MIT-WPU

।। विश्वशान्तिर्ध्रुवं ध्रुवा ।।

Dr. Vishwanath Karad
MIT WORLD PEACE UNIVERSITY | PUNE
TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

MINI PROJECT (4-1)
MID-TERM PRESENTATION

Real Estate based Recommendation System (Machine Learning)

Guide name: Prof. Bhakti Paranjape
Batch Coordinator: Prof. Shamla Mantri

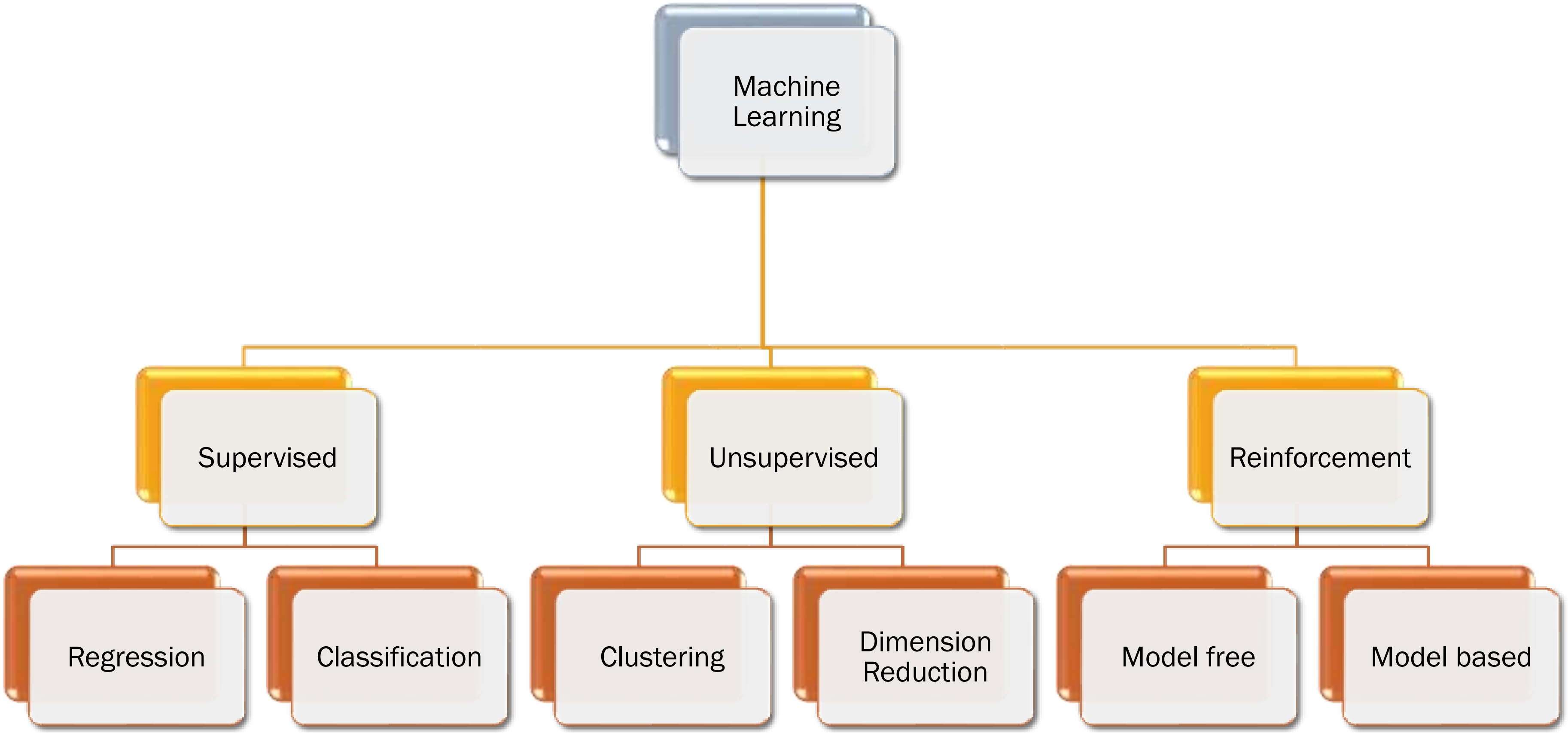
MACHINE LEARNING:

- Arthur Samuel (1959). Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.
- Tom Mitchell (1998) Well-posed Learning Problem: A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.
- Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task T in this setting?

○ Classifying emails as spam or not spam.

○ Watching you label emails as spam or not spam.

○ The number (or fraction) of emails correctly classified as spam/not spam.



