BRIN
BADAN RISET
DAN INOVASI NASIONAL

IC3INA

G20
INDONESIA
2022

PROCEEDING

The 2022
IC3INA
The 9th INTERNATIONAL CONFERENCE ON COMPUTER, CONTROL, INFORMATICS AND ITS APPLICATIONS

Digital Transformation Towards Sustainable Society For Post Covid-19 Recovery

Online Conference :
22-23 November 2022

ISBN 978-1-4503-8524-4

ICPS
Published by ACM

The Association for Computing Machinery
1601 Broadway, 10th Floor New York
New York 10019-7434

IC3INA 2022
ACM ISBN: 978-1-4503-9791-9

The 2022 9th International Conference on Computer, Control, Informatics and Its Applications (IC3INA 2022)

Table of Contents

Preface

On behalf of the IC3INA 2022 organizing committee, I am honored and delighted to welcome you to the ninth of 2022 International Conference on Computer, Control, Informatics, and its Applications (IC3INA 2021) with the theme: "Digital Transformation Towards Sustainable Society For Post Covid-19 Recovery". The theme is perfectly in line with these two important reasons. First, the acceleration of digital transformation caused by change of lifestyle due to Covid-19 pandemic. This momentum is perfect to forge partnerships and cooperation, and to share research progress between research institutions, campuses, and small to large scale industries in order to optimize the management and the use of the available resources into multi-disciplinary science, to attain sustainable development. And second, with the world recession is around the corner, the theme never becomes more important than today for us to exchange ideas on how digital transformation could also tackle future challenges and ensure that everyone has the best (access) to participate and benefit from the digital transformation of our economy.

Our technical program is rich and varied with 5 keynote speakers and 76 technical papers split between 15 parallel oral sessions in two days' virtual conference. From 112 reviewed papers, we have accepted 76 papers, which means 67.9% of acceptance rate and 32.1% rejection rate. Both accepted paper number and rejection rate increase compared to IC3INA 2021 (45 accepted papers, 10% of rejection rate). IC3INA is committed to maintain and improve the quality of accepted papers. The accepted papers are separated into Computer and Networking (10 papers), Control (6 papers), Informatics and Data Processing (51 papers), and Information Communication Technologies (ICTs) Applications (9papers). The IC3INA 2022 has been approved and sponsored by the Association for Computing Machinery (ACM) as the global proceeding which is indexed by the Scopus.

Regarding the conference, this is the second year that we are holding the IC3INA annual event in a virtual format, with no audience physically in the

room. Following our previous success in 2021, this year we use a conference management system (Indico) to ensure that we all could optimally engage and present the latest innovations and developments, and to exchange ideas and various aspects of advances in computer technology, control, informatics and their applications.

As a conference chair of IC3INA 2022, I know that the success of the conference depends ultimately on the many people and the committees who have worked with us in planning and organizing both the technical program and the supporting team arrangements. We thank the steering committees for their wise advice and brilliant suggestions from the organizing committee. The technical program committee (TPC) for their thorough and timely reviewing of the papers, and our ad-hoc teams who have helped us to keep down the costs of IC3INA2021 for all participants. Recognition should go to the organizing committee members who have all worked extremely hard both individually and team efforts for the details of important aspects of this virtual conference program and the other related activities. Last but not least, thank you very much for your participation and support of our International Conference on Computer, Control, Informatics, and Its Applications (IC3INA 2022).

General Chair of IC3INA 2022,

Dr. Purnomo Husnul Khotimah M.T.

Conference Committees

The 2022 International Conference on Computer, Control, and Informatics, and Its Applications (IC3INA 2022)

"Digital Transformation Towards Sustainable Society For Post Covid-19 Recovery"

Steering Committee

Chair

Dr. Budi Prawara, Chairman of Research Organization for Electronics and Informatics (ROEI), BRIN

Vice Chair

Dr. Esa Prakasa, Head of Research Center for Data and Information Sciences (RCDIS), BRIN
Dr. Rifki Sadikin, Head of Research Center for Computing (RCC), BRIN
Dr. Anto Satriyo Nugroho, Head of Research Center for Artificial Intelligence and Cyber Security (RCAICS), BRIN

Organizing Committee

Chair

Dr. Purnomo Husnul Khotimah, M.T. (RCDIS)

