International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Nexus Landholdings: Predicting Indian Real Estate Prices

Shubhangi Sharma[1], Hemant Dhawan[2], Vaibhav Arora[3], Dr. SC Gupta[4]

[1]*Computer Science and Engineering Department, Panipat Institute of Engineering and Technology*

[2]*Computer Science and Engineering Department, Panipat Institute of Engineering and Technology*

[3]*Computer Science and Engineering Department, Panipat Institute of Engineering and Technology*

[4]*Computer Science and Engineering Department, Panipat Institute of Engineering and Technology*

---***---

Abstract – Real Estate holds great significance as it is the basic building block of society; the same society that has adopted technology at such a fast pace that they seem intertwined to the point of no return. But the same cannot be said for real estate because a substantial quantum of work still happens manually, without employing the emerging technologies. Previous research shows that bringing the emerging technologies to this field could revolutionize it, thus increasing the satisfaction for all the stakeholders of real estate, but it also brings to light the fact that some websites that do exist for real estate fail to provide proper information and leave the consumers regretful after purchase. This research project aims to provide a solution for this issue by creating a website that employs machine learning to provide the predicted prices for real estate properties in Bangalore, India. This is implemented by drawing on the previous research that provides a Real Estate Technology Acceptance Model, a part of which is practically realized in this project.

Key Words – Machine Learning, RESTAM, Technology Acceptance Model, Software-as-a-Service (SaaS), Price Prediction, Real Estate Price Prediction.

1. INTRODUCTION

Technology has seen its uses in most of the big fields. However, Indian Real Estate is still largely lagging in the use of technology, especially the emerging technology, for bringing the customers in an advanced, more transparent world. This should not be the case because real estate is one of the most important pillars of human society. It represents home, wealth and prestige, and is the place where human societies are formed. A few search keywords away are thousands of aphorisms and words of wisdom about financial strength and investments, a huge number of which are dependent on Real Estate. To exemplify, a Hebrew Proverb says, “He is not a full man that does not own a piece of land.” A renowned American Industrialist, Andrew Carnegie, quoted, “Ninety percent of all millionaires become so through owning Real Estate.” Even the famous English writer and poet, William Shakespeare, added to the plethora of such statements by saying, “I would give a thousand furlongs of sea for an acre of barren ground.” Such statements make the significance of Real Estate quite conspicuous. Thereby, it should be included in the technical revolution that is happening around the world. This research project aims to bring technology closer to Real Estate by creating a web application for the stakeholders of real estate. The aim of this research is twofold:

1) Working on the implementation of the SAAS (Software as a Service) aspect as elaborated in the RESTAM (Real Estate Stakeholders Technology Acceptance Model) given by [1].
2) Working on the big data aspect of RESTAM by bringing Machine Learning into the field of Real Estate to bring more customer satisfaction.

2. Literature Review

Globally, Real Estate is subjected to large amounts of financial investments. According to latest research [2] investments in real estate technologies have shown constant growth since 2015 (when the Global Real Estate Value was at $217 Trillion [3]) , and have increased four times in the span of 2015 to 2019. The largest growth was seen from 2018 to 2019, when the fundings grew by an amount of 65%. Only due to COVID-19 were there losses observed in 2020. This is shown in the graph as follows:

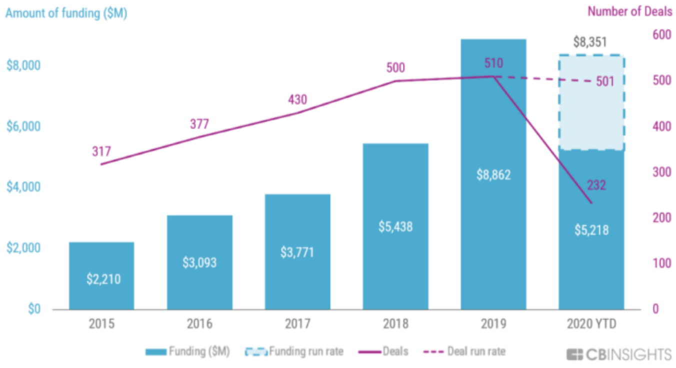


Fig 1: Growth Trends in Fundings to Real Estate Tech Companies [2]

Real Estate Technology refers to the software tools and software platforms which are being used by the players in Real Estate Industry, like investors, managers, brokers, mortgage lenders and property owners [3]. Technology has advanced to a great level in other fields, but many researchers claim that either the technology is not well adopted into the Real Estate Industry [4], or it is not giving the consumers much satisfaction where it is adopted [5] [6] [7]. Hence, the two aspects of this research come into play.

2.1. Software as a Service (SaaS)

In the times bygone when software development was still a novelty, software products were distributed to the customers/users as standalone entities which needed to be set up on their personal computers [8]. With the advent of internet, this started to change and now a new concept known as SaaS has materialized. It means giving the software as a service to the customers, without needing to set it up for every individual. Hence, many large businesses are shifting towards this concept by providing software capabilities as services through the use of web based communication. For consumers, this is a great advantage because everyone can use the internet. There is no cost of maintenance at the user side and all bugs can be resolved by the technical team on the admin side [9].