Machine Learning

Supervised

Unsupervised

Reinforcement

Regression

Classification

Clustering

Dimension Reduction

Model free

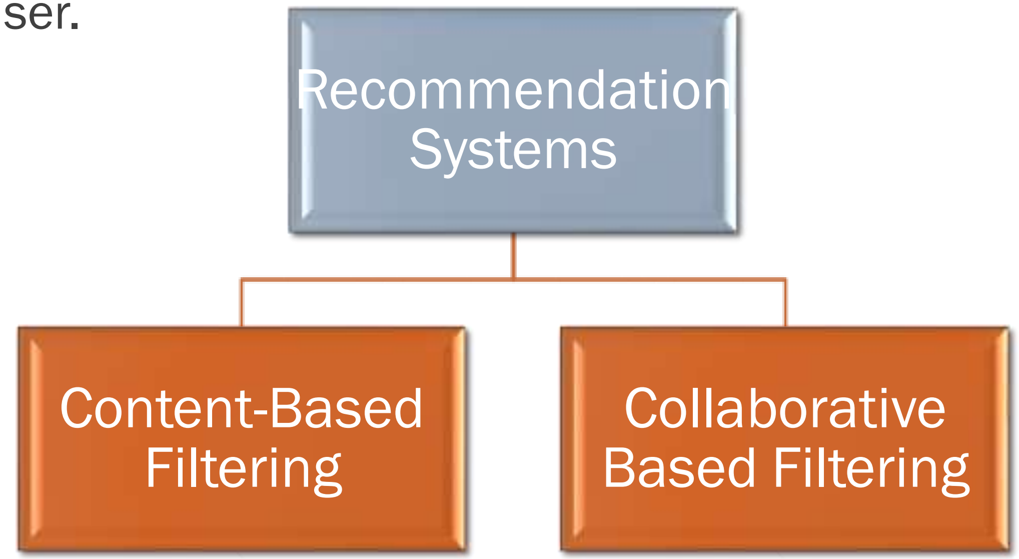
Model based

DATA SCIENCE

- Data science is a broad term for multiple disciplines, machine learning fits within data science.
- The main difference between the two is that data science as a broader term not only focuses on algorithms and statistics but also takes care of the entire data processing methodology.
- Data science, data analytics, and machine learning are some of the most in-demand domains in the industry right now.

RECOMMENDATION SYSTEMS:

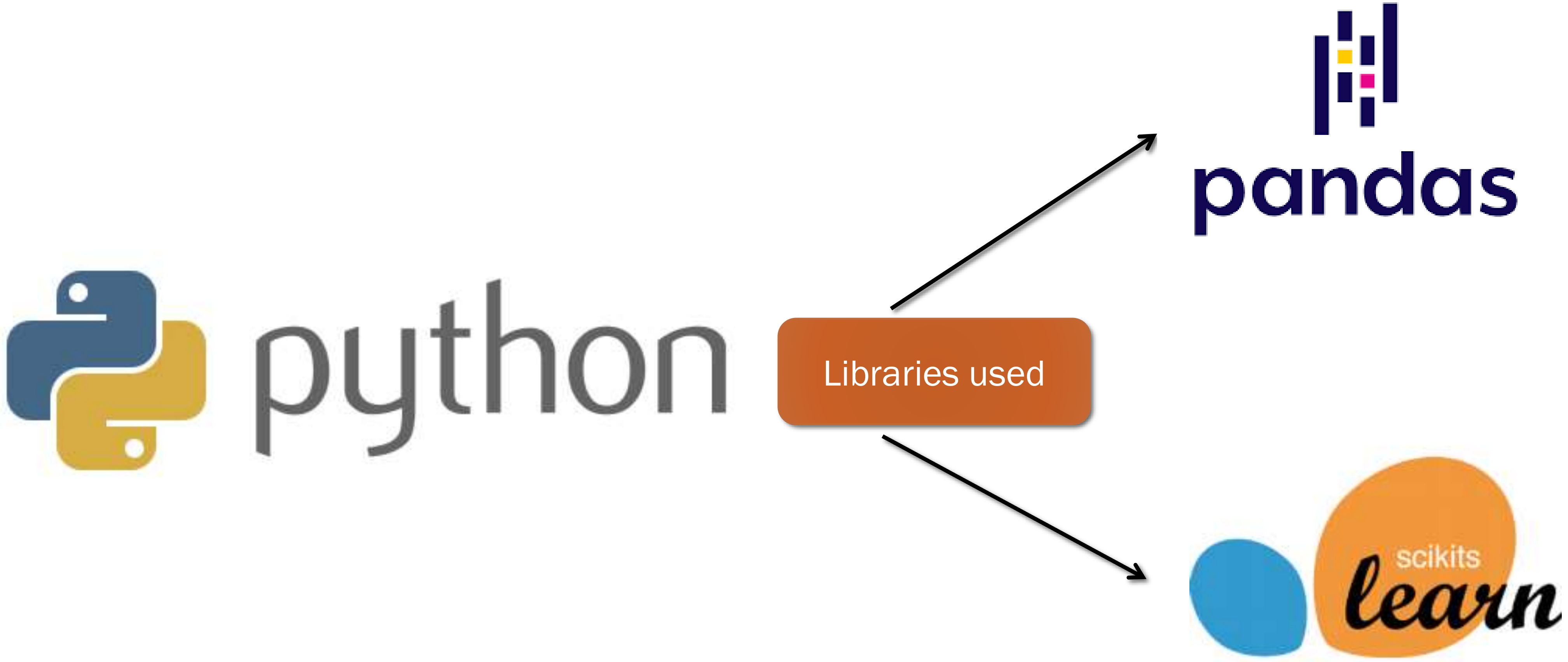
- Recommendation Systems are composed by algorithms that generate recommendations of a given type of item for users based on the information they provide
- Recommender engines (REs) also known as recommender systems are software tools and techniques providing suggestions to a user.



