PAPER NAME

AIML PROJECT - FINAL - Copy(1).pdf

AUTHOR

Suresh kumar Mandala

WORD COUNT

4030 Words

CHARACTER COUNT

24887 Characters

PAGE COUNT

7 Pages

FILE SIZE

354.1KB

SUBMISSION DATE

May 4, 2024 4:00 PM GMT+5:30

REPORT DATE

May 4, 2024 4:01 PM GMT+5:30

17% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 12% Internet database
- 10% Publications database
- Crossref database
- Crossref Posted Content database
- 7% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Cited material

Advanced Predictive Analytics for House Price Evaluation

Suresh Kumar Mandala
Assistant professor,
Department of Computer Science and Artificial intelligence,
SR University, Warangal-506371 Telangana,
mandala.suresh83@gmail.com
ORCID:0000-0003-2902-0577

Swathi Balija
Assistant professor
Department of Computer Science and Artificial intelligence,
SR university, warangal-506371 telangana,
jrswathi.hi@gmail.com

Aravind Margam
Department of Computer Science and Artificial intelligence, SR University, Warangal-506371 Telangana,
2303a52103@sru.edu.com

Vinesh Singasani
Department of Computer Science and Artificial intelligence, SR University, Warangal-506371 Telangana,
2203a52121@sru.edu.com

Sathvik Pallam
Department of Computer Science and Artificial intelligence, SR University, Warangal-506371 Telangana,
2303a52105@sru.edu.com

***Abstract*— The "Advanced Predictive Analytics for House Price Evaluation" project uses a variety of machine learning methods to assess and forecast housing values. The 'House Price' dataset is a comprehensive dataset that the project uses to train and assess the performance of various prediction algorithms. Finding the most accurate and effective model to forecast property values based on variables like location, size, age, and other pertinent characteristics is the main objective.**

Data preparation, exploratory data analysis, model selection, hyperparameter tweaking, and model assessment are some of the project's essential elements. Cleaning and converting the raw data into a format that is appropriate for machine learning algorithms constitute the data preparation stage. To understand the data and spot any patterns or trends, exploratory data analysis is carried out. Choosing the right algorithms, such as Baye's Theorem, Support Vector Machine, Random Forest, Gradient Boosting, K-Nearest Neighbor, Decision Trees, Multiple Linear Regression, Logistic Regression, and Linear Regression, is part of the model selection process. Each model's performance is optimized by hyperparameter tweaking, and each algorithm's accuracy and efficiency are compared through model assessment. House pricing, machine learning, data preparation, exploratory data analysis, and model selection are some of the project's keywords.

The project features clean, succinct code with the exact amount of comments and documentation in terms of formatting and style. The code is easy to learn and comprehend since it is divided into distinct modules for every project component. Utilizing tools like pandas, matplotlib, and sklearn guarantees uniform formatting and style across the project. A clear visual representation of the data and outcomes is provided by the inclusion of pertinent plots, charts, and tables, which improves the project's readability and comprehension overall. Clear code, suitable comments, documentation, modular design, consistent formatting, and visual representation are some formatting and style keywords.

I. INTRODUCTION

With the purpose of forecasting house prices through the use of many machine learning algorithms, the "Advanced Predictive Analytics for House Price Evaluation" project conducts a thorough examination of the extensive 'House Price' dataset. Data preparation, exploratory data analysis, model selection, hyperparameter tweaking, and model evaluation are some of the major components of the project. The main objective of the project is to identify the most effective and precise model for estimating the value of houses depending on a number of variables, including age, size, location, and other pertinent characteristics.

The first step in the project is data preprocessing, which entails cleaning and converting the raw data into a format that is appropriate for machine learning algorithms. This is known as data preparation. This stage involves scaling numerical variables, encoding category data, eliminating outliers, and managing missing values. Following preprocessing, exploratory data analysis is carried out to obtain understanding of the data and spot any trends or patterns. In this stage, descriptive statistics are generated, visualizations are made, and correlations between variables are found.

