A Hybrid Approach for Product Price Prediction

Rola M. Elbakly, Magda M. Madbouly, and Shawkat K. Guirguis

Abstract — In alignment with today’s online market needs, this study is concerned with a major topic present in any purchasing interaction, namely price prediction. It is of critical importance to both the buyer and seller to be able to estimate the proper price of the prospective merchandise with accuracy to ensure maximum profit and avoid any possible fraud situations. The purpose of our work is to test the deep learning network used in previous literature to predict prices from image data only, on an image data set of a more complex application, namely real estate listings. Also, a hybrid model is designed to improve price prediction accuracy by combining both numerical and image data predictions. The proposed model has achieved a Mean Squared Logarithmic Error (MSLE) of 0.05 and a R² of 0.91.

Key words — Deep Neural Networks, Price Prediction, Regression.

I. INTRODUCTION

This study provides an exciting opportunity to advance our knowledge of the usage of image data and its importance in solving the classic problem of price prediction. Especially, given the current abundance of image data due to the available high-definition cameras and continuous connection to the internet, the search for better exploitation methods of image data in different applications is becoming more necessary than ever. The findings should make an important contribution to the current literature by providing a hybrid solution to the price prediction problem that leverages the usage of image data to gain higher prediction accuracy.

The aim is to build on top of the latest research to explore the possibility of including image data in the process of price estimation, with the intent of improving the current price estimation accuracy. Currently, studies have shown that tree-based approaches deliver highest accuracy and best stability for numerical and categorical data. The objective is to use the aforementioned technique for numerical and categorical data in combination with Deep Learning and Computer Vision Techniques on image data to obtain a hybrid model that accomplishes the task of price estimation with the desired high accuracy and precision.

After this introduction the remaining parts of this paper are structured as follows: part two provides an extensive literature review of the latest research and studies done in the field of interest. The next part is concerned with the methodology used in this study followed by a presentation of the findings of the thesis, focusing on the analysis behind the results obtained. Finally, the conclusion gives a brief summary and critique of the findings and identifies areas of further research.

II. LITERATURE REVIEW

Several researches have shown how using various combinations of regression trees, support vector machines, and deep networks can effectively contribute to a more accurate price prediction discipline; [1]-[3] are only a few papers to mention. Models consisting of regression tree ensembles like Random Forest have been proven to predict prices with the highest rate of accuracy as in [4], [3] has demonstrated that using a complex Machine Learning algorithm usually yields better results than a linear regression model. This is due to smaller difference values between the actual commodity prices and the price values predicted by a complex Machine Learning Architecture. Yet, the drawback of using the latter-mentioned approach is a high time complexity.

Convolution Neural Network (CNN) is the main engine that allows using image data in the prediction of product images. Convolution is the repeated application of a set of filters on the input data and results in a feature map, which indicates the location and strength of a feature in an input image. CNN has the ability to learn a large number of filters in parallel specific to a training dataset, enabling the detection of highly specific features anywhere on the input image. Thus, implementing numerical as well as image data in a Hybrid Deep Learning Architecture is believed to attain more accurate prediction results.

Previous studies in price prediction mostly used Random Forest (RF) as the main technique due to its accuracy, robustness, universality, and feature importance extraction property. Research carried out by [4] uses Linear Regression and Random Forest to predict car prices in three different models: model for a certain car make, model for a certain car series and a universal model. The results showed that Random Forest achieved higher performance for the universal model and had stable but less than ideal results for certain car make and certain car series models compared to Linear Regression. Another study by [5] compares the following three machine learning techniques: RF, Gradient Boosting Machine (GBM), and Support Vector Machine (SVM) and again RF achieves more accurate results together with GBM.

Other studies have taken a hybrid approach to price prediction, in [1] for instance, Artificial Neural Networks (ANN), SVM, and RF are used individually and as an ensemble to predict car prices. And the results revealed that any single technique scored an accuracy less than 50%, while the combination of the machine learning methods had an
average accuracy of 87.38%. The paper by [6], highlights the power of hybrid approaches as it uses a combination of Lasso Regression and GBM to attain the best results. The study in [2], proved that a hybrid regression accomplishes better accuracy even better than RF at the cost of higher time complexity and a more complicated architecture.

In the 2017 study of [4], a deep learning architecture called “PriceNet” was designed to predict cars and bikes prices using images only. The model consists of four blocks of convolutional and pooling filters, followed by fully connected layers. The results of this architecture were compared by the results of other techniques like Linear Regression and Transfer Learning and “PriceNet” outperformed both of the mentioned methods.

Another interesting study conducted in [7], collected a new data set of California Houses, which will be used later in our experiments and will be discussed further in the next section, and proved that using a one 4 node layer CNN on image and numerical data enhances the prediction accuracy of a numerical data only model. Also, it was proved that NN achieves better results than SVM using the exact data set.

