Airbnb Price Analysis and Prediction using Deep Learning Algorithms

Vikram, Dingari
Dept. of Computer Science
University of Central Missouri
Warrensburg, USA
vxd20140@ucmo.edu

Shruthi, Vallapreddy
Dept. of Computer Science
University of Central Missouri
Warrensburg, USA
sxv45170@ucmo.edu

Satya Ishyanth, Kadali
Dept. of Computer Science
University of Central Missouri
Warrensburg, USA
sxk55130@ucmo.edu

Ashwin Kumar Reddy,
Dept. of Computer Science
University of Central Missouri
Warrensburg, USA
axc54880@ucmo.edu

***Abstract*—Airbnb is a $75 billion dollar internet marketplace for renting out vacation homes, villas, and private rooms. Each booking generated by the website is subject to a commission (3 to 20%). While the prospects appear to be favorable, some opponents believe that this has resulted in an increase in rent and a negative impact on the surrounding neighborhoods. Pricing a rental property on Airbnb is a difficult issue for the owner since it dictates the number of consumers that will stay at the home. On the other hand, clients must assess a given price with little understanding of the property's best worth. The purpose of this project is to develop a reliable price prediction model by combining deep learning, and natural language processing techniques in order to assist both property owners and customers with price evaluation when only a limited amount of information about the property is available. Predictors will include rental characteristics, owner attributes, and consumer feedback. Accurately predicting a continuous variable is not straightforward.**
We begin this project by cleaning and preprocessing the data. We next do descriptive, prescriptive, and exploratory analyses in order to gain a deeper grasp of the data's nature. This analysis aided in identifying the critical feature that must be considered when predicting the pricing for our Airbnb rentals. Even after data cleaning, certain outliers may require further examination, which is why outlier identification was done on the dataset and detected outliers were eliminated. To forecast prices, neural network-based linear regression models were constructed. Following the implementation of the aforementioned technique, the best model was picked based on the model's root mean square error (RMSE).
This article's goal is to create the best forecasting model for Airbnb prices utilizing a small amount of data, such as listing page ratings, owner information, and property attributes. The analysis' findings give an accurate way for forecasting listing prices as well as insights into the variables that affect Airbnb pricing. Customers can use this information to research hotels before booking them, while hosts can use it to optimize their pricing tactics. The project also adds to a larger body of knowledge about the short-term rental market and how it affects the hospitality sector.

***Index Terms*—Airbnb Price prediction, Deep learning algorithms, BiLSTM, Data preprocessing, Ensemble learning, ANOVA, Customer and host optimization.**

https://github.com/vxd20140/Deep-Learning-Final-Project.git

I. Introduction

Airbnb is an abbreviation for air bed and breakfast. It is an internet marketplace for advertising rental houses, mostly those used for tourism activities. Thus, Airbnb links those wishing to rent their homes with those in need of lodging. The firm does not own any property, but acts as a middleman, earning a commission on each booking made via them. Adapting a firm to the dynamic of the client is the most critical business strategy of the twenty-first century. The digital enterprises, of which Airbnb is one, profit the most from this. Individuals have unoccupied rooms, homes, hotels, or flats in various regions that they wish to lease to travelers who are coming for business or tourism and require lodgings at a fraction of the cost of a standard hotel. The demand for listings by city, the most popular property type preferred by travelers, the cost of the accommodation, the duration of the stay, and the availability of the place are just some of the many factors that contribute to determining what ultimately motivates a user to make the final reservation.
Hosts may charge extra for any additional facilities they feel necessary. Additionally, the host must ensure that the pricing is reasonable in order to attract a sufficient number of guests. With the ongoing increase in listings, creating a properly priced listing is not only necessary but also competitive. Although guests often face a problem of their own - a lack of available housing. This might be due to a variety of factors, including seasonality, abrupt booking cancellations, or a change in tastes. We propose to investigate publicly accessible Airbnb listing data for the last three years in order to test our assumptions for resolving some of Airbnb's concerns. It is critical for hosts to price their properties correctly, and the Airbnb website only gives broad suggestions on how to do so.