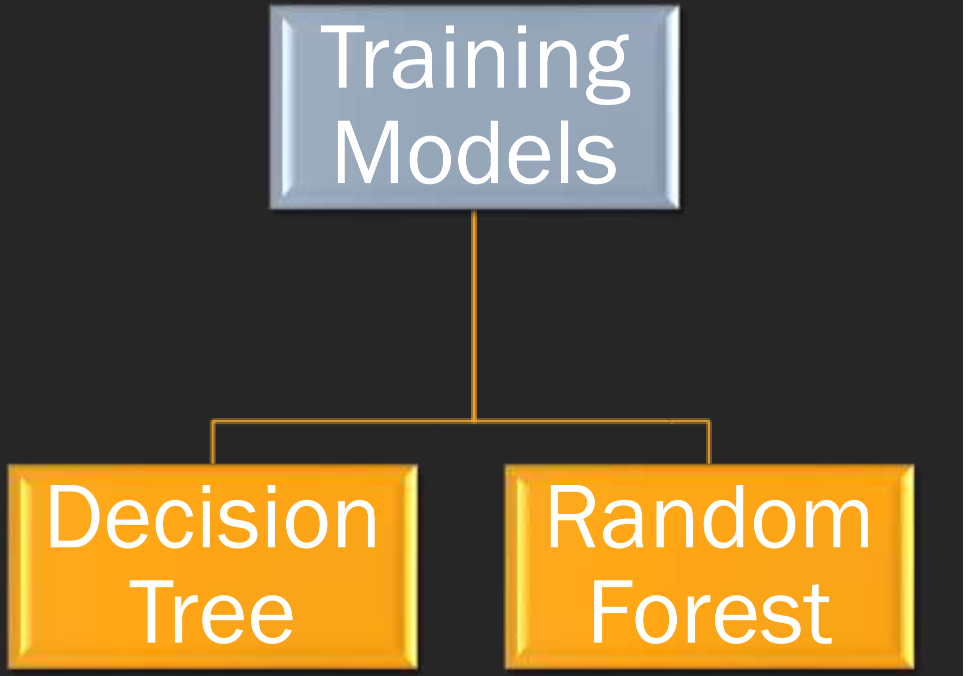
INTRODUCTION:

- This project outlines how the Recommendation System for Real Estate Project will address current real estate platform which has become a major priority for buyers & sellers.
- The project will compare and evaluate the results from two training models i.e. Decision Tree and Random Forest to predict the price.
- The motivation behind is to show how a recommender system can be applied in Real estate scenario.
- The main contributions of this project can be summarized as follow:
  - ❖ To analyze data of the real estate from the dataset
  - ❖ To evaluate and predict the results by using training models (Decision Tree and Random Forest)
  - ❖ To recommend the sectors of real estate based on data and compare the results.

SOFTWARE USED:



Libraries used



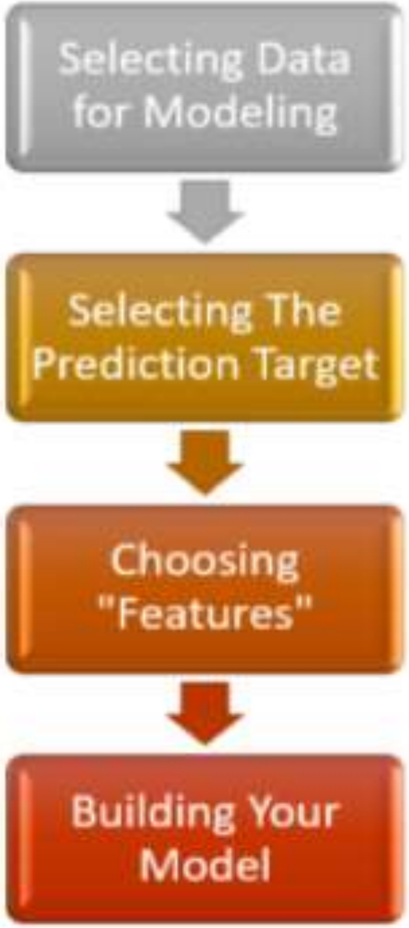
DECISION TREE

- Decision trees are constructed via an algorithmic approach that identifies ways to split a data set based on different conditions. It is one of the most widely used and practical methods for supervised learning.
- Decision Trees are a non-parametric **supervised learning** method used for both **classification** and **regression** tasks.

