Tracking Property Ownership Variance and Forecasting Housing Price with Machine Learning and Deep Learning

Junfeng Jiao, Ph.D.
Urban Information Lab
The University of Texas at Austin
Austin, Texas, U.S.
jjiao@austin.utexas.edu

Seung Jun Choi
Urban Information Lab
The University of Texas at Austin
Austin, Texas, U.S.
jun.choi@utexas.edu

Weijia Xu, Ph.D.
Texas Advanced Computing Center
The University of Texas at Austin
Austin, Texas, U.S.
xwj@tacc.utexas.edu

2021 IEEE International Conference on Big Data (Big Data) | 978-1-6654-3902-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/BigData52589.2021.9671298

***Abstract*—** Big data and its production, management, and utilization are essential components in smart city planning. This paper presents a research framework for applying machine learning and deep learning using multiple big data sets on real estate. We built ensemble machine learning models to track property ownership variance in Austin, TX, USA. Then, the study employed the long short term memory (LSTM) model as a deep learning approach to forecasting property value in the same area. For model validation, root mean squared errors were calculated in both models. To avoid underfitting or overfitting of LSTM, we experimented with specific parameters settings. Bagging-based random forest machine learning model outperformed other ensemble machine learning models. Regarding property ownership variance, the Random Forest model's highest feature importance generally comprised the race, age of residents, land use and built environment factors, number of schools, and neighborhood location. Our LSTM model predicted Austin to retain a rising curve in housing prices and identified which part of Austin experiences an increase or decrease in property value. The predictive models may help city planners to quantify and gain insights on future impacts of developing neighborhoods.

Keywords: Property Ownership, Property Value, Machine Learning, Deep Learning, Big Data

I. INTRODUCTION

Advances in new technology have stimulated a new wave of infrastructure deployments around the world. Although new and updated infrastructure deployment is crucial for economic flourishing and residents' wellbeing improvement for local communities, it can also lead to adverse effects such as accelerating the gentrification process and increasing displacement for vulnerable residents [1]. The rapid gentrification process is an urgent issue for any community undergoing rapid growth and calls for attention.

Past research generally conflated gentrification and displacement; however, they do have clear distinctions [2]. Gentrification is neighborhood revitalization, which results in pressing displacement [2]. However, displacement can occur in the absence of gentrification [2].

Displacement occurs when households are forced to leave the neighborhood as it becomes increasingly expensive, either out of choice or necessity. Despite satisfying pre-imposed occupancy conditions, it occurs beyond the household's ability to prevent displacement, or forced pressure makes continued occupancy unaffordable [2]. There is a clear, traditional pattern of development that drives displacement.

Gentrification happens when higher-income households move into a neighborhood, drive up housing prices, and change the neighborhood's character or by cities' renewal projects. Although it is hard to quantify the number of residents forced out due to gentrification, past research has shown that low-income and minority residents are often the most at risk [3]. In the meantime, displacement is a central concern in gentrification [2]. Studies have documented the effect of displacement that follows from gentrification in a diverse range of US urban areas [4-8]. However, existing studies are often after-the-fact reports that describe comparative metrics over time.

The main goal of this paper is to investigate viable computational approaches to predict the gentrification process using machine learning and big real estate property data. Too often, the adverse consequences of gentrification are identified after the process has begun, and opportunities for early intervention and mitigation are missed. Even when equity is flagged as a significant or primary concern during the infrastructure planning process, adverse outcomes still arise [9-11]. Here we focus on predicting two critical indicators of gentrification: real estate property value and ownership changes. Property value and owner-occupancy reflect current values and mortgage market conditions which is a crucial indicator for estimating the affordability of housing [12].

The first stage of gentrification is the displacement of residents. When regional economies flourish, the price of residential and commercial real estate rises. Existing property owners have an incentive to sell their property and leave their neighborhood to profit from the appreciation of their real estate, avoid higher real estate taxes, and offset increased living costs. In addition to these economic inducements, they may no longer feel comfortable given the changing character of the neighborhood (e.g., demographic shifts, changing of local businesses). Therefore, looking into property ownership variance and real estate value trends are vital indicators for early identification of the gentrification analysis.

In the case of Austin, TX, housing price has been steadily increasing over the years. Moreover, Austin ranks as top ten cities in the nation that experiences rapid urban gentrification [13]. Early detections of gentrification neighborhoods and identifications of high-risk residents enable early intervention opportunities for city leaders. For instance, Project Connect (https://www.capmetro.org/project-connect), a major new light rail project recently approved in the City of Austin, seeks innovative, practical solutions to the anti-displacement challenges.

Towards this goal, we constructed both machine learning and deep learning models to analyze big real estate property data in Austin.

The machine learning approach is used to create a predictive model for property ownership changes over time. The study traced property ownership variation in Austin, TX, USA, by comparing individual property owners to the former year.

We obtained sophisticated time series housing sales data under the agreement. The data was used to construct a deep learning model to diagnose housing price trends from an artificial intelligence perspective. We adopted a Long Short-Term Memory (LSTM) neural network, which is a Recurrent Neural Network (RNN) architecture that resolves vanishing, and exploding gradient problems of conventional RNNs [14]. Compared to conventional feedforward neural networks, RNN uses cyclic connection sequences [14]. The predicted housing sales prices in Austin were compared to actual prices during the study period.

Major research questions include the following:

RQ1: Which variables are found to be having the highest feature importance on property ownership variance?

RQ2: Is deep learning applicable to forecast housing prices?

RQ3: What is the ongoing housing market trend in Austin from an AI perspective?

In the subsequent sections, we give additional background and related work in Section II. We describe data and methods used in Section III. Section IV presents the results from our analysis and model validation. We discuss the results of the research questions in Section V and summarize conclusions in Section VI.

II. Background and Related Work

The advance of smart city technology and its application has gained attention in recent years [15]. Though the definition may vary, smart city technologies include installing hardware, management, and appliance in software that can be melted into daily living. Behind these three components, producing, managing, and analyzing big data remain essential tasks.