The literature review above shows that tree-based approaches, RF for instance, ensure robust and accurate results given that the available dataset consists of numerical and categorical data only. On the other hand, in order to achieve satisfying outcomes using image data a fully connected CNN has to be applied. Hence, this conclusion opens up the opportunity to examine the possibility of combining both above mentioned techniques in a hybrid model to achieve more accurate price estimations.

III. METHODOLOGY

A. Data Analysis

To conduct the experiments two different data sets were used, the first is a collection of 535 records of California houses collected by [7] in 2016 paper. The second data set is a set of more than 12 million records of Brazilian housing advertisements provided online for a data science competition in 2019 [8]. In the California Houses data set each house is represented by four images: frontal view of the house, bedroom, kitchen, and bathroom. And five numerical attributes: number of bedrooms, number of bathrooms, area of the house, zip code of the house location, and price of the listing. The data set has no missing values and the images are consistent, which is favorable for our experiments. The downside, though, is that the size of the data set is relatively small for deep learning applications, so to compensate for this deficiency we augment our data by using crop, flip, scale, rotate and gaussian blur transformation randomly on the images resulting in a total of 2000 data points. To prepare this data set, first the visual data was separated from the numerical data. Second, the four images of each house listing must be pasted together to form a single tiled image, as shown in Fig. 1. This step is essential to avoid unnecessary complications of the network, which can lead to inaccurate results. Lastly the tiled image is resized to 224 × 224 pixels.

As for the Brazilian Houses data set, it contains the following twenty-four features: id, floors, rooms, created on, collected on, property id, operation, property type, place name, place with parent name, country name, state name, geonames id, currency, description, title, lat lon, lon, lat, surface covered in m², surface total in m², expenses, price, image thumbnail. In contrast to the first data set, the volume of data here is immense, as mentioned before more than 12 million records are available, but there is also a great amount of missing, redundant, and inconsistent data. Hence, as a start a random sample of 5000 records was chosen to represent our data set. Next, columns containing text (place name, place with parent name, description, title), or GPS coordinates (lat lon, lon, lat, geonames id), or date of advertisement (created on, collected on) were removed as these features are not of concern to this study. Afterwards, features with constant values like country name, currency and operation were also removed, as these do not contain any useful information and can cause unexpected outcomes. Lastly, columns with exceptionally high rates of missing data, which in this case were floors and expenses, and rows with missing values or repeated property id were dropped. As a result, the final data set used is composed of the following 8 features: id, rooms, state name, property type, surface covered in m², price, and image thumbnail. And again, as done before for the California Houses data set the image data was separated from the numerical and categorical data.

B. Research Design

This study consists of two main experiments, the goal of the first experiment is to test the price estimation accuracy achieved by the images only of each of the data sets used. While the second experiment is to test the overall accuracy of the newly created hybrid system combining both numerical and visual features. The first experiment uses transfer learning to predict the price of real estate directly without the help of any numerical attributes. More specifically, the SqueezeNet introduced by [9] in 2016 is used for this experiment with the modifications explained in [10], as Fig. 2 shows. SqueezeNet is a deep learning architecture designed as a more compact replacement for AlexNet [11].
In fact, it has 50 fewer parameters and performs 3 times faster, which makes it ideal for the data sets at hand. The network consists of “squeeze” and “expand” layers described inside a fire module, which is repeated several times inside the network. In order to fit our application, the output layer of the network is replaced by a fully connected layer followed by a single layer activation output to achieve a single output for the regression model.

Building on top of the first experiment, the second one uses the trained SqueezeNet regression network in combination with Random Forest network to form a hybrid system for predicting the price of real estate. Combining regression predictions usually involves using statistical methods. The two main ones used are Mean Predicted Value and Median Predicted Value, in weighted or unweighted form. The mean Predicted Value is suitable when the distribution of predictions is gaussian or nearly gaussian, while the Median Prediction Value is suitable when the distribution of predictions is unknown. But as the hybrid model at hand has only two predictions outputted from the two base models; the Mean Predicted Value was used to combine the results into the final prediction Fig. 3.

IV. EXPERIMENTAL RESULTS

A. Evaluation Metrics

As in any Machine Learning and Deep Learning experiment, the choice of the loss function is a crucial step. In this study, the selected loss function is the Mean Squared Logarithmic Error (MSLE), which is given by (1).

$$L(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^{N} (\log(y_i + 1) - \log(\hat{y}_i + 1))^2$$  \hspace{1cm} (1)

The main advantage of using MSLE is that it is concerned with the percentile difference between log-transformed actual and predicted values, which in turn makes it less punishing for a model that predicts large values as is the case in this study.

Also, for the sake of completeness, better understanding of the system performance, and easier comparison to former regression systems the $R^2$ score was as well calculated and included in the results tables. The $R^2$ score also known as the coefficient of determination measures the amount of variance in the predictions as explained by the formula below. The values of the $R^2$ score range from 0 to 1, where 1 is the perfect score. In other words, the closer the $R^2$ value is to 1 the better the performance of the system at hand.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$  \hspace{1cm} (2)

B. First Experiment

As opposed to all expectations, the first data set which has the larger amount of data points, actually yielded the worst results possible. The MSLE was around 0.9 and the R-squared score was almost zero. Hence, this data set was discarded, and no further experimentation was made using it. All the possible causes of this high error outcome will be analyzed further in the discussion section.

