



# INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 6.078

(Volume 6, Issue 3)

Available online at: [www.ijariit.com](http://www.ijariit.com)

## Smart E-commerce hub for real estate web application

Tanmay Kumar

[kumartanmay1998@gmail.com](mailto:kumartanmay1998@gmail.com)

Don Bosco Institute of Technology,  
Bangalore

Abhishek Kumar

[abshk6399@gmail.com](mailto:abshk6399@gmail.com)

Don Bosco Institute of Technology,  
Bangalore

Hardik Koul

[creatinggalaxy@gmail.com](mailto:creatinggalaxy@gmail.com)

Don Bosco Institute of Technology,  
Bangalore

Hemant Sharma

[hemantsmarts@gmail.com](mailto:hemantsmarts@gmail.com)

Don Bosco Institute of Technology,  
Bangalore

Dr. Ashok Kumar P. S.

[ashokdbit2017@gmail.com](mailto:ashokdbit2017@gmail.com)

Don Bosco Institute of Technology,  
Bangalore

### ABSTRACT

*Real estate is known as one of the most important sectors of the economy, hence Machine Learning Techniques outperforms a real estate professional when it comes to conducting demographic market research, environmental and financial analysis such as price predictions and loan approvals. The main objective of this paper is to find the best model to predict the prices and loan approvals and apply them to develop a smart E-Commerce application where users can search and inquire about a property, post their property Ads and can use the smart feature to predict the eligibility for their loan approvals using Machine Learning Model. The first thing that we have taken into account is the dataset of the apartment prices and different representative cases of loan applications. The dataset was pre-processed and tuned up for real analysis. Hence, our paper will also focus on data preprocessing of the raw dataset. Secondly after processing the data, we will review the use of Neural Network and Random Forest and the outcome that it generates. In addition, the proposed paper examines the use of the prediction system in real world settings applied into the smart real estate web application. The successful predictions made by the models will be a great asset for the real estate sectors and will provide real-life solutions to the problems that real estate investors face.*

**Keywords**— Real Estate, Machine Learning, Smart Web Application, AJAX, Neural Network, Random Forest, Dataset, Data Pre-processing

### 1. INTRODUCTION

In recent decades, whenever searching is done for a new house, the main focus is on the location with good basic services available from chosen location like displaying the distance of nearest bus stop, railway station, hospital etc. The application provides the user to select any particular location and get information appropriately along with the smart features to predict the loan approvals. Technology has always been

indicator of smartness. Real estate needs to improve its adoption of disruptive technologies to move from traditional to smart real estate. [1]

The main aim of this project is to design and develop Smart web application that can be used to keep track of all the properties in the country for the purpose of sell, share and rent along with all the relevant data related to the users in a Centralized database. The availability of website makes the process more user friendly and makes it more effective. The user can search the property by entering the keywords in the search tab after which it shows the property listings along with the listings nearby to those keywords. It is required for the user to register on the web application to avail all the features. Once after the successful registration user receives an email for the same. There are some important issues in developing the real estate web application [3]. First, the search time should be minimum. Second, the web application should give the services that both buyer and seller want. Third, the web application should have a friendly and understandable user interface.

The project is developed based on Django which is a high level Python Web framework encouraging rapid development and pragmatic, clean design. The goal here is to allow developers to instead of implementing the same solutions over and over again, focus on the parts of their application that are new and unique to their project. The project uses PostgreSQL as its database engine.

The Dynamic and Smart Web application uses different machine learning models such as Random Forest Regressor to predict the prices and Neural Network Classifier to predict the loan approvals based on the users data.

### 2. PROPOSED SYSTEM

As we know that, the Data security and Data accessing is very slow in the existing system and also the existing system does not perform well when there is a change in the operating

environment. The proposed system is a Smart E-Commerce application which is completely related to online system, which provides the centralized database.

The Web enabled application uses Neural Network to predict for the loan approvals based on the data given by the users. The system uses Tensor Flow, a python module which acts as a higher level API to build the model. A multi-layer feed-forward neural network with backpropagation learning algorithm was used to build up the proposed model.

