NAME OF THE PROJECT:

HOUSE PRICING PREDICTION PROJECT

Submitted by:

UNGWADA EMMANUEL, OGBECHE

CONTENTS

INTRODUCTION

Thousands of homes are bought every day. There are a few questions each client asks himself like: What is the real charge that this residence deserves? Am I paying a honest charge? In this paper, a device studying version is proposed to expect a residence charge primarily based totally on information associated with the residence (its size, the yr it turned into constructed in, etc.). During the improvement and assessment of our version, we can display the code used for every step accompanied through its output. This will facilitate the reproducibility of our work. In this study, Python programming language with some of Python applications may be used.

The foremost goals of this study are as follows:

- To
- practice statistics pre-processing and coaching strategies with a purpose to reap smooth statistics
- To
- construct system studying fashions capable of are expecting residence charge primarily based totally on residence features
- To

- examine and examine fashions overall performance with a purpose to pick the satisfactory model

LITERATURE REVIEW

In this section, we study some current research which might be associated with our subject matter and notice how fashions have been constructed and what outcomes have been carried out in that research.

STOCK MARKET PREDICTION USING BAYESIAN-REGULARIZED NEURAL NETWORKS

In a study accomplished with the aid of using Ticknor (2013), he used Bayesian regularized artificial neural community to predict the future operation of economic market. Specifically, he constructed a version to predict future inventory prices. The enter of the version is preceding inventory records similarly to a few economic technical records. The output of the version is the next–day ultimate rate of the corresponding stocks.

The version proposed in the study is constructed using Bayesian regularized neural community. The weights of this kind of networks are given a probabilistic nature. This lets in the community to penalize

very complicated models (with many hidden layers) in an automated manner. This in flip will lessen the overfitting of the version.

The version includes a feed forward neural community which has 3 layers: an enter layer, one hidden layer, and an output layer. The creator selected the wide variety of neurons withinside the hidden layer primarily based totally on experimental methods. The enter records of the version is normalized to be between –1 and 1, and this operation is reversed for the output, so the expected rate seems in the correct scale.

STOCK MARKET PREDICTION USING MACHINE LEARNING MODEL

In every other study carried out via way of means of Hegazi, Soliman, and Salam (2014), a device become proposed to predict everyday inventory marketplace prices. The device combines particle swarm optimization (PSO) and least rectangular aid vector machine (LS–SVM), wherein PSO become used to optimize LV–SVM. The authors declare that during maximum cases, synthetic neural networks (ANNs) are challenge to the overfitting problem. They country that aid vector machines algorithm (SVM) become evolved as an opportunity that does not be afflicted by overfitting. They characteristic this benefit to SVMs being primarily based totally at the stable foundations of VC–theory. They are similarly problematic that LS–SVM technique

become reformulation of conventional SVM technique that makes use of a regularized least squares feature with equality constraints to acquire a linear device that satisfies Karush–Kuhn–Tucker situations for purchasing an best solution. The authors describe PSO as a famous evolutionary optimization technique that become stimulated via way of means of organism social conduct like fowl flocking. They used it to discover the best parameters for LS–SVM. These parameters are the value penalty C, kernel parameter γ , and insensitive loss feature ϵ . The version in proposition in the examined become primarily based totally at the evaluation of historic records and technical economic signs and the usage of LS–SVM optimized via way of means of PSO to are expecting destiny every day inventory prices. The version enters become six vectors representing the historic records and the technical economic signs.

DATA PREPARATION/DISCRIPTION

In this study, we used a housing dataset contains 1460 entries each having 81 variables.

Data contains Null values that needed to be treated using the domain knowledge and self-understanding. And it was treated according, refer to python file. The dataset contains many variables that are involved in determining a house price.

EXPLORATORY DATA ANALYSIS

In this section, we explore the data using visualizations. This allows us to understand the data and the relationships between variables better, which helps us build a better model.

PREDICTION TYPE AND MODELING TECHNIQUE

In this section, we pick out the kind of system getting to know prediction this is appropriate to our hassle. We need to decide if that is a regression hassle or a type of hassle. In this project, we need to expect the rate of a residence given statistics approximately it. The rate we need to expect is a non-stop value; it can be any actual number. This may be visible via way of means of searching on the goal vitiable in our dataset

Model Building and Evaluation

In this part, we are able to construct our prediction model: we are able to select algorithms for every of the strategies we stated withinside the preceding section. After we construct the model, we are able to compare its overall performance and results.

Splitting the Dataset

As standard for supervised gadget getting to know problems, we need a training dataset to teach our version and a look at dataset to assess the version. So, we are able to break up our dataset randomly into parts, one for schooling and the opposite for testing. For that, we are able to use any other characteristic from Scikit–Learn called train_test_split()

ANALYSIS AND COMPARISON

We created a few models: for every version, we looked for excellent parameters then we built the version the usage of the ones parameters, then trained (fitted) the version to our education data (X_train and y_train), then examined the version on our take a look at data (X_test) and finally, we evaluated the version overall performance through evaluating the version predictions with the genuine values in y_test. We used the imply absolute error (MAE) to assess version overall performance. Using the consequences, we were given withinside the preceding section, we gift a desk that suggests the imply absolute error (MAE) for every version whilst carried out to the take a look at set X_test. The desk is looked after ascendingly in line with MAE score.

CONCLUSION

In this paper, we constructed several regression fashions to predict the rate of a few residences given a number of the residence functions. We evaluated and compared every version to decide the only with maximum performance. We additionally checked out how a few fashions rank the functions in step with their importance. In this paper, we accompanied the records technological know–how system beginning with getting the records, then cleansing and preprocessing the records, accompanied through exploring the records and constructing fashions, then comparing the outcomes and speaking to them with visualizations. As a recommendation, we suggest applying this version (or a model of it skilled with greater current records) through individuals who need to shop for a residence withinside the place protected through the dataset to have an concept approximately the real rate. The version may be used additionally with datasets that cowl exclusive towns and regions furnished that they include the identical functions. We additionally advocate that human beings think about the functions that have been deemed as maximum essential as visible withinside the preceding section; this could assist them estimate the residence rate better.

REFERENCES:

Alkhatib, K., Najadat, H., Hmeidi, I., & Shatnawi, M. K. A. (2013). Stock price prediction using k-nearest neighbor (kNN) algorithm. International Journal of Business, Humanities and Technology, 3(3), 32-44.

de Abril, I. M., & Sugiyama, M. (2013). Winning the kaggle algorithmic trading challenge with the composition of many models and feature engineering. IEICE transactions on information and systems, 96(3), 742-745.

Feng, Y., & Jones, K. (2015, July). Comparing multilevel modelling and artificial neural networks in house price prediction. In Spatial Data Mining and Geographical Knowledge Services (ICSDM), 2015 2nd IEEE International Conference on (pp. 108-114). IEEE.

Hegazy, O., Soliman, O. S., & Salam, M. A. (2014). A machine learning model for stock market prediction. arXiv preprint arXiv:1402.7351.

Ticknor, J. L. (2013). A Bayesian regularized artificial neural network for stock market forecasting. Expert Systems with Applications, 40(14), 5501-5506.