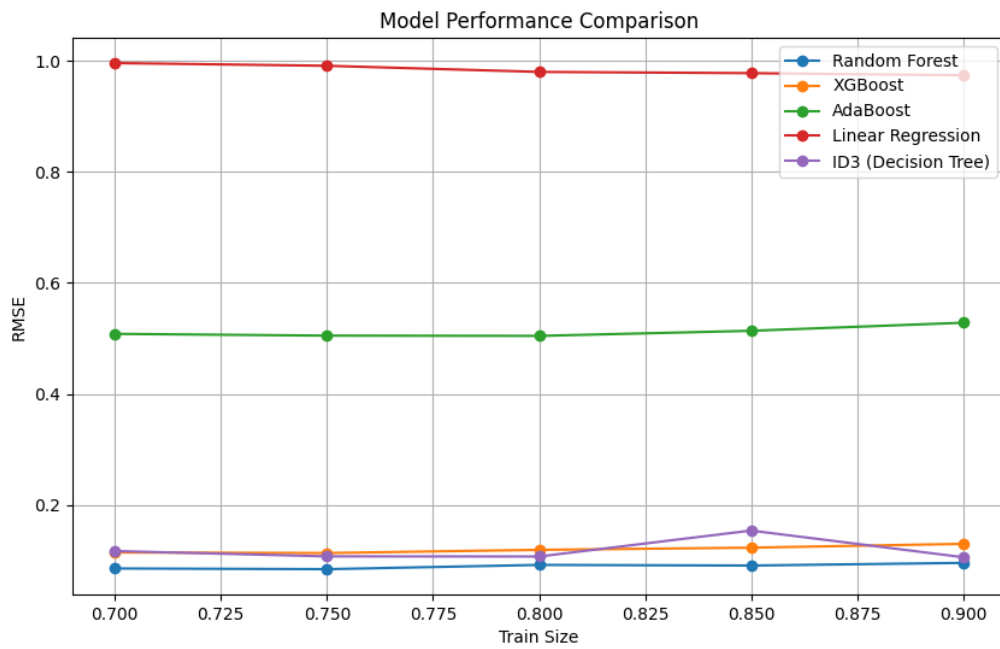


Figure 1: Compararea RMSE între algoritmi pe setul de date de testare.

## Ierarhia Categoriilor pe Țări

Loc pentru grafic: **Ierarhia categoriilor pentru fiecare țară.**

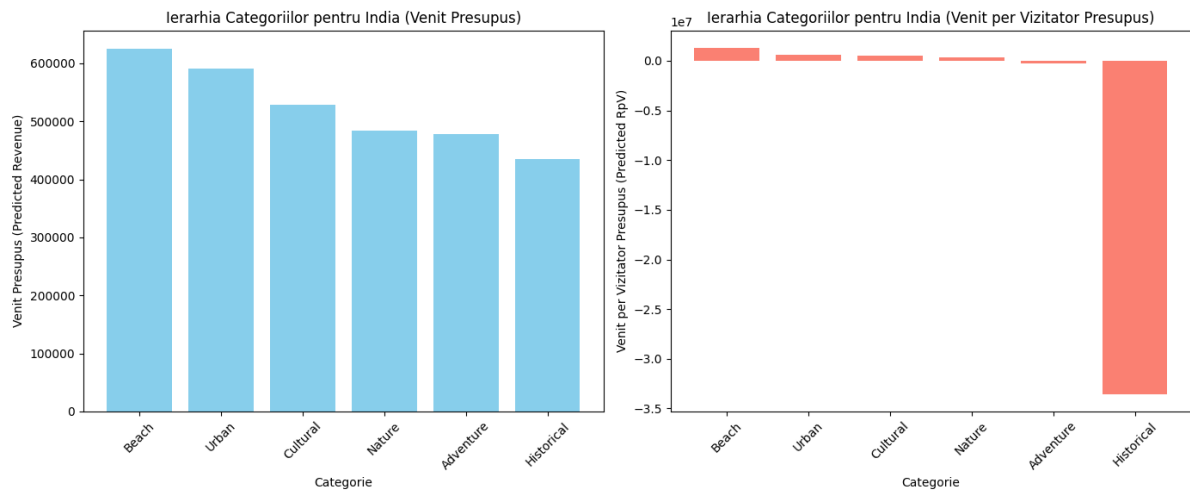


Figure 2: Ierarhia categoriilor pentru India (cel mai mare venit la categoria Beach).