Vice Chair

Arafat Febriandirza, Ph.D (RCDIS)

Technical Program Committee

General Conference Chair

Dr. Purnomo Husnul Khotimah, M.T. (RCDIS)

Vice Conference Chair

Arafat Febriandirza, Ph.D (RCDIS)

Secretary

1. Raden Sandra Yuwana S.T., MT (Main Secretary) (RCDIS)
2. Wachyuni A.Md (Secretary I) (RCDIS)
3. Vicky Zilvan, S.Kom., M.T. (Secretary II) (RCDIS)
4. Yulia Aris Kartika, M. Kom. (Secretary III) (RCC)
5. Ariani Indrawati, M.TI (RCC)
6. Aris Yaman, M.Si (RCC)

Treasurer

1. Ekasari Nugraheni M.T. (Treasurer I) (RCDIS)
2. Riyo Wardoyo, M.T. (Treasurer II) (RCDIS)

Public Relation

1. M. Yudhi Rezaldi, Ph.D (PR Coordinator) (RCDIS)
2. Andre Sihombing, S. Kom., M. Sc. (RCDIS)
3. Ir. Dianadewi Riswantini, M. Sc., M.B.A. (RCDIS)
4. Ayu Shabrina,M.Sc. (RCC)
5. Dewi Suci Rafianti, S.Kom. (RCAICS)

IT Support

1. Andria Arisal, MEDC (IT Coordinator) (RCDIS)
2. Andri Fachrur Rozie, S.Kom. (IT Support) (RCDIS)
3. Aang Gunawan Sutyawan, M.T (RCDIS)

TPC & Keynote Speaker Chair

1. Dr. Eng Hilman Ferdinandus Pardede, S.T., MEICT (TPC Coordinator) (RCDIS)
2. Dr. Shidiq Al Hakim (RCDIS)
3. Akbari Indra Basuki, M.T. (RCDIS)

Publication, Documentation, & Program Book Chair

1. Dr. Foni Agus Setiawan, M.Kom. (Coordinator) (RCDIS)
2. Dr.Tr. Lindung Parningotan Manik (Publication, Documentation, & Program Book) (RCDIS)
3. Dr. Inna Syafarina, M.Si (RCC)
4. Siska Pebiana (RCAICS)
5. Muhammad Thufaili Imdad, S.T, S.Si, M.T (RCAICS)
6. Auliati Nisa (RCAICS)

Scientific Committee

1. Nguyen Phong Hoang, PhD., Department of Computer Science, The University of Chicago
2. Agung Toto Wibowo, PhD., Cyber Physical Systems Research Group, School of Computing, Telkom University
3. Dr. Kemas Muslim Lhaksmana, Master of Computing, School of Computing, Telkom University
4. Samar Helou, PhD., Global Center for Medical Engineering and Informatics, Osaka University, Japan
5. Luciano Santos, PhD., Fitting Cloud, Japan
6. Dr. Eng. Ahmad Afif Supianto (RCDIS)
7. Rezzy Eko Caraka ,S.Si.,M.Sc(RES).,PhD. (RCDIS)
8. Dr. Heru Susanto, S.Komp., MSc (RCDIS)
9. Dr. Didi Rosiyadi, M.Kom (RCDIS)
10. Dr. Wiwin Suwarningsih, M.T. (RCDIS)
11. Dr. DIPL. ING (FH) Anne Parlina, M.P. (RCDIS)
12. Dr. Eng. Lia Sadita, S.Kom, M.Eng. (RCDIS)
13. Dr. Bambang Setiadi, ST, MEng (RCDIS)
14. Risnandar, PhD (RCDIS)
15. Dr Dian Andriana (RCDIS)

16. Dr. Foni Agus Setiawan(RCDIS)
17. Al Hafiz Akbar Maulana Siagian, Ph.D.(RCDIS)
18. Dr.Tr. Lindung Parningotan Manik (RCDIS)
19. Dr. Shidiq Al Hakim (RCDIS)
20. Arafat Febriandirza, Ph.D (RCDIS)
21. Dr. Ambar Yoganingrum (RCDIS)
22. Dr. Eng. Budi Nugroho (RCDIS)
23. M. Yudhi Rezaldi, Ph.D (RCC)
24. Dr. Arnida L. Latifah(RCC)
25. Dr. Inna Syafarina, M.Si(RCC)
26. Zaenal Akbar, M.Kom., Ph.D (RCC)
27. Dr Asril Jarin (RCAICS)

