
# Week-3 AAM IPL Boston Housing Price Prediction

B.Tech – CSE(AIML)
V Semester - ML and AIUP, Aug-Oct 2024
Department of Computer Science Engineering – AI and ML (CSM)
G.Pulla Reddy Engineering College (Autonomous), Kurnool, AP

## Algorithm of Application

Linear Regression

## Project Title

Boston Housing Price Prediction

## Objective

The Boston Housing dataset is a classic dataset used for regression tasks, particularly in the domain of housing price prediction.

The dataset contains information collected by the U.S. Census Service concerning housing in the Boston suburbs. The dataset has been widely used to illustrate the workings of machine learning algorithms, particularly linear regression

## Dataset

- Description
  - The Boston Housing dataset is a famous dataset commonly used for regression exercises in machine learning. It consists of various features about houses in the Boston suburbs, and the task is to predict the median value of owner-occupied homes (in $1000s). The dataset has been used to demonstrate concepts like linear regression, decision trees, and other supervised learning algorithms.
- Dataset Details:
  - Number of Instances: 506
  - Number of Features: 13
  - Response Variable (Target): Median value of owner-occupied homes (MEDV, in $1000s)
- Features:
  - 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 oxide concentration (parts per 10 million).
  - RM — Average number of rooms per dwelling.
  - AGE — Proportion of owner-occupied units built before 1940
  - DIS — Weighted distances to five Boston employment centres.

C-501, Salarpuria Serenity, 5th Main, Sector 7, HSR Layout, Bengaluru 560102 KA India
Mobile: +91 97012 22130, Email: vmelachervu@gmail.com, vmela23@iitk.ac.in, Website: www.linkedin.com/in/vmelachervu

  - RAD — Index of accessibility to radial highways.
  - TAX — Full-value property tax rate per $10,000.
  - PTRATIO — Pupil-teacher ratio by town.
  - B — $1000(Bk - 0.63)^2$ where Bk is the proportion of Black residents by town.
  - LSTAT — Percentage of lower status of the population.
- Target Variable:
  - MEDV — Median value of owner-occupied homes in $1000s.
- Usage in Machine Learning:
  - The dataset is used to predict the MEDV (median value of homes) using the other 13 features. Typical machine learning exercises include:
    - Linear Regression: To establish a relationship between the features and MEDV.
    - Polynomial Regression: Extending linear regression to capture non-linear relationships.
    - Ridge/Lasso Regression: To deal with multicollinearity or overfitting.
    - Decision Trees and Random Forests: For non-linear, tree-based models.
    - Feature Importance Analysis: Understanding the most influential factors impacting housing prices.
    - Residual Analysis: Evaluating model performance and predictions.
- Data Source and Published By:
  - The Boston Housing dataset was originally collected by the U.S. Census Bureau and was used in a study by Harrison, D. and Rubinfeld, D.L. in 1978 to analyze the relationship between housing prices and various economic factors.
  - It was made available through the UCI Machine Learning Repository and has been widely used in the machine learning community.
- Data Download Link
  - Kaggle - Boston housing dataset (kaggle.com)
  - Directly from Python sklearn

```
from sklearn.datasets import load_boston
data = load_boston()
# Deprecated in some versions, consider using alternative datasets
```

## Implementation Steps

1. Perform Exploratory Data Analysis (EDA)
   a. Print any missing values in the provided dataset
   b. Print total data samples/points in the data set
   c. Print first 5 rows of the data in the set
   d. Plot histograms of all features with continuous values
   e. Plot correlation heatmap of features
2. Build a Linear Regression Model
   a. Standardize the dataset and train the model (test size – 20%)
   b. Predict the target variable (MEDV) using the independent features
   c. Plot Actual Vs Predicted Home Prices
3. Evaluate the Model - Use evaluation metrics to assess model's performance
   a. $R^2$ Score
   b. Mean Squared Error (MSE)
   c. Root MSE
   d. Mean Absolute Error (MAE)
   e. Mean Absolute Percentage Error (MAPE)

