Mini project Report

A “Paris Housing Prediction” Report submitted to

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, ANANTAPURAMU

In Partial Fulfillment of the Requirements for the Award of the degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE AND SYSTEMS ENGINEERING
BY

GANGARAPU DANUSH 21125A1504

Under the Guidance of

Mrs. K Pujitha

Assistant Professor

Department of Computer Science and Systems Engineering
SREE VIDYANIKETHAN ENGINEERING COLLEGE
(Affiliated to JNTUA, Anantapuramu)
Sree Sainath Nagar, Tirupathi – 517 102
2022-2023

TABLE OF CONTENTS

Institute Vision and Mission

VISION

To be one of the Nation's premier Engineering Colleges by achieving the highest order of excellence in Teaching and Research.

MISSION

- To foster intellectual curiosity, pursuit and dissemination of knowledge.
- To explore students' potential through academic freedom and integrity.

DEPARTMENT OF COMPUTER SCIENCE AND SYSTEMS ENGINEERING

VISION AND MISSION

VISION

- To become a Center of excellence in Computer Sciences and Systems Engineering through Teaching, Training and Innovation to produce high quality engineering professionals who can solve the growing complex problems of the society and industry.

MISSION

- Established with cause of development of Technical education in advanced Computers Sciences and Systems Engineering with applications to systems thereby serving the society and Nation.
- Transfer of knowledge through contemporary curriculum and fostering faculty and student development.
- Create keen interest for research and innovation among students and faculty by understanding the needs of the society and industry.
- Skill Development among diversity of students in technical domains and profession for development of systems and processes to meet the demands of the industry and research.
- Imbibing values and ethics in students for prospective and promising engineering and develop a sense of respect for all.

Program Educational Objectives (PEO's)

After few years of graduation, the graduates of B.Tech(CSSE) will:

1. Demonstrate competencies in the Computer Science domain and Management with an ability to comprehend, analyze, design and create software systems for pursuing advanced studies in the areas of interest.
2. Evolve as entrepreneurs or be employed by acquiring required skill sets for developing computer systems and solutions in multi-disciplinary areas.
3. Exhibit progression and professional skill development in Computer programming and systems development with ethical attitude through life-long learning.

Program Outcomes (PO's)

1. Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems (**Engineering knowledge**).

2. Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences (**Problem analysis**).

3. Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations (**Design/development of solutions**).

4. Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions (**Conduct investigations of complex problems**).

5. Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations (**Modern tool usage**)

6. Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice (**The engineer and society**)

7. Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development (**Environment and sustainability**).Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice (**Ethics**).

8. Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings (**Individual and team work**).

9. Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions (**Communication**).

10. Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments (**Project management and finance**).

11. Recognize the need for, and have the preparation and ability to engage in independent and life long learning in the broadest context of technological change (**Life-long learning**).

Program Specific Outcomes (PSO's)

On successful completion of the Program, the graduates of B. Tech (CSSE) program will be able to:

PSO1: Employ Systems Approach to model the solutions for real life problems, design and develop software systems by applying Modern Tools.

PSO2: Develop solutions using novel algorithms in High Performance Computing and Data Science.

PSO3: Use emerging technologies for providing security and privacy to design, deploy and manage network systems.

PROBLEM STATEMENT

The Prediciton of Prices of houses in Paris using Machine Learning algorithms(ML) are Decision Tree and NaiveBayes.

ABSTRACT

This project focuses on the classification of Paris housing using machine learning techniques. The dataset used consists of a variety of features such as housing size, location, age, and price. The goal of this project is to develop a model that can accurately classify Paris housing based on these features. The model will use supervised learning algorithms such as Decision Tree(DT) and NaiveBayes to classify the data into categories such as high-end, mid-range, and low-end housing to improve the accuracy of the model. The results of this project will provide insight into the housing market in Paris, as well as a model that can be used to accurately classify the city's housing.

INTRODUCTION

The Paris housing classification mini project aims to identify the characteristics of the different housing types within Paris and to classify them accordingly. Through this project, we hope to gain a better understanding of the city's residential landscape and how it has evolved over time. We will examine various factors that influence the housing type, such as location, age of the building, number of occupants, and building amenities. We will also look at the various prices associated with the different housing types to gain insight into the Paris real estate market. Finally, we will use a combination machine learning algorithms to classify the different housing types. By doing so, we hope to gain a better understanding of the different housing types in Paris and to provide useful information to real estate investors, tenants, and other stakeholder.