Big data in cities refer to data linked to time and space [16]. The production, monitoring, and regulation of these data and analytics methods are essential aspects of the smart city [17] and codependent with each other. For instance, the new installation of sensors in cities provides big real-time data from designated space throughout their lifetime. Then managed big data is used to construct a robust artificial intelligence (AI) model or analyze trends and characteristics. It sounds simple; however, due to the fuzziness of big data, the data mining approach, required management, and analysis skills vary by different types of big data [18].

An atmosphere of applying AI in a functional normative perspective has been formed. In other words, we apply them because it is deemed as a good way of doing it. Though the definition and types of AI vary [19], AI is already applied in our daily lives, from mobile phone applications to autonomous vehicles. Soon, AI will help planners forecast demands from multi-perspectives like housing, transportation, and public services concerning sustainable smart city planning.

To model robust AI, machine learning and deep learning models should be built prior. Two models are a subset of AI and closely intertwined with one another. Among the vein of thought, urban planning discourse began exploring these approaches in recent years as big data named urban informatics became available. Moreover, machine learning or deep learning models are already found to be outperforming traditional statistical models.

Ensemble learning uses multiple learners to resolve the joint problem by constructing varying hypothesis and combining them [20]. It is widely adopted machine learning approaches that are deep-rooted from Decision Tree models. It can be classified into mainly two types; bagging or boosting approach. The bagging uses the same algorithm; however, it runs dissimilar sampling during the training, referring to bootstrapping [20-22]. RandomForest (RF) is a widely adopted bagging approach [21]. In the meantime, boosting put weight continuously during the training to fix the incorrectness from the prior prediction. XGB (eXtreme Gradient Boosting), and LGBM (LightGBM) is the primary boosting approach [20].

RF is an effective tool that usually resolves overfitting issues by creating sufficient randomness, "making them accurate classifiers and regressors." [21] It is more flexible and faster than boosting approach because traditional boosting took a long time to set proper parameters. However, recent gradient ensemble learning has constantly improved. For instance, LGBM came out as the next generation boosting approach, improving the accuracy and cutting down training duration. However, in either case, ensemble learning models' 'generalization' ability is still better than the single learner model [20].

Similarly, a former study proved that ensemble-based learning is systematically better and always had the lowest mean absolute error even in the different number of estimators compared to other models regarding housing price prediction [22]. However, though property value gained much attention [22-24], property ownership variance has not been deeply explored.

In terms of using big real estate property data, LSTM has been used to forecast housing prices [25-28]. The previous studies admit that LSTM is an adequate tool to be used in the time series model. Some conclude LSTM to produce reliable housing price estimates with acceptable error [27]. However, as deep learning performance is based on inputted parameters and the fact that result changes every time after training do not result in unequivocal consensus to argue LSTM outperforms other time series models consistently [28].

III. DATA AND METHODOLOGY

A. Study Area

The study area includes the portion of the city of Austin located within Travis County, Texas, USA. The study area composes 103 neighborhoods. Fig. 1 describes the areal interpolated 2010 census population from census tract to neighborhoods. The result describes that the census population is mostly evenly distributed across boundaries.

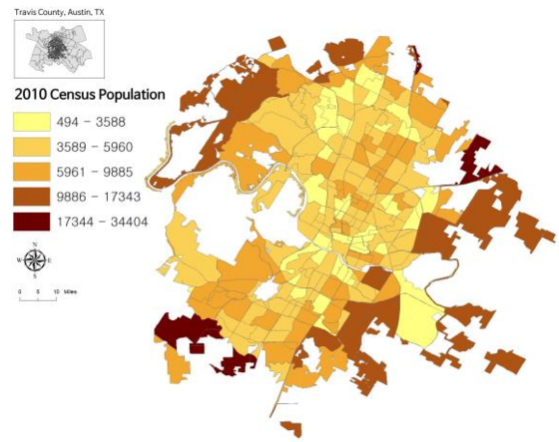

Fig. 1. Study Area

B. Property Big Data

The study used two sets of big property data obtained from Travis Central Appraisal District (TCAD) and home sales data given from Austin/Central Texas Realty Information Services through a data-sharing agreement. Each data set includes records over ten years period. TCAD data is from 2010 to 2020. Home sales data is from 2008 to 2018.

TCAD data consists of 6.8 million property records with unique parcel IDs from 1993 to 2020. Other features include state property type, owner ID, the total area of the parcel, the estimated value of the property, the year built, the interior area of the building, and the number of building permits issued [29]. Home sales data include actual sales price, listed price, duration in the housing market, and other property information such as exterior features, sales restrictions, property types. Individual home sales records also include the latitude and longitude of the property.

TCAD data was used to trace property ownership variance. Data curation went through five steps. First, data were split into twelve groups to include data from 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, and 2020. The 2009 data was used to trace 2010 property ownership variance. Second, using a geographic information system (GIS), we areal interpolated building parcels to census tract in Austin to associate individual building parcels' with census tract information. 362,216 property parcels were used and given their geographic information. Third, we traced the owner ID variation of individual property parcels by comparing ownership of the current year with the previous year. As a result, we created a dummy variable. In specific, parcels in 2010 were compared with parcels in 2009 using parcel ID and owner ID. Parcels in 2011 were compared with parcels in 2010. Other years went through the same procedure. Value 1 is used to indicate the owner of the property changed from the previous year. For example, if a property was built and owned in 2013, and the owner changed in 2014, it is counted as one ownership change. On the other hand, if the property owner were the same or untraceable, we marked the value as 0. Fourth, the ownership variation variable was grouped by tracts with average and count values computed per group using Python packages. Fifth, the property ownership ratio was calculated by dividing the total ownership variation counts by the total property counts in each Austin census tract. Lastly, the results were areal interpolated to neighborhoods in GIS.

Home sales data was used to forecast housing prices throughout the study area. Each record was spatially joined with neighborhoods in Austin to add neighborhood information. We normalized sales price to price per square feet (sqft) by using the assigned individual property square feet information. To remove outliers and to focus on low price single-family housing sqft sales price was clipped to be greater than \$80 and less than \$200 inclusive. In total, 55,334 home sales data were used. Also, unknown property types and condominium was removed.