RANDOM FOREST

- Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble.
- Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model’s prediction

BLOCK SCHEMATIC:



PROCESS METHODOLOGY

1. Import data from the dataset

2. Read the data and store data in DataFrame

3. Choose variables/columns, with the columns property of the DataFrame

4. Use the dot notation to select the column we want to predict

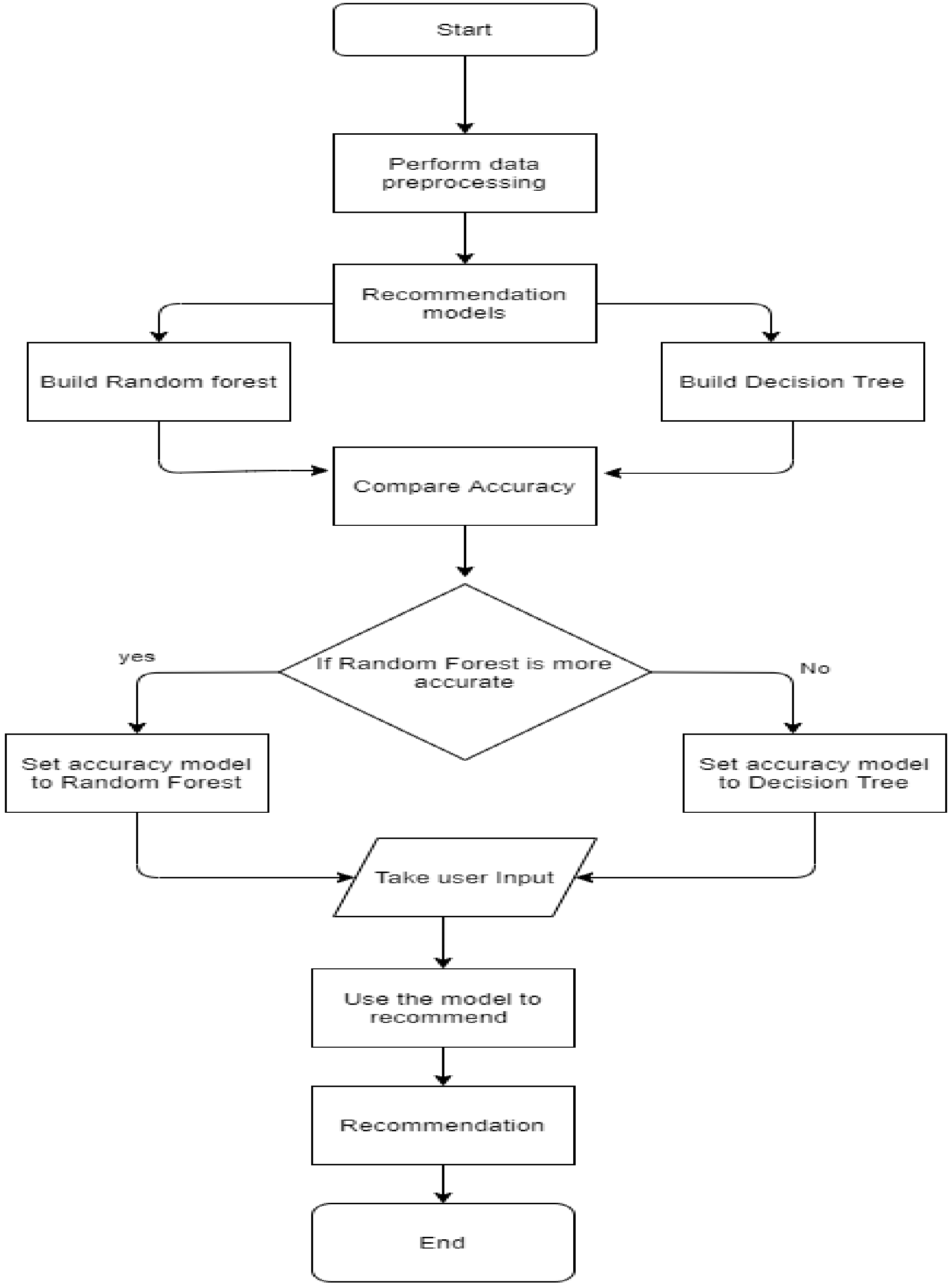
5. Select multiple features by providing a list of column names

6. Use the scikit-learn library to create your models.

7. Define a decision tree model and a Random Forest with scikit-learn and fitting it

with the features and target variable.