1. CLASSIFICATION

DATASET : Paris Housing Classification

1.1 DECISION TREE

About Dataset

Context

This is a set of data created from data of house prices in an urban environment - Paris. I recommend using this dataset for educational purposes, for practice and to acquire the necessary knowledge. The dataset name is given as ParisHousingClass.csv and it consists of 14 attributes and 998 values.

Description

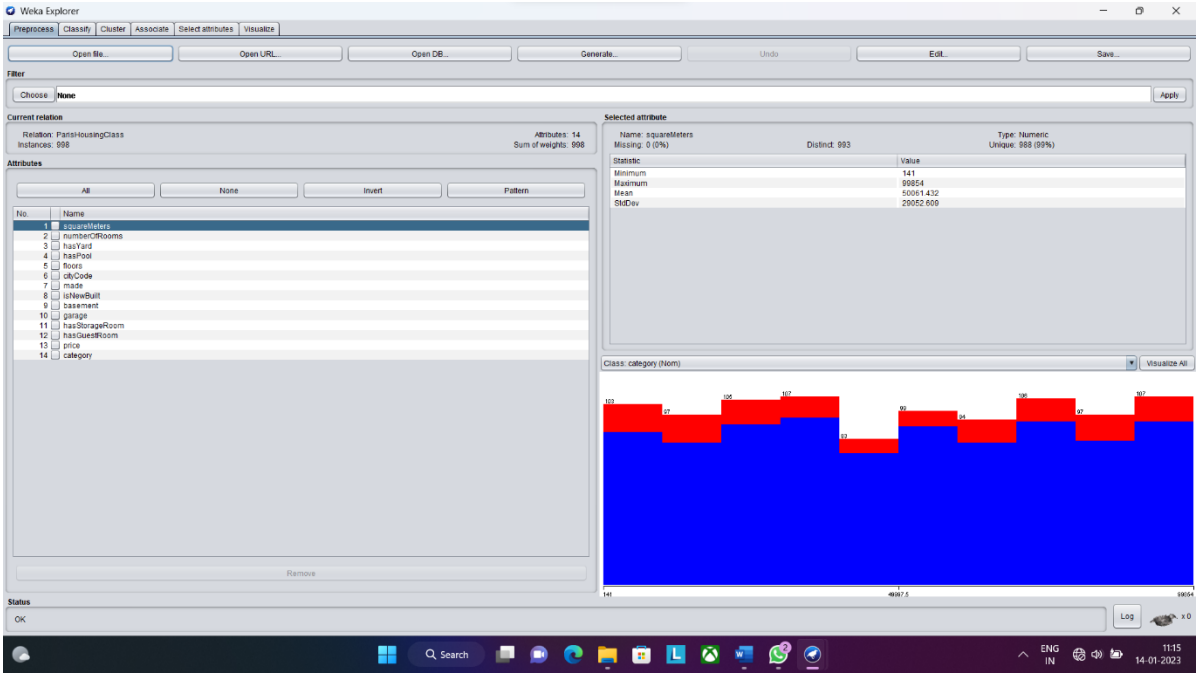
All attributes are numeric variables and they are listed below:

- squareMeters
- numberOfRooms
- hasYard
- hasPool
- floors - number of floors
- cityCode - zip code
- made – year
- isNewBuilt
- basement - basement square meters
- garage - garage size
- hasStorageRoom
- hasGuestRoom - number of guest rooms
- price - price of a house
- category - Luxury or Basic

IMPLEMENTATION ON WEKA:

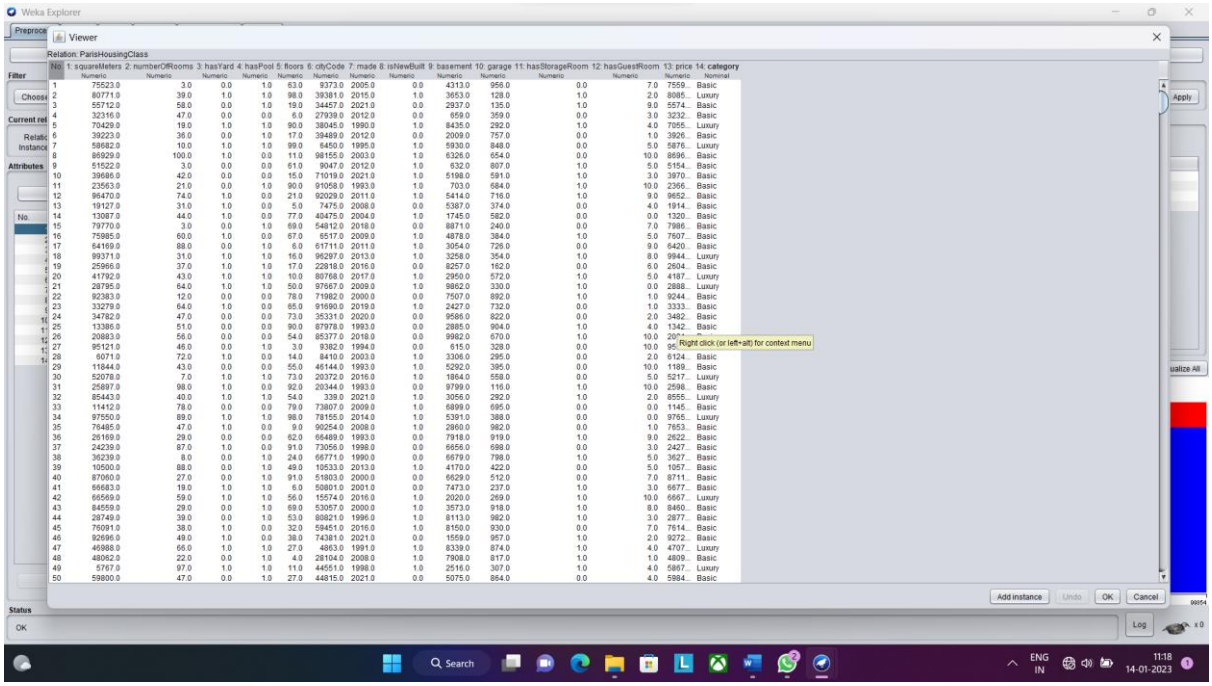
Procedure for Constructing **Decision Tree:**

1) Open Start -> Programs -> Weka
2) Open **explorer**.
3) Click on open file and select dataset.



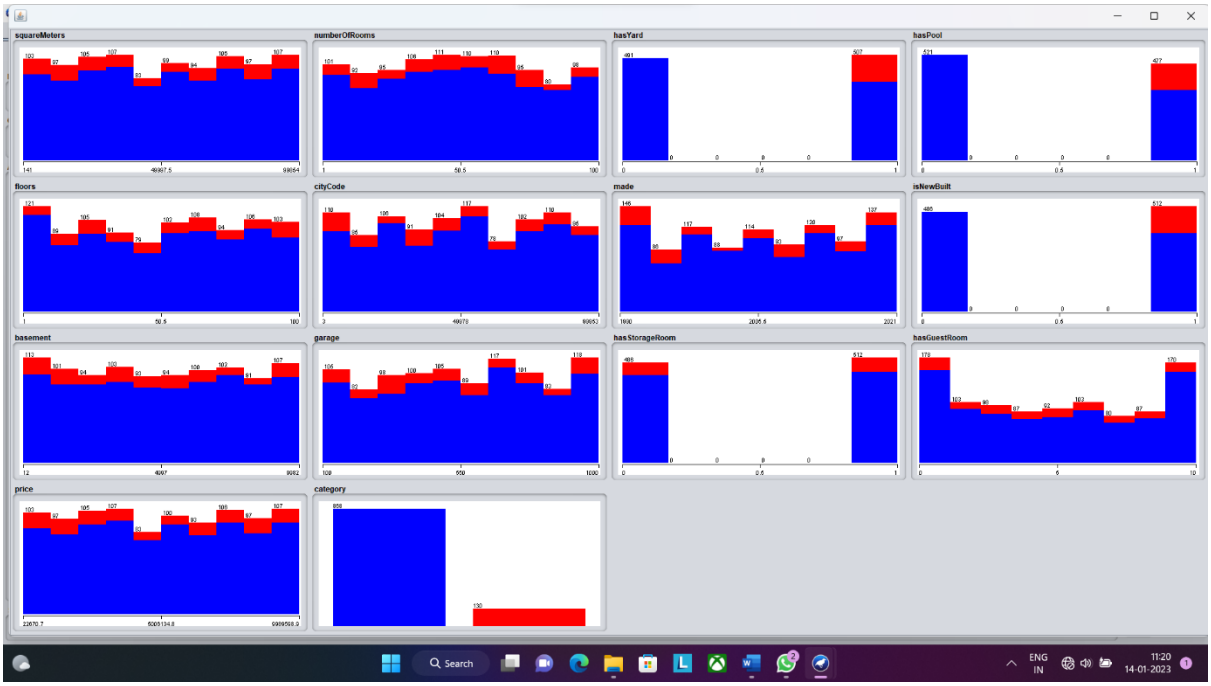
1.1.1. On importing data to weka

4) Click on Edit button to view the table.