The adoption of SaaS is increasing in every industry by the day, especially after the coronavirus outbreak. The world is rapidly shifting onto a digital platform with the SaaS market witnessing a growth of 18% per annum [10]. Be that as it may, real estate is still mostly trapped in the confines of traditional methodologies and there is a great need for it to adapt to the changing technological world.

However, adopting a SaaS business model does not guarantee the success of a business or a project, and this fact is more pronounced for the real estate market. There are various parameters that dictate the desirability and productivity of a website and User is at the core of all those parameters [4]. Various versions of Technology Adoption Models (TAM) have been constructed (that include technologies like Website Management and SaaS) to conceptually define how the different factors influence the adoption of these technologies into a field, thus making it successful [4] [6]. There have been many literature mentions of the various characteristics that make a real estate website more efficient and friendly for the user, like user experience, design of the website, speed of loading the contents, quality of service provided, customer support, information accuracy etc. [11] [12] [13] [14] [15].

2.2. Big Data and Machine Learning

Machine Learning (ML) is one of the most widely used emerging technologies which could bring a plethora of new possibilities into the world of real estate. It is a part of Artificial Intelligence that depends on data to make predictions or to create classifications [16]. Conway defines Machine Learning as an intersection of Hacking Skills (which means algorithmic thought process) and Math and Statistics Knowledge [17]. Machine Learning is an umbrella of many algorithms that are broadly classified into 3 categories:

- **Supervised Learning Algorithms:** These are the most commonly used algorithms for the application of property price prediction [18]. The training in these algorithms happens through a ‘teacher.’ The output that should be predicted or classified is already known (called labels) and the model is trained based on that [19]. These algorithms have two subcategories, one for regression problems and one for classification problems. Price Prediction in this research is taken as a Regression Problem.
- **Unsupervised Learning Algorithms:** These are less commonly used and do not have the concept of a ‘teacher.’ They are generally faster but are also more error prone than supervised learning algorithms [19]. They are used for finding the insights that are not previously known from the data.
- **Reinforcement Learning Algorithms:** These algorithms have a reward-and-punishment system. It also has a teacher which gives a reward if the output is correct and punishment if not.

The literature has a shortage of research that focuses on price prediction in India. [20] compared the Hedonic Price Model (HPM) with Artificial Neural Networks (ANN) to find out which works best for price prediction. Hedonic Price Theory states that a property can be thought of as an amalgam of various sub-constituents like number of bedrooms, bathrooms, parking, etc. which ultimately become the determinants of the implicit price that a property should have. ANN is a network of artificial computational neurons that are ideated from biological neural networks. Limsombunchai found that ANN trained by trial-and-error strategy was a better approach in predicting real estate prices as opposed to HPM.

[18] did a literature review to find the most used models for the scope of house price prediction. They brought to light the fact that Valuation Method is used for determining the market value. Their review showed that the five most used algorithms are (by ranking) Random Forests, Gradient Boosting, Support Vector Machine, Decision Tree and Linear Regression.

[21] pointed out that although research has been done in price prediction for real estate, there have not been any concrete solutions to the over-valuation and under-valuation issues. In the Boston Dataset, Random Forest turned out to be the best algorithm, closely followed by Decision Trees. Other algorithms that worked were Ridge and Linear Regression. However, their research used two datasets to show that Extreme Gradient Boosting gave better accuracy than Random Forest, followed by Decision Trees.

[22] stated that Random Forest and Gradient Boosting works best for price prediction. Random Forest was further backed by [23], also claiming that Indian Real Estate is lagging behind the US and European Real Estate. [24] performed their research to compare Linear Regression, Lasso Regression and

Decision Trees for price prediction, out of which Linear Regression gave the best accuracy.

3. Proposal and Hypothesis

A variation of the traditional TAM, called RESTAM, has been provided by [1], which defines 9 big technologies that need to be adopted in the real estate world. These technologies are Drones, IoT, clouds, SaaS, Big Data, 3D Scanning, Wearable Technologies, AR and VR, Artificial Intelligence and Robotics. As a solution to the problems mentioned in the literature review, we propose to make a website that would keep User Satisfaction as the central point while development. The website would be the implementation of SaaS as suggested by the RESTAM to bring more transparency for the consumers and to reduce post-purchase regrets.