Published in the ACM ICPS

ISBN 978-1-4503-8524-4

BRIN
BADAN RISET
DAN INOVASI NASIONAL

ICPS
Published by ACM

Housing Price Prediction Using a Hybrid Genetic Algorithm with Extreme Gradient Boosting

Maria Ulfah Siregar*
Pahlevi Wahyu Hardjita†
Farhan Armawan Asdin‡
maria.siregar@uin-suka.ac.id
UIN Sunan Kalijaga
Sleman, Daerah Istimewa Yogyakarta, Indonesia

Dewi Wardani§
Ardhi Wijayanto¶
dww_ok@uns.ac.id
ardhi.wijayanto@staff.uns.ac.id
Universitas Negeri Sebelas Maret
Surakarta, Indonesia

Yessi Yunitasari
Universitas PGRI
Madiun, Indonesia
yessi@unipma.ac.id

Muhammad Anshari
Universiti Brunei Darussalam
Brunei, Brunei Darussalam
anshari.ali@ubd.edu.bn

ABSTRACT

Predicting property prices provides a better service for customers to evaluate and estimate price movement before their purchases. Some features including OverallQual and GrLivArea, which were selected when applying GA, become important features that can influence property prices. This research proposes a hybrid Genetic algorithm combined with the Extreme Gradient Boosting algorithm to predict real estate housing prices. The proposed scheme is evaluated by Root Mean Square Error, processing time, and the number of deleted features. The proposed scheme has been compared with the sole Extreme Gradient Boosting. The experimental results show that the proposed scheme produces the smallest root mean square error value of 0.129 compared to 0.133 of the sole Extreme Gradient Boosting. Furthermore, the predicted time of the proposed scheme is much better than the sole method.

CCS CONCEPTS

• **Computing methodologies → Machine learning algorithms**.

KEYWORDS

feature selection, RMSE, crossover, mutation

ACM Reference Format:
Maria Ulfah Siregar, Pahlevi Wahyu Hardjita, Farhan Armawan Asdin, Dewi Wardani, Ardhi Wijayanto, Yessi Yunitasari, and Muhammad Anshari. 2022. Housing Price Prediction Using a Hybrid Genetic Algorithm with Extreme Gradient Boosting. In *The 2022 International Conference on Computer, Control, Informatics and Its Applications (IC3INA 2022), November 22–23, 2022, Virtual Event, Indonesia.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3575882.3575939

*All authors contributed equally to this research.

IC3INA 2022, November 22–23, 2022, Virtual Event, Indonesia

ACM ISBN 978-1-4503-9790-2/22/11…$15.00
https://doi.org/10.1145/3575882.3575939

1 INTRODUCTION

The real estate and property sector could be categorized as quite profitable but with a high risk because sales are pegged to market prices with high and low buyers' interest. Real estate and property businesses consist of land, buildings, and everything attached to the land [1].

During the Covid-19 pandemic, many companies were downsizing their firms due to economic uncertainties and business turmoil. These facts have encouraged people to seek alternative jobs. Real estates and housing businesses promise high returns for business and primary needs for almost everyone [2].

On the other hand, this pandemic is also blamed for the slow increase or even decrease in house prices. Bank Indonesia (BI)/ Indonesian Bank recorded that the prices of real properties, especially residential ones, relatively stagnate in the fourth quartile of 2020 [3]. This slowdown is predicted to stay until the first quartile of 2021.

By knowing the interest of the market, the company could predict these prices. There exist algorithms and tools performing prediction. However, it is always a necessity to get accurate results. Therefore, this paper examined the combination of the eXtreme Gradient Boosting (XGBoost) algorithm and the Genetic Algorithm (GA) and compared both algorithms independently. By doing so, it is expected to find the better accuracy of those algorithms for this prediction. We aim to get a house that suits the wishes of the buyer based on price and other criteria.