1.1.2. View of the dataset

5) Click on visualize tab to visualize the attributes.



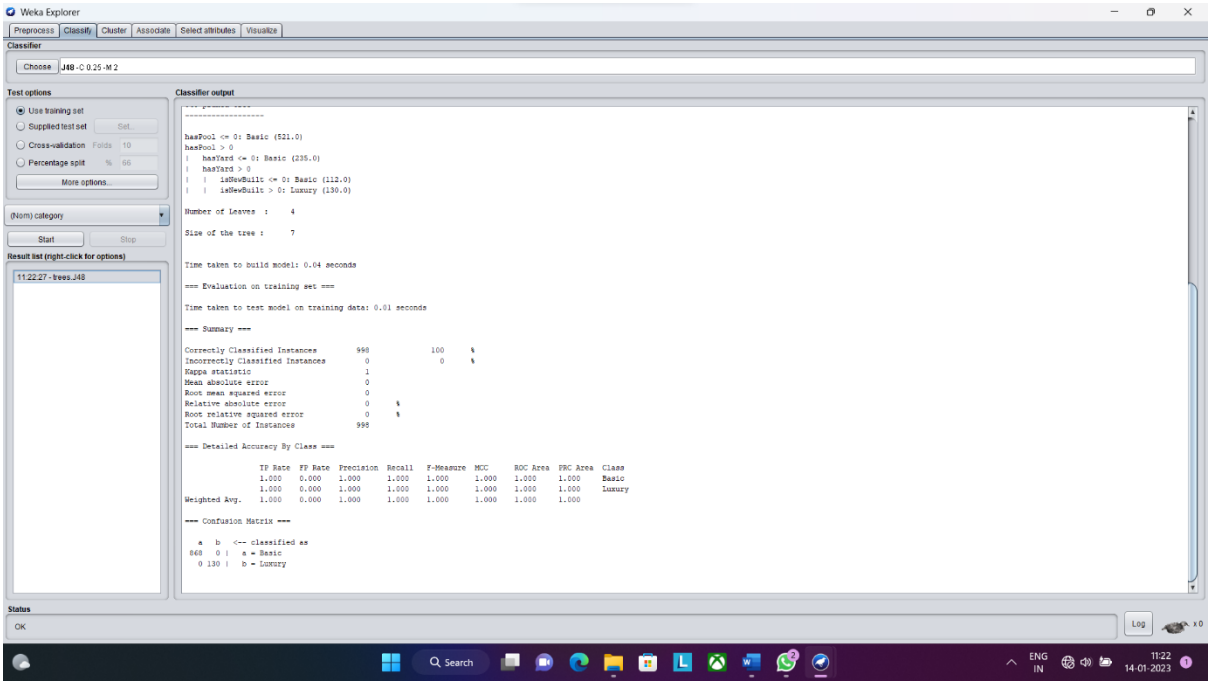
1.1.3. Visualization of the dataset

6) Select **classifier** option on the top of the Menu bar.

7) Select **choose** button and click on Tree option.