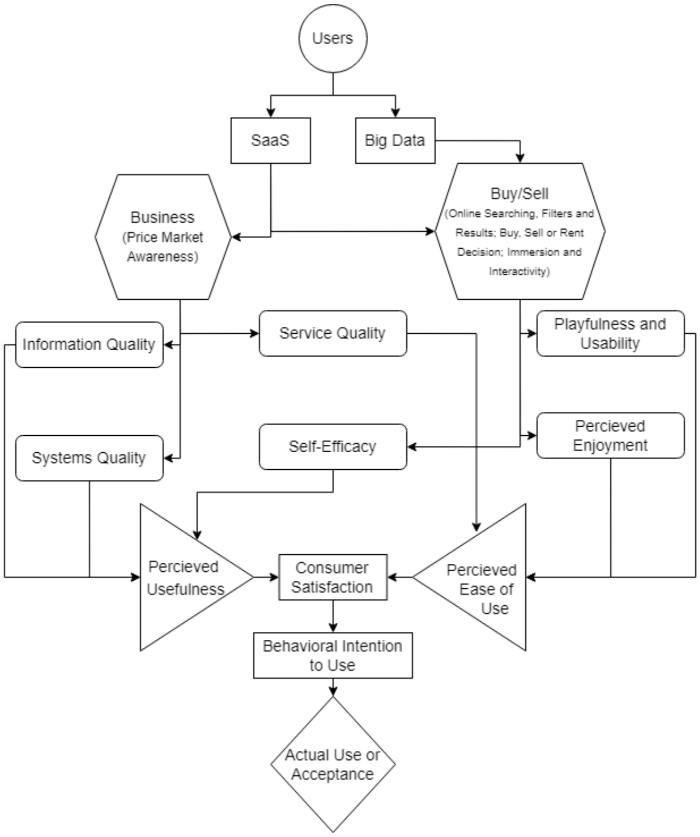


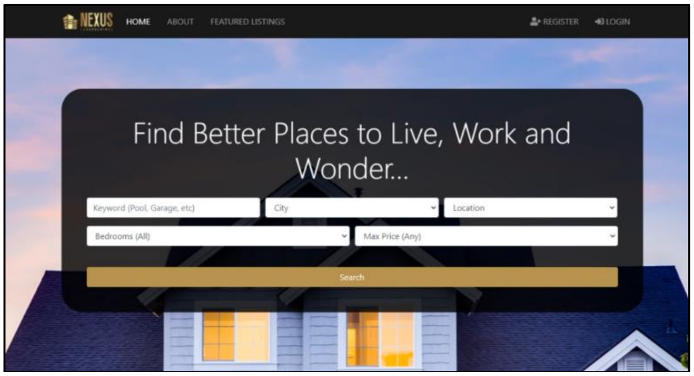
Fig 2: The aspects of RESTAM for our website

We will increase the transparency through machine learning, by training a model and using it for prediction of property prices, which will be displayed along with the property listing so that the customer can get an idea of what the price should be. This will solve the problem of overpricing. Moreover, the website would have a section for prediction based on parameters that would be entered by the seller in order to know the approximate price that should be set for a property. This would solve the problem of underpricing.

We will compare the accuracy of various popular models that are mentioned in the literature review and test if the mostly used models provide better accuracy. We propose the hypothesis that a custom ensemble of two or more machine learning models would provide better accuracy than a single model, and we will try to find such ensemble.

4. Implementation

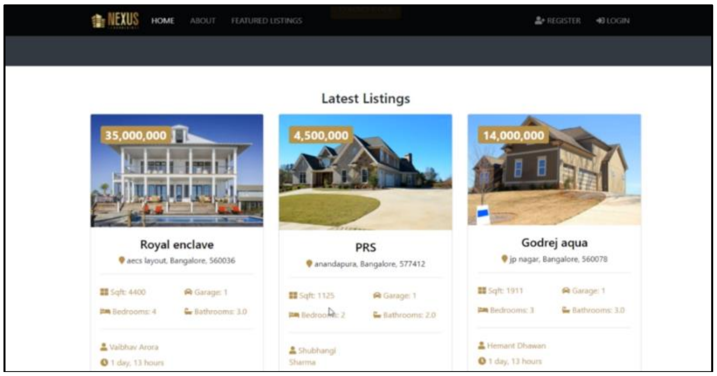
The website is called Nexus Landholdings and is created through Python and Django Framework and use the Heroku platform for SaaS implementation through live hosting. The user interface is easy to use and is created by keeping in mind all the features mentioned in the diagram given in the previous section, in essence, online searching, filtering the search, immersion, interactivity etc. Some of the screenshots of the website are:



Find Better Places to Live, Work and Wonder...