FLOWCHART



Start

Perform data preprocessing

Recommendation models

Build Random forest

Build Decision Tree

Compare Accuracy

If Random Forest is more accurate

yes

No

Set accuracy model to Random Forest

Set accuracy model to Decision Tree

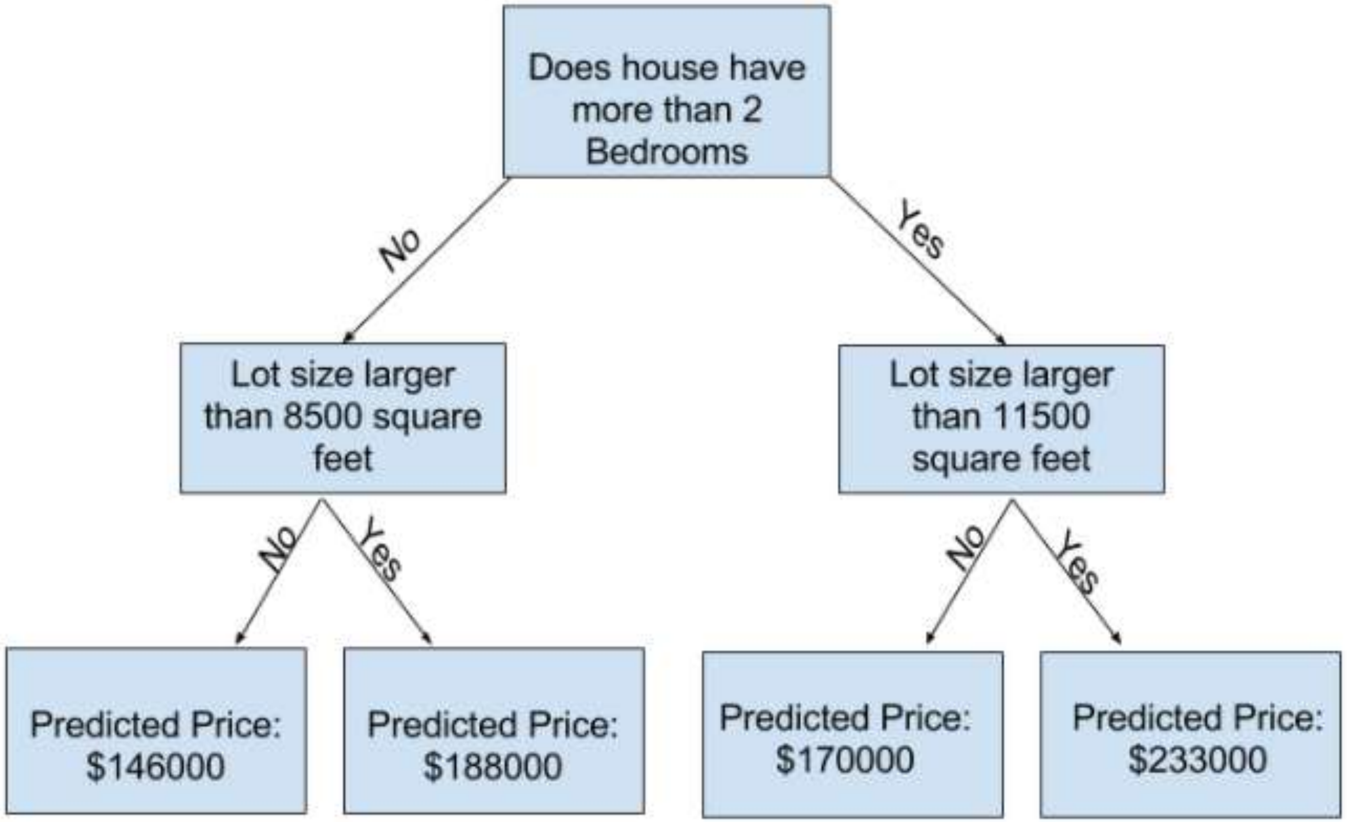
Take user Input

Use the model to recommend

Recommendation

End

DECISION TREE FLOWCHART



Does house have more than 2 Bedrooms

No

Yes