The GA is a prediction method which has advantages such as dealing with non-stationary data [4]. Another advantage is a modelled phenomenon which is not necessary to define since this model will be enhanced gradually during the process. This method is a kind of adaptive heuristic algorithm that can be integrated to select the best features for prediction, rather than spend much time searching for the ideal features using the greedy algorithm [5]. It is because GA can be used for global optimization, which determines the optimal features [6]. GA is a search technique used in computer science to find approximate solutions for optimization and search problems which were created to mimic some of the processes observed in natural evolution [7]. Furthermore, GA is

a strong stochastic algorithm that is quite successful in machine learning relating to those problems.

XGBoost is one of the methods in machine learning, which uses the ensemble method for a regression and classification algorithm. This method is a variant of the algorithm of the Tree Boosting System, a scalable end-to-end of this algorithm [8]. Tree boosting is a greatly effective and broadly used machine learning method.

XGBoost has been implemented or reviewed in many research. For example, XGBoost has been a winner of challenges hosted by Kaggle, 17 out of 29 solutions which eight these solutions were obtained from the implementation XGBoost alone [8].

Another application of XGBoost is on medical problems such as in [9]. In their research, this algorithm is used to classify the fMRI (functional MRI) in epileptic patients and healthy subjects. The research also aims at determining the best combinations of features that show the most predictive power. Features here are region, hemisphere, and task which focused on filter and wrapper methods. The finding of this research is a special mix of features, the SEM_LH BA_47-21 was the most compelling to identify patients. XGBoost is also applied to predict gene expression value [10]. The profiling of these values is used to reflect the health of the body, diagnose genetic diseases, and others. By using the XGBoost method, the cost to profile genome-wide expression could be reduced. The performance of XGBoost was compared to the D-GEX algorithm, linear regression, and KNN methods. The result is that the XGBoost model has a lower overall error than the three models.

XGBoost has also been applied to predict environmental resources. One of these applications is the prediction of the levels of groundwater in Selangor, Malaysia [11]. There, XGBoost, ANN, and Support Vector Regression have been tested and evaluated in order to obtain accurate groundwater level prediction. The precise input variables are critical for the model's achievement; thus cross-correlation is used to find the match between the target variable and the delayed values of inputs. The results of this research are XGBoost has the highest mean absolute error (MAE) and root mean square error (RMSE). Another application is the prediction of PM2.5 concentration every hour in Tianjin, China [12]. This prediction was modelled using data-driven models. The application of XGBoost was compared with other established data-mining models, such as the Random Forest, Multiple Linear Regression, Decision Tree Regression, and Support Vector Machine for Regression. The result is the RMSE and MAE of XGBoost which is better than the four compared models. It is claimed that the XGBoost is more efficient to predict the hourly PM2.5 concentration.

Although much research has been conducted using XGBoost, we have never found any of these that combined with GA and relate to house price prediction. Thus, in this paper, we present our research on applying a combination of XGBoost and GA to this problem. Algorithm implementation will be processed in the Python programming language. Following this, our contribution is to identify the better features of the problem of house pricing prediction by implementing the combination of GA and XGBoost using the Kaggle dataset and machine. Features in our study are factors regarding the House Price Index (HPI), which is commonly used to estimate changes in house prices. Based on this index, house prices are highly correlated with other factors such as location, area, population, and other information.

Table 1: The XGBoost's Parameter

Parameter	Value
n_estimators	500
learning_rate	0.05
subsample	0.5
colsample_bytree	0.5
max_depth	3
objective	reg:linear
gamma	0
alpha	0
lambda	2

2 MATERIAL AND METHOD

There are several stages of our research. A brief description is given as follows.

2.1 Data Collection

The data is from Kaggle.com in the form of a dataset of Ames Housing compiled by Dean De Cock. The following is the data source link https://www.kaggle.com/c/house-prices-advanced-regression-techniques [13]. This dataset consists of 4377 entries for houses in Ames, Iowa. A number 0f 2917 data if for training, and 1460 data is for testing. There are 23 nominal, 23 ordinal, 13 discrete, and 20 continuous variables for each property.

2.2 Conducting Pre-processing

The next step is the implementation of pre-processing over the data. This stage includes the implementation of replacing the missing values data and cleansing the data.

2.3 Performing Feature Selection using GA

The feature selection stage is carried out to reduce complexity and avoid unused/ irrelevant features so as to achieve a better accuracy rate of the result [14]. In the feature selection phase, the method which is used is the GA. Combining GA and XGBoost differs our research considerably from the implementation of the sole XGBoost algorithm.