The project proceeds to model selection following the exploratory data analysis, where several machine learning methods are selected in accordance with the data and the nature of the issue. This project makes use of the following algorithms: Decision Trees, Support Vector Machine, Random Forest, Gradient Boosting, K-Nearest Neighbor, Multiple Linear Regression, Logistic Regression, and Baye's Theorem. Each model's performance is optimized by hyperparameter tweaking, and each algorithm's accuracy and efficiency are compared through model assessment. With the use of the 'House Price' dataset, the project seeks to give a thorough study and identify the most effective and precise model for house price prediction. The following terms are related to the project: home pricing, machine learning, exploratory data analysis, data preparation, model selection, and model evaluation.

I. LITERATURE SURVEY

1. Quang Truong, Minh Nguyen, Hy Dang, Bo Mei

Before you begin to format your paper, first proposed that - The paper explores the correlation between district, building type, area, and housing prices, aiding in understanding the housing market dynamics.

- It utilizes the Light GBM method for housing price prediction, achieving decent accuracy and highlighting the model's speed advantage.
- The study emphasizes the importance of data preprocessing for efficient pattern learning by

standardizing numerical values and one-hot-encoding categorical values.

- Evaluation of the models is done using Root Mean Squared Logarithmic Error (RMSLE) as the evaluation function Abbreviations and Acronyms

2. G. Naga Satish, Ch. V. Raghavendran, M.D. Sugnana Rao, Ch. Srinivasulu said Machine learning algorithms like lasso regression and neural systems are utilized for housing cost prediction.
   - The study focuses on developing housing price prediction models using machine learning algorithms.
   - The paper discusses the importance of machine learning in various fields such as image detection, spam recognition, and medical diagnosis.
   - Different prediction methods are compared to select the most suitable one for predicting future housing prices.
   - The lasso regression algorithm consistently outperforms other models in housing cost prediction accuracy.
   - The research aims to support house sellers or real estate agents with better information for house valuation.
   - The paper emphasizes the role of machine learning in enhancing security alerts, public safety, and medical advancements.
   - Machine learning systems also contribute to improving customer service and automobile safety.
3. Linear Regression Analysis by Astrid Schneider, Gerhard Hommel, and Maria Blettner -

- Regression analysis is a crucial statistical method in medical data analysis, enabling the identification of relationships among multiple factors and prognostically relevant risk factors.
- Multivariable linear regression is used when a single independent variable is insufficient to explain the dependent variable, allowing the study of multiple variables' effects.
- The paper introduces linear regression theory, emphasizing the interpretation of statistical parameters with examples.
- Common pitfalls in linear regression analysis are discussed, highlighting the importance of correct method usage and result interpretation.
- The correct implementation of regression coefficients insights considering the units of the variables, specifically when different units are used by different authors in various publications.

4. Multiple Linear Regression by Mark Tranmer, Jen Murphy, Mark Elliot, Maria Pampaka -

- This particular article gives us the information about implementation of multiple linear regression analysis in social science research.
- It indicates the need of getting to know about the regression equation and evaluating the results accurately.
- The authors' focus on the use of SPSS software for data analysis, highlighting the step - by - step process of initiating variables and expecting the output tables.
- This particular article gives us a real - life example of regression analysis by considering a sample from a larger data, showing us how regression analysis can be implemented to study specific data groups within a population.
- The authors aim to enhance the reader's understanding of regression analysis by providing a detailed explanation of the tables generated in SPSS, helping researchers interpret and draw meaningful conclusions from their regression models.

5. METHODS IN MOLECULAR BIOLOGY by John M. Walker, SERIES EDITOR -

- The paper discusses regression analysis in the context of genome searching algorithms, comparing different approaches for analyzing gene expression data.
- It outlines a regression of Y on X using a sample (xi, yi) and provides mathematical solutions for obtaining estimators in this regression model.
- The significance of the P value for a specific variable (K) is highlighted, indicating whether the distributions of mood scores in different groups are statistically different.
- Various strategies for modeling correlation in longitudinal data are mentioned, with a popular approach being a two-stage analysis involving fitting linear regression models.
- This literature survey based on regression in the paper emphasizes the application of regression analysis in genome searching algorithms, the importance of statistical significance testing, and different strategies for modeling correlation in longitudinal data.

