Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

# Integration Of CRISP-DM And Machine Learning In Residential Sales Decision Making In The Middle And Upper Middle Class At Pt XYZ

[1]Patrik Mario Agata, [2]Hendra Achmadi
[1,2]Faculty of Economy, Universitas Pelita Harapan, Jakarta

Article Info
Keywords:
Random Forest
CRISP-DM
Gender
Age Group
Martial Status
Occupation
Monthly Expenditure
Monthly Income

ABSTRACT
This study was conducted to predict the effect of monthly income on 6 other demographic variables. Research data was obtained through internal questionnaire data of PT XYZ involving 393 respondents of middle residential class and 47 respondents of upper middle residential class. Data collection done with a field questionnaire containing 6 demographic questions. The data was analyzed using the CRISP-DM integration method and the Random Forest analysis method. The results of this study state that middle-class residential targets with variables such as Monthly Expenditure and Age Group have a significant influence on Monthly Income and upper-middle-class residential targets with variables such as Age Group, Employment Status and Occupation have a significant influence on Monthly Income.



Corresponding Author:
Patrik Mario Agata
Faculty of Economy, Universitas Pelita Harapan, Jakarta
patrikmario16@gmail.com

## INTRODUCTION

This thesis examines the crucial issue of Indonesia's economic growth, which is closely related to the development of the middle class. Data shows a significant decline in the size of the middle class, from 57.33 million in 2019 to 47.85 million in 2024 (Universitas Airlangga, 2024). While the government points to the COVID-19 pandemic as a major factor in this decline (BBC News Indonesia, 2024), this study also considers deeper structural factors. One of the factors identified was premature industrialization that led to reduced employment in the manufacturing sector, even before the pandemic occurred (BBC News Indonesia, 2024). Another factor exacerbating the situation is the significant rise in the cost of living, which has eroded the purchasing power of the middle class (BBC News Indonesia, 2024). Nonetheless, the market potential remains large with 140 million aspiring middle class by 2024, indicating economic opportunities if economic conditions improve (BBC News Indonesia, 2024).

The middle class plays an important role in Indonesia's economy, contributing significantly to state revenue and social stability. Based on LPEM UI data, the middle class accounts for around 50.7% of the country's total tax revenue, which is crucial for funding public development programs and infrastructure investment (CNN Indonesia, 2024). However, the decline of the middle class from 21.45% in 2019 to 17.44% in 2023 (Antara News, 2024) indicates the risk of a potential economic crisis that could affect their purchasing power and tax contributions. Furthermore, the middle class accounts for around

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

81.49% of total public consumption, with the main focus on basic needs such as food, housing, and education (Infobanknews, 2024). The decline in household consumption due to the decline in the middle class has the potential to slow economic growth, a serious concern for the government (ekon.go.id, 2024).

In this context, PT XYZ, as a property developer in Indonesia, focuses on residential development for the middle and upper-middle class. The company recognizes the challenges faced due to the declining middle class and seeks to implement various strategies to increase sales and market penetration. One of the key strategies is to understand consumer needs and preferences through in-depth market research, as emphasized by the CEO of PT XYZ, John Riady (Muhammad Rizki Vauzi, 2022). To that end, PT XYZ launched new residential products for first-time buyers and introduced price variations, with the aim of reaching a wider market segment (Muhammad Rizki Vauzi, 2022). The company is also focusing on developing midrise apartments and ready-to-occupy apartment units to meet increasing demand (Muhammad Rizki Vauzi, 2022), targeting significant revenue growth from IDR 5.2 trillion in 2022 to IDR 18.33 trillion (Muhammad Rizki Vauzi, 2022). In addition, PT XYZ supports the government's commitment to address the housing shortage in Indonesia (BM Lukita Grahadyarini, 2019) and is recognized for its innovative design, development, and implementation of sustainability and social responsibility principles (Indopremier, 2024). Data from cariproperti.com shows a significant difference in housing prices between the middle and upper-middle class, demonstrating the importance of proper market segmentation (cariproperti.com).