2.4 Algorithm Implementation

The next stage is to process data from the XGBoost algorithm and GA. Parameters are depicted in Table 1.

For the XGBoost model, we used XGBRegressor so the training model and testing data could use all the features or some of the features as the result of applying GA. Thus, we fed the selected features obtained by GA to the XGBoost. After training the data model, the sale price is predicted over the testing data. These sale prices are then sent to Kaggle to obtain scores of RMSE.

Our research could be categorized as research which describes technical results. Among the four types of technical results, the fitting type is a system construct. In this case, we set a research method by feeding the XGBoost algorithm on the results of GA.

2.5 Evaluation of Results

After implementation, the accuracy testing phase of the prediction results using RMSE is conducted. In this case, better accuracy will be found for the prediction of the price of each house.

3 RESULTS AND DISCUSSION

This section displays the results obtained after carrying out the steps described in the method section. Through the data collection stages, the training model data is on 81 columns and the testing data is on 80 columns.

We applied three methods of replacing missing values. We replaced the missing values from data that do not number/integers using their modus. Meanwhile, for the number/ integer data values, we replaced the missing values with their mean or median. Furthermore, values of "NaN" will be replaced with "0" or "None".

The next process is data cleansing. There were as many as 78 training model data features and 78 data testing features within 329 columns. These 329 columns were obtained after pre-processing the data using One-hot Encoding method in which every categorical data will be changed into several columns based on their categorical values. We used this method to ease us in the selection phase. In the cleansing step, we deleted columns with many "NaN" values or columns with only one cluster since it indicates the less important or we could say that they are outliers. First, we plotted data using a Scatter plot, then we delete the outliers. Figure 1 shows a correlation among features. Based on Fig. 1, the two features that influence the price of houses most are OverallQual and GrLivArea. Previous research has found that by performing clustering, outliers could be detected and removed [15][16]. In this research, we found that there are two outliers.

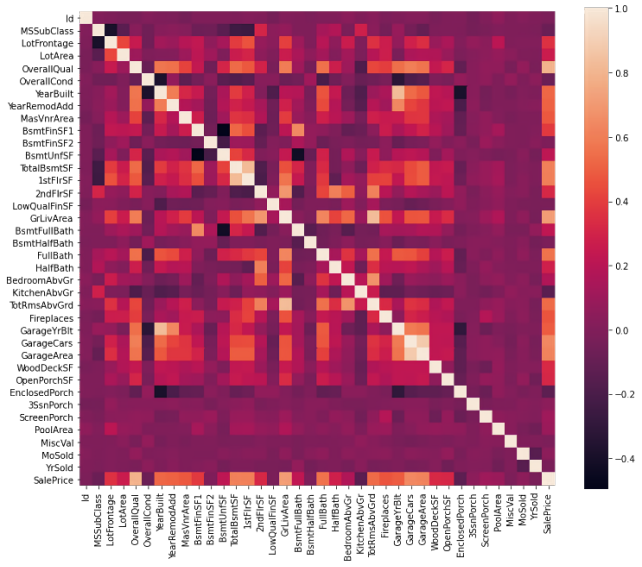

Figure 1: A heat map correlation among columns

The correlation that is shown in Fig. 1 is to correlate the features of the numerical types to other types of features. It ranges from -0.5 to 1, from darker colour to brighter colour, respectively. The brighter colour represents the strong correlation between the feature in the X ordinate and to feature in the Y ordinate. For example, the correlation between SalePrice and OverallQual is a bright colour tile. It means that both have a strong correlation with each other.

The next process is the feature selection using GA, which is a single-point crossover and flip mutation to reduce columns in the data and speed up the training model process. The mutation rate is defined as 30%. Our GA accepts input: all columns from the training model data except the target column (the column that contains phenotypes/ features of the house), targets (SalePrice), number of populations, and number of generations. A chromosome will represent an individual binary string whose length of this chromosome suits the number of features. Each chromosome represents a feature. Fitness values are obtained from the RMSE values of each population. Thus, the one with the smallest value of RMSE will be the optimum solution for the problem. After several iterations and this number reach the maximum number of iterations, the global optimum is reached [17][18]. For training the data, we used all the training datasets, then, we split the training dataset using the 80:20 method; 80% of the data will be for training, and 20% of it will be for validation.