The application also uses Random Forest Regressor to predict the apartment prices which is useful for the accurate predictions based on the model.

This application helps user to find the property listings based on their requirements and sends notification to the particular Realtors for the same. The Smart E-Commerce application also uses the AJAX technique to run the system dynamically and smoothly. Powerful software based on ML which can be designed to cater consumers desire to get to know whether a particular place could turn into their "Dream Home".

### **3. IMPLEMENTATION**

#### **3.1 Web Framework**

**3.1.1 Django:** Django is a free and open source web application framework written in Python. A framework is a collection of modules that make development easier. They are grouped together, and allow you to create applications or websites from an existing source, instead from scratch. It is a collection of python libraries allowing you to quickly and efficiently create quality web application which is suitable for both frontend and backend.

It offers a big collection of modules which you can use in your own projects. Mainly, frameworks exist to save developers a lot of wasted time and headaches and Django is no different. Additionally, the Django framework enables you to model your domain and code classes, and before you know it, you already have an ORM.

#### **3.2 Front End Implementation**

**3.2.1 HTML and CSS:** A HTML (the Hyper Text Markup Language) and CSS (Cascading Style Sheets) are two of the core technologies for building Web pages. HTML provides the *structure* of the page, CSS the (visual and aural) *layout*, for a variety of devices. Along with graphics and scripting. HTML and CSS are the basis of building Web page and Web Applications.

"Hypertext" refers to links that connect web pages to one another, either within a single website or between websites. Links are a fundamental aspect of the Web. By uploading content to the Internet and linking it to pages created by other people, you become an active participant in the World Wide Web.

CSS is the language for describing the presentation of Web pages, including colors, layout, and fonts. It allows one to adapt the presentation to different types of devices, such as large screens, small screens, or printers. CSS is independent of HTML and can be used with any XML-based markup language. The separation of HTML from CSS makes it easier to maintain sites, share style sheets across pages, and tailor pages to different environments. This is referred to as the separation of structure (or: content) from presentation.

**3.2.2 Javascript:** JavaScript is a dynamic computer programming language. It is lightweight and most commonly used as a part of web pages, whose implementations allow client-side script to interact with the user and make dynamic pages. It is an interpreted programming language with object-oriented capabilities.

**3.2.3 AJAX:** AJAX stands for **A**synchronous **J**avaScript and **X**ML. AJAX is a new technique for creating better, faster, and more interactive web applications with the help of XML, HTML, CSS, and Java Script. With AJAX, when you hit submit, JavaScript will make a request to the server, interpret the results, and update the current screen. In the purest sense, the user would never know that anything was even transmitted to the server.

### **3.3 Back End Implementation**

#### **3.3.1 Python**

A Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

Python can be used to build server-side web applications. While a web framework is not required to build web apps, it's rare that developers would not use existing open source libraries to speed up their progress in getting their application working.

#### **3.4 Machine Learning Models**

**3.4.1 Neural Network:** Neural networks are an emerging artificial intelligence technology that imitates the human brain on the computer. These techniques are based on the parallel, distributed processing design. Neural networks are an emerging artificial intelligence technology that imitates the human brain on the computer. These techniques are based on the parallel, distributed processing design. The study develops a proposed model that identifies artificial neural network as an enabling tool for evaluating credit applications to support loan decisions in the commercial banks.

**3.4.2 Random Forest:** Random forest algorithm is being used for the apartment price prediction. Since it has been termed as one of the easiest to use and flexible machine learning algorithm, it gives good accuracy in the prediction.

This is usually used in the classification tasks. Because of the high volatility in the apartment prices, the task of predicting is quite challenging. In price prediction we are using random forest classifier which has the same hyperparameters as of a decision tree. The decision tool has a model similar to that of a tree. It takes the decision based on possible consequences, which includes variables like event outcome, resource cost, and utility.

