

Predicția prețurilor caselor



Boston Housing Dataset

Number of Instances: 506

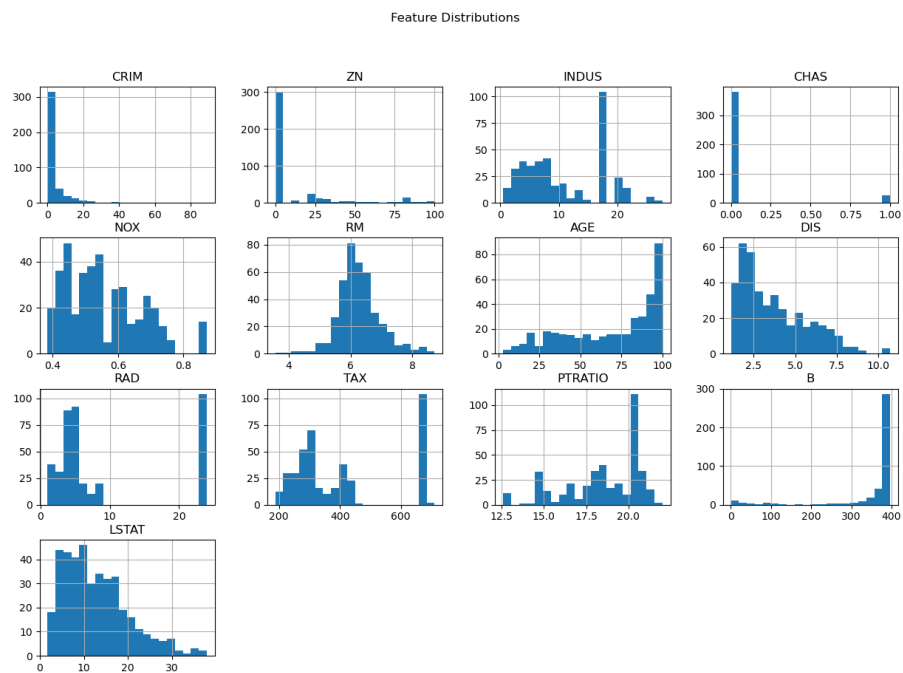
Number of Attributes: 13 numeric/categorical predictive. Median Value (attribute 14) is usually the target.

Attribute Information (in order):

- CRIM per capita crime rate by town
- ZN proportion of residential land zoned for lots over 25,000 sq.ft.
- INDUS proportion of non-retail business acres per town
- CHAS Charles River dummy variable (1 if tract bounds river; 0 otherwise)
- NOX nitric oxides concentration (parts per 10 million)
- RM average number of rooms per dwelling
- AGE proportion of owner-occupied units built prior to 1940
- DIS weighted distances to five Boston employment centres
- RAD index of accessibility to radial highways
- TAX full-value property-tax rate per \$10,000
- PTRATIO pupil-teacher ratio by town
- B $1000(B_k - 0.63)^2$ where B_k is the proportion of blacks by town
- LSTAT % lower status of the population
- MEDV Median value of owner-occupied homes in \$1000's

1. Încărcați setul de date Boston Housing folosind API-ul Keras cu TensorFlow.
2. Reprezentați grafic distribuțiile variabilelor predictive.

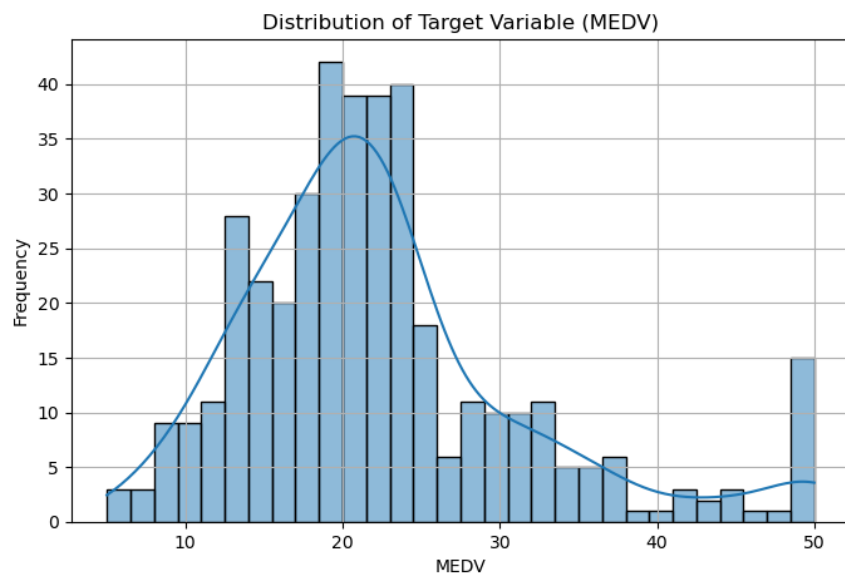
Comentați rezultatele obținute.