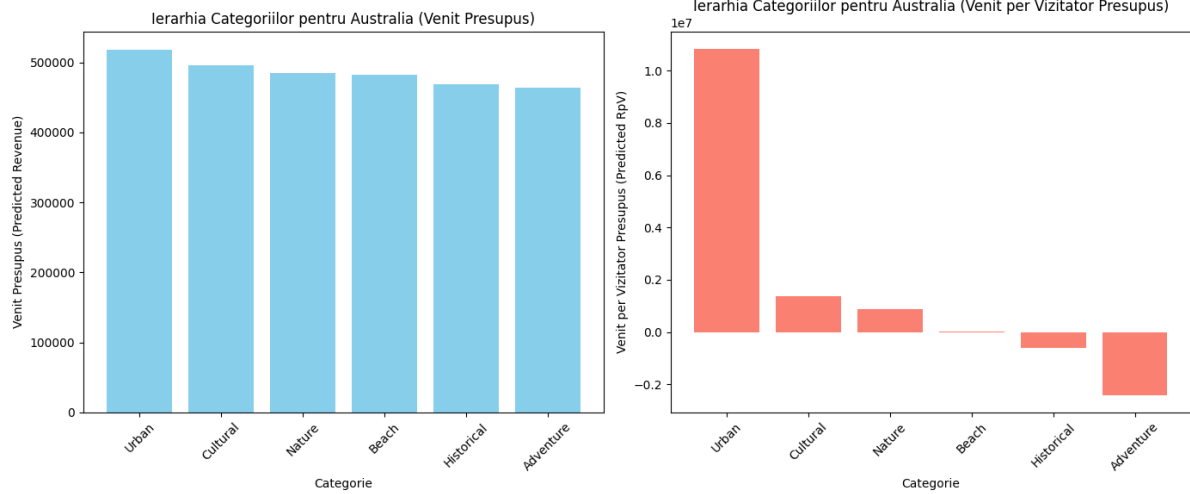


Figure 3: Ierarhia categoriilor pentru Australia.

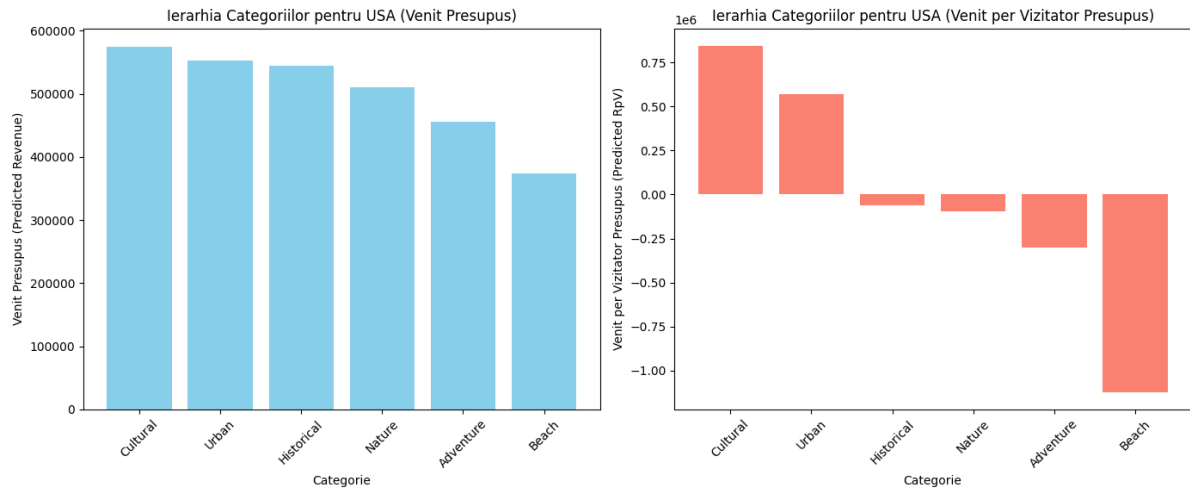


Figure 4: Ierarhia categoriilor pentru USA.