Keyword (Pool, Garage, etc)
City
Location
Bedrooms (All)
Max Price (Any)

Fig 3: Home Page of Website



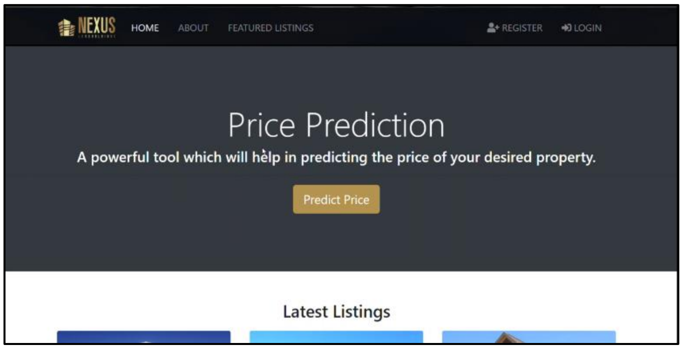
Latest Listings

35,000,000
Royal enclave
aecs layout, Bangalore, 560036
Sqft: 4400 Garage: 1
Bedrooms: 4 Bathrooms: 3.0
Vaibhav Arora
1 day, 13 hours

4,500,000
PRS
anandapura, Bangalore, 577412
Sqft: 1125 Garage: 1
Bedrooms: 2 Bathrooms: 2.0
Shubhangi Sharma

14,000,000
Godrej aqua
jp nagar, Bangalore, 560078
Sqft: 1911 Garage: 1
Bedrooms: 3 Bathrooms: 3.0
Hemant Dhawan
1 day, 13 hours

Fig 4: Home Page of Website



Price Prediction

A powerful tool which will help in predicting the price of your desired property.

Fig 5: Link to the Price Prediction from Home Page

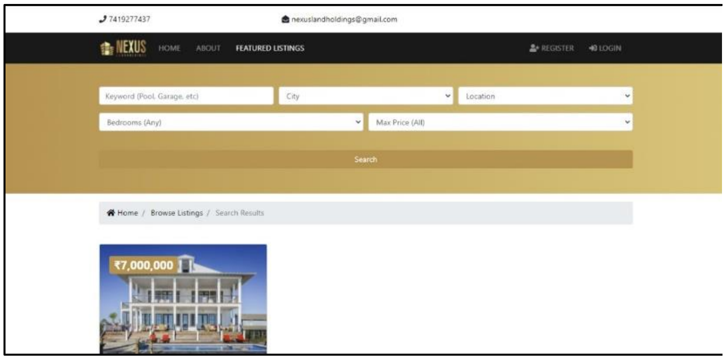
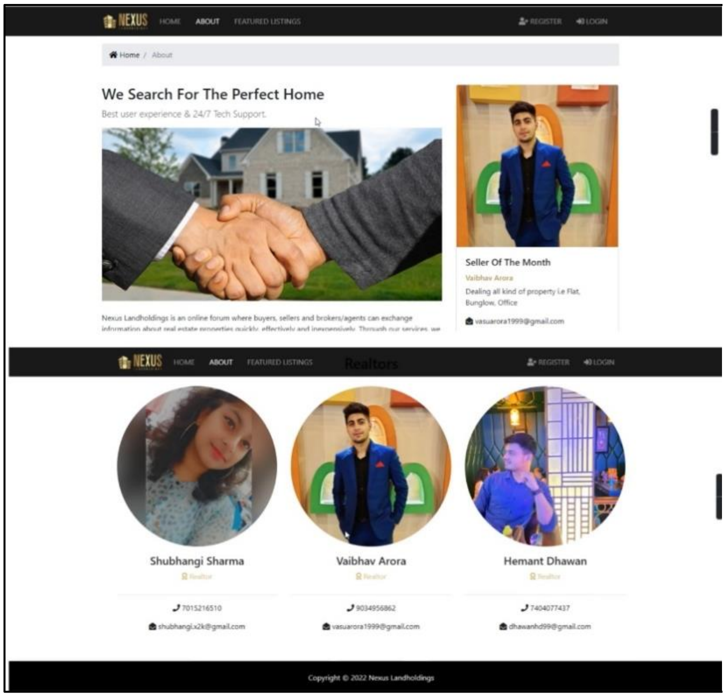


Fig 6: Results from Search



We Search For The Perfect Home

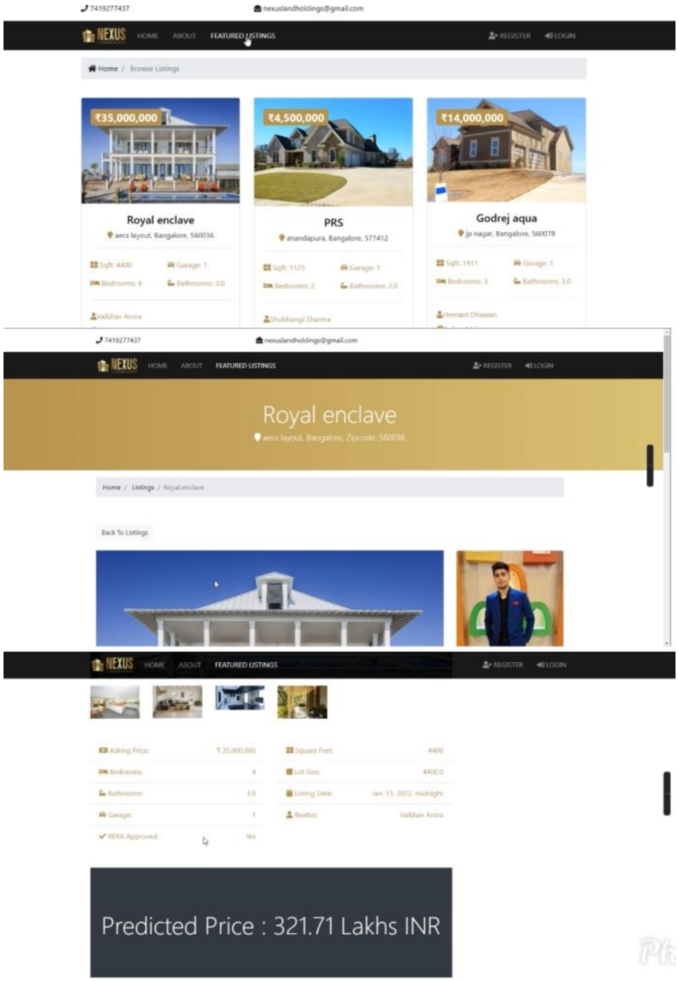
Best user experience & 24/7 Tech Support.