For machine learning modeling, social demographic factors including population density per sqft, race (people of color to total population), gender (male population to female population), aging index (age under 18 to age over 64), built year, stories, average property ownership variance from 2010 to 2020 were used. For deep learning modeling, neighborhood name, subdivision (property type), unit style of property, sales restriction, occupant type, property condition at the sale were one-hot encoded. Also, listed price, sales price, listed and sales price per sqft, and price gap were used to improve prediction performance. The price gap is the difference between the listed housing price and the actual sales price. Generally, the sales price is lower than the listed price.

Both models include land use factors such as land use inventory ratio and building footprints, including average max height of buildings, elevation, number of hospitals and schools, and other property-related factors like duration in the market, built year, and stories.

In total, 498,754 records and 310 features were used for machine learning modeling. Moreover, 55,334 records and 6,656 features were used for deep learning modeling. In sum, 522,916,844 data values were used. The study period is from 2008 to 2020. The geographic unit of analysis is a neighborhood in the City of Austin. The dependent variable is traced property ownership variance and sqft property sales price throughout the study area. For technological support, we used various Python libraries, including Scikit-learn, ArcGIS-Pro, and Tableau Desktop.

We used the Frontera supercomputer in TACC (Texas Advanced Computing Center) to train and test data. Frontera has 39PF (Petaflops Peak Performance with over 8,000 compute nodes, each equipped with Intel Cascade Lake CPU and 128GB DDR4 memory. Frontera is equipped with 50+ PB disk, 3PB of Flash with 1.5TB/sec peak I/O rate. The training and testing of deep learning models used GPU nodes on Frontera. Each GPU node has four RTX5000 GPUs. All computation conducted in this paper is limited to a single compute node at a time.

C. Ensemble Learning ML Model Building

The present study ran four ensemble machine learning regressor models; RandomForest (RF), GradientBoostingModel (GBM), XGB (eXtreme Gradient Boosting), and LGBM (LightGBM). For bagging, we used RF. For boosting, we used GBM, XGB, and LGBM. Ensemble learning can smoothly handle nonlinear features and does not require feature scaling. However, to reduce the losses, we took a log on the dependent variable as described in Fig. 2.

Fig. 3 describes the counted total property ownership change over time in Travis county. From 2010 to 2020, property ownership change incrementally increased and hit its peak in the first year (2020) of COVID-19 with 39,667 cases (12.5%). Year by year, the first year (2020) of COVID-19 had the greatest property ownership change by 39,667 cases (12.5%). On average, 26,908 of the property goes through ownership changes annually, which compose 9.1% of all properties.

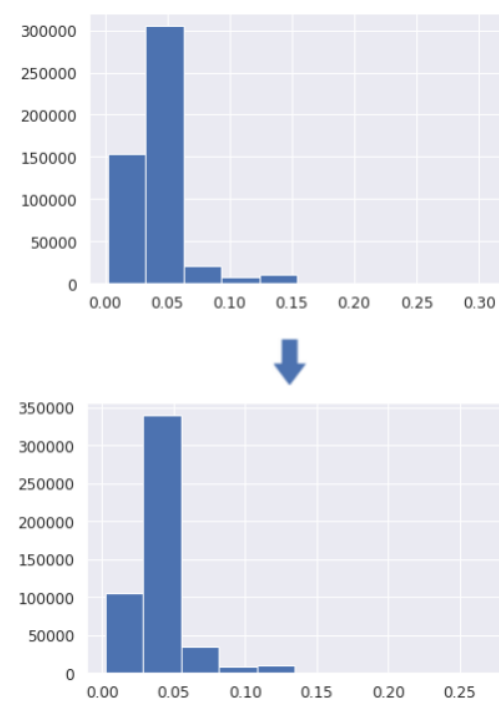

Fig. 2. Ownership Variance Taken Log

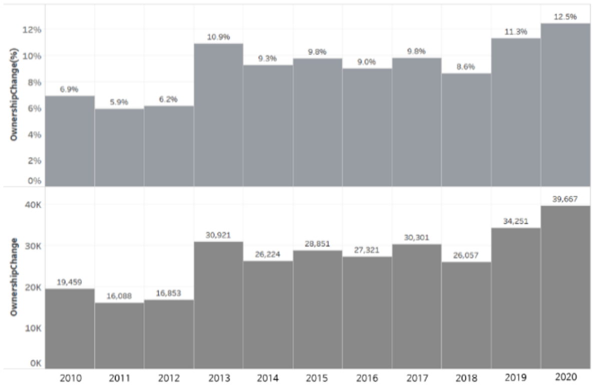

Fig. 3. Ownership Variance (%) & Ownership Change

D. LSTM (Long Short-Term Memory)

We used a Long Short-Term Memory (LSTM) neural network and a recurrent neural network (RNN) algorithm to forecast housing prices. Traditional neural networks have vanishing gradient or long-term dependency (LTD) issues in which RNNs gradually forget previous inputs due to the sensitivity of input values decaying over time [30].

LSTM resolves the vanishing gradient problem by adding one additional cell (called a "forget gate") that lets the model decide whether to add or exclude information in its cell state [31]. In addition to the forget gate, LSTM uses two other gates (input and output gate), as shown in Fig. 4. The input gate determines whether to update the information to the cell state, and the output gate updates attribute.

An equation for each stage is shown below in (1) through (6). f_t, i_t, C_t, C_t, o_t, h_t are forget gate, input gate, cell input activation, cell state, output gate, hidden layer output in the time sequence t, respectively. All three gates (f_t, i_t, o_t) use the sigmoid function (σ), and tanh also refers sigmoid function. W are weight matrices, b is bias, x_t is input in time sequence t. The variables were scaled using MinMaxScaler.

$$\text{Forget gate: } f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$\text{Input gate: } i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\text{Cell input activation: } C_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

$$\text{Cell state: } C_t = f_t \cdot C_{t-1} + i_t \cdot C_t \quad (4)$$

$$\text{Output gate: } o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$\text{Hidden layer output: } h_t = o_t \cdot \tanh(C_t) \quad (6)$$