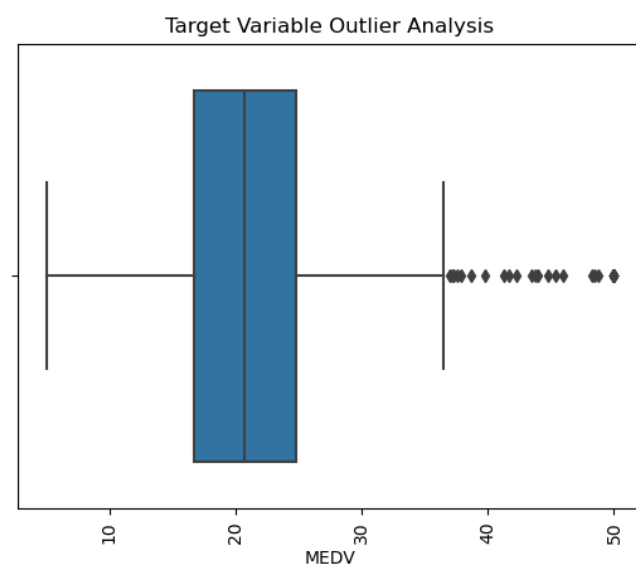
3. Reprezentați grafic distribuția variabilei țintă.

Verificați dacă variabila țintă are valori anormale.

Comentați rezultatele obținute.



Number of outliers: 32



4. Creați un set de validare din setul de antrenare (20%).
5. Folosind modelul secvențial, construiți următoarea rețea neuronală de referință:

```
# Baseline Model
model = Sequential([
    Dense(64, activation='relu', input_shape=(13,)),
    Dense(1)
])
model.compile(optimizer='adam', loss='mse', metrics=['mae'])
```

6. Experiment:

- scalarea datelor de intrare
- timp de antrenare: 100 de epoci

Comentați rezultatele obținute.

```
Unscaled Input - Loss (MSE): 33.99, MAE: 4.66  
Scaled to [0, 1] - Loss (MSE): 29.76, MAE: 4.13  
Scaled to [-1, 1] - Loss (MSE): 14.41, MAE: 2.88  
Standardized - Loss (MSE): 21.59, MAE: 3.48
```

7. Experiment:

- funcții de cost diferite
- date de intrare scalate în intervalul $[-1, 1]$
- timp de antrenare: 100 de epoci

Comentați rezultatele obținute.

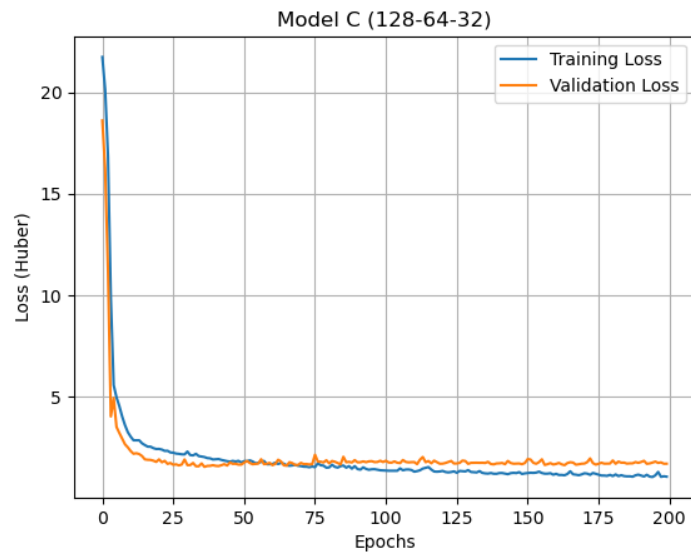
```
Loss function comparison (with scaled inputs):  
MSE Loss: 14.41, MAE: 2.88  
MAE Loss: 2.54, MAE: 2.54  
Huber Loss: 2.08, MAE: 2.52
```

8. Experiment:

- creșterea capacității rețelei
- date de intrare scalate în intervalul $[-1, 1]$
- timp de antrenare: 200 de epoci
- funcție de cost: Huber

Comentați rezultatele obținute.

```
Model A (64) - Loss (Huber): 1.77, MAE: 2.22  
Model B (128-64) - Loss (Huber): 1.69, MAE: 2.13  
Model C (128-64-32) - Loss (Huber): 1.70, MAE: 2.15
```



9. Experiment:

- tehnici de regularizare
- date de intrare scalate în intervalul $[-1, 1]$
- timp de antrenare: 200 de epoci
- funcție de cost: Huber
- model: C (128-64-32)

Comentați rezultatele obținute.

```
None: Val Loss = 1.70, MAE = 2.15
L1: Val Loss = 2.22, MAE = 2.13
L2: Val Loss = 1.84, MAE = 2.12
L1 + L2: Val Loss = 2.37, MAE = 2.25
Dropout: Val Loss = 1.83, MAE = 2.27
```

10. Evaluați modelul final pe setul de test.

Comentați rezultatele obținute.

```
Test Set Performance → Huber Loss: 2.13, MAE: 2.57
```

11. Reprezentați grafic predicțiile pentru setul de test în funcție de valorile țintă corespunzătoare.

Comentați rezultatele obținute.