Dealing all kind of property i.e Flat, Bunglow, Office

vasuarora1999@gmail.com

Nexus Landholdings is an online forum where buyers, sellers and brokers/agents can exchange information about real estate properties quickly, effectively and inexpensively. Through our services, we

Fig 7: About Page of the Website Displaying all the Sellers and the Best Seller



Asking Price:	₹ 35,000,000	Square Feet:	4400
Bedrooms:	4	Lot Size:	4400.0
Bathrooms:	3.0	Listing Date:	Jan. 13, 2022, midnight
Garage:	1	Realtor:	Vaibhav Arora
RERA Approved:	Yes		

Predicted Price : 321.71 Lakhs INR

Fig 8: Featured Listings along with Predicted Price for the selected property

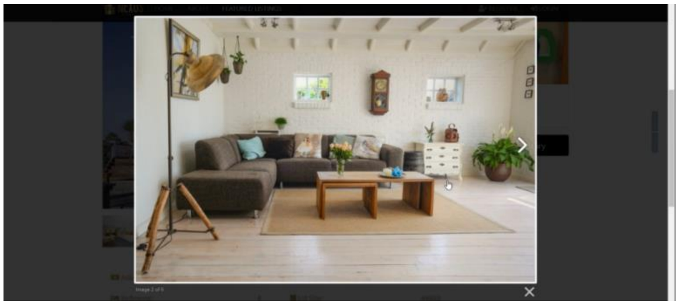
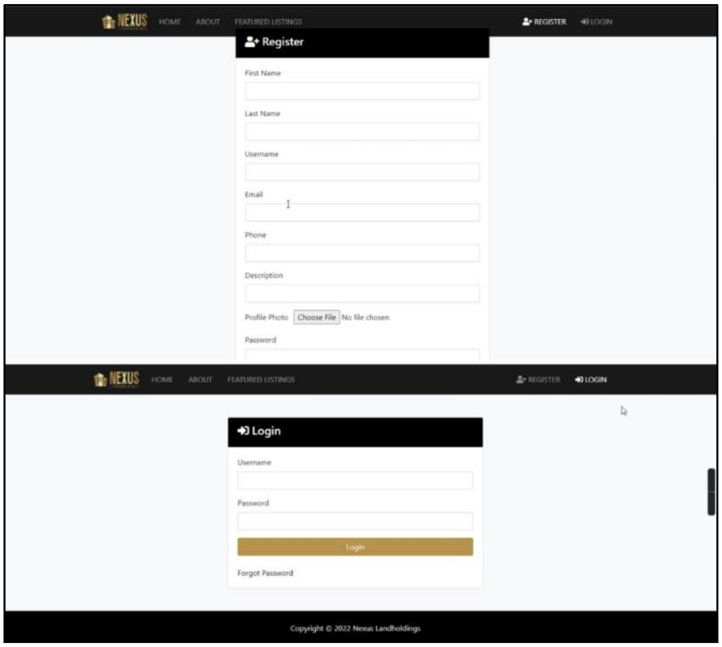


Fig 9: Image Gallery for the selected property



First Name

Last Name

Username

Email

Phone

Description

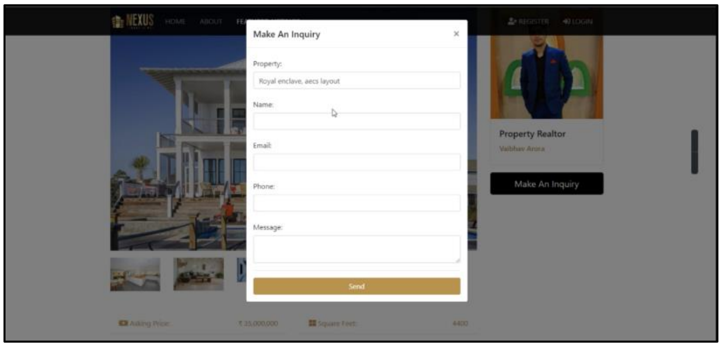
Profile Photo No file chosen

Password

Username

Password

Fig 10: Register and Login Pages



Property: Royal enclave, aecs layout

Name:

Email:

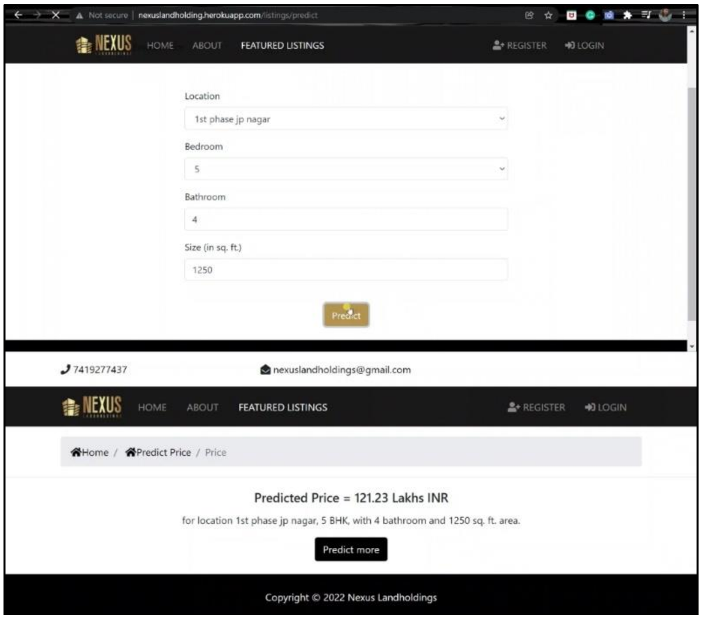
Phone:

Message:

Vaibhav Arora

Asking Price: ₹ 35,000,000 Square Feet: 4400

Fig 11: Contact Form for Making Inquiries from the Seller about the Property


Not secure | nexuslandholding.herokuapp.com/listings/predict
NEXUS
HOME
ABOUT
FEATURED LISTINGS
REGISTER
LOGIN
Predict
Home / Predict Price / Price
Predict more
Copyright © 2022 Nexus Landholdings

Location: 1st phase jp nagar

Bedroom: 5

Bathroom: 4

Size (in sq. ft.): 1250

7419277437 nexuslandholdings@gmail.com

Predicted Price = 121.23 Lakhs INR

for location 1st phase jp nagar, 5 BHK, with 4 bathroom and 1250 sq. ft. area.

Fig 12: Price Prediction through ML (after clicking the Predict Price button on Homepage)

The different features of our website are elaborated as follows:

- **Homepage:** The Nexus Landholdings Homepage (Figure 1) displays the search area in which the various parameters can be filled to obtain Search Results (Figure 6). It also contains the Latest Listings that have been added by the realtors of the website (Figure 4).
- **About Page:** This page contains the realtors that have registered on our website and have added listings (Figure 7). It also contains the best seller of the month which will be selected by the admin.
- **Inquiry Form:** To inquire about a property or to make purchases, the buyer can contact the seller by filling out the enquiry form (Figure 11).
- **Featured Listings:** The popular properties will be displayed on the featured listings page. Users can click on those listings to get more information for that property, the predicted price of that property and the seller details also (Figure 8) Users can also click on the image of that property to get an image gallery of that property (Figure 9).
- **Price Prediction:** Users can go on the "Predict Price" Section (Figure 5) of the website to open the price prediction page (Figure 12). It is not necessary for a property to be listed on our website to give the predicted price. Users can just enter the parameters to predict the price. The values of those parameters will be fed into a trained custom ensemble machine learning model to predict the prices. The method of model creation and training is mentioned as follows.

The dataset is open-source, taken from Kaggle.com and contains data of Bangalore Properties. It has the following features:

- **Area-Type:** Nominal Variable with 4 distinct values – Plot Area, Built-up Area, Super Built-up Area, Carpet Area.
- **Availability**: When will the property be available for buying
- **Location**: Nominal variable denoting the location of the property
- **Size**: Ordinal variable to denote the size of the property in terms of BHK.
- **Society**: Nominal variable for name of society in the given location
- **Total_sqft**: Quantitative variable for area of the property in square feet
- **Bath**: Quantitative variable for number of bathrooms in the property
- **Balcony**: Quantitative variable for number of balconies in the property
- **Price**: Quantitative variable for price of the property in lakhs INR

Data Exploration, Data Cleaning and Feature Engineering was performed on this dataset. Thereafter, 4 variables were selected as independent variables, namely location, total_sqft, bath and bhk (which is a new variable derived from size and contains the numerical values of bhk). Price was the dependent variable that needed to be predicted. The sample size was 7239 after data preprocessing. Of all the independent variables selected, location was nominal and hence was subjected to one hot encoding. The dimensions of training matrix X were [7239 x 243] and training vector y was [7239 x 1].