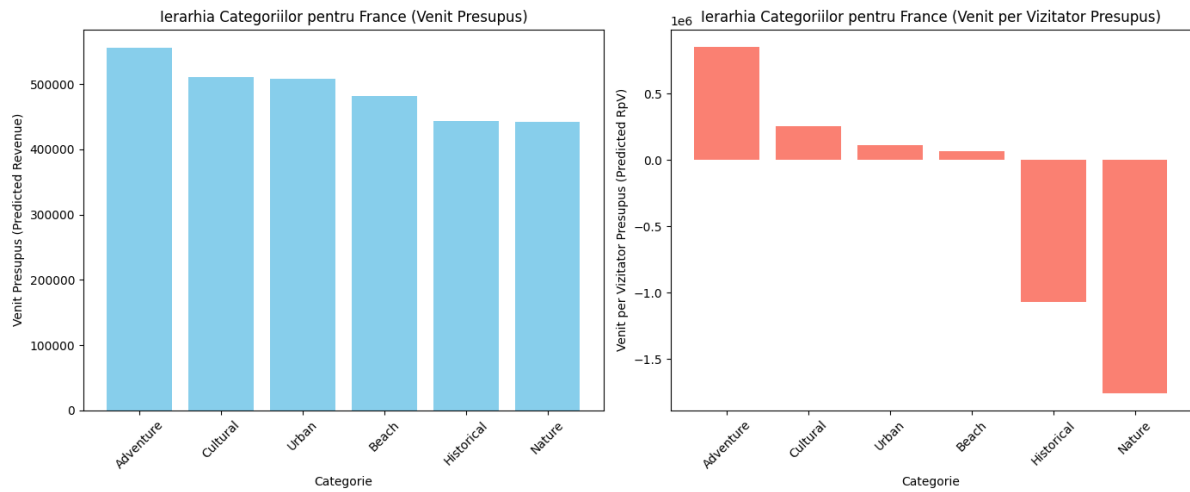


Figure 5: Ierarhia categoriilor pentru Franta.

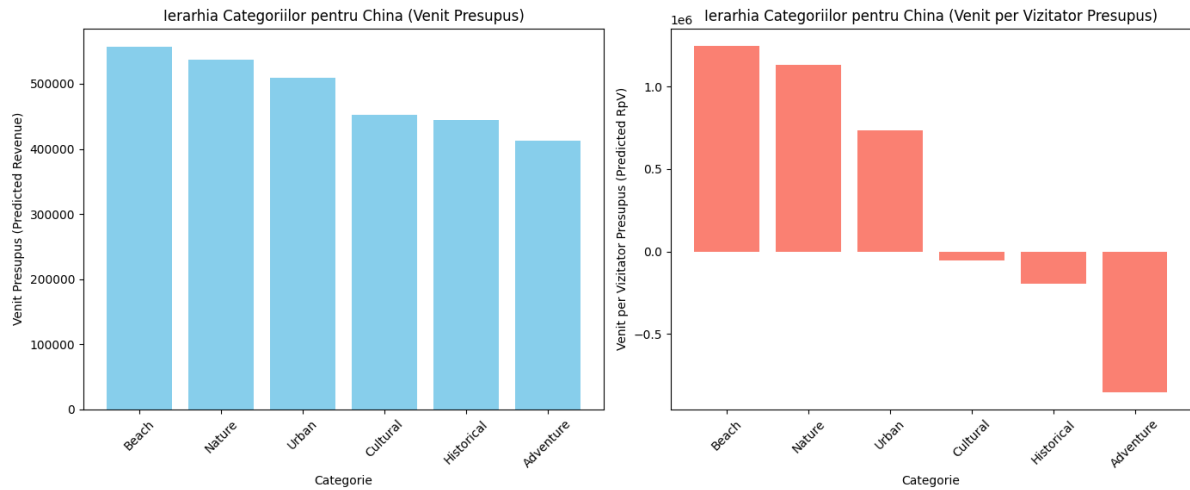


Figure 6: Ierarhia categoriilor pentru India (cel mai mare venit la categoria Beach).