The random forest algorithm represents an algorithm where it randomly selects different observations and features to build several decision tree and then takes the aggregate of the several decision trees outcomes. The data is split into partitions based on the questions on a label or an attribute.

### **4. SYSTEM DESIGN AND FLOW CHARTS**

There is an online community for data analysis and predictive modeling known as Kaggle. It also contains dataset of different fields, which is contributed by data miners. It allows the users to use their datasets so that they can build models and work

with various data science engineers to solve various real-life data science challenges. The dataset used in the proposed project has been downloaded from Kaggle. Which is present in what we call raw format. The first step is the conversion of this raw data into processed data which is done by using feature extraction and is followed by a classification process. The training data set is used to train the model whereas the test data is used to predict the accuracy of the model.

Once the model is trained then it is pickled and applied into the Smart web application to work accordingly. The Model is then used to predict for the loan approvals and for price predictions as per users HTTP request which return the result with a HTTP response.

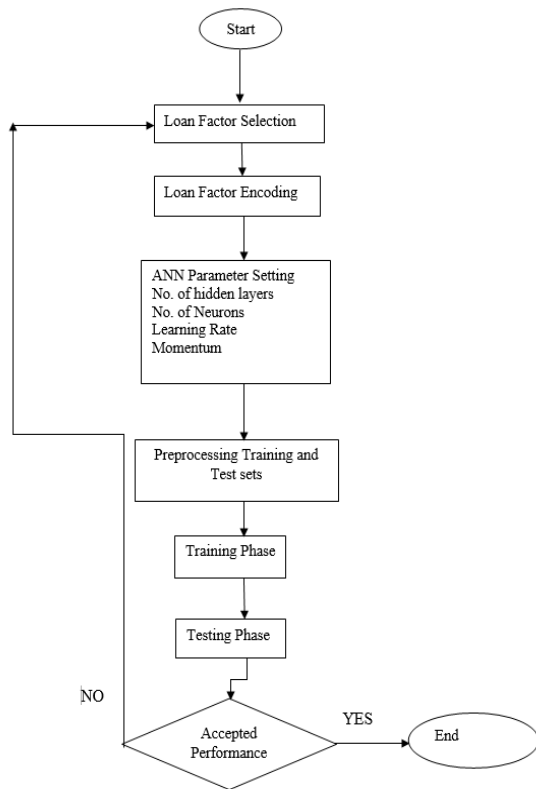


Fig. 1: Flow chart of Neural Network model

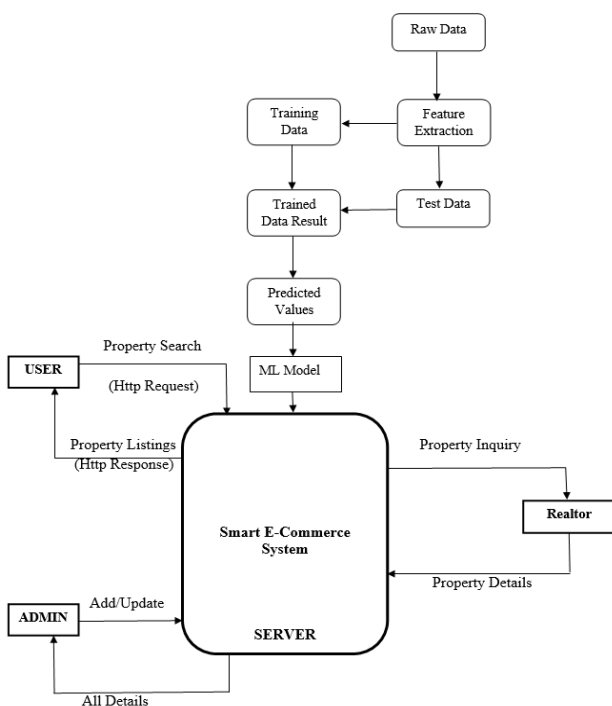


Fig. 2: System Architecture

## 5. EXPERIMENTAL ANALYSIS AND RESULTS

The dataset file consists of raw data based on which we are going to analyze the models. The first dataset that we took is of Loan status and the another is of Apartment prices. We found that Neural Network best suits for the Loan Approval Model whereas Random Forest best suits in predicting the accuracy of apartment prices.