The most commonly used algorithms for price prediction as found during literature review were:

1) Random Forest (RF)
2) Gradient Boosting (GB)
3) Decision Tree (DT)
4) Linear Regression (LR)
5) Extreme Gradient Boosting (XGB)
6) Support Vector Machine (SVM)

So these models were trained individually at first and then ensembles were made for different combinations of these models to test the accuracy. The training and testing data was divided using Repeated K-Fold technique so that cross validation could be applied to test the accuracy. The K-value was kept at 10 and repetitions were 3, which implies that 30 accuracy scores were obtained for different parts of training data. The mean of these 30 scores was calculated to find the average accuracy of the models. The highest and lowest accuracy scores were also compared for these models.

The accuracy scores of different models are summarized in the following table:

Table 1: Cross Validation Accuracy Scores for 30 tests

MODEL	BEST ACCURACY SCORE	AVERAGE ACCURACY SCORE	WORST ACCURACY SCORE
SINGLE MODELS			
Linear Regression	0.9028	0.8415	0.7536
Gradient Boosting	0.9009	0.7732	0.3618
Random Forest	0.8762	0.7756	0.5586
Decision Tree	0.8855	0.6892	0.1976
Extreme Gradient Boosting	0.9260	0.8005	0.3111
Support Vector Machine	0.6939	0.5687	0.2609
ENSEMBLE MODELS			
Linear Regression + Random Forest	0.9298	0.8436	0.6890
Linear Regression + Decision Tree	0.9193	0.8238	0.6479
Linear Regression + Gradient Boosting	0.9317	0.8354	0.6931
Linear Regression + Extreme Gradient Boosting	0.9299	0.8482	0.6927
Random Forest + Decision Tree	0.8980	0.7476	0.4170
Linear Regression + Random Forest + Gradient Boosting	0.9292	0.8322	0.6674
Linear Regression + Random Forest + Decision Tree	0.9175	0.8184	0.6497
Random Forest + Decision Tree + Gradient Boosting	0.9038	0.7782	0.4420

Linear Regression + Random Forest + Extreme Gradient Boosting	0.9324	0.8389	0.6602

5. Results

A cursory glance on the accuracy table elucidates that the most commonly stated algorithm in the literature review, in essence, Random Forest, did not perform very well with this dataset. Gradient Boosting and Extreme Gradient Boosting also fell short in precision. The only stand-alone algorithm that worked well with this dataset is Linear Regression which gave an accuracy of 84.1%. Out of the 30 tests performed, the best accuracy of linear regression was 90.3% and worst was 75.4%.

Another observation that can be derived from the table is that some algorithms, like XGB, RF, GB and DT did not give as much accuracy alone, but when they were combined in ensembles with other algorithms, their accuracy improved greatly. This is the first proof that validates our hypothesis.

In the ensemble of different algorithms:

- A combination of LR and RF gave an average accuracy score that was just 0.2% better than LR. Even though the best accuracy for this combination was 3% higher than that of LR, the worst accuracy score was 6% lower than that of LR.
- Combination of LR and XGB gave an average accuracy of approximately 85%, which is 1 percent more than that of LR, despite the worst accuracy score being lower than that of LR. The best accuracy score in this case was also 3% higher as compared to that of LR.
- The combination of LR, RF and XGB gave the highest accuracy (93.2%) for the best case, but on the other hand, it also gave the lowest accuracy for worst case (66%), thus decreasing its average accuracy to 83.8%. It is close to that of LR but not quite the same.
- The combinations of [LR + RF + XGB], [LR + RF + GB], and [LR + GB] remain in close competition with almost similar values for best and average accuracy scores. However, the worst accuracy score of [LR + GB] is approximately 3% higher than the former two combinations, thus making it better.

6. CONCLUSION

From all the models tested, the best accuracy was given by the ensemble model that was constituted from Linear Regression and Extreme Gradient Boosting Algorithm, thus proving our hypothesis correct. Hence, we trained this model and deployed it on the website which we created by keeping in mind the various SaaS characteristics as given in the RESTAM. Moreover, Random Forest, which is the most popular algorithm did not work well with this data. This shows that the algorithm to use for price prediction really depends on the dataset.

The ensemble combination mentioned above is followed in accuracy value by Linear Regression, and a combination of Linear Regression and Gradient Boosting.

7. REFERENCES

[1] F. Ullah, S. Sepasgozar and T. H. Ali, "Real Estate Stakeholders Technology Acceptance Model (RESTAM): User-focused Big9 Disruptive Technologies for Smart Real Estate Management," in *International Conference on Sustainable Development in Civil Engineering*, MUET, Pakistan, 2019.

[2] CBINSIGHTS, "Research Briefs," 19 August 2020. [Online]. Available: https://www.cbinsights.com/research/real-estate-tech-funding-trends/.