Lot size larger than 8500 square feet

Lot size larger than 11500 square feet

No

Yes

No

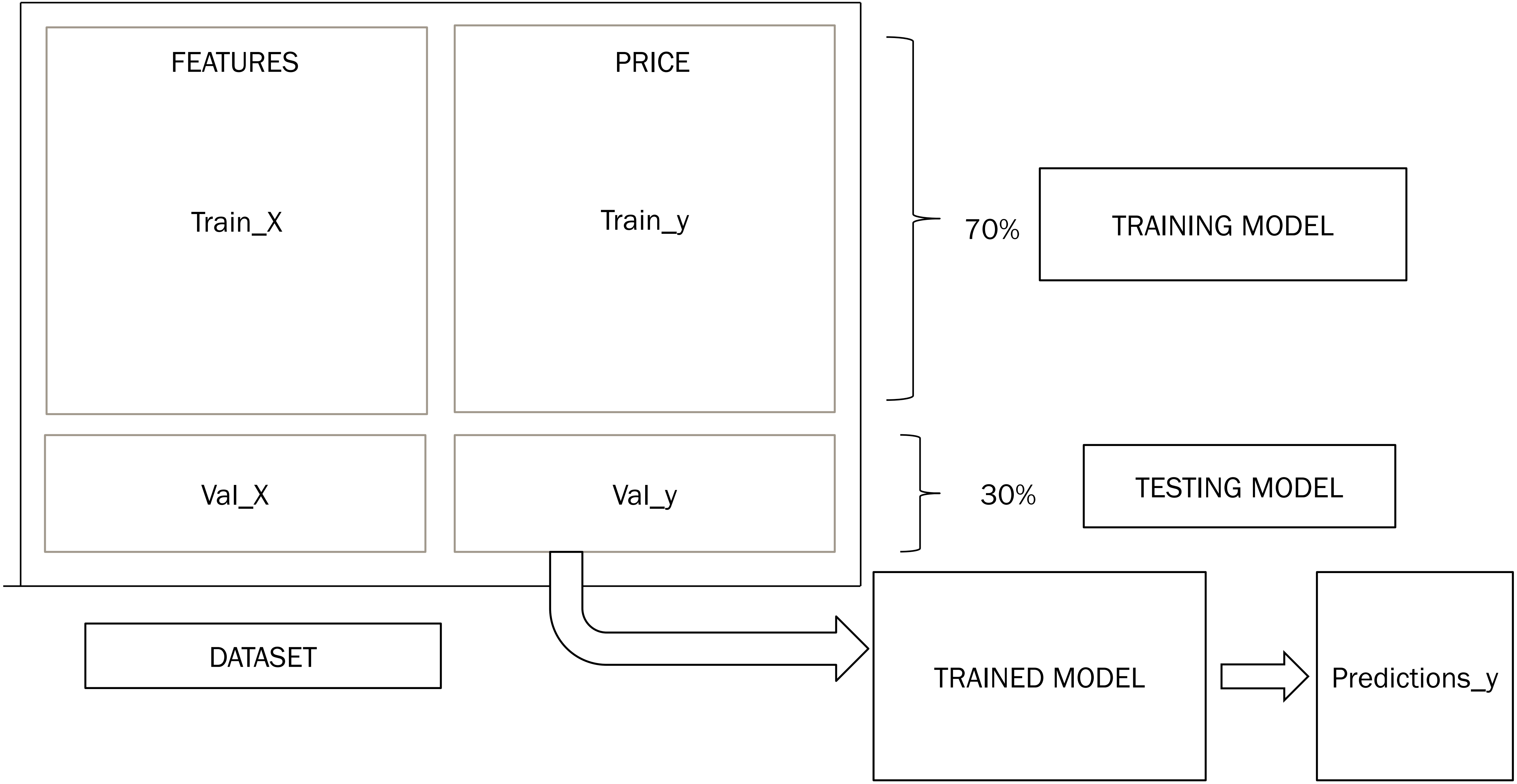
Yes

Predicted Price: $146000

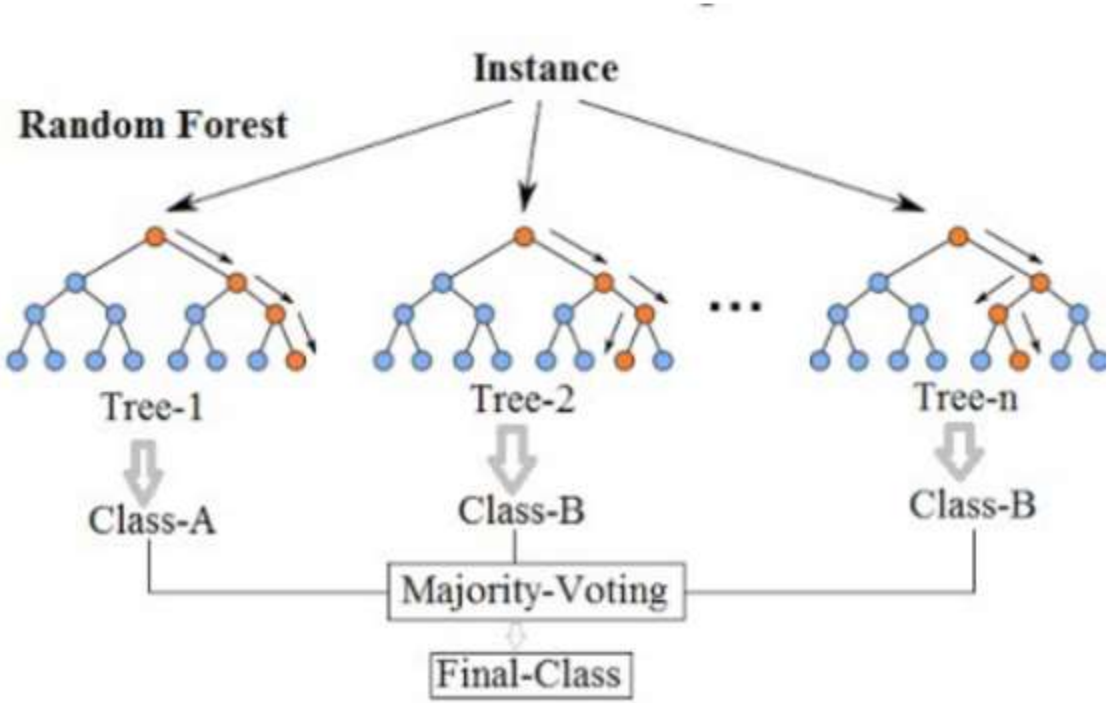
Predicted Price: $188000

Predicted Price: $170000

Predicted Price: $233000

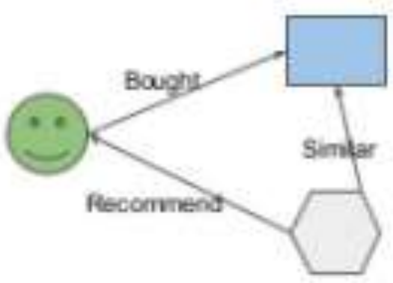

FEATURES
PRICE
Train_X
Train_y
70%
TRAINING MODEL
Val_X
Val_y
30%
TESTING MODEL
DATASET
TRAINED MODEL
Predictions_y