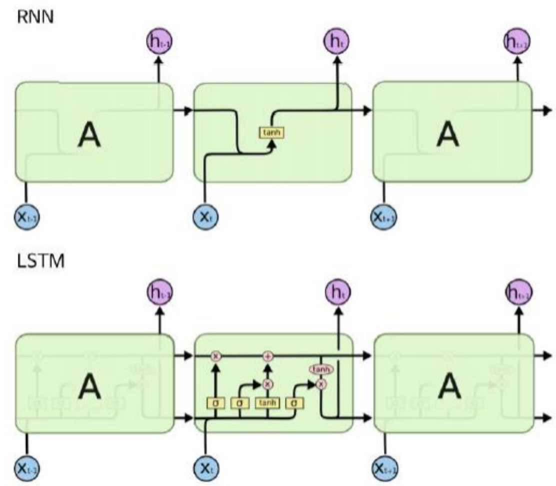

Fig. 4. Data Transmission in RNN & LSTM [31]

E. Model Validation

For model validation, RMSE (Root Mean Square Error) was measured as summarized in Table 1. RMSE is calculated by determining the residuals [32]. Residuals are the differences between the actual values (y_i) and predicted values ($\hat{y}_i$). The equation of RMSE is described below and n is a number of validation samples.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}}$$

The number of estimators for ensemble learning was set as 30. The data were split into 80% for training and 20% for machine learning and deep learning testing. Then, the feature importance was pulled out based on the RF model for having the lowest RMSE.

For the deep learning model, train loss and validation loss were compared. For a proper model, train loss should always be lower than validation (test) loss. It resolved either underfitting or overfitting issues as described in Fig. 5 when using specific parameters; the batch size of 2000, 80 nodes, the dropout rate of 0.05, and 100 epochs.

TABLE I. COMPARSION OF ROOT MEAN SQUARE ERROR FOR DIFFFERENT MODEL PREDICTIONS

DV	Model		RMSE
Ownership Variance	ML	RF	0.000
		GBM	0.010
		XGB	0.002
		LGBM	0.003
Per Sqft. Sales Price	DL	LSTM	31.163

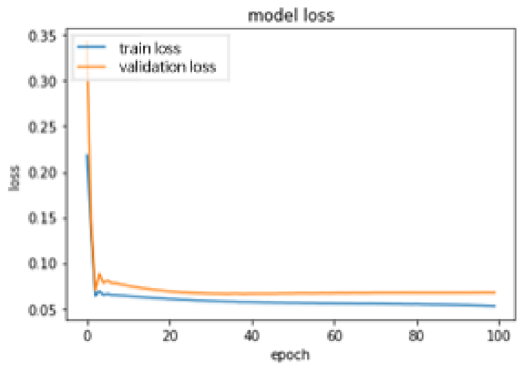

Fig. 5. LSTM Validation Loss Compared to Train Loss

IV. RESULT

A. Machine Learning: Regressor Model Result

Fig. 6 describes the top 20 feature importance from RF. Social demographic factors like race, aging index, gender were found with high importance. Also, one hot encoded neighborhood, including Avery Ranch-Lakeline, River Place, Barton Hill, Boggy Creek, were included as top 20. Land use and built environment factors such as average elevation and average max building heights, zoning ratio (Office, Civic, Single-family, Commercial, Mobile Homes, Multi-family, Open Space, Industry), and a number of schools were included.

B. Deep Learning: LSTM Forecast Result

Fig. 7 illustrates the average sqft listed housing price and sqft sales housing price based on home sales data. Both lines show ups and downs. And average sqft sales price is always lower than sqft listed price. The gap is about $15 on average. Starting from 2014, average housing prices drastically increased than previous years.

The prediction began in 2016. Fig. 8 Compares predicted sqft home sales point to original ones. Similar to original records, predicted sqft home sales values showed an incrementally rising curve. The prediction records do overlap with to original values. However, underestimation was observed as having fewer red dots on the right upper side.

We calculated the increase rate by dividing the 2017 average sqft sales price or prediction by 2016 average sqft sales price or prediction and compared them. The model predicts that there will be a decrease in housing prices, mainly in central Austin. These neighborhoods do overlap with actual sqft sales price average increase rate ones as described in Fig. 9.

V. DISCUSSION

RQ1: Which variables are found to be having the highest feature importance on property ownership variance?

Ensemble learning is proven to outperform traditional models in estimating the housing market price [33]. We used four Decision Tree based ensemble regressor models to estimate property ownership variance. The bagging-based RF outperformed than boosting models by having the lowest error score.

RF model defined several social demographic factors, land use, and built environment factors, and one hot encoded neighborhoods dummy to have the highest importance than other 290 features. Surprisingly, the race and aging index ranked as having the highest feature importance by taking the importance of nearly 45%. As Austin has suffered contentious tensions of gentrification between East and West Austin concerning race, poverty, and inequity [3], it seems reasonable for these two factors to have the highest importance on understanding property ownership variance.

Land use and built environment factors like office, civic, commercial, open space land use ratio, location of neighborhoods, and the number of schools seem to play an essential role in property ownership variance. This result suggests that urban amenities and schooling may play a significant indicator for buying and selling property.

RQ2: Is deep learning applicable to forecast housing price?

We used a neural network based deep learning method named LSTM to forecast sqft housing sales prices for individual neighborhoods. LSTM forecasted the average sales price to show a rising curve in line with the rising pattern of the concurrent sales price in Austin. From a growth rate perspective, the model predicted the central part of Austin to experience a decrease in housing prices than neighborhoods in the edges. Surprisingly this forecast was quite similar to the actual growth rate calculated by original home sales data. A TCAD data that only compose single-family housing showed a similar pattern, for having a decrease in average housing price in central in 2017 than 2016. These results suggest that LSTM can be a valuable tool to forecast housing price when combined with geographic information for locating increase or decrease in property value.

RQ3: What is the ongoing housing market trend in Austin in AI perspective?

Machine learning and deep learning are the basis for modeling artificial intelligence. They are linked to one another. We used four ensemble learning models to track property ownership variance and LSTM as a deep learning model to forecast housing prices. Among four ensemble machine learning models. RF scored 0.000 RMSE, which is quite an astounding performance. RF viewed the race, age, land use and built environment, school, and neighborhood location as having greater effects on property ownership variance.

