Cogs 209 Peer Reviews- Group 3

Lead Author: Yanghong Wei

Co-authors: Yunyi Huang, Yuhan Fu，Xiaoyu Liu

Peer Review 1 - Deciphering House Price Dynamics Using Predictive Modeling on Housing Data (Group 8)

Summary of the Study:

The study "Deciphering House Price Dynamics Using Predictive Modeling on Housing Data" dives into the complex world of house prices to figure out what really drives them. Using the Ames Housing dataset, the study built predictive models, especially focusing on decision trees, to uncover which property features matter most. They found that factors like the size of the property, open porch space, and basement area play significant roles in determining house prices. This study offers valuable insights for home buyers, sellers, and investors, helping them make smarter decisions in the real estate market. Despite some data challenges, the study sheds light on key trends and suggests directions for future research to further enhance our understanding of housing prices.

Major Concerns and Flaws:

1. Data Preprocessing:
    a. The study mentions removing features and rows with significant missing values, leading to a loss of over half the dataset. This reduction might have impacted the robustness and generalizability of the model. As a potential solution, the report could benefit from exploring advanced imputation techniques to handle missing data more effectively.
2. Model Overfitting
    a. Models like decision trees are prone to overfitting, and even though the study used grid search for hyperparameter tuning, there is limited discussion on how overfitting was explicitly addressed. The study could incorporate techniques such as pruning or using ensemble methods like random forests to reduce overfitting.
3. Generalization
    a. The dataset is specific to Ames, IA, and findings may not generalize to other regions. The study should explicitly discuss the limitations of geographic specificity and suggest directions for validating the model with diverse datasets from different locations.

Minor Concerns:

1. Visualization and Interpretation:
    a. It is noticeable that there is a lot of information about the analysis results on the figures included in the project report. It will definitely help the readability of axis labels and figure titles if the font size is increased. Overleaf or LaTex scale the figures while not adjusting the font size on the figures.

b. The figures and tables are informative, but adding more detailed captions below them could enhance understanding. Providing explanations about what the results in each figure represent or what data is being presented can help the reader better grasp the context. Rather than simply labeling the figures, more descriptive captions would improve clarity.

2. Discussion and Future Directions:
    a. The definition of some factors could be explicitly explained in the report so that readers can understand each feature better. For example, what is included in overall quality?
    b. The author is well aware of the limitations of the project and makes suggestions to improve their model. It may be worthwhile to discuss, among all the measurements that could help, what method is the most effective way of improving the robustness of the model.
    c. Except for working with collected data, it may also be a concern that how the model and data set can be updated or retrained periodically as response to the dynamic changings of the society.

Constructive Feedback:

Overall, the study tackles an interesting topic with practical implications for the residential real estate market in Ames, IA. By addressing the concerns outlined above, particularly those related to data preprocessing and model overfitting, and by increasing the geographic coverage of the data, the study can significantly enhance its robustness and applicability. Implementing more advanced modeling techniques and expanding the feature set will likely yield deeper insights into the industry.

Peer Review 2 - Predicting San Francisco Airbnb Prices (Group 12)

Summary of the Study:

The study "Predicting San Francisco Airbnb Prices" investigates the factors influencing Airbnb prices in San Francisco. It utilizes two linear regression models to analyze how listing features such as the number of bedrooms, neighborhood, and availability on weekends affect Airbnb prices. By examining these relationships, the study aims to provide insights that can help both tourists and hosts. Tourists can use this information to find affordable accommodations, while hosts can optimize their pricing strategies. The study's findings contribute to understanding the factors driving accommodation prices in the evolving tourism landscape.

Major Concerns and Flaws:

1. Model Complexity and Feature Selection:
    a. The study's use of linear regression models may not fully capture the complexity of the relationships between the variables. The low R-squared values suggest that the models do not explain a significant portion of the variance in Airbnb prices.
    b. The addition of polynomial features and neighborhood variables improved the model's performance, but the R-squared values remained relatively low. This indicates that other important factors might be missing from the model.
    c. Linear regression models are often insufficient for capturing complex, non-linear relationships in data. Exploring advanced models such as decision trees, random forests, or neural networks could provide better predictive performance.
    d. Feature engineering is crucial for model performance. The study could benefit from a more comprehensive exploration of potential features, such as seasonal trends, proximity to landmarks, or host-specific factors.
2. Data Preprocessing and Handling:
    a. The study mentions data cleaning but does not provide detailed information on how missing values were handled. It is crucial to ensure that the imputation methods used do not introduce biases.
    b. The removal of outliers is noted, but the criteria for identifying outliers are not discussed. This step can significantly impact the model's performance and the generalizability of the findings.
    c. Clear documentation of data preprocessing steps, including how missing values are handled and the rationale for outlier removal, is essential. This transparency ensures that the results are reproducible and that the preprocessing steps do not bias the findings.
3. Evaluation Metrics:
    a. The study uses R-squared and RMSE as evaluation metrics. While these are standard metrics, additional metrics such as Mean Absolute Error (MAE) and adjusted R-squared could provide more insights into the model's performance and help account for model complexity.
    b. Incorporating additional evaluation metrics can provide a more nuanced understanding of the model's performance. For instance, MAE is less sensitive to

outliers than RMSE, and adjusted R-squared accounts for the number of predictors in the model, providing a better measure of model fit.

Minor Concerns:

1. Visualization and Interpretation:
    a. The visualizations provided in the study, such as histograms and scatter plots, are helpful. However, adding more descriptive captions and annotations can enhance the interpretability of these figures.
    b. The study could include more visualizations comparing the actual and predicted prices, such as residual plots or parity plots, to assess model performance visually.
2. Discussion and Future Directions:
    a. The discussion section could benefit from a more detailed exploration of the implications of the findings for different stakeholders, such as Airbnb hosts and city planners.
    b. The suggested future directions, such as incorporating advanced models and additional features, are valuable. Providing a concrete plan or timeline for these future research efforts would strengthen the study's contribution.

Constructive Feedback:

Overall, the study addresses an important topic with practical implications for the Airbnb market. By addressing the concerns outlined above, particularly related to model complexity and data preprocessing, the study can enhance its robustness and applicability. Exploring advanced modeling techniques and expanding the feature set will likely yield more accurate and actionable insights. Additionally, thorough documentation of the methods and a broader range of evaluation metrics will improve the transparency and credibility of the findings.