RANDOM FOREST FLOWCHART

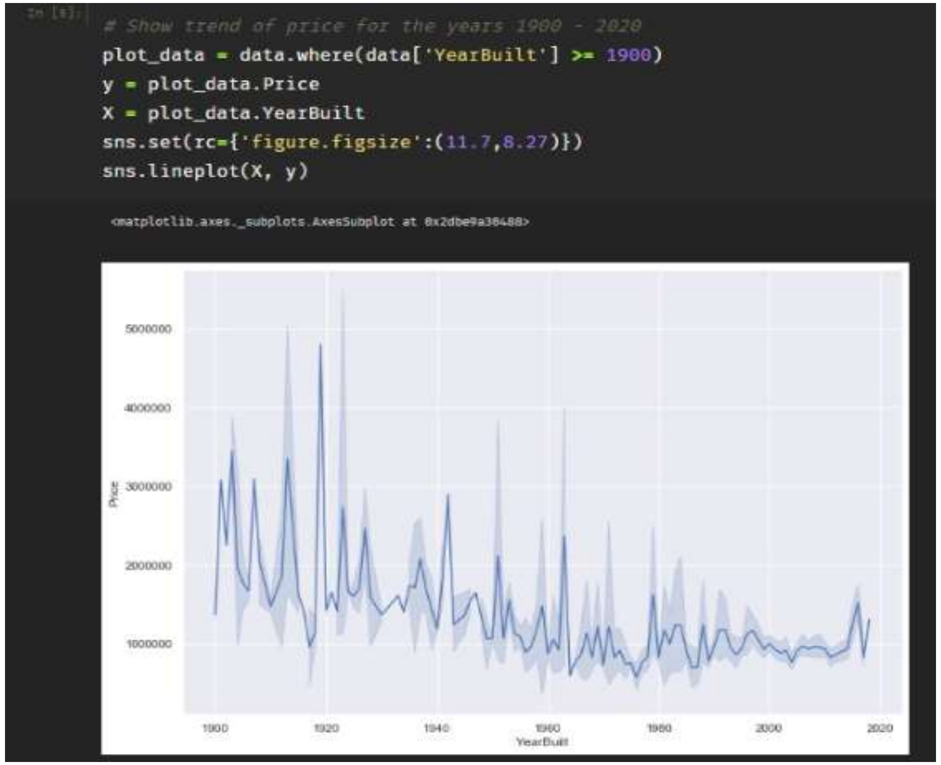


ADVANTAGES OF RECOMMENDER SYSTEMS

- A recommendation system uses data analysis techniques to figure out the items that match the users' taste & preferences.
- The ultimate aim of any recommendation engine is to stimulate demand and engage users.
- Recommendation engines can have many use cases like in entertainment, e-commerce, mobile apps, education, etc.