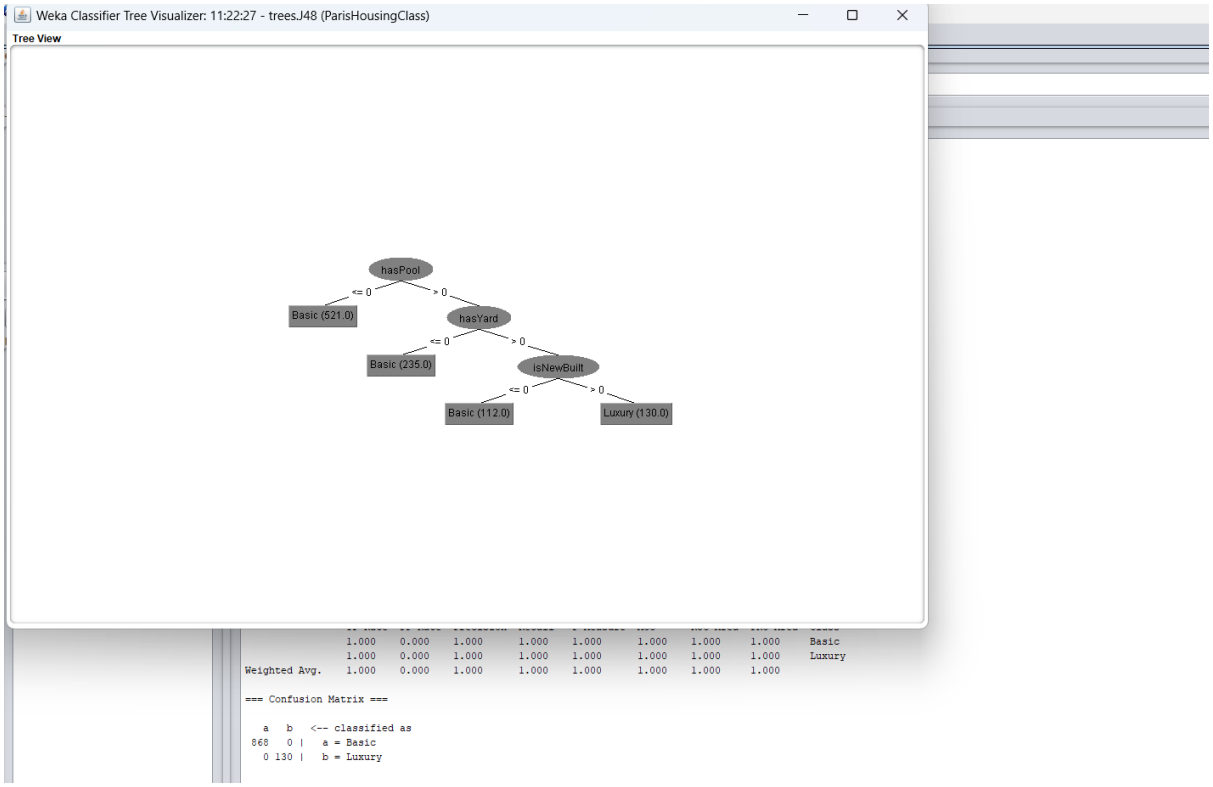
8) Click on **J48.**

9) Select the result list and right click on result list and select Visualize Tree option.



1.1.4. The result of the j48

10)Then **Decision Tree** will be displayed on new window.



1.1.5. Tree Visualization

RESULT:

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: ParisHousingClass

Instances: 998

Attributes: 14

squareMeters

numberOfRooms

hasYard

hasPool

floors

cityCode

made

isNewBuilt

basement

garage

hasStorageRoom

hasGuestRoom

price

category

Test mode: evaluate on training data

=== Classifier model (full training set) ===

J48 pruned tree

----

hasPool <= 0: Basic (521.0)

hasPool > 0

| hasYard <= 0: Basic (235.0)

| hasYard > 0

| | isNewBuilt <= 0: Basic (112.0)

| | isNewBuilt > 0: Luxury (130.0)

Number of Leaves : 4

Size of the tree : 7

Time taken to build model: 0.04 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.01 seconds

=== Summary ===

Correctly Classified Instances	998	100 %
Incorrectly Classified Instances	0	0 %
Kappa statistic	1	
Mean absolute error	0	
Root mean squared error	0	
Relative absolute error	0	%
Root relative squared error	0	%
Total Number of Instances	998	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Basic
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Luxury

Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000

=== Confusion Matrix ===

a	b	<-- classified as
868	0	a = Basic
0	130	b = Luxury

Decision Tree has been successfully constructed for dataset using Weka.