LSTM forecasted the average sqft sales price to go up over the years and retain the rising curve. It identified central Austin to suffer a decline in home sales value and edges to experience a rise in home sales value. Fig. 10 describes the highest and lowest housing price increase rate of neighborhoods in order. RMMA, South Lamar, Upper Boggy Creek, Hancock, Spicewood, Zilker, Bull Creek, Holly, South River City, and Windsor Park were estimated to experience the highest housing price increase rate. Among this list of top ten neighborhoods, Spicewood and Bull Creek intersects with Northwest neighborhoods, where will have a grand new opening of Apple Campus in 2022 [34]. Policymakers can closely look into these neighborhoods for having a high likelihood of suffering gentrification with steadily rising home values.

On the other hand, Chestnut, Highland Park, Johnston Terrace, Old West Austin, Robinson Ranch, North Shoal Creek, Windsor Road, Wooten, Walnut Creek—Pioneer Hill, and Del Valle East are forecasted to have the lowest housing price increase rate and suffer a decrease in housing price. Here, Del Valle is well known for accommodating affordable housing and having low-income households. And Johnston Terrace ranks as among the top dangerous places to live in, having a 1 in 15 chance of getting involved in crime [35]. One thing of note is that Highland Park, where ranks as the 7th affluent place in the states [36], was also predicted to experience a decrease in home values. These neighborhoods can be closely watched out for possible urban decline or flee. Perhaps, using a testing dataset concerning Project Connect and fitting the trained LSTM models to predict future home values would contribute to understanding more pending concerns.

VI. CONCLUSION

Big data is an essential component in a smart city. As a methodological framework, machine learning and deep learning analytics strategies enable planners to detect the patterns that hide in big data. As two methods are closely intertwined as a subset of AI, intermingling machine learning and deep learning would help planners have both human structured and computational artificial frameworks.

Our study has explored the potential of using ensemble machine learnings and LSTM deep learning on big real estate property data. By using regressor machine learning models, we can use big data to uncover new patterns—moreover, the bagging-based model outperformed boosting models in tracking property ownership variance. We suggest using LSTM when having sophisticated time-series data such as property value data. Adding assigned geographic information such as neighborhood, property type, sales restriction, and building features could be a valuable way to improve LSTM performance. Here, spatiotemporal geographic information is crucial to spot a change in home values from a given time to space.

The present study suggests the feasibility of applying machine learning and deep learning methods with big property data. However, we would like to point out that there is a tendency to apply these new tools because they are deemed suitable. This is similar to previous planning traditions of implementing zoning and building code regulation in cities. Machine learning and deep learning applications need to be carefully validated. Some studies skip explaining this part. Previous studies defined LSTM to produce a pretty accurate result with minor errors than the traditional times series model regarding housing price forecast [26].

Nevertheless, it is not clear whether the former studies validated the underfitting and overfitting of their LSTM models. Instead, they only discern error scores like RMSE. Either machine learning or deep learning both require an optimization process regarding hyper parameters. Explaining the structure of components and the optimization process to reduce the error score is an essential part. These facts should be wisely incorporated.

For future studies mixture of qualitative and 'new' quantitative analysis methods should be intertwined. First, understanding the process of gentrification concerning displacement should be prioritized. Second, home assessment, including property value, should incorporate contentious racial segregation pertinent to planning histories like redlining, racial covenants, and contract homebuyers. Categorizing neighborhoods by different race types and low-income households could be one option. Third, linking the prediction to newly rising property finance options and diverse tenure options that accommodate both renters and owners like ADU (accessory dwelling unit) and CLT (community land trust) may bring innovative insight [37].

VII. LIMITATION

The study has several limitations. First, due to confidential agreement, the explanation of obtained property big data is only briefly introduced. Second, there was a slight overfitting issue in LSTM. Thus, the performance could differ when using different parameters. Third, home sales data do not have the records from 2018 to study points. Fourth, areal interpolated factors would be more accurate if it was given by the individual neighborhood unit. Fifth, optimization on a number of estimators in Ensemble Learning modeling has not taken place. Future studies should consider these limitations.