Advances in machine learning have had a major impact on the property sector, including in price prediction, market trends, and strategic location selection (Galuh Mafela Mutiara Sujak, 2023). The analytical capabilities of machine learning enable more accurate price predictions and the identification of important factors that affect property prices (Galuh Mafela Mutiara Sujak, 2023). Furthermore, machine learning increases efficiency for consumers and developers in finding and offering suitable properties (DQ Lab, 2023). The steady growth of the property market post-pandemic (DQ Lab, 2023), along with an increase in demand for commercial property (DQ Lab, 2023) makes the application of machine learning increasingly important to process complex data and generate deep insights. A Bank Indonesia report noted a 1.19% increase in commercial property demand in the first quarter of 2022 (DQ Lab, 2023), supported by government policies such as VAT DTP tax incentives. The increasing trend of “Machine learning” searches on Google Trends (Google Inc., 2017) and its significant projected market share growth (Goodstats 2024) indicate the widespread acceptance and utilization of this technology.

Therefore, this research aims to harness the power of machine learning in analyzing demographic data and prospective buyer behavior to identify more accurate market trends and assist PT XYZ in crafting effective sales strategies in the middle and upper-middle class segments. This research focuses on how demographic variables such as age, income, and lifestyle preferences, affect monthly income (Property Lounge, 2024) and how monthly income prediction can be optimized by utilizing data analysis and machine learning.

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

## Literature Review

### The CRISP-DM Method and its Applications

The literature review section starts by describing the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology as a systematic framework for managing data mining projects. CRISP-DM consists of six main stages: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment (Provost & Fawcett, 2020). Each stage is described in detail, emphasizing the importance of business understanding and solution relevance (Han et al., 2019), the use of data visualization for initial understanding (Han et al., 2019), feature engineering and dimensionality reduction techniques in data preparation (Aggarwal, 2020), the importance of algorithm selection and hyperparameter settings (Aggarwal, 2020), cross-validation techniques for evaluation (Provost & Fawcett, 2020), and automation in implementation (Provost & Fawcett, 2020). The relevance of CRISP-DM in this research is explained as a structured framework for analyzing middle and upper-middle class data, facilitating the effective processing of demographic and economic variables to generate insights that support data-based decision making.

### Random Forest Algorithm and its Characteristics

The literature review then discussed the Random Forest algorithm, a popular ensemble learning algorithm for classification and regression (Breiman, 2001). Random Forest builds multiple decision trees and combines the results to improve accuracy and reduce overfitting (Breiman, 2001; James et al., 2021). The working process includes bootstrap sampling, decision tree formation with feature bagging, incorporation of predictions through voting or averaging, and model evaluation (Hastie et al., 2021; Kuhn & Johnson, 2019; Zhang et al., 2020). The advantages of Random Forest include good generalization ability, tolerance for unimportant features, robustness to missing data and unbalanced data, and the ability to quantify the importance of features (James et al., 2021; Hastie et al., 2021; Zhang et al., 2020; Kuhn & Johnson, 2019). Its weaknesses include long computation time, difficulty in model interpretation, and sensitivity to redundant features (Aggarwal, 2020; James et al., 2021; Hastie et al., 2021; Zhang et al., 2020). This research chose Random Forest because of its ability to handle complex data, produce accurate predictions, overcome overfitting, and provide important information about the contribution of each variable to the prediction.