### 5.1 Neural Network using Tensor Flow for loan model

There are 13 columns or 13 attributes that describes the Loan Status of a person. Some of them are Loan\_Id, Gender, Married, Dependents, Education, Self\_Employed, Applicant\_Income, CoapplicantIncome, LoanAmount, Loan\_Amount\_Term, Credit\_History, Property\_Area, Loan\_Status. We dropped off the first column i.e Loan\_Id which doesn't play crucial role in predicting the status.

A	B	C	D	E	F	G	H	I	J	K	L	M	
1	Loan_Id	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
2	LPO01002	Male	No	0	Graduate	No	5849	0	360	360	1	Urban	Y
3	LPO01003	Male	Yes	1	Graduate	No	4583	1508	128	360	1	Rural	N
4	LPO01005	Male	Yes	0	Graduate	Yes	3000	0	66	360	1	Urban	Y
5	LPO01006	Male	Yes	0	Not Graduate	No	2583	2358	120	360	1	Urban	Y
6	LPO01008	Male	No	0	Graduate	No	6000	0	141	360	1	Urban	Y
7	LPO01011	Male	Yes	2	Graduate	Yes	5417	4196	267	360	1	Urban	Y
8	LPO01013	Male	Yes	0	Not Graduate	No	2333	1516	95	360	1	Urban	Y
9	LPO01014	Male	Yes	3	Graduate	No	3036	2504	158	360	0	Semiurban	N
10	LPO01018	Male	Yes	2	Graduate	No	4006	1526	168	360	1	Urban	Y
11	LPO01020	Male	Yes	1	Graduate	No	12841	10968	349	360	1	Semiurban	N
12	LPO01024	Male	Yes	2	Graduate	No	3200	700	70	360	1	Urban	Y
13	LPO01027	Male	Yes	2	Graduate	No	2500	1840	109	360	1	Urban	Y

Fig. 3: Raw Data of Bank Loan

After that we used One Hot Encoder method to encode the Categorical values or features. This encodes the categorical feature as a one-hot numeric array This creates a binary column for each category and return a sparse matrix or dense array.

Gender_Female	Gender_Male	Married_No	Married_Yes	Education_Graduate	Education_Not Graduate	Self_Employed_No	Self_Employed_Yes	Property_Area_Rural	Property_Area_Semi
0	1	0	1	1	0	1	0	1	0
0	1	0	1	1	0	0	1	0	0
0	1	0	1	0	1	1	0	0	0
0	1	1	0	1	0	1	0	0	0
0	1	0	1	1	0	0	1	0	0
0	1	0	1	0	1	1	0	0	0
0	1	0	1	1	0	1	0	0	0
0	1	0	1	1	0	1	0	0	0
0	1	0	1	1	0	1	0	0	0

Fig. 4: One Hot Encoded Data

The method was applied on the columns Gender, Married, Education, Self\_Employed and Property\_Area which contains categorical data.

Parameters	Values
Number of Training Data	615
Number of Testing Data	369
Number of Hidden Neurons	200 (Condensed to 1)
Number of Hidden Layer	4
Learning Rate	0.2
Epochs	100
Momentum	0.9
Threshold	0.05

Fig. 5: The Optimal Set of Parameter

The training stopped as it gives us the best accuracy for the above dataset. The accuracy classified is between 86-90% was

obtained using Neural Network Algorithm as compared to SVM model which yielded 81% of accuracy for the same.

apartment prices in a particular area.