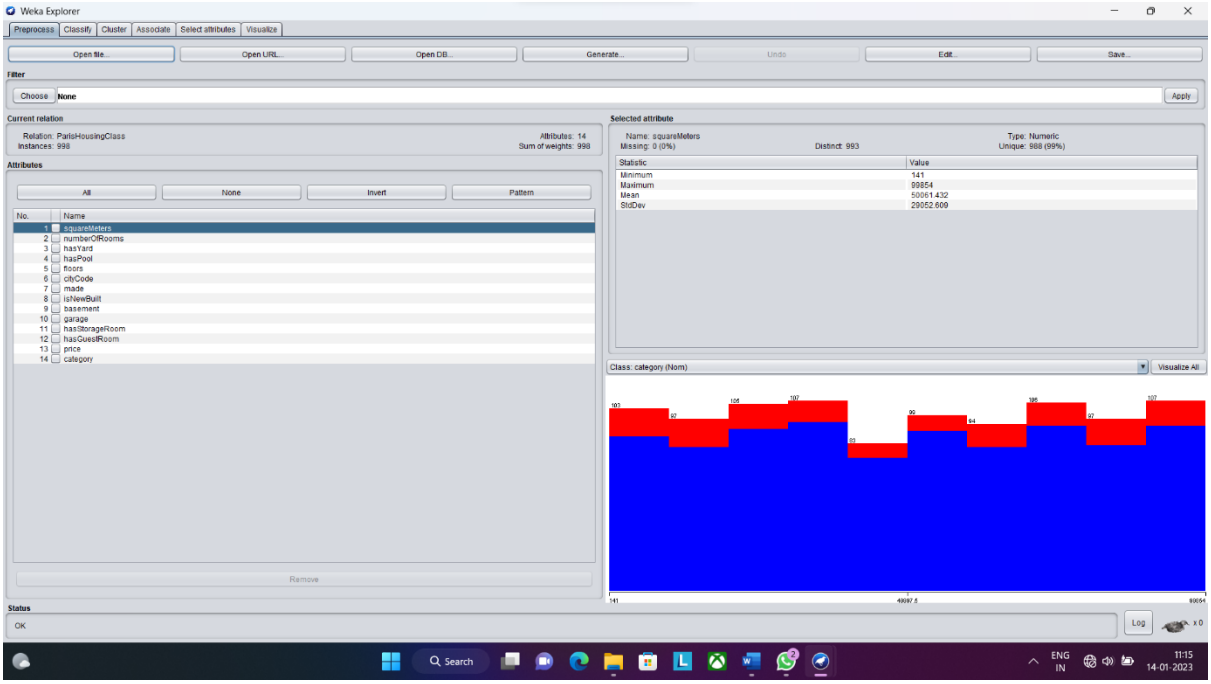
1.2.NAIVE BAYES

Naive Bayes Implementation On Weka:

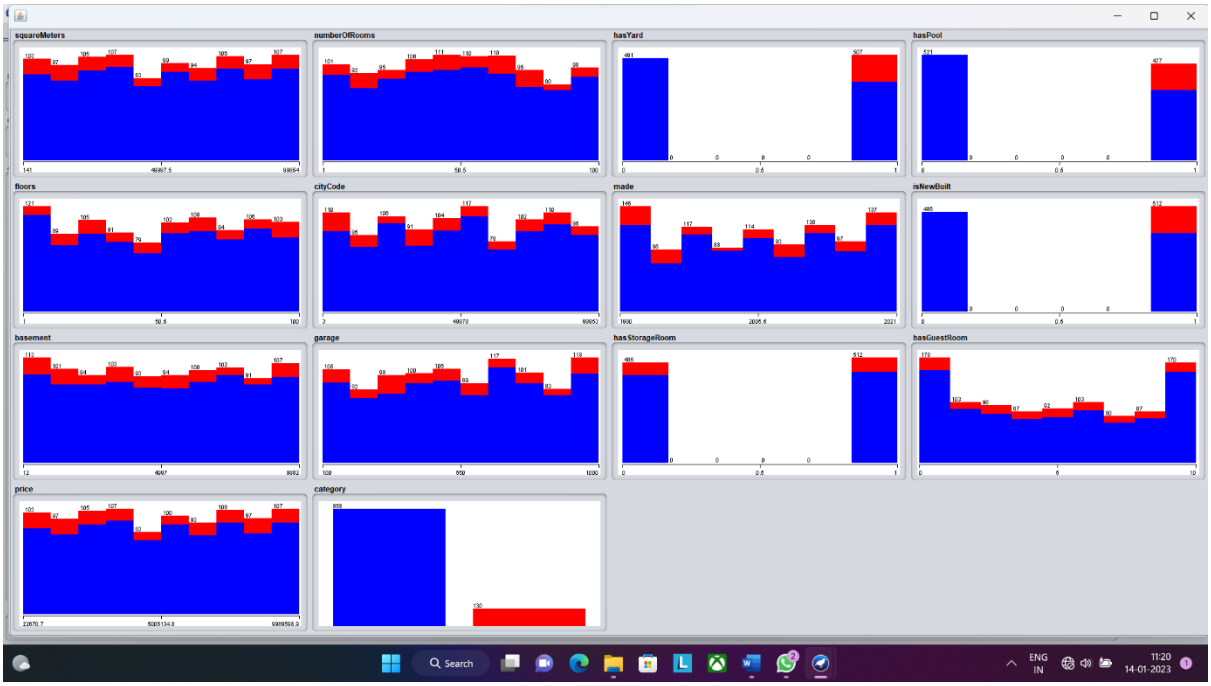
1) Open Start → Programs → Weka

2) Open explorer.