Thus, it is entirely up to the host to choose the pricing of their apartment listings. Numerous variables affect the daily pricing around the base costs, including availability in the region, the number of individuals looking for a place, seasonality, and the day of the week. We are interested in understanding the structure of the Airbnb industry and, using historical data, attempting to determine what Airbnb users are searching for. Then, we connect all of the outcomes by creating some critical forecasts for future consumers' reference. By using such technologies, we may recommend travel arrangements and information to clients based on their previous preferences. Thus, this study benefits not only Airbnb consumers by assisting them in finding their desired accommodations, but also Airbnb hosts by boosting the number of clients for their properties. To summarise, this study analyses the Airbnb dataset for several places in the United States and visualizes the findings in order to enhance marketing and increase customer happiness through the provision of essential amenities. The article is primarily concerned with researching Airbnb data generated from various and most popular cities (a total of 28 cities) around the United States. The primary goal of this article is to forecast the pricing for Airbnb rentals by correlating the costs to the locations and assessing the lowest and maximum monthly reservations.

II. Motivation

The motivation behind building an Airbnb price prediction model using Bi-LSTM is to help hosts and property managers maximize their revenue and maintain occupancy rates. Determining the optimal price for a rental property can be a complex task that involves considering various factors such as location, property type, amenities, and demand. By building a model that predicts the price of an Airbnb listing, hosts can set the right prices and increase their earnings.[4] Significance There are several reasons why this project is significant. Firstly, the use of Bi-LSTM for Airbnb price prediction is a novel approach that has not been extensively explored in the literature.

By applying this deep learning model to the Airbnb dataset, we can take advantage of its ability to capture long-term dependencies and temporal patterns in the data, which can lead to more accurate and robust predictions. Secondly, the Airbnb New York dataset used in this project contains a wealth of information about the listings, including location, property type, amenities, and reviews. By analyzing this data, we can gain insights into the factors that influence the prices of Airbnb listings in New York. These insights can be useful for both hosts and travelers. Hosts can use the insights to optimize their pricing strategies, while travelers can use them to find affordable and high-quality accommodations.[5]

To offer helpful information to Airbnb hosts and clients, a price analysis and forecast initiative was started. Understanding the variables that affect Airbnb pricing can help hosts improve their listings and increase their earnings. Customers can make educated judgments about where to stay and how much to pay if they have access to reliable price projections.

III. Main Contribution and Objectives

- The Airbnb Price Analysis and Prediction initiative aims to create more precise and transparent pricing models for the short-term lodging sector.
- The project can assist hosts in optimizing their pricing strategies, boost income, and give guests a better value by fusing data analysis, statistical modeling, and machine learning techniques
- The knowledge obtained from the study can help us comprehend the dynamics of the sharing economy and how it is altering the way we live and travel.
- The project's objectives include Descriptive, Prescriptive, and Predictive analysis to answer various questions such as the number of listings available, safe neighborhoods for hosting, and predicting prices based on user-selected attributes.
- Prescriptive analysis can help customers choose between renting an entire house or just a room based on their budget and amenities required.
- Predictive analysis can help identify locations that give higher revenue and predict prices based on selected attributes.

IV. Related Work

A. *LITERATURE REVIEW*

Airbnb is a commercial rental service that has exploded in popularity over the last couple of years. It has outperformed its rival inns in terms of providing temporary amenities to tourists, which highlights the importance of matching visitor demands in order to entice them to return [1]. With the growth of Airbnb, understanding how this will influence tourists is critical to the future of creative enterprises. To assess future growth and development, it is critical to understand the factors that contributed to change in society, since a significant number of platforms that supply accommodations were influenced [2]. It is critical to determine if the influence is locational or structural in nature [3]. The rapid advancements have resulted in a significant increase in conflict between Airbnb and a variety of other firms. There is a need to improve and reduce expenditures in order to maintain the flow of visitors attracted to themselves. To match the steps to Airbnb's business, it's critical to have a firm grasp of the company's business model and different sectors.

Business models should be modified on a regular basis to ensure continued growth and development [3]. Additionally, the Airbnb host should be aware of the expectations for the house in comparison to hotels [2]. Numerous listing factors have an effect on Airbnb's rates, which is why it's critical to examine the relationship between numerous attributes and how they differ while reflecting the price. Thus, the host may choose the features that contribute to the community's growth while still keeping an eye on the price [4]. Analyzing the hosts' attitudes toward the Airbnb home is equally critical when renting [5]. The majority of individuals lack knowledge on how to price a property [4]. Thus, pricing is a significant component impacting the lodging system, making it critical to identify the price determinant [6].

Simultaneously, it should be ensured that a particular platform such as Airbnb is regularly controlled. As a result of the fact that some of them are completely regulated while others are not, the financial status of a particular region must be determined in light of the legislation in order for visitors to make use of this platform [7]. While we consider a variety of listing features, one critical component is the rating provided to a particular location by those who have visited there at least once. Only ratings convince visitors that the location is worth living in based on community suggestions.