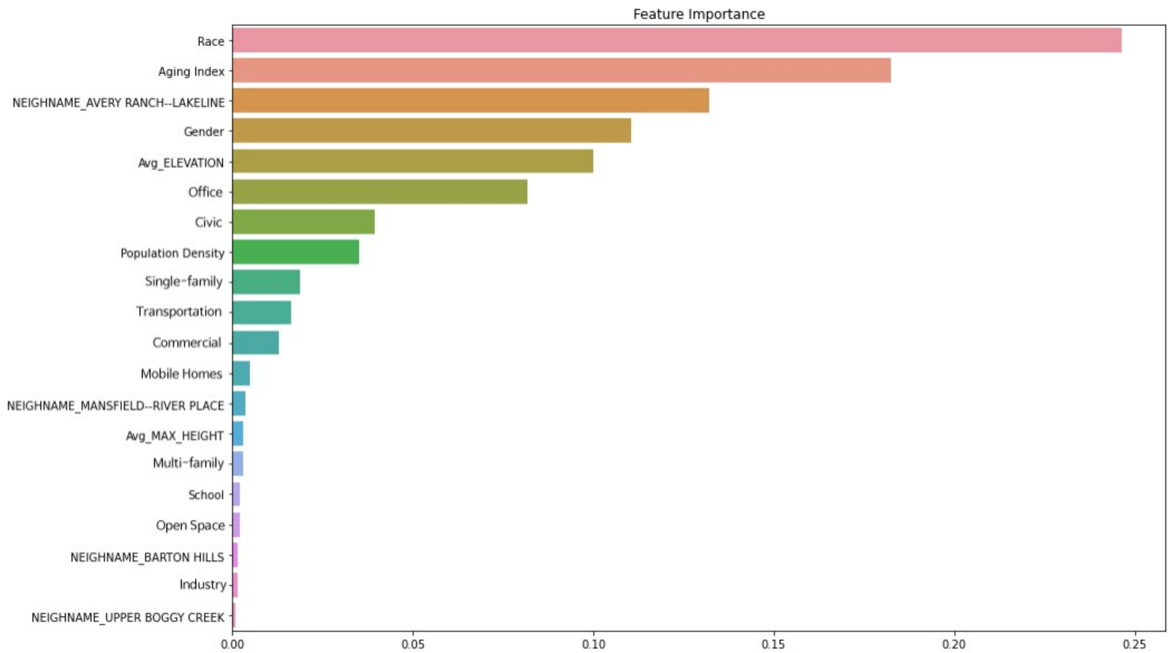

Fig. 6. Feature Importance of RF

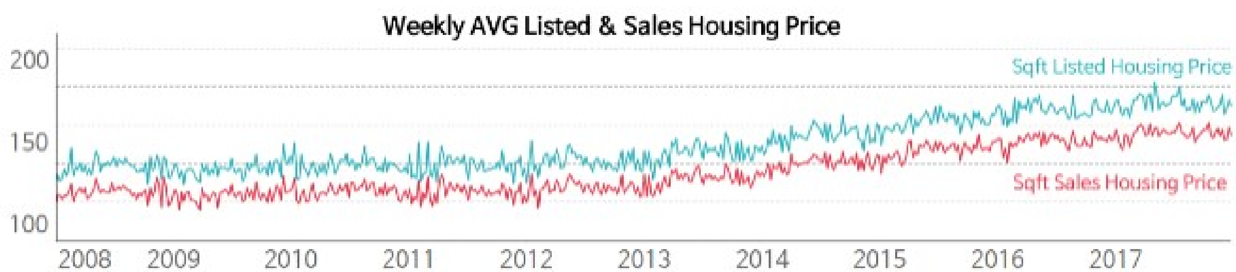

Fig. 7. Weekly Average Housing Price

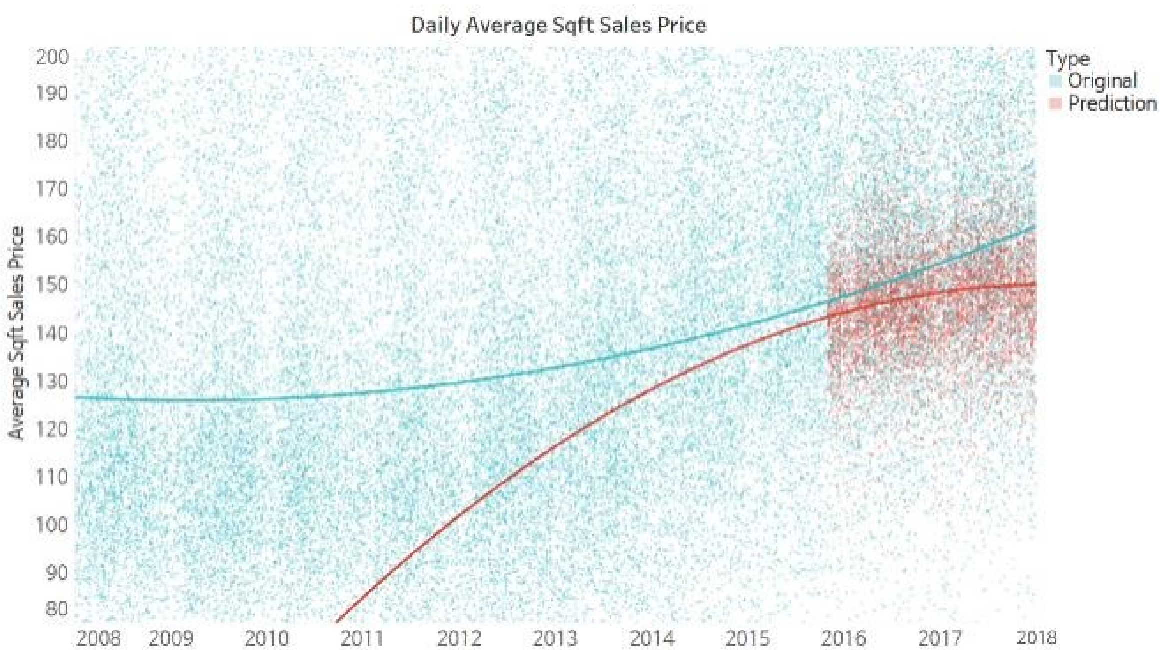

Fig. 8. LSTM Forecast Result

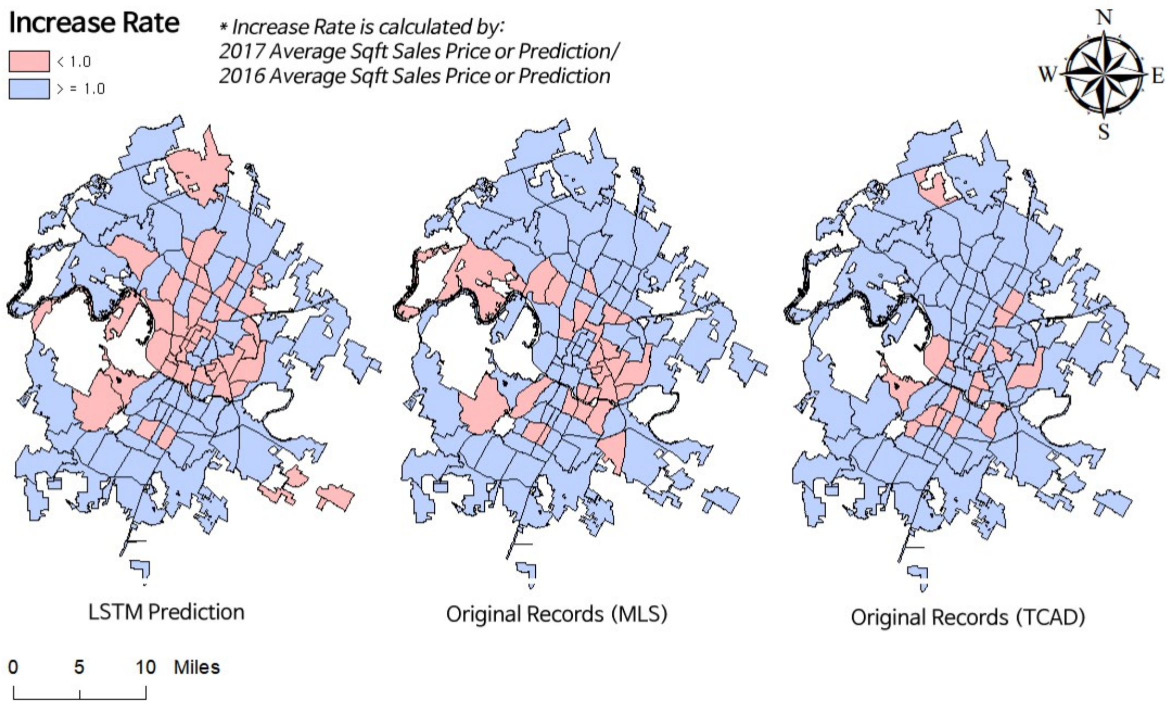

Fig. 9. Spatial Validation of LSTM Forecast

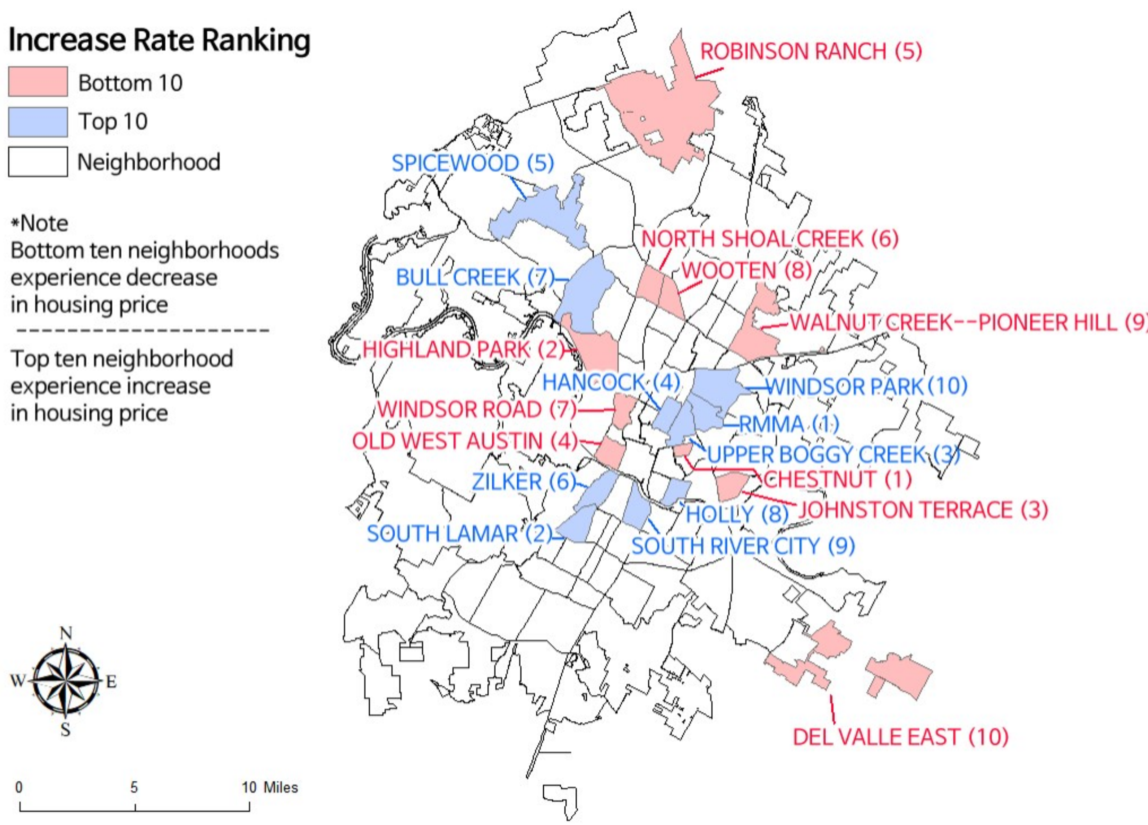

Fig. 10. Top 10 & Bottom 10 of Predicted Increase Rate

ACKNOWLEDGMENTS

This work has been funded through The Good System project by The University of Texas at Austin.

REFERENCES

[1] Calderón, César, and Luis Servén. "Infrastructure, Growth, and Inequality: An Overview." Policy Research Working Papers WPS7034 (2014).

[2] Zuk, M., Bierbaum, A.H., Chapple, K., Gorska, K., Loukaitou-Sideris, A., 2018. Gentrification, Displacement, and the Role of Public Investment. Journal of Planning Literature 33, 31–44. https://doi.org/10.1177/0885412217716439

[3] Way, Heather, Elizabeth Mueller, and Jake Wegmann. 2018. "Uprooted: Residential Displacement in Austin's Gentrifying Neighborhoods and What Can Be Done About It." The University of Texas Center for Sustainable Development in the School of Architecture and Community Development Clinic in the School of Law. https://sites.utexas.edu/gentrificationproject/.

[4] Lance Freeman, "Displacement or Succession? Residential Mobility in Gentrifying Neighborhoods," Urban Aairs Review 40, no. 4 (March 2005): 463-91.

[5] Bates, Lisa K. 2013. "Gentrification and Displacement Study: Implementing an Equitable Inclusive Development Strategy in the Context of Gentrification." Urban studies and planning faculty publications and presentations. Paper 83. URL: http://pdxscholar.library.pdx.edu/usp_fac/83.

[6] Denver Office of Economic Development. 2016. "Gentrification Study: Mitigating Involuntary Displacement." URL:https://www.denvergov.org/content/dam/denvergov/Portals/690/Reports%20and%20Studies/GENT%20STUDY%20052316.pdf.

[7] Goetz, Edward G., Brittany Lewis, Anthony Damiano, and Molly Calhoun. 2019. "The Diversity of Gentrification: Multiple Forms of Gentrification in Minneapolis and St. Paul." University of Minnesota Center for Urban and Regional Affairs. URL: https://www.cura.umn.edu/research/diversity-gentrification-multiple-forms-gentrification-minneapolis-and-st-paul.