The value predictive between XGBoost and GA is obtained after processing the trained data model and followed by the tested data. We used also all the test datasets. Here we used the RMSE value, processing time, and the number of eliminated features as the benchmark to find the best algorithm of the two algorithms. The RMSE value is obtained by first inputting the predicted results of the data test for each model into Kaggle's machine. The number of eliminated features and GA's processing Time are gained from the training process in the GA, whereas the Predicted Time is from the test process.

We conducted 26 experiments with several combinations of numbers of population and generation, in which one of the experiments is the model of XGBoost solely. The comparison is given in Table 2. GA_P25G25 means that the GA parameters used in the experiment are: 25 for the number of populations and 25 for the number of generations.

Based on our experiments, both XGBoost and GA_P20G15, and XGBoost and GA_P25G25 have a better result of RMSE than the other algorithms after being examined by using the Kaggle machine. The best predicted time is obtained on the XGBoost and GA_P15G25 algorithm. The best GA processing time is on XGBoost + GA_P5G5. The last evaluation is the number of eliminated features. The algorithm that has many deleted features is the one that has the smallest predicted time.

Based on research [19] it was found that the quality of the house, house_area, median_house_neigh, and the house age were the feature importance components in predicting urban residential housing prices. It is found also that the weighted generated model had the best error value of RMSE which is 25,000. As compared to our research, our RMSE is better, and the number of the important features is different.

From previous research regarding house prediction in [20], this research found the smallest errors which are 16208.5 for MAE, 0.878 for R^2 (coefficient of determination), 10135.9 for MedAE (median absolute error), and 754362031.6 for MSE, Those values were obtained using Random Forest algorithm with values of hyperparameters

Table 2: Comparison Between XGBoost and Genetic Algorithm

Algorithm	Population	Iteration	Number of Eliminated Features	GA's Processing Time (second)	Predicted Time (second)	RMSE
XGBoost	-	-	-	-	6.825	0.133
XGBoost+GA_P5G5	5	5	172	49.607	3.569	0.152
XGBoost+GA_P5G10	5	10	165	85.021	2.362	0.142
XGBoost+GA_P5G15	5	15	148	123.204	2.068	0.142
XGBoost+GA_P5G20	5	20	164	161.126	1.930	0.138
XGBoost+GA_P5G25	5	25	175	199.863	1.818	0.136
XGBoost+GA_P10G5	10	5	163	92.574	1.935	0.133
XGBoost+GA_P10G10	10	10	174	168.733	1.803	0.140
XGBoost+GA_P10G15	10	15	155	246.618	2.027	0.136
XGBoost+GA_P10G20	10	20	174	322.058	1.857	0.144
XGBoost+GA_P10G25	10	25	176	396.661	1.832	0.133
XGBoost+GA_P15G5	15	5	159	137.853	1.989	0.141
XGBoost+GA_P15G10	15	10	165	251.729	1.978	0.139
XGBoost+GA_P15G15	15	15	174	368.672	1.808	0.135
XGBoost+GA_P15G20	15	20	166	484.766	1.896	0.137
XGBoost+GA_P15G25	15	25	181	642.557	1.750	0.130
XGBoost+GA_P20G5	20	5	165	185.300	1.931	0.143
XGBoost+GA_P20G10	20	10	169	338.597	1.899	0.134
XGBoost+GA_P20G15	20	15	172	487.316	1.855	0.129
XGBoost+GA_P20G20	20	20	159	646.379	1.997	0.138
XGBoost+GA_P20G25	20	25	154	800.272	2.033	0.133
XGBoost+GA_P25G5	25	5	153	232.330	2.030	0.137
XGBoost+GA_P25G10	25	10	151	421.556	2.070	0.132
XGBoost+GA_P25G15	25	15	150	615.438	2.075	0.138
XGBoost+GA_P25G20	25	20	160	804.718	1.984	0.143
XGBoost+GA_P25G25	25	25	163	987.446	1.939	0.129

are 41 for the number of estimators, 63 for maximum features, and criterion is mse. These values are bigger than the value of the best RMSE of our research.

Another research found 8.31% for MDAPE (median absolute percent error), 14 minutes of CPU-time training, 7.3 minutes of CPU-time predicting using the cluster aggregation model in the lazy learning model of kNN and 5000 objects strong Uppsala municipality sample [21]. This research obtained a better error value rather than ours. However, for the predicted times and processing time, our research is better than this research.