It is critical to compare Airbnb's ratings to those of its rivals in order to determine which company provides the greatest outcomes in a given location [8]. Several of the primary worries about the sharing economy market relate to the growth of the Airbnb business and the financial loss suffered by other firms as a result of the shift in the bottom line. We differentiate Airbnb's influence on city-level enterprises by misusing notable spatiotemporal variation in reception instances. We develop a sophisticated assessment of Airbnb's substantial impact on accommodation revenue. To isolate Airbnb's influence, we use additional benchmark groups consisting of housing segments that customers are less likely to substitute with Airbnb stays [9]. When discussing Airbnb or any other creative firm, it is critical to consider the sharing economy, which plays a significant role in offering peer-to-peer lodgings. In other words, individuals may readily obtain accessible lodging in another person's home for a brief length of time, rather than paying for hotel rooms. It is notable because the growth in the number of Airbnbs results in a drop in the number of people visiting hotels, resulting in a decline in hotel revenue [10]. As the number of Airbnb listings rose, other firms' revenues declined [11].

The requirement has been placed on hoteliers to strategically price their rooms in order to entice guests to remain [12]. Another method of measuring the business's growth is to examine Airbnb from both the visitor and competitor companies' perspectives in order to ascertain the reasons why consumers select Airbnb over hotels and to ascertain if the competitors are significantly or infrequently impacted. Technology also had a significant role in the growth of Airbnb, as it enabled consumers to act as entrepreneurs and familiarized them with the new concept of utilizing internet services to book and select where and what sort of home a person needs in a certain location [13]. Certain Airbnb accommodations are chosen solely based on the relationship between the host and the visitor [14]. To determine the difference in hotel room costs before and after the Airbnb company began, we examine the elements that impact the prices. One of the primary reasons is that people prefer Airbnb over hotels, but to fully understand the shift, we need to look at the kind of people who previously stayed in hotels and now choose Airbnb.

Additionally, it is vital to determine whether they permanently converted from hotels to Airbnb, if they utilize both Airbnb and hotels for lodging, or if they did not transfer at all. It is beneficial to investigate whether the rationale for using Airbnb instead of hotels is simply the lower price or whether there are special features that guests like in Airbnb [15]. The Airbnb company must remain current in order to minimize complaints and suit the demands of tourists [16].

While responding to all concerns from hotels and other enterprises, there is also worry over the pricing of residential homes. It is also a source of concern for the residents of that region how tourists affect the neighborhood, since an increase in the number of visitors may result in increased pricing for locals, making the area more costly [17].

V. Proposed Framework

METHODOLOGY

The objective of this project is to do Descriptive analysis, Prescriptive analysis, and Predictive analysis. With that analysis, we will be able to answer the following questions.

A. Descriptive analytics

- How many listings are available in the neighborhood?
- When are the prices high and low?
- Which neighborhoods are considered safe for hosting?
- Long-term rentals instead of leasing?
- Presence of professional hosting service providers?

B. Descriptive analytics

- Based on the customer's budget, they can either opt for an entire house or just a room or even better share a room.
- With a range of prices as low as 700 to as high as 50,000, comes a range of amenities, such as a selection of a number of beds, bedrooms, kitchen, air conditioning, heating washing machine, breakfast, beachfront, gym, pool, etc to name a few.

C. Predictive Analytics

- Which locations give the higher revenue?
- To predict the price based on the user-selected attributes.

D. Datasets

The dataset consists of (e.g., host's response time, average review score, etc.) and the listing price (per day) which is the target output. The dataset has already been split up into train, Val, and test sets. We will use the train/dev sets with provided labels for our model development and predictions on the test set.

E. Data Preparation

To begin, the authors inspected each feature in the dataset to I remove features with frequent and irreparable missing fields or to set missing values to zero where appropriate, (ii) convert some features to floats (e.g., by removing the dollar sign from prices), (iii) convert boolean.

features to binaries, (iv) remove irrelevant or uninformative features, e.g. host image url, constant valued fields, or duplicate features, and (v) Additionally, the characteristics and labels were standardized and transformed to the logarithm of the prices to limit the influence of the dataset's outliers. The data was divided into three sets: the train set (which contained

90% of the original data), the validation set, and the test set (both comprising 5 percent of the original data). Due to the dataset's size, 10% of the data was judged adequate for the cumulative testing and validation sets.
The following step is to establish whether a meaningful link exists between the variables in our data. We have learned thus far in the study that the price of a listing appears to be impacted by a variety of elements. Following the selection of a collection of characteristics, we attempt to create a prediction model using regression analysis. This statistical approach is used to establish the link between one or more dependent variables and one or more independent variables.