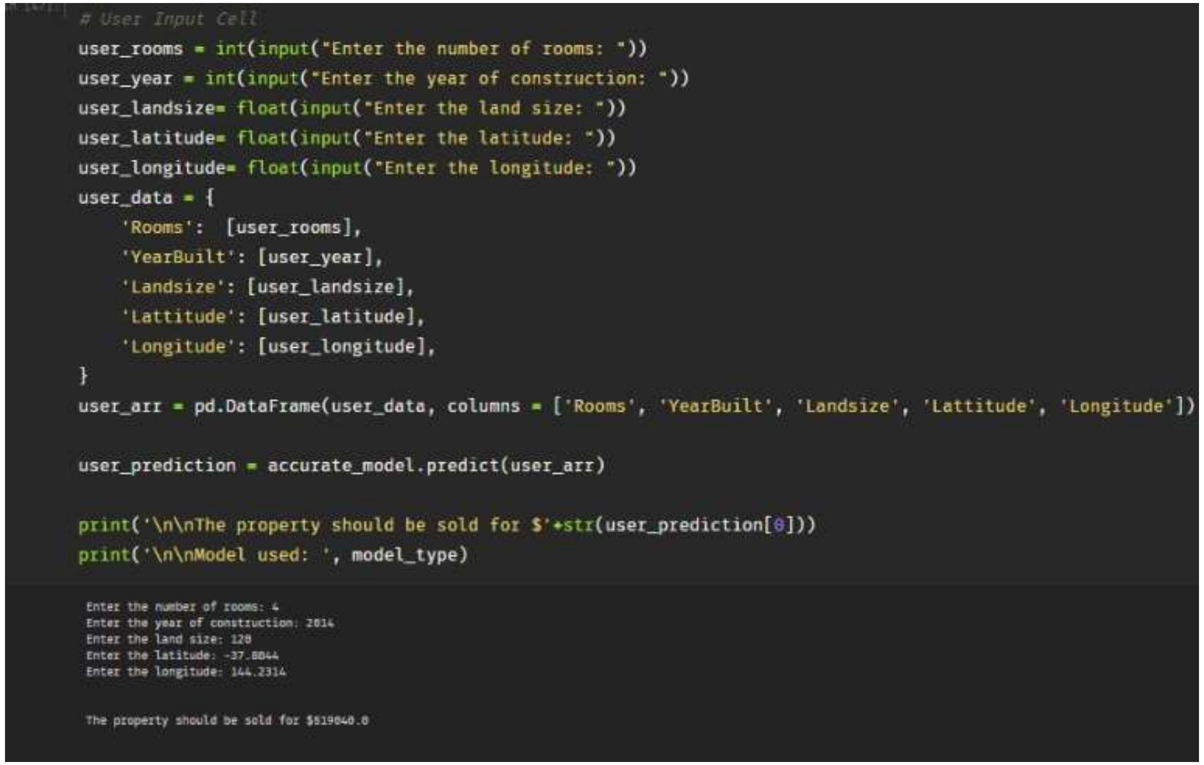


RESULTS



```
In [8]:
# Show trend of price for the years 1900 - 2020
plot_data = data.where(data['YearBuilt'] >= 1900)
y = plot_data.Price
X = plot_data.YearBuilt
sns.set(rc={'figure.figsize':(11.7,8.27)})
sns.lineplot(X, y)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x2dbe9a38488>
```



```
# User Input Cell
user_rooms = int(input("Enter the number of rooms: "))
user_year = int(input("Enter the year of construction: "))
user_landsize= float(input("Enter the land size: "))
user_latitude= float(input("Enter the latitude: "))
user_longitude= float(input("Enter the longitude: "))
user_data = {
    'Rooms':  [user_rooms],
    'YearBuilt': [user_year],
    'Landsize': [user_landsize],
    'Lattitude': [user_latitude],
    'Longitude': [user_longitude],
}
user_arr = pd.DataFrame(user_data, columns = ['Rooms', 'YearBuilt', 'Landsize', 'Lattitude', 'Longitude'])

user_prediction = accurate_model.predict(user_arr)

print('\n\nThe property should be sold for $'+str(user_prediction[0]))
print('\n\nModel used: ', model_type)
```

```
Enter the number of rooms: 4
Enter the year of construction: 2014
Enter the land size: 120
Enter the latitude: -37.8044
Enter the longitude: 144.2314


The property should be sold for $519040.0
```

CONCLUSION

- Project has been developed successfully.
- Throughout this project I have explored the possibility of integrating a recommender system by using Machine learning concepts.
- In this project I have used two training models decision tree and Random Forest to implement and recommend the price of the real-estate system by extracting the features from the dataset.

THANK YOU!

SANSKRUTI RAUT (ECE14)

FINAL YEAR ECE-C