[3] CBINSIGHTS, "RESEARCH WEBINAR: The State of Real Estate Tech," 23 December 2018. [Online]. Available: https://www.cbinsights.com/research/briefing/real-estate-trends/recording/.

[4] F. Ullah, "A Study of Information Technology Adoption for Real-Estate Management: A System Dynamic Model," in *Innovative Production and Construction*, World Scientific, 2019, pp. 469-486.

[5] J. H. Lee, S. J. Kim and X. Yuan, "Toward a User-oriented Recommendation System for Real Estate Websites," *Information Systems*, vol. 38, pp. 231-243, 2013.

[6] F. Ullah, S. M. Sepasgozar and C. Wang, "A Systematic Review of Smart Real Estate Technology: Drivers of, and Barriers to, the Use of Digital Disruptive Technologies and Online Platforms," *Sustainability*, vol. 10, no. 9, 2018.

[7] F. Ullah, S. Sepasgozar and T. H. Ali, "Real Estate Stakeholders Technology Acceptance Model (RESTAM): Userfocused Big9 Disruptive Technologies for Smart Real Estate Management," *International Conference on Sustainable Development in Civil Engineering*, 2019.

[8] E. A. Teracino and D. Seo, "Conceptualization of the Convergence Phenomenon to Develop an Applicable and Integrated Framework for the Emergence of Software-as-a-Service," *Journal of Global Information Management*, vol. 21, no. 4, pp. 1-16, December 2013.

[9] A. Benlian, M. Koufaris and T. Hess, "Service Quality in Software-as-a-Service: Developing the SaaS-Qual Measure and Examining Its Role in Usage Continuance," *Journal of Management Information Systems*, vol. 28, no. 3, pp. 85-126, January 2002.

[10] L. Shiff and C. Kidd, "The State of SaaS in 2022: Growth Trends & Statistics," 17 September 2021. [Online]. Available: https://www.bmc.com/blogs/saas-growth-trends/#.

[11] M. Agrebi and A. L. Boncori, "What Makes a Website Relational? The Expert's Viewpoint," *European Management Journal*, vol. 35, pp. 617-631, 2017.

[12] J. Ainsworth and P. W. Ballantine, "Consumers' Cognitive Response to Website Change," *Journal of Retailing and Consumer Services*, vol. 37, pp. 56-66, 2017.

[13] A. Arndt, D. M. Harrison, M. A. Lane, M. J. Seiler and V. L. Seiler, "Real Estate Agent Target Marketing: Are Buyers Drawn Towards Particular Real Estate Agents?," *Journal of Housing Research*, vol. 26, pp. 39-52, 2017.

[14] H. Richardson and L. Zumpano, "Further Assessment of the Efficiency Effects of Internet Use in Home Search," *Journal of Real Estate Research*, vol. 34, pp. 515-548, 2012.

[15] A. J. F. Yang, Y. C. Huang and Y. J. Chen, "The Importance of Customer Participation for High-Contact Services: Evidence from a Real Estate Agency," *Total Quality Management & Business Excellence*, pp. 1-17, 2017.

[16] F. Kamalov and I. Gurrib, "Financial Forecasting with Machine Learning: Price Vs Return," *Journal of Computer Science*, vol. 17, no. 3, pp. 251-264, 2021.

[17] D. Conway, "THE DATA SCIENCE VENN DIAGRAM," 30 September 2010. [Online]. Available: http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram.

[18] N. S. Ja'afar, J. Mohamad and S. Ismail, "Machine Learning for Property Price Prediction and Price Valuation: A Systematic Literature Review," *Planning Malaysia: Journal of the Malaysian Institute of Planners*, vol. 19, no. 3, pp. 411-422, 2021.

[19] L. Fu, "Supervised and Unsupervised Learning," in *Neural Networks in Computer Intelligence*, Delhi, Tata McGraw-Hill, 2003, pp. 67-68.

[20] V. Limsombunchai, "House Price Prediction: Hedonic Price Model vs Artificial Neural Network," in *NZARES Conference*, Blenheim, New Zealand, 2004.

[21] S. Dabreo, S. Rodrigues, V. Rodrigues and P. Shah, "Real Estate Price Prediction," *International Journal of Engineering Research & Technology*, vol. 10, no. 4, pp. 644-649, 2021.

[22] A. S. Ravikumar and T. Lust, "Real Estate Price Prediction Using Machine Learning," 2017.

[23] S. Putatunda, "PropTech for Proactive Pricing of Houses in Classified Advertisements in the Indian Real Estate Market," 2019.

[24] P. Mali, S. Patil, P. Gujar and M. Tiwari, "Prediction of House Sales Prices Using Machine Learning Algorithm," *International Research Journal of Modernization in Engineering Technology and Science (IRJMETS)*, vol. 3, no. 3, pp. 638-642, March 2021.