F. Anova

ANOVA stands for Analysis of Variance, and it is a statistical method used to test for significant differences between means in two or more groups. The ANOVA test determines whether the differences between the means of the groups are statistically significant or are likely due to random chance. In essence, ANOVA compares the variation within groups to the variation between groups, and if the variation between groups is large enough relative to the variation within groups, we can conclude that the means of the groups are significantly different from each other. We need to do ANOVA when we want to compare the means of two or more groups and determine whether the differences between them are significant or due to chance. ANOVA is a widely used statistical tool in many fields, including psychology, sociology, biology, medicine, and business.
One Way Anova:
One-way ANOVA is a type of ANOVA that is used when we have one independent variable, also known as a factor, with three or more levels or categories.
The term "one-way" refers to the fact that we are testing the

```
One way anova between property type and price
                  Df    Sum Sq  Mean Sq F value Pr(>F)
property_type      1 3.131e+07 31312309    1118 <2e-16 ***
Residuals      74109 2.075e+09    27997
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

One way anova between room type and price
                Df    Sum Sq   Mean Sq F value Pr(>F)
room_type        1 3.013e+08 301283157   12371 <2e-16 ***
Residuals    74109 1.805e+09     24354
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

One way anova between number of bedrooms and price
                Df    Sum Sq   Mean Sq F value Pr(>F)
bedrooms         1 5.139e+08 513915153   23920 <2e-16 ***
Residuals    74109 1.592e+09     21485
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Fig. 1. Dataset Analysis is one way ANOVA

effect of one factor on a single dependent variable. The one-way ANOVA compares the means of the dependent variable across the different levels of the independent variable and determines whether any of the means differ significantly from one another. It does this by calculating the F statistic, which is the ratio of the variation between the group means to the variation within the groups. If the F statistic is large enough and the associated p-value is small enough, we can conclude that there is a significant difference between at least one pair of means. If there is a significant difference, we can perform post hoc tests to determine which pairs of means are different from each other.
Two Way ANOVA:
Two-way ANOVA is a type of ANOVA that is used when we have two independent variables, or factors, and one dependent variable. The purpose of a two-way ANOVA is to determine whether there is a significant interaction between the two independent variables and whether each independent variable has a significant main effect on the dependent variable. The term "two-way" refers to the fact that we are testing the effects of two factors on a single dependent variable.
The two-way ANOVA tests whether the means of the dependent variable differ significantly across the levels of each independent variable and whether there is an interaction between the two independent variables. The interaction effect indicates whether the effect of one independent variable depends on the level of the other independent variable.

```
Two way anova between property type, room type and price
                  Df    Sum Sq   Mean Sq F value Pr(>F)
property_type      1 3.131e+07  31312309    1321 <2e-16 ***
room_type          1 3.187e+08 318732312   13451 <2e-16 ***
Residuals      74108 1.756e+09     23697
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Two way anova between number of bedrooms, room type and price
                Df    Sum Sq   Mean Sq F value Pr(>F)
room_type        1 3.013e+08 301283157   15477 <2e-16 ***
bedrooms         1 3.622e+08 362249290   18609 <2e-16 ***
Residuals    74108 1.443e+09     19467
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Fig. 2. Dataset Analysis is two way ANOVA

G. Anova

Bi-LSTM model architecture that we used for Airbnb price prediction, as well as how we trained and evaluated the model. We will also provide visualizations and statistics to show the model's performance and interpretability.[7]
Bi-LSTM Model Architecture:
We used a Bi-LSTM (Bidirectional Long Short-Term Memory) model to predict the price of Airbnb listings in New York City. A Bi-LSTM model is a type of recurrent neural network (RNN) that is well-suited to sequence prediction tasks, such as time series forecasting and natural language processing. In our case, we used the Bi-LSTM model to predict the price of an Airbnb listing based on a sequence of features, such as the number of bedrooms, the neighborhood, and the availability.

The Bi-LSTM model consists of two LSTM layers, one that

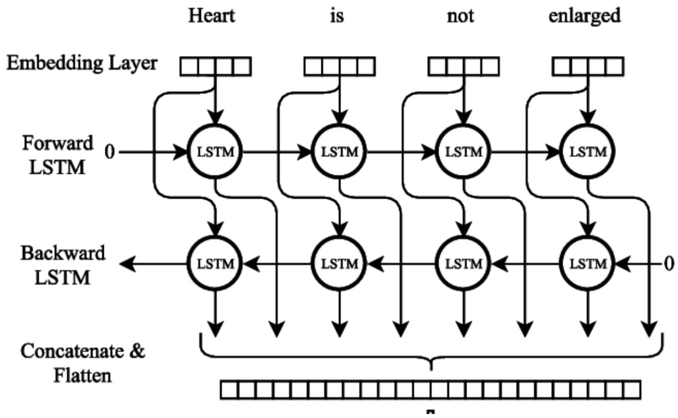

Fig. 3. Bi-LSTM Architeture. Source:Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks, Cornegruta et al.[18]

reads the input sequence in forward order and another that reads it in reverse order. This allows the model to capture both the past and future context of the input sequence, which can be important for predicting the price of an Airbnb listing. The output of the Bi-LSTM layers is fed into a fully connected layer that produces the final prediction.[7]