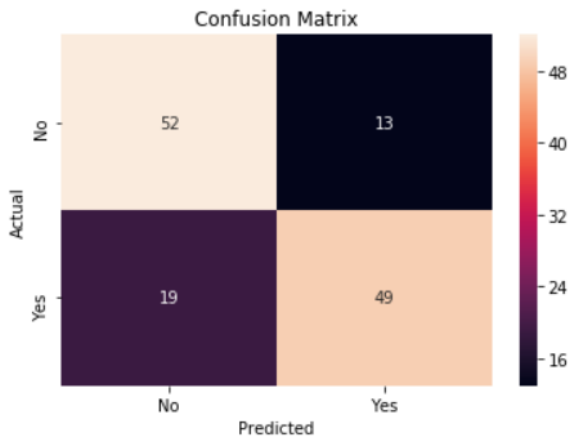


Fig. 6: Confusion Matrix

The above given Fig. 6 shows Confusion Matrix states that: When it was Predicted as NO it was Actually correct 52 times. When it was Predicted as YES it was Actually correct 49 times. When it was Predicted as NO it was Actually YES 19 times. When it was Predicted as YES it was Actually NO correct 13 times.

5.2 Random Forest for Price Prediction

There are 9 columns or 9 attributes that describes the different apartments data. Some of them are area\_type, availability, location, size, society, total\_sqft, bath, balcony, price.

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Comee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00

Fig. 7: Raw Dataset of Apartments

After splitting training and testing data we have used Label Encoder and after changing the categorical features into numeric array the data is as shown below in Fig 8.

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	0	2	430	14	487	1056.0	2.0	1.0	39.07
1	2	0	325	20	2569	2600.0	5.0	3.0	120.00
2	1	0	1220	17	1528	1440.0	2.0	3.0	62.00
3	0	0	778	17	2303	1521.0	3.0	1.0	95.00
4	0	0	736	14	1528	1200.0	2.0	1.0	51.00

Fig. 8: Encoded Training Data using Label Encoder

After that we moved to Preprocess the columns and we found that there are many missing values in different columns such as 'bath' and 'balcony'.

So, the missing values are filled by grouping the rows based on location and taking the mean of the column 'bath' and 'balcony' in that location. Even after doing this, there is a missing value. This is because there is a row in which the location is unique (occurred only one time) and the value is NaN. So, it cannot fill as there are no other values for bathrooms to find mean. In this case, the missing values are filled with the mean of the whole column. In the next step we dropped off the price column from test data so that we can use the testing data on trained data result to predict the prices. The model was trained using Random Forest Regressor which is best suited to predict the

5.3 Pickling the model to incorporate into the Web Application

Once we get the trained data result of both the algorithms, we pickled the model using python's pickle module. Python pickle module is used for serializing and de-serializing a Python object structure. Any object in Python can be pickled so that it can be saved on disk. What pickle does is that it "serializes" the object first before writing it to file. Pickling is a way to convert a python object (list, dict, etc.) into a character stream. The idea is that this character stream contains all the information necessary to reconstruct the object in another python script.

The pickled model is then loaded into the web application in the Django framework to iterate over the data given by the user and to predict the values.