4 CONCLUSION

Our research found that the hybrid GA with XGBoost resulted in the best RMSE, processing times, and predicted times compared to the sole XGBoost. It is also found that there are two features that most contributed to housing price prediction: OverallQual, which represents the overall material and finish quality of the house, and GrLivArea, which represents above-grade (ground) living area square feet of the house. Based on the comparison to three previous research, our research could perform better error values than two of those research. Thus, the hybrid method could be developed for housing price prediction.

ACKNOWLEDGMENTS

To Universitas Negeri Sebelas Maret, for the funding on this paper.

REFERENCES

[1] Jean Murray (2020) "Why Business Property is Important to Your Business," The Balance Small Business. https://www.thebalancesmb.com/why-business-property-is-important-to-your-business-397499
[2] Ray White Indonesia (2021) "7 Cara Berbisnis Properti 2021," 2021. https://www.raywhite.co.id/news/cara-berbisnis-properti-2020
[3] Vincent F. Thomas (2021) "BI Prediksi Kenaikan Harga Properti akan Terus Melambat di Q1 2021," tirto.id, 2021. https://tirto.id/bi-prediksi-kenaikan-harga-properti-akan-terus-melambat-di-q1-2021-galj
[4] Krzystof Drachal and Michal Pawlowski (2021) "A Review of the Applications of Genetic Algorithms to Forecasting Prices of Commodities," economies, vol. 9, no. 6, 2021, [Online]. Available: https://www.mdpi.com/2227-7099/9/1/6/pdf
[5] Yu-Sheng Kao, Kazumitsu Nawata, and Chi-Yo Huang (2020) "Predicting Primary Energy Consumption Using Hybrid ARIMA and GA-SVR Based on EEMD Decomposition," mathematics, vol. 8, 2020, [Online]. Available: https://www.mdpi.com/2227-7390/8/10/1722/pdf
[6] Qian Zhang, Dandan Deng, Wenting Dai, Jixin Li, and Xinwen Jin (2020) "Optimization of culture conditions for diferentiation of melon based on artifcial neural network and genetic algorithm," Sci. Rep., vol. 10, 2020, [Online]. Available: https://www.nature.com/articles/s41598-020-60278-x.pdf
[7] Risa Wati (2016) "Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Baye," Evolusi, vol. 4, no. 1, pp. 25–31, 2016, [Online]. Available: https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/604/495
[8] Tianqi Chen and Carlos Guestrien (2016) "XGBoost: A Scalable Tree Boosting System," in KDD '16: Proceedings of the 22nd ACM SIGKDD International

Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794. doi: https://doi.org/10.1145/2939672.2939785.

[9] L. Torlay, Perrone-Bertolotti, E. Thomas, and M. Baciu (2017) "Machine learning–XGBoost analysis of language networks to classify patients with epilepsy," Brain Informatics, vol. 4, pp. 159–169, 2017, doi: 10.1007/s40708-017-0065-7.

[10] Wei Li, Yanbin Yin, Xiongwen Quan, and Han Zhang (2019), "Gene Expression Value Prediction Based on XGBoost Algorithm," Front. Genet., 2019, doi: https://doi.org/10.3389/fgene.2019.01077.

[11] Ahmedbahaaldin Ibrahem Ahmed Osman, Ali Najah Ahmed, Ming Fai Chow, Yuk Feng Huang, and Ahmed El-Shafie (2021) "Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia," Ain Shams Eng., 2021, doi: https://doi.org/10.1016/j.asej.2020.11.011.

[12] Bingyue Pan, "Application of XGBoost algorithm in hourly PM2.5 concentration prediction," IOP Conf. Ser. Earth Environ. Sci., vol. 113, 2018, [Online]. Available: https://iopscience.iop.org/article/10.1088/1755-1315/113/1/012127/pdf

[13] Kaggle, "kaggle." https://www.kaggle.com/c/house-prices-advanced-regression-techniques

[14] Sari Ayu Wulandari, Sutikno Madnasri, Ratih Pramitasari, and Susilo Susilo (2020) "Feature Selection Method to Improve the Accuracy of Diabetes Mellitus Detection Instrument," IJID (International J. Informatics Dev., vol. 9, no. 2, pp. 72–79, 2020, doi: https://doi.org/10.14421/ijid.2020.09203.