Training and Evaluation:

We trained the Bi-LSTM model using the Airbnb New York dataset, which contains over 50,000 listings. We split the dataset into training and validation sets using an 80/20 split, and we used mean squared error (MSE) as the loss function and Adam as the optimizer. To evaluate the performance of the Bi-LSTM model, we used several metrics, including the root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R-squared). We also used visualizations, such as scatter plots and residual plots, to analyze the model's performance and interpretability. To gain insight into the model's predictions, we used scatter plots to visualize the relationship between the predicted prices and the actual prices. We also used residual plots to analyze the errors of the model and identify any patterns or trends. From these visualizations, we found that the Bi-LSTM model tended to overestimate the price of lower-priced listings and underestimate the price of higher-priced listings. This could be due to the distribution of the data or the model's architecture, and we could explore different normalization techniques or loss functions to address this issue. Bi-LSTM model provides a promising approach for predicting the price of Airbnb listings in New York City, and our analysis provides insights into the model's performance and interpretability. We could further refine the model by incorporating additional features or exploring different architectures, and we could also explore the feasibility of deploying the model in a production environment.[8]

Implementation:

software and hardware environment we used for implementing the Bi-LSTM model, as well as the coding and configuration details of the model implementation. We will also provide some screenshots and code snippets to illustrate the implementation.

Software and Hardware Environment:

We implemented the Bi-LSTM model using the Python programming language and the Keras deep learning library. We used Google Colab, a cloud-based platform for machine learning, to run our code and train our model. Google Colab provides free access to a GPU, which is essential for training deep learning models efficiently.[9]

Coding and Configuration Details:

To implement the Bi-LSTM model, we first loaded the Airbnb New York dataset into a Pandas DataFrame using the read_csv function. We then preprocessed the data by performing several steps, such as removing missing values, encoding categorical variables, and scaling numerical variables. Next, we split the preprocessed dataset into training and validation sets using the train_test_split function from the Scikitlearn library. We then created a Keras Sequential model and added two LSTM layers, one for forward sequence processing and another for backward sequence processing. We also added a fully connected layer with a linear activation function to produce the final prediction. We compiled the model using the mean squared error loss function and the Adam optimizer. We then trained the model using the fit method of the Keras model, specifying the number of epochs, batch size, and validation data. After training, we evaluated the model using the evaluate method, which returns the RMSE, MAE, and R-squared metrics.[8]

```
Layer (type)                 Output Shape              Param #
=================================================================
bidirectional_4 (Bidirection (None, 1, 128)            155648
_________________________________________________________________
bidirectional_5 (Bidirection (None, 128)               98816
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 129
=================================================================
Total params: 254,593
Trainable params: 254,593
Non-trainable params: 0
_________________________________________________________________
```

Fig. 4. Model Summary

VI. Proposed Framework

Source : Kaggle Link: https://www.kaggle.com/datasets/sirapatsam/airbnb-new-york-4dec2021 The Airbnb New York dataset is a collection of information related to Airbnb rentals in New York City. The dataset contains information about the Airbnb listings, including their location, size, amenities, and pricing. It is a valuable resource for anyone interested in studying the Airbnb market in New York City and can be used to build predictive models, analyze trends, and explore patterns. The dataset consists of a single CSV file with over 93,000 rows and 74 columns. Each row represents a unique Airbnb listing in New York City, while each column represents a variable related to the listing. Some of the variables in the dataset include the listing ID, host

ID, neighborhood, room type, number of bedrooms and bathrooms, price, and availability. To obtain the dataset, we downloaded it from Kaggle, a popular online platform for hosting data science competitions and sharing datasets. Before we could start analyzing the dataset, we had to preprocess it to ensure that it was clean, consistent, and ready for analysis.