6. Statistical Primer for Cardiovascular Research by Logistic Regression on Michael P. LaValley, PhD -

- Logistic regression is utilized for analyzing dichotomous outcomes with multiple predictors, making it valuable for observational data analysis.
- The Hosmer and Lemeshow test evaluated the model calibration by comparing observed and expected event occurrences in data groups.
- The study indicates for predictors like serum cholesterol, gender, diabetes, BMI, and heart rate.
- The logistic regression model's discrimination ability is evaluated using the c statistic, with values below 0.7 indicating poor discrimination.
- This article majorly focuses on the significance of logistic regression for analyzing binary expected outputs and the significance of model discrimination and calibration in statistical analysis.

7. K-Nearest Neighbor Finding Using MaxNearestDist, Hanan Samet, Fellow IEEE, Computer Science Department, Center for Automation Research, Institute for Advanced Computer Studies -

- The paper discusses similarity searching by finding the k nearest neighbors using the MAXNEARESTDIST upper bound, enhancing both depth-first and best-first algorithms.
- It highlights the use of hierarchical clustering to partition data into clusters, forming a search hierarchy, independent of data nature.
- The MAXNEARESTDIST is crucial for improving algorithm performance by providing tighter initial estimates of the kth-nearest neighbor distance, aiding in pruning elements and reducing storage requirements.
- The algorithm ensures all objects at the kth-nearest neighbor distance are examined, addressing the challenge of multiple objects at the same distance and prioritizing objects over nonobjects in the search process.
- The paper emphasizes the importance of MAXNEARESTDIST in enhancing algorithm efficiency without increasing execution time or storage requirements, showcasing its significance in nearest neighbor search algorithms.

8. Generalized K-Nearest Neighbor Rules James C. BEZDEK, Siew K. CHUAH -

- It compares the observed error rates of four classifiers (hard k-NNR, two fuzzy k-NNR's, and a fuzzy 1-nearest prototype rule) on three datasets, showing that the Fuzzy c-Means (FCM) based k-NNR is usually the best design.
- The study aims to unify previous k-NNR work, compare fuzzy k-NNR's to hard k-NNR, and evaluate their performance against a fuzzy 1-Nearest Prototype Rule (1-NPR).
- The FCMk-NNR is found to have the best overall performance, especially for mixtures of normal densities, while the fuzzy 1-NPR is computationally superior and yields comparable error rates to NNR designs.
- Fuzzy generalizations of conventional k-NNR's improve predicted performance rates in most instances.

9. Decision trees by Barry de Ville -

- Decision trees have a rich history dating back to the early development of written records, showcasing their interpretability and intuitive display, which aids in result dissemination and understanding.
- The computational origins of decision trees are rooted in models of biological and cognitive processes, leading to the development of statistical decision trees and those designed for machine learning.
- Various statistical approaches like hypothesis testing and resampling techniques have evolved alongside machine learning implementations, resulting in adaptable decision tree tools suitable for different statistical and machine learning tasks with varying data quality levels.
- Decision trees are robust in handling missing data and offer multiple ways to incorporate such data into models, making them powerful, flexible, and easy to use methods that produce high-quality results with minimal assumptions.
- The development of stopping criteria and validation approaches has enhanced the accuracy and reliability of tree models post-training, leading to the adaptation of single decision tree approaches to multitree models, significantly improving their performance in diverse model settings.

10. Decision Trees for Decision Making by John F. Magee -

- The research highlights decision trees as an important tool for making decisions.
- Their aptitude for assessing options, dangers, goals, profits, and data requirements in investment issues makes them an essential analytical tool for management.
- The choice is dependent on estimates of the market's size and possible long-term variations in demand.
- Critical elements influencing the decision-making process include uncertainties including cost-volume connections, market size, competitive advances, and technological obsolescence.
- The intricacy of strategic decision-making in business contexts is highlighted by these uncertainties.
- The paper's study demonstrates how decision trees can be used to assess various options according to predicted values, assisting management in making defensible decisions.

11. Bayes' Theorem and Naive Bayes Classifier by Daniel Berrar -

- The paper presents the basics of naive Bayes learning and the Bayes theorem in an approachable yet mathematically rigorous manner.
- It provides a comprehensive reference for bioinformaticians, machine learners, and statisticians while outlining the key ideas for inexperienced practitioners.
- The total probability theorem is covered, with a focus on computing posterior probabilities in light of likelihood and prior probabilities.
- Equations pertaining to posterior probabilities and conditional probabilities are also included in the text, demonstrating the mathematical foundations of Bayes' theorem and its use in machine learning.