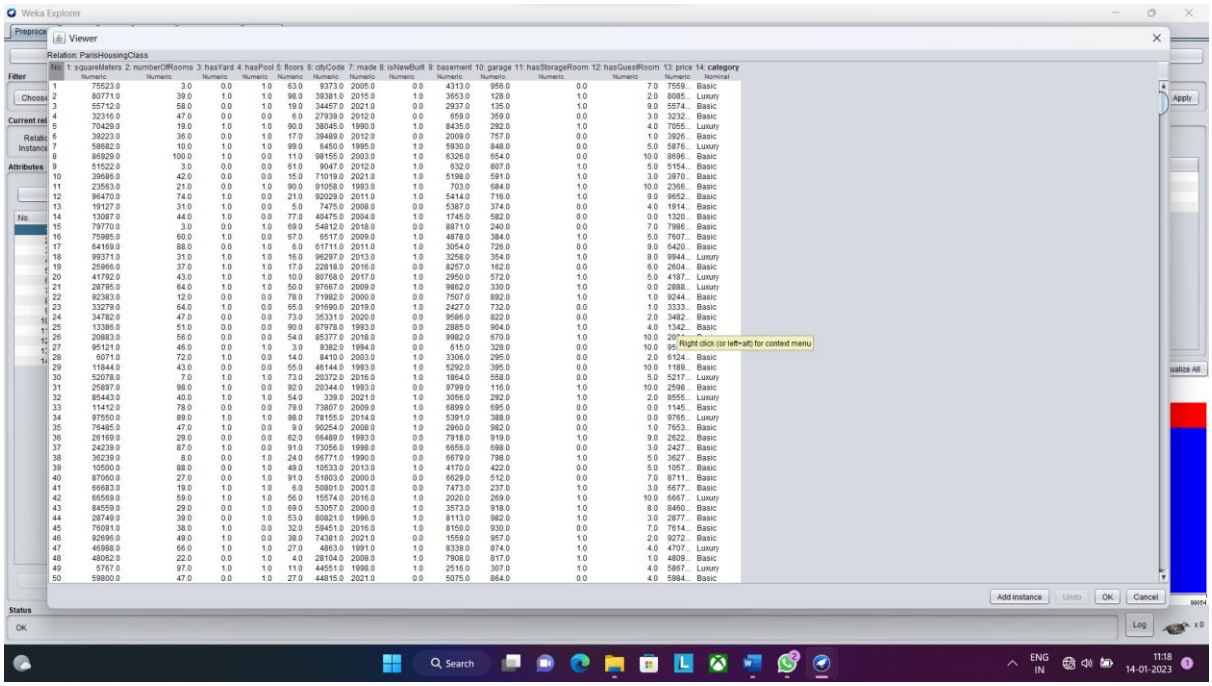
3) Click on open file and select data set.