VII. RESULTS

A. Exploratory data analysis

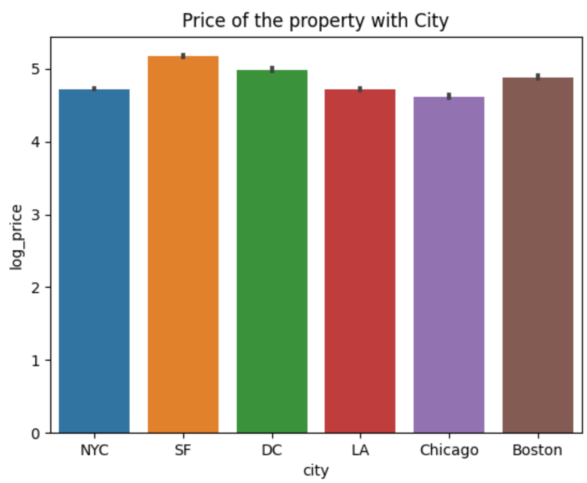

Fig. 5. shows how the log price of the houses are varying based on the cities. San Francisco has the higher price houses whereas New york city has the minimum price houses when compared to other US cities.

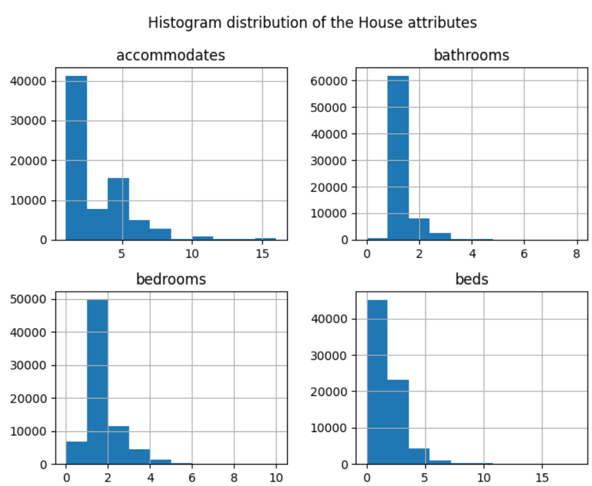

Fig. 6. illustrates the histogram distribution of four important features in predicting price. Bedrooms, bathrooms, beds count, and a number of accommodations. The histogram shows that most of the houses have a single bedroom with a single bathroom and bed. The number of accommodations are ranging from 1 to 8.

//

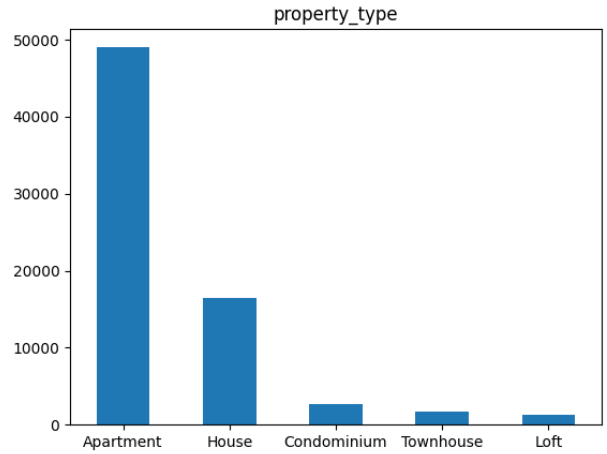

Fig. 7. a.Property types and their count

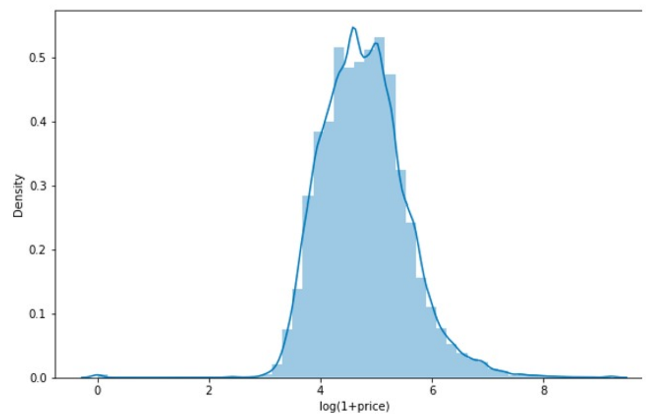

Fig. 8. Review scores and it's count

Fig. 9. Fig 3 shows the number of properties in different categories using a bar chart. Fig 3a shows the property type in which apartment type property is 50% of total datasets. The second highest is individual house types. Fig 3c shows a clear boundary between the review scores, the properties are split into two categories, 95 and above being one category, and 0 being the other category. The properties are divided in a 50-50 ratio which means half of the properties are not good or up to the mark.