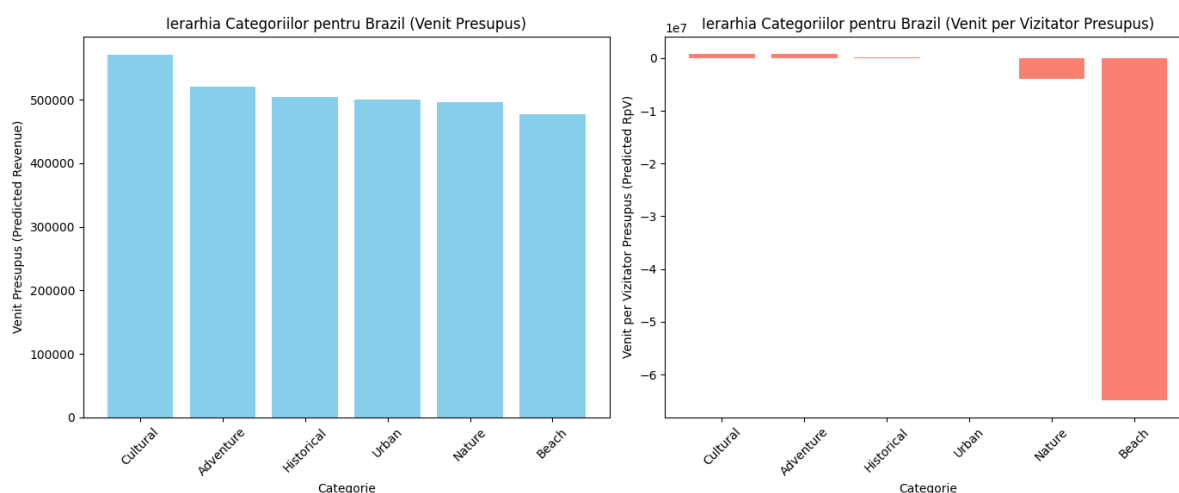


Figure 7: Ierarhia categoriilor pentru Brazilia.

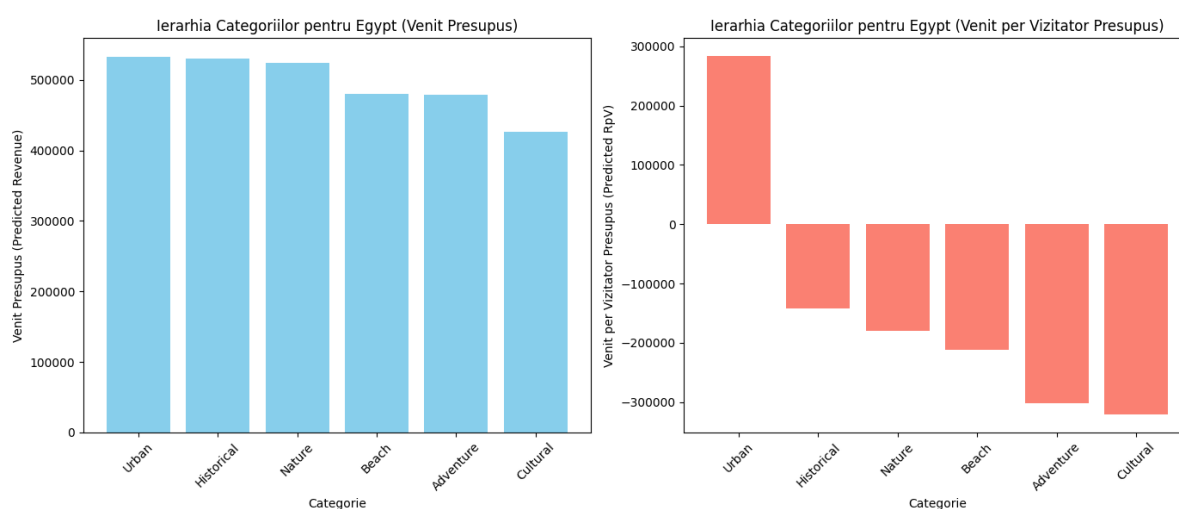


Figure 8: Ierarhia categoriilor pentru Egipt.

## Concluzii

Rezultatele experimentale arată că ID3 oferă o performanță superioară (dintre algoritmi învățați) pentru această problemă datorită flexibilității și robusteții sale. În continuare, graficele atașate pot fi utilizate pentru a analiza în detaliu venitul estimat și venitul per vizitator pentru fiecare țară și categorie. Acestea arată că țara cu cel mai mare venit este India cu 625511, la categoria 'Beach', iar venitul per vizitator are aceeași ierarhie cu venitul general.