1.2.1. On importing data to weka



1.2.2. Visualization of the dataset

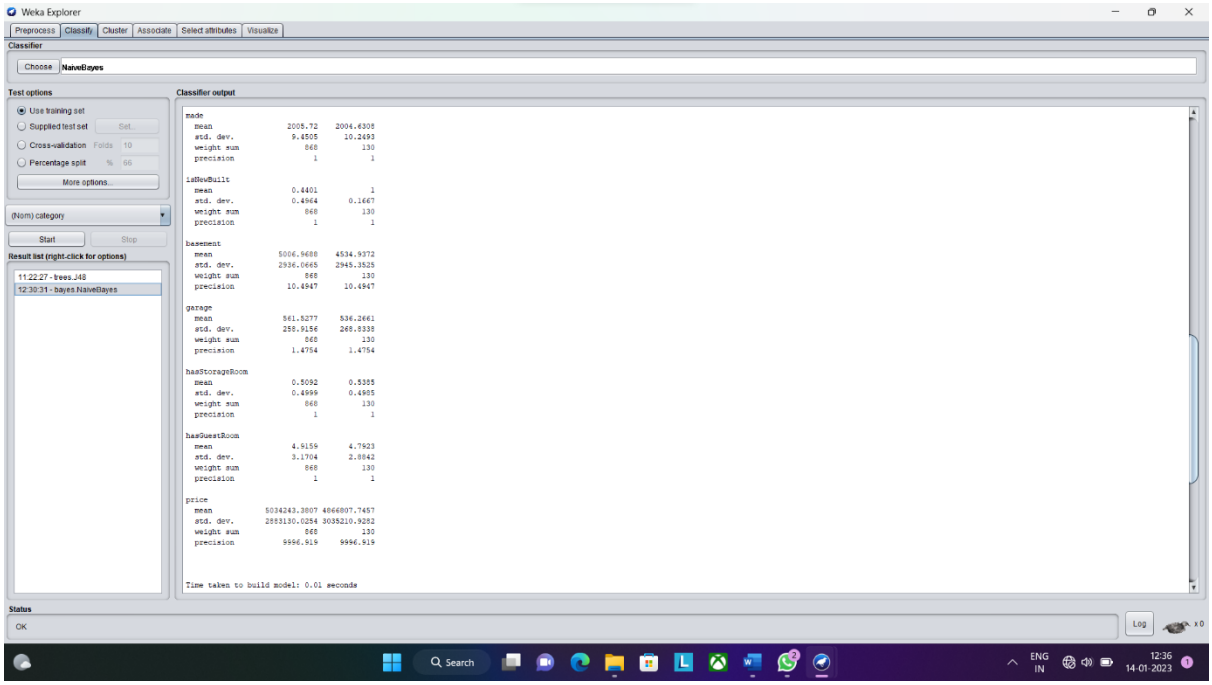


1.2.3. View of the dataset

4) Select **Classify** option on the top of the Menu bar.

5) Select **Choose** button and click on **wekaclassifierbayesNaiveBayes**.

6) Click on **Start** button and output will be displayed on the right side of the window.



1.2.4. Result of NaiveBayesClassification for the dataset

RESULT :

```
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     ParisHousingClass
Instances:    998
Attributes:   14
              squareMeters
              numberOfRooms
              hasYard
              hasPool
              floors
              cityCode
              made
              isNewBuilt
              basement
              garage
              hasStorageRoom
              hasGuestRoom
              price
              category
Test mode:    evaluate on training data

=== Classifier model (full training set) ===

Naive Bayes Classifier

                    Class
Attribute             Basic       Luxury
                     (0.87)      (0.13)
============================================
squareMeters
  mean            50284.8559   48569.1074
  std. dev.       28831.4968   30349.0444
  weight sum             868          130
  precision          100.5171     100.5171

numberOfRooms
  mean                50.1878      49.8846
  std. dev.            28.516      26.8339
  weight sum              868          130
  precision                 1            1

hasYard
  mean                 0.4343            1
  std. dev.            0.4957       0.1667
  weight sum              868          130
  precision                 1            1

hasPool
  mean                 0.3998            1
  std. dev.            0.4899       0.1667
  weight sum              868          130
  precision                 1            1

floors
```

mean	50.1198	52.4923
std. dev.	29.5852	29.3004
weight sum	868	130
precision	1	1

cityCode

mean	50402.7657	45891.3308
std. dev.	28694.7625	29821.4409
weight sum	868	130
precision	101.0617	101.0617

made

mean	2005.72	2004.6308
std. dev.	9.4505	10.2493
weight sum	868	130
precision	1	1

isNewBuilt

mean	0.4401	1
std. dev.	0.4964	0.1667
weight sum	868	130
precision	1	1

basement

mean	5006.9688	4534.9372
std. dev.	2936.0665	2945.3525
weight sum	868	130
precision	10.4947	10.4947

garage

mean	561.5277	536.2661
std. dev.	258.9156	268.8338
weight sum	868	130
precision	1.4754	1.4754

hasStorageRoom

mean	0.5092	0.5385
std. dev.	0.4999	0.4985
weight sum	868	130
precision	1	1

hasGuestRoom

mean	4.9159	4.7923
std. dev.	3.1704	2.8842
weight sum	868	130
precision	1	1

price

mean	5034243.3807	4866807.7457
std. dev.	2883130.0254	3035210.9282
weight sum	868	130
precision	9996.919	9996.919

Time taken to build model: 0.01 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.02 seconds

=== Summary ===

Correctly Classified Instances	998	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.0432		
Root mean squared error	0.1199		
Relative absolute error	19.0036 %		
Root relative squared error	35.6273 %		
Total Number of Instances	998		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Basic
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Luxury
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

```
  a   b   <-- classified as
868   0 |   a = Basic
  0 130 |   b = Luxury
```

Bayesian Classifier has been successfully constructed for dataset using Weka.

CONCLUSION

The Paris housing market is a complicated and dynamic system. While there are many factors that determine the price and classification of housing in Paris, it is clear that location, size, condition, and amenities are all important considerations to keep in mind when assessing a home's value. Additionally, the socio-economic makeup of the neighborhood, the availability of public transportation, and other local amenities can have a significant impact on the classification of the housing market. Ultimately, the classification of the Paris housing market is determined by a variety of factors and will be different depending on the specific location and situation.