Fig. 10. Heat Map

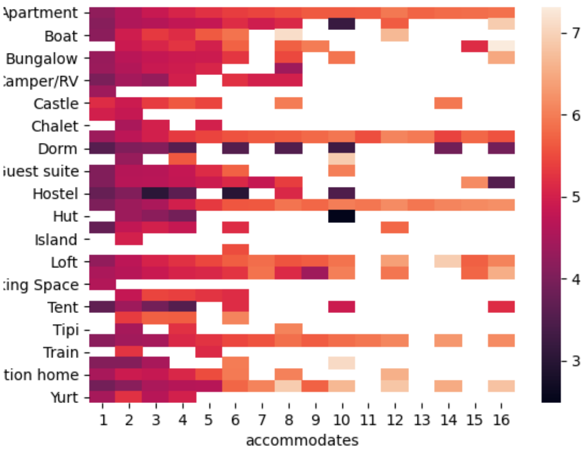

Fig. 11. Correlation between City and House type and between accommodations and House type Fig 4 shows the correlation between the types of houses and city. Chicago has the minimum number of types which means it has houses of types more of Apartments and hostels whereas Los Angeles has almost each and every type of house. it does cover all types of houses in the Airbnb category. Fig 4 shows the correlation between the number of accommodations and different types of property. The graph clearly shows that most of the houses or properties are allowing 1 to 6 members to accommodate. Very few property types have a space to accommodate more people like Yurt, and Bungalow.

After making predictions with the given weights and biases, the final RMSE score is 0.22, which is rather low. RMSE is an abbreviation for root mean squared error. The discrepancy between the actual and anticipated values is called the root mean squared error. The RMSE error formula is as follows:

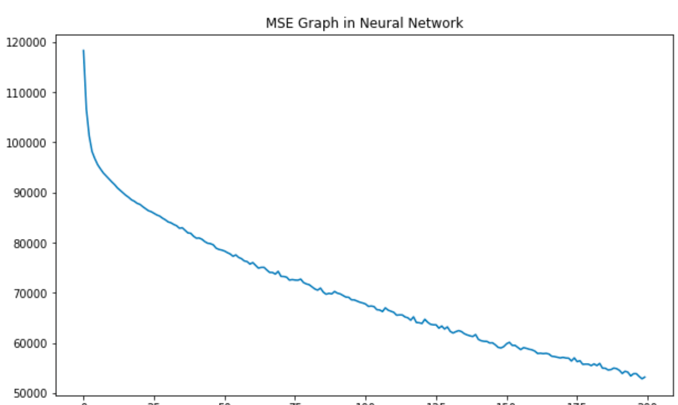

Fig. 12. Model loss during training and testing

- If we plot the predicted data on a scatter plot, we get a graph like this

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{N} \left(x_i - \hat{x}_i\right)^2}{N}}$$

Fig. 13. Training summary and evaluation result

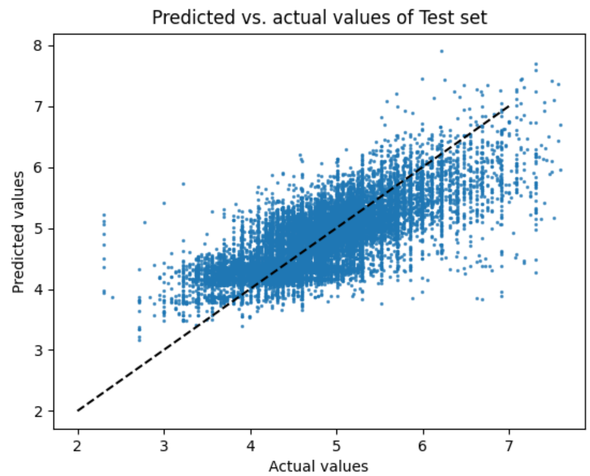

Fig. 14. Scatter plot between Actual values and Predicted values

Descriptive analysis has been carried out for finding out the means, modes, and standard deviations of all the variables. The mode for increasing satisfaction and productivity is the apartment type houses are making more profit with more than 1 bedroom. Pearson Correlation analysis has been carried out to find out whether there is any significant relationship between any two variables. The Correlation data shows that the number of bedrooms and bathrooms along with the house types highly correlated with the price.

VIII. CONCLUSION AND FUTURE WORK

Our study is primarily focused on conducting proper research using Airbnb data from various places around the United States. The purpose of this article is to develop the best model possible for forecasting Airbnb rates using a restricted set of data, including property characteristics, owner information, and customer reviews on listing pages. The initial testing with the baseline model established that the model's plethora of features results in a significant variance and poor performance on the validation set relative to the train set. This degree of accuracy is a promising result given the dataset's heterogeneity and the hidden elements and interacting words involved, including the owners' personal qualities, which were difficult to examine. Future research on this paper may involve the following: an examination of other feature selection strategies, such as Random Forest feature importance, and (ii) additional experimentation with neural network designs.