[8] Armstrong, Jacob Benjamin. "Equity Impacts of Central Link Light Rail in Seattle, WA." PhD diss., 2019.

[9] Bullard, R. (2004). Equity in the United States *. Fordman Urban Law Journal, 31, 27.

[10] Lo, Puck, and Marcy Rein. "San Francisco Bay Area Transit Justice Movement Emerges." Race, Poverty & the Environment, vol. 17, no. 2, 2010, JSTOR, pp. 45–52.

[11] Karner, Alex, Jonathan London, Dana Rowangould, and Kevin Manaugh. "From transportation equity to transportation justice: within, through, and beyond the state." Journal of Planning Literature 35, no. 4 (2020): 440-459.

[12] Joice, P., 2014. Measuring Housing Affordability. Cityscape, 16, 1, 299-308. US Dept of Housing and Urban Development.

[13] Austin is the 10th 'fastest-gentrifying' city in the nation, says new report - Curbed Austin [WWW Document], n.d. URL https://austin.curbed.com/2017/1/31/14460906/east-austin-gentrification (accessed 11.14.21).

[14] Sak, H., Senior, A., Beaufays, F., 2014. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. arXiv:1402.1128 [cs, stat].

[15] N. Komninos, C. Kakderi, A. Panori & P. Tsarchopoulos (2019) Smart City Planning from an Evolutionary Perspective, Journal of Urban Technology, 26:2, 3-20, DOI: 10.1080/10630732.2018.1485368

[16] R. Kitchin, "The real-time city? Big data and smart urbanism," GeoJournal, vol. 79, no. 1, pp. 1–14, Feb. 2014, doi: 10.1007/s10708-013-9516-8.

[17] M. Batty, "Big data, smart cities and city planning," Dialogues in Human Geography, vol. 3, no. 3, pp. 274–279, Nov. 2013, doi: 10.1177/2043820613513390.

[18] K. Zhou, C. Fu, and S. Yang, "Big data driven smart energy management: From big data to big insights," Renewable and Sustainable Energy Reviews, vol. 56, pp. 215–225, Apr. 2016, doi: 10.1016/j.rser.2015.11.050.

[19] Bartneck, C., Lütge, C., Wagner, A., Welsh, S., 2021. What Is AI?, in: An Introduction to Ethics in Robotics and AI, SpringerBriefs in Ethics. Springer International Publishing, Cham, pp. 5–16. https://doi.org/10.1007/978-3-030-51110-4_2.

[20] Zhou ZH., 2009 Ensemble Learning. In: Li S.Z., Jain A. (eds) Encyclopedia of Biometrics. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-73003-5_293.

[21] Breiman, L. (2001) Random Forests. Machine learning. [Online] 45 (1), 5–32.

[22] Baldominos, A., Blanco, I., Moreno, A.J., Iturrarte, R., Bernárdez, Ó., Afonso, C., 2018. Identifying Real Estate Opportunities using Machine Learning. Applied Sciences 8, 2321. https://doi.org/10.3390/app8112321.

[23] Chen, X., Wei, L., Xu, J., 2017 House Price Prediction Using LSTM.

[24] Yu, L., Jiao, C., Xin, H., Wang, Y., Wang, K., 2018 Prediction on Housing Price Based on Deep Learning 12, 10.

[25] Y.-R. Lin and C.-C. Chen, "House Price Prediction in Taipei by Machine Learning Models," INTERNATIONAL JOURNAL OF DESIGN, vol. 8, no. 1, p. 7, 2019.

[26] H. Kim, Y. Kwon, and Y. Choi, "Assessing the Impact of Public Rental Housing on the Housing Prices in Proximity: Based on the Regional and Local Level of Price Prediction Models Using Long Short-Term Memory (LSTM)," Sustainability, vol. 12, no. 18, p. 7520, Sep. 2020, doi: 10.3390/su12187520.

[27] X. Chen, L. Wei, and J. Xu, "House Price Prediction Using LSTM," p. 6.

[28] S.-W. Bae, J.-S. Yu, and Dankook University, "Predicting the Real Estate Price Index Using Machine Learning Methods and Time Series Analysis Model," Korean Assoc Hous Polic Stud, vol. 26, no. 1, pp. 107–133, Feb. 2018, doi: 10.24957/hsr.2018.26.1.107.

[29] TRAVIS CENTRAL APPRAISAL DISTRICT [WWW Document], n.d. . TRAVIS CENTRAL APPRAISAL DISTRICT. URL https://www.traviscad.org/ (accessed 6.1.21).

[30] C.-J. Huang and P.-H. Kuo, "A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities," Sensors, vol. 18, no. 7, p. 2220, Jul. 2018, doi: 10.3390/s18072220.

[31] "Understanding LSTM Networks -- colah's blog." http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (accessed Apr. 10, 2021).

[32] "RMS Error." https://statweb.stanford.edu/~susan/courses/s60/split/node60.html (accessed Sep. 04, 2021).

[33] A. Baldominos, I. Blanco, A. Moreno, R. Iturrarte, Ó. Bernárdez, and C. Afonso, "Identifying Real Estate Opportunities Using Machine Learning," Applied Sciences, vol. 8, no. 11, p. 2321, Nov. 2018, doi: 10.3390/app8112321.

[34] "Apple expands in Austin," Apple Newsroom. https://www.apple.com/newsroom/2019/11/apple-expands-in-austin/ (accessed Sep. 04, 2021).

[35] Texas, 2020. 6 Most Dangerous Places to Live in Austin, Plus Safe Alternatives | Home By Four [WWW Document]. URL https://homebyfour.com/texas/6-most-dangerous-places-to-live-in-austin-plus-safe-alternatives/ (accessed 11.14.21).

[36] Locke, T., 2020. These are the top 10 richest places in America [WWW Document]. CNBC. URL https://www.cnbc.com/2020/02/21/bloomberg-richest-places-in-america.html (accessed 11.14.21).

[37] Wegmann, J., Schafran, A., Pfeiffer, D., 2017. Breaking the Double Impasse: Securing and Supporting Diverse Housing Tenures in the United States. Housing Policy Debate 27, 193–216. https://doi.org/10.1080/10511482.2016.1200109