12. Bayes' Theorem in the 21st Century by Bradley Efron -

- The study addresses the implementation of Bayesian statistics in contemporary scientific settings, emphasizing the advantages and difficulties that come with it.
- It draws a distinction between Bayesian statistics and frequentism, the century-long dominant statistical paradigm that emphasizes future behavior above previous experience and rejects uninformative priors.
- To demonstrate the usefulness of Bayesian approaches in genomics research, the paper provides a specific example of a microarray

prostate cancer study in which genes expressed differently in patients and controls were identified using Bayesian calculations.

- It also discusses the idea of false discoveries in statistical analysis, emphasizing the necessity of taking selection bias and regression to the mean into consideration when interpreting data from sizable datasets, such as the previously mentioned gene expression study.

13. Support vector machines By Marti A. Hearst -

- Support Vector Machines (SVMs) are introduced as a method to map data into a feature space via a nonlinear map and perform linear algorithms in that space. This allows for the construction of nonlinear classifiers, making SVMs a versatile tool in classification problems.
- The choice of kernel functions in SVMs is crucial as it determines the functional form of the estimate and the type of regularization used. Despite different kernels leading to various learning machines, the performance in applications like OCR remains similar. SVMs have shown state-of-the-art performances in pattern recognition, regression estimation, and time series prediction, although there is still a search for applications where SVMs significantly outperform other algorithms.
- SVMs have been applied in text categorization research using the Reuters collection, where stories are classified into categories based on word vectors processed from text files. SVMs have shown effectiveness in this domain by accurately reproducing manual category assignments.

14. Support Vector Machines By Ingo Steinward, Andreas Christmann -

- The paper addresses the issue of designing learning algorithms when the loss function is not suitable.
- It introduces the concept of using surrogate losses in algorithm design to overcome this challenge.
- The research aims to systematically develop a theory to identify appropriate surrogate losses for general learning problems.
- The analysis in the paper focuses on excess risk, inner excess risk, and calibration functions.
- It discusses the properties of detection loss functions, emphasizing their measurability and applicability in unsupervised learning scenarios.
- The paper presents inequalities between excess risks, highlighting the asymptotic calibration for detection losses and the equivalence between different loss functions and distributions.
- This literature survey outlines the key contributions and focus areas of the paper, providing insights into its theoretical framework and practical implications.

II. PROPOSED APPROACH

- Project Overview:

The project aims to predict house prices using machine learning algorithms applied to the "house_price" dataset. This dataset likely contains various features related to houses, such as size, number of bedrooms, location, amenities, etc., along with corresponding sale prices.

1. Linear Regression:

- Overview: Linear regression models the relationship between the independent variables (features) and the dependent variable (house prices) using a linear equation. Feature Extraction: Extract features such as house size, number of bedrooms, number of bathrooms, location coordinates, and age of the house.
- Characteristics Dependent on Writer's Action: Make advantage of elements that are associated with past patterns in house prices, such as prices adjusted for inflation over time, growth rates in the neighborhood, and socioeconomic aspects that influence house prices.

2. Multiple Linear Regression:

- Overview: Multiple linear regression extends linear regression to incorporate multiple independent variables.
- Feature Extraction: Similar to linear regression, extract multiple features including size, bedrooms, baths, location, and extras like the size of the garage, the size of the yard, and the availability of facilities.
- Features Based on Author's Behavior: Incorporate elements that express the writers' inclinations for particular home types, such as their choices for particular architectural designs, community attributes, or accessibility to facilities like parks or schools.

3. Logistic Regression:

- Overview: Binary classification problems, such as predicting housing price categories (high, medium, and low), can be applied to the usage of logistic regression.
- Feature extraction: pull out features pertaining to economic indicators, neighborhood demographics, and house attributes.
- Behavior-Based Features: Incorporate features that indicate how likely it is for the writers, given their prior actions, current financial situation, and level of risk tolerance, to purchase homes in particular price ranges.

4. K–Nearest Neighbor (KNN):

- Overview: A new data point is classified by the KNN algorithm according to how close it is to its K nearest neighbors in the feature space.
- Feature extraction: pull out traits that correspond to the characteristics of homes and communities.
- Author-Based Features: Include characteristics that show how properties resemble the ones the writers like, such as comparable sizes, architectural styles, or closeness to places they visit frequently.