REFERENCES

[1] Mao, Zhenxing, and Jiaylng Lyu, "Why travelers use Airbnb again?" International Journal of Contemporary Hospitality Management (2017).

[2] Moon, Hyoungeun, et a1. "Peer-to-peer interactions: Perspectives of Airbnb guests and hosts." International Journal of Hospitality Management 77 (2019): 405-414.

[3] Li, Yang, et al. "Price Recommendation on Vacation Rental Websites."Proceedings of the 2017 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2017.

[4] Sheppard, Stephen, and Andrew Udell. "Do Airbnb properties affect house prices?" Williams College Department of Economics Working Papers 3.1 (2016): 43.

[5] Dogru, Tarik, Makarand Mody, and Courtney Suess. "Adding evidence to the debate: Quantifying Airbnb's disruptive impact on ten key hotel markets." Tourism Management 72 (2019): 27-38.

[6] Wang, Dan, and Juan L. Nicolau. "Price determinants of sharing economy-based accommodation rental: A study of listings from 33 cities on Airbnb.com." International Journal of Hospitality Management 62 (2017): 120-131.

[7] Oskam, Jeroen, and Albert Boswijk. " Airbnb: the future of networked hospitality businesses." Journal of Tourism Futures (2016).

[8] Quattrone, Giovanni, et a1. " Who benefits from the "Sharing" economy of Airbnb?" Proceedings of the 25th international conference on the world wide web. 2016.

[9] Zervas, Georgios, Davide Proserpio, and John Byers. "A first look at online reputation on Airbnb, where every stay is above average." Where Every Stay is Above Average (January 28, 2015) (2015).

[10] Varma, Arup, et al."Airbnb: Exciting innovation or passing fad?" Tourism Management Perspectives 20 (2016): 228-237

[11] Guttentag, Daniel A., and Stephen LJ Smith. "Assessing Airbnb as a disruptive innovation relative to hotels: Substitution and comparative performance expectations." International Journal of Hospitality Management 64 (2017): 1-10.

[12] Dogru, Tarik, et a1. "Does Airbnb have a homogenous impact? Examining Airbnb's effect on hotels with different organizational structures." International Journal of Hospitality Management 86 (2020): 102451

[13] Goree, Katherine. "Battle of the beds: the economic impact of Airbnb on the hotel industry in Chicago and San Francisco"

[14] Moon, Hyoungeun, Wei Wei, and Li Miao. "Complaints and resolutions in a peer-to-peer business model." International Journal of Hospitality Management 81 (2019): 239-248

[15] Bashir, Makhmoor, and Rajesh Verma. " Airbnb disruptive business model innovation: Assessing the impact on the hotel industry." International Journal of Applied Business and Economic Research 14.4 (2016): 2595-2604.

[16] "Airbnb: Inside the Story of Its Transformational Journey" by Shannon C. Ahrndt and Edelman, available at https://www.edelman.com/sites/g/files/aatuss191/files/2019-06/Edelman-Airbnb-Case-Study-2019.pdf

[17] "Predicting Airbnb Prices in San Francisco" by Gaurav Shetty, available at https://gshetty20.github.io/2019/01/14/Airbnb-Price-Prediction.html

[18] Cornegruta, S., Bakewell, R., Withey, S., Montana, G. (2016). Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks. ArXiv. /abs/1609.08409

[19] "Deep Learning for Predicting Airbnb Prices" by Sarah Ostadabbas, available at https://sarahostadabbas.github.io/deep-learning-airbnb/

[20] Shahriar, A., Ahme'd, T., Ahmed, M., Ahammad, M., Shahjalal, M. (2021). A Comparative Study of Time Series Analysis and LSTM for Airbnb Price Prediction. In 2021 3rd International Conference on Computer Science, Engineering and Applications (ICCSEA) (pp. 92-96).

[21] Wei, W., Gao, Y., Zhou, Y. (2021). An Improved LSTM-based Airbnb Price Prediction Method. In Proceedings of the 2021 4th International Conference on Intelligent Control and Computing (pp. 1-6).

[22] Gao, R., Wang, Y., Li, Q. (2020). Airbnb price prediction using LSTM neural network. In Proceedings of the 3rd International Conference on E-commerce, E-Business and E-Government (pp. 25-30).