[15] Sugriyono Sugriyono and Maria Ulfah Siregar (2020) "Prapemrosesan klasifikasi algoritme kNN menggunakan K-means dan matriks jarak untuk dataset hasil studi mahasiswa," J. Teknol. dan Sist. Komput., vol. 8, no. 4, pp. 311–316, 2020, [Online]. Available: https://jtsiskom.undip.ac.id/index.php/jtsiskom/article/download/13874/12640

[16] Ganjar Alfian et al. (2020) "Blood glucose prediction model for type 1 diabetes based on artificial neural network with time-domain features," Biocybern. Biomed. Eng., vol. 40, no. 4, pp. 1586–1599, 2020, doi: https://doi.org/10.1016/j.bbe.2020.10.004.

[17] Ganjar Alfian et al. (2020) "Deep Neural Network for Predicting Diabetic Retinopathy from Risk Factors", doi: 10.3390/math8091620.

[18] Muhammad Syafrudin et al. (2020) "A Self-Care Prediction Model for Children with Disability Based on Genetic Algorithm and Extreme Gradient Boosting", doi: 10.3390/math8091590.

[19] Joshua M. Ngobia. (2021). "Predicting Model for Urban Residential Housing Prices: An Application of Ensemble Learning", Research Project of School of Computing and Information Management, KCA University.

[20] Alan Ihre, and Isak Engstrom. (2019). "Predicting house prices with machine learning methods", Thesis of Bachelor Degree in Computer Science of School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Swedia.

[21] Johan Oxenstierna. (2017). "Predicting house prices using Ensemble Learning with Cluster Aggregations", Degree Project in Department of Information Technology, Uppsala Universitet, Swedia.

CERTIFICATE

OF APPRECIATION | No. P-IC3INA22-10

is presented to:

Maria Ulfah Siregar

Title: Housing Price Prediction Using a Hybrid Genetic Algorithm with Extreme Gradient Boosting

as : Presenter

The International Conference 2022

The 9th International Conference on Computer, Control, Informatics and its Applications (IC3INA 2022)

with theme: "Digital Transformation Towards Sustainable Society for Post Covid-19 Recovery"

virtually through Zoom meetings on November 22 - 23, 2022

Dr. Purnomo Husnul Khotimah, M.T.

Conference Chair

Dr. Edi Kurniawan S.T, M.Eng

General Chair

IC3INA 2022

The International Conference on Computer, Control, Informatics and its Applications (IC3INA 2022)
- Digital Transformation Towards Sustainable Society for Post Covid-19 Recovery –

Paper Number : 1570854900
Subject : Letter of Acceptance (LoA)

Bandung, November 10, 2022

Dear Mr./Ms./Mrs./ Maria Ulfah Siregar,

Congratulations.

The Organizing Committee of the IC3INA 2022 is pleased to inform you that your paper:

ID	:	**1570854900**
Title	:	**Housing Price Prediction Using a Hybrid Genetic Algorithm with Extreme Gradient Boosting**
Author(s)	:	Maria Ulfah Siregar; Pahlevi Hardjita; Farhan Asdin; Muhammad Anshari; Dewi Wardani; Yessi Yunitasari; Ardhi Wijayanto

Has been **accepted** for proceeding publication and conference presentation in 9th International Conference on Computer, Control, Informatics and its Applications (IC3INA) 2022.

We are looking forward to seeing you at the conference between 22-23 November, 2022. For further information, please do not hesitate to contact us.

Warm regards,

Dr. Purnomo Husnul Khotimah

Please, follow the ACM template Sigconf carefully (http://www.acm.org/publications/article-templates/proceedings-template.html/). The publication of the IC3INA 2022 paper will be carried out through The ACM Publishing System (TAPS) system. So please check your email (registered at EDAS) regularly and make sure that you whitelist tapsadmin@aptaracorp.awsapps.com.

The General Chair of IC3INA 2022

Please, follow the ACM template Sigconf carefully (http://www.acm.org/publications/article-templates/proceedings-template.html/). The publication of the IC3INA 2022 paper will be carried out through The ACM Publishing System (TAPS) system. So please check your email (registered at EDAS) regularly and make sure that you whitelist tapsadmin@aptaracorp.awsapps.com.