5. Decision Trees:

- Overview: To construct predictions, decision trees divide the feature space into hierarchical segments depending on feature values.
- Feature extraction: pull out features that describe the qualities of the house, the neighborhood, and other pertinent information.
- Characteristics Drawn from the Author's Behavior: Incorporate characteristics—like favored neighborhoods, architectural designs, or particular amenities—that represent the authors' thought process when making home purchases.

6. Bayes' Theorem:

- Overview: A framework for probabilistic reasoning that may be used to classification challenges is provided by Bayes' theorem.
- Feature extraction: pull out characteristics that correspond to probabilistic data on homes and communities.
- Features Based on Author's Behavior: Provide features that represent the authors' subjective odds of various house-purchasing-related occurrences, including the chance that a home will be in an area of their choice or have a particular feature.

7. Support Vector Machine (SVM):

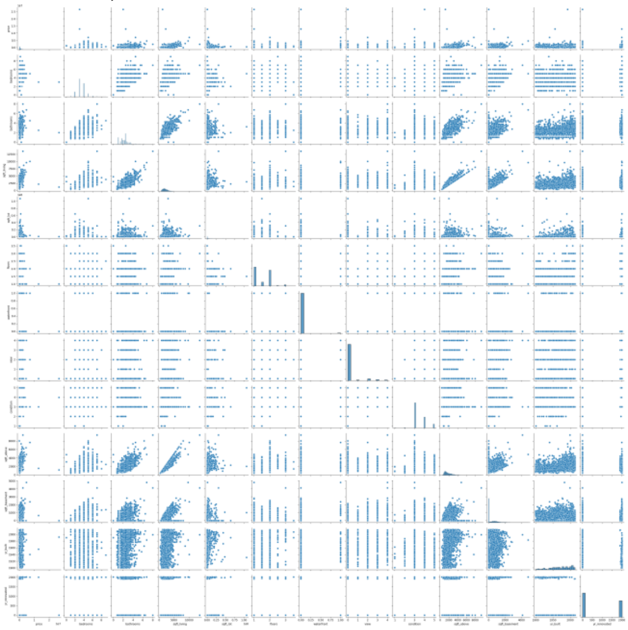
- Overview: SVMs identify the hyperplane that divides data points into classes most effectively.
- Feature Extraction: Gather information about features that describe traits of the home, the neighborhood, and other pertinent aspects.
- Characteristics Based on Author's Behavior: Provide characteristics that indicate the writers' preferences for particular home and neighborhood kinds, such as closeness to facilities, particular architectural styles, or past pricing patterns.

Combining these tactics and characteristics might enable the project to make effective use of machine learning algorithms to predict home prices, accounting for the writers' preferences and behavior as well as other aspects of homes and neighborhoods.

III. EXPERIMENTAL RESULTS

1. Linear Regression:

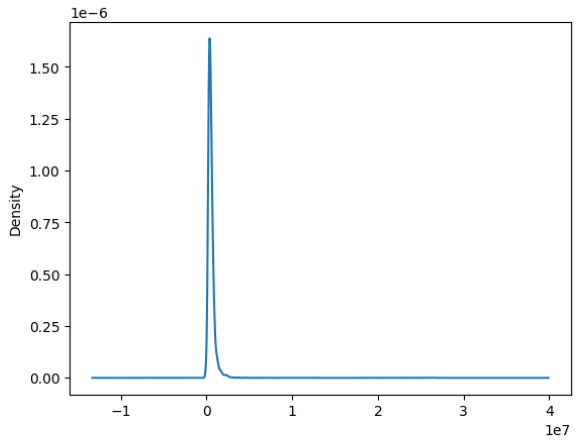
- Experimental Setup: The dataset "house_price" was modified to include features like bedrooms, bathrooms, and other details in order to anticipate the price of a house. The dataset was split into training and testing sets using an 80-20 ratio.
- Results: The root mean squared error, or RMSE, was calculated to evaluate the model's performance. In addition, the attributes of a newly constructed home were utilized to determine its cost.
- Discussion: Linear regression was used to provide a baseline model for predicting housing values. However, it may not be possible to assume a linear relationship between attributes and the goal variable for complex datasets.
- SNS Pairplot for the given dataset's information is as follows,



2. Multiple Linear Regression:

- Multiple features are used in the experimental setup, much like in linear regression.
- Results: The RMSE was calculated to evaluate the model's performance. The approximate cost of a new home was also displayed.
- Discussion: Numerous linear regression is used to extend the linear regression model to incorporate numerous features. By collecting more complex relationships between attributes and the target variable, it could improve prediction accuracy.