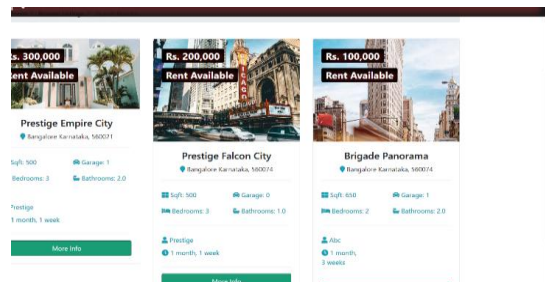


Fig. 9: Smart Web app showing the results based on user's Search

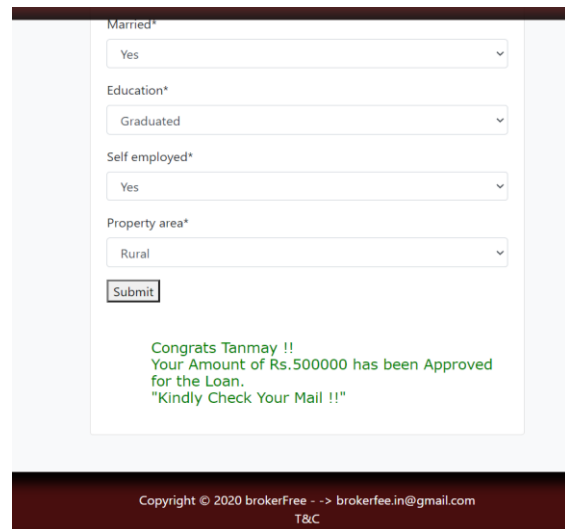


Fig. 10: An Approved Loan Status showing to a user

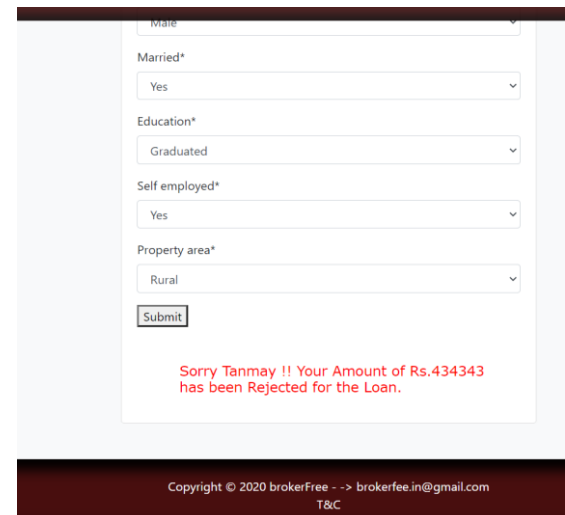


Fig. 11: A Rejected Loan Status showing to a user



## **6. CONCLUSION**

Smart E-Commerce application is user-friendly and flexible which will eventually save the time of the user in searching property listings and get right business solutions. By measuring the accuracy of the different algorithms such as SVM and Neural Networks, we found that the most suitable algorithm for predicting the Loan Approvals based on various data points from the historical data is the Neural Network Algorithm. The algorithm will be a great asset for the associated banks and users since it is trained on a huge collection of historical data and has been chosen after being tested on a sample data, whereas the project uses Random Forest Regressor to predict the accurate prices of apartments, which in turn will be a great asset to the Real estate owners to set the apartment price accurately to get right business solutions.

The project demonstrates the machine learning model to predict the loan approvals and price predictions with more accuracy as compared to previously implemented machine learning models.

## **7. FUTURE ENHANCEMENT**

Future scope of this project will involve adding more parameters and factors like the financial ratios, multiple instances, etc. The more the parameters are taken into account more will be the accuracy. The algorithms can also be applied for analyzing the contents of public comments and thus determine patterns/relationships between the customer and the

corporate employee. The use of traditional algorithms and data mining techniques can also help predict the corporation's performance structure as a whole.

## **8. ACKNOWLEDGMENT**

We extend our sincere thanks to our project guide Dr. Ashok Kumar P.S and our Computer Science and Engineering Dept. for their encouragement. We would also like to thank our institution Don Bosco Institute of Technology.

## **9. REFERENCES**

- [1] Sabina-Cristiana Necula, Vasile-Daniel Pavaloaia, Caralin Strimbei and Octavian. "Enhancement of E-Commerce Websites with Semantic Web Technologies", MDPI 2018.
- [2] Jitendra Mahajan, Ganesh Mahajan, Vivek bhuvad, Prof Reena Chaudhari "Real Estate Management System", IJARCCCE 2016.
- [3] Xu Wei, Lu Guiying, Jade Ling Qing Xiong Yanfu, "the new e-business models in the modern enterprise applications – based on the example of Guangxi Sugar Network Social Survey". Coastal Enterprises and Science & Technology, 2008 (09).
- [4] Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, K Vikas" Loan Prediction by using Machine Learning Models", IJET 2019.
- [5] Joylin Zeffora.A, R. Shobarani " Statistical Analysis of Random Forest on Real Estate Prediction", IJITEE 2019.