4. Print the Regression Coefficients of the Model
5. Generate the PDF of Code and Output

## Project Files Provided

- Project shell code file - ***AAM-IPL-Wk-3-LinearReg-Boston-Housing-Shell-Code-V3.ipynb***
- Boston Housing Data - ***BostonHousing.csv***
- Watermark image for plots - ***AAM-IPL-Header-6.png***

## Project Overview, Implementation and Submission Timeline

**28-09-2024 – Saturday 10:30 AM – Next Project Details Announcement – Topic, Data Set, Shell Code etc. Announcement Channels – Google Class, Industry Projects WhatsApp Group.**

| | | | |
|---|---|---|---|
| **Internal Exams Break – 17-09-2024 to 28-09-2024** | | | |
| **4** | 28-09-2024 - Saturday<br>Duration: 1.5 Hrs | Linear Regression – Model Overview/Recap, Project Description, and Interactive Q&A | Online – Google Class |
| **5** | 29-09-2024 - Sunday<br>Duration: 1.5 Hrs | Linear Regression – Model Building, Output Demonstration, Q&A | Online – Google Class |

**03-09-2024 – Thursday 11:59 PM - Deadline to upload the project code submission by all students in Google Class.**

**Guest Lecture Timings:**
Saturdays: 10:30 AM IST – 12:00 Noon IST
Sundays: 10:30 AM IST – 12:00 Noon IST
Mondays: 6:30 PM IST – 8:00 PM IST

## Development Environment

- Computing Language – Python
- IDE – Visual Studio Code with Jupyter Notebook

## Instructor

| **Instructor** | |
|---|---|
| Venkateswar Reddy Melachervu (alumnus of GPREC, ECE Class of '92)<br>CTO, Brillium Technologies, Bengaluru<br>Email: **venkat.reddy.gf@gprec.ac.in**<br>Profile: LinkedIn | Visiting Faculty |

## Coordination

All the activities of this programme – lecture venues, weekly projects details announcements, general announcements, changes in lecture timings, etc. will be coordinated by CSM faculty member Sri V.Suresh.

Channels of Communication and Announcements

- Google Classroom
- Whatsapp group **- Applied AI & ML Industry Projects Lab**
- Emails (Strictly GPREC email addresses only)

## Programme Coordinator



© "...dare to dream; care to win..."

| | |
|---|---|
| Prof. V.Suresh<br>Email: **vsuresh.ecs@gprec.ac.in** | Faculty Member, CSM |

## Reference Books

- Pattern Recognition and Machine Learning by Chris Bishop, 2006 – PDF Link
- Machine Learning using Python by Manaranjan Pradhan and U Dinesh Kumar, Wiley 2019 – PDF Link

## Policies

- **Attendance:** All sessions are expected to be attended by all the enrolled students. In case of inability to attend, prior information is expected to be provided by the student to the coordinator with a copy to the visiting faculty
- **Project Submissions:** Duly completed projects (Jupyter NB file and a PDF of the Jupyter NB file) are expected to be submitted through google class prior to the deadline. In case of inability to complete due various unforeseen circumstance, students are expected to seek extension for the submission deadline.
- **Academic Integrity:** Students are expected to uphold the highest standards of academic integrity in all assignments for the Applied AI & ML Industry Projects Lab. Each assignment must be the student's own work, and all sources and collaborators must be properly acknowledged. By submitting their completed project source code, students confirm that they have adhered to this integrity policy and completed their work in an honest and ethical manner.

## Additional Information

For students interested in engaging with special projects in the field of Gen AI, please reach out to the visiting faculty at **venkat@brillium.in** for further details and opportunities.

## Contact Information

For any questions or concerns or further details on this programme, please contact **Program Coordinator** during office hours or via email.

---------- End of the Document----------



© "...dare to dream; care to win..."