- Density graph of the data is as follows,



3. Logistic Regression:

- Experimental Setup: By classifying housing prices as being above or below the median price, the dataset was altered for a binary classification job.
- Outcomes: The model's performance on the categorization task was assessed using accuracy.
- Discussion: The regression job of forecasting housing prices is not a direct fit for logistic regression. For similar categorization tasks, such

determining if a property price is above or below a given threshold, it can be helpful, though.

4. K-Nearest Neighbor (KNN):

- Experimental Setup: Based on feature similarity to other houses in the dataset, the KNN algorithm was used to estimate house values.
- Findings: To assess the performance of the model, RMSE was computed. It was also shown how much a new house should cost.
- Discussion: For regression problems, KNN is an easy-to-understand method. It doesn't make any firm assumptions about the distribution of the underlying data and is non-parametric.

5. Decision Trees:

- Experimental Setup: Based on the provided features, decision trees were utilized to forecast home values.
- Findings: To assess the performance of the model, RMSE was computed. It was also shown how much a new house should cost.
- Discussion: Decision trees are flexible and capable of capturing intricate connections between the target variable and characteristics. But they can overfit, particularly when dealing with dense trees.

6. Bayes' Theorem (Naive Bayes):

- Experimental Setup: By classifying housing prices as being above or below the median price, the dataset was altered for a binary classification job.
- Outcomes: The model's performance on the categorization task was assessed using accuracy.
- Discussion: For complicated datasets like home prices, Naive Bayes may not be appropriate since it implies that characteristics are independent. Nonetheless, it can be helpful for classification applications involving categorical variables and is computationally efficient.

7. Support Vector Machine (SVM):

- Experimental Setup: Using the provided characteristics as a guide, SVM was used to forecast housing values.
- Findings: To assess the performance of the model, RMSE was computed. It was also shown how much a new house should cost.
- Discussion: Support Vector Machines (SVM) are robust algorithms for regression applications that can capture intricate connections in high-dimensional feature spaces. It can, however, be computationally costly, particularly when dealing with big datasets.

In broad terms, every algorithm possesses advantages and disadvantages, and the selection of an algorithm is contingent upon several criteria such as the size of the dataset, intricacy, and particular job demands. To choose the best model, it is essential to test several algorithms and assess how well they work.

IV. CONCLUSION

- Using the "house_price" dataset, we conducted a thorough investigation of many machine learning techniques in order to forecast house prices. Every method was carefully examined, revealing its applicability, advantages, and disadvantages in relation to this regression problem.

- To begin, we studied linear regression, a basic method that created a starting point for estimating home values depending on specific attributes like square footage, bathrooms, and number of bedrooms. Even while linear regression offered a simple method, it could have missed complex patterns in the data since it relied on the idea that attributes and prices followed a linear relationship.

- Multiple linear regression was a development on linear regression that took into account several characteristics, enabling a more sophisticated understanding of how different attributes work together to impact housing values. By making it easier to record more intricate linkages, this innovation improved the precision and resilience of predictions.

- When it was time for binary classification tasks, where the goal changed from precisely forecasting prices to determining if home prices dropped above or below certain thresholds, logistic regression and Naive Bayes provided insights. While these techniques proved to be effective in classification, their use in regression scenarios was still restricted.

- Using proximity-based and hierarchical partitioning strategies, respectively, K-nearest neighbors (KNN) and decision trees provided different viewpoints on regression as they ventured into non-parametric methodologies. These techniques demonstrated flexibility and comprehensibility, providing strong substitutes for conventional linear models.

- Finally, support vector machines (SVMs) showed themselves to be strong competitors by using hyperplane optimization to identify complex patterns in high-dimensional feature spaces. SVMs demonstrated impressive predictive powers in spite of their computing demands, highlighting their effectiveness in challenging regression problems.