The second data set, on the other hand, gave satisfying results. The MSLE during the training phase and the R-squared score were on average 0.06 and 0.89, respectively. While at the testing phase the results were 0.075 and 0.78. As shown in Table I. A comparison of the two results reveals the importance of the quality of data in the prediction of prices using images only.
C. Second Experiment

The hybrid system gave the prediction error shown in the table below. The outcome consists of a MSLE of 0.05 and $R^2$ score of 0.91 as Table II shows, which is an outstanding result, given the complexity of the problem at hand. These findings suggest that each of the sub-models extracted different features than the other, which in turn lead to a better accuracy when combined together in the hybrid model.

<table>
<thead>
<tr>
<th>TABLE I: THE RESULTS OF FIRST EXPERIMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error</td>
</tr>
<tr>
<td>MSLE</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE II: THE RESULTS OF SECOND EXPERIMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error</td>
</tr>
<tr>
<td>MSLE</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE III: COMPARISON WITH EMAN AHMED MODEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
</tbody>
</table>

V. DISCUSSION

A. First Experiment Analysis

As mentioned before, the first data set resulted in a very high prediction error rate despite the abundance of data records. Possible reasons for this occurrence have been investigated and the obvious explanation is that the image data provided was inconsistent and random. For example, a record will give a view of the building from the outside, another will show the living room Fig. 4. there is absolutely no condition on what part of the listing this image could represent. Therefore, the deep learning network was not able to detect useful features during training to be able to predict correct prices in future testing. This interpretation was further reinforced when the second more organized and consistent data set yielded a much lower error rate.

The images provided by the second data set strictly represented four parts of the house: the bathroom, the bedroom, the living room, and the outside of the house in a frontal view, in that exact order tiled together in one image. This led to much less confusion of the neural network and consequently better prediction accuracy. This experiment shows that the network represented in [10] does achieve price prediction results with high accuracy, given that the input image data set is consistent. Which is well explained by the fact that the experiments conducted in the paper used two different data sets both with solid white backgrounds, and side view images. Obviously, this was not the case in the first data sets used in this study, which justifies the low prediction accuracy of the first data set and the better accuracy of the second data set.

B. Second Experiment Analysis

The hybrid system, on the other hand, generated generally exceptional results, namely as Table II shows above both the SqueezeNet part and the RF part of the system, each have a lower accuracy than the hybrid system that combines them both. This proves the hypothesis presented in this thesis to be correct. The reason for this outcome is that the image data and the numerical data each hold different features about the product on sale, so by training the two different neural networks on each data type separately and then combining them together a better accuracy is reached. In comparison to previous research conducted on the exact same data set our hybrid system achieves an $R^2$ value of 0.91 and the model presented in [7] archives an $R^2$ value of 0.92 as shown in Table III. Despite the slight decrease in accuracy compared to [7], our model is believed to have better scalability, as given a large enough data set the squeezeNet used in our model is more capable of producing accurate results than a simple four node layer CNN.

VI. CONCLUSION

In summary, our study was designed first to investigate the possibility of predicting the prices of real estate listings using deep learning networks on image data only. And second, to build a hybrid deep learning model, which uses numerical and image data to predict the prices of the houses with more accuracy than a model that uses numerical data only or image data only. The study has shown that with the usage of SqueezeNet deep learning network the price of a house listing can indeed be predicted with very good accuracy from image data. The second finding was that the quality of the used data set plays a major role in the results obtained by the network, meaning that a well-prepared data set with images taken from the same view in the same order is critical in attaining the desired prediction accuracy. Thus, the limitation of this scientific work is that the experiment is made under a controlled environment, as explained before the data set was carefully prepared before training and testing. And the other data set used containing inconsistent and miscellaneous images yielded the worst accuracy possible, as previously discussed in detail.

The hybrid model presented in this scientific work, confirms previous findings and contributes additional evidence that suggests the immense potential included in image data, which can increase the accuracy of a deep learning network prediction. As the prediction accuracy of the hybrid model accomplished better results than when either of the subsystems is used singularly for predictions.

Further research is required to test the model against
different data sets which might include a different product than the one used in this study, multiple various products at the same time, or different views of the same products. The latter is highly encouraged because it is a positive step to finding ways to improve the limitations that appeared in this research. Also, it may act as a preparatory stage to allow the user to estimate a price of an item simply by sending a phone-taken picture or pictures of it. Furthermore, the data sets available suitable for this type of experimentation are limited to only two, which was one of the challenges faced during this work. So, it would be of great value to collect a new large data set that includes both image data and numerical data. Then it would be highly expected to achieve better results and higher accuracy.

**FUNDING**

No Sponsoring or financial support were received to cover the costs to do this research.

**CONFLICT OF INTEREST**

Authors declare that they do not have any conflict of interest.

**REFERENCES**