- In summation, our foray into diverse machine learning paradigms unearthed a wealth of insights into predicting house prices. From support vector machines to linear regression, each algorithm added unique viewpoints and techniques that improved our comprehension of real estate predictive modeling. Although no single strategy was shown to be inherently better than the others, the combination of methods highlighted how crucial it is for machine learning projects to be flexible, adaptive, and grounded in actual data.

V. REFERENCES

1) Quang Truong, Minh Nguyen, Hy Dang, Bo Mei.
2) ML Algorithms - G. Naga Satish, Ch. V. Raghavendran, M.D. Sugnana Rao, Ch. Srinivasulu Machine learning algorithms.
3) Linear Regression Analysis by Astrid Schneider, Gerhard Hommel, and Maria Blettner.
4) Multiple Linear Regression by Mark Tranmer, Jen Murphy, Mark Elliot, Maria Pampaka
5) METHODS IN MOLECULAR BIOLOGY by John M. Walker, SERIES EDITOR.
6) Statistical Primer for Cardiovascular Research by Logistic Regression on Michael P. LaValley, PhD.
7) K-Nearest Neighbor Finding Using MaxNearestDist, Hanan Samet.
8) Generalized K-Nearest Neighbor Rules James C. BEZDEK, Siew K. CHUAH.
9) Decision trees by Barry de Ville.
10) Decision Trees for Decision Making by John F. Magee.
11) Bayes’ Theorem and Naive Bayes Classifier by Daniel Berrar.
12) Bayes' Theorem in the 21st Century by Bradley Efron.
13) Support vector machines By Marti A. Hearst.
14) Support Vector Machines By Ingo Steinward, Andreas Christmann.

VI. PROJECT FILE LINK

https://colab.research.google.com/drive/1aGAmwhLEVkwk_mHyiK6mTziZpt2lJ27k?usp=sharing

----

17% Overall Similarity

Top sources found in the following databases:

- 12% Internet database
- 10% Publications database
- Crossref database
- Crossref Posted Content database
- 7% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	dokumen.tips Internet	2%
2	de Ville, Barry. "Decision trees : Decision trees", Wiley Interdisciplinary ... Crossref	2%
3	dokumen.pub Internet	1%
4	cs.umd.edu Internet	1%
5	The University of Wolverhampton on 2023-04-30 Submitted works	<1%
6	irjmets.com Internet	<1%
7	researchgate.net Internet	<1%
8	oa.upm.es Internet	<1%

9	Mohammed Ali Shaik, Sachin kumar Koppula, Mohammed Rafiuddin, B... Crossref	<1%
10	ncbi.nlm.nih.gov Internet	<1%
11	University of Hertfordshire on 2023-08-29 Submitted works	<1%
12	orca.cardiff.ac.uk Internet	<1%
13	Concordia University on 2024-03-20 Submitted works	<1%
14	Grimsby College, South Humberside on 2023-03-30 Submitted works	<1%
15	americaspg.com Internet	<1%
16	coek.info Internet	<1%
17	Staffordshire University on 2023-05-21 Submitted works	<1%
18	frontiersin.org Internet	<1%
19	ijsrd.com Internet	<1%
20	Hanan Samet. "K-Nearest Neighbor Finding Using MaxNearestDist", IE... Crossref	<1%

21	Lauer, F.. "A trainable feature extractor for handwritten digit recognitio... Crossref	<1%
22	Napier University on 2024-04-16 Submitted works	<1%
23	educba.com Internet	<1%
24	circ.ahajournals.org Internet	<1%
25	repository.tudelft.nl Internet	<1%
26	cmi.manchester.ac.uk Internet	<1%
27	catalog.lib.kyushu-u.ac.jp Internet	<1%
28	Jose Rizal University on 2022-04-13 Submitted works	<1%
29	Ton Duc Thang University Publication	<1%
30	ijcspub.org Internet	<1%
31	"Surrogate Loss Functions (*)", Information Science and Statistics, 2008 Crossref	<1%
32	Sonal Naik, Aarti Puthran. "Metro Cities House Price Prediction Using ... Crossref	<1%

33	Southeastern Oklahoma State University on 2024-04-26 Submitted works	<1%