### Research Variables and Previous Research

The literature review further describes the research variables, namely Gender, Age Group, Marital Status, Occupation, Monthly Expenditure, and Monthly Income, and explains in detail the theoretical influence of each variable on Monthly Income. References to previous research using Random Forest and CRISP-DM are comprehensively reviewed. Research by Adetunji et al. (2021), Valecha et al. (2018), and Ullah et al. (2019) are discussed and compared with this study, showing similarities and differences in methodology and research focus. The research of Jaggia et al. (2020), Schröer et al. (2021), and Wiemer et al. (2019) using CRISP-DM are also reviewed, demonstrating the flexibility and relevance of CRISP-

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

DM in various contexts. Conclusions from previous research emphasize the effectiveness of Random Forest in prediction and classification (Adetunji et al., 2021; Valecha et al., 2018; Ullah et al., 2019) and the flexibility and relevance of CRISP-DM (Jaggia et al., 2020; Schröer et al., 2021; Wiemer et al., 2019). This study fills the research gap by focusing on individual income prediction based on socioeconomic variables and integration of Random Forest with CRISP-DM.

### Conceptual Framework

This section describes the conceptual framework that connects the independent variables (Gender, Age, Marital Status, Employment, Expenditure Per Month) with the dependent variable (Monthly Income). The reason for choosing Random Forest as the analysis method is explained based on its ability to handle multivariate data, overcome overfitting, provide feature importance information, and its compatibility with CRISP-DM.

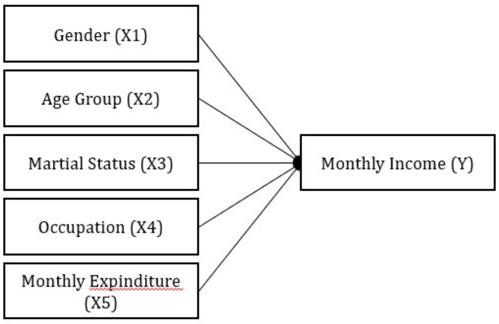



**Figure 1.** Research Model

## RESEARCH METHOD

The research method in this thesis follows the CRISP-DM framework, running in stages from business understanding to implementation. The Business Understanding stage starts by identifying the main objective: predicting the monthly income of prospective buyers of middle and upper-middle class residences at PT XYZ to develop a more effective marketing strategy. Data Understanding includes collecting PT XYZ's internal data through field surveys, initial exploration of the data using Python (Pandas and Matplotlib) to understand the distribution and relationships between variables, and data quality checks to ensure completeness and accuracy. Data Preparation includes data cleaning (removal of duplicates and outliers), data transformation (converting categorical variables to numeric with Label Encoding and StandardScaler), feature engineering (addition of 'Status_Employment' feature), and feature selection using feature importance from Random Forest. Modeling using Random Forest algorithm with hyperparameter optimization through RandomizedSearchCV. Evaluation uses several metrics (accuracy, precision, recall, F1-score), confusion matrix, ROC-AUC curve, learning curve, and feature importance analysis to assess model performance and detect overfitting. Finally, Deployment focuses on the implementation of the prediction model into PT XYZ's CRM system for real-time use, regular monitoring of model performance, and comprehensive model documentation to support business decision making.

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

## RESEARCH RESULT

### Demographic Profile Respondent

Descriptive statistical analysis was conducted to provide an overview of the research variables and respondent demographics. Descriptive statistics that will be used include frequencies and percentages. The variables to be analyzed include gender, age group, marital status, occupation, Expenditure Per Month, and Income Per Month. This analysis aims to understand the profile of respondents and identify differences in characteristics between the middle and upper middle groups.

**Table 1.** Demographic Profile Respondent Middle Class

| | Category | Frequency | Percentage (%) |
|---|---|---|---|
| Gender | Man | 178 | 45.29% |
| | Woman | 215 | 54.71% |
| Total | | 393 | 393 |
| Age Group | 17-24 years | 83 | 21.12% |
| | 25-34 years | 173 | 44.02% |
| | 35-44 years | 94 | 23.92% |
| | 45-54 years | 32 | 8.14% |
| | 55-60 years | 11 | 2.80% |
| Total | | 393 | 393 |
| Marital status | Divorced | 7 | 1.78% |
| | Bachelor | 138 | 35.11% |
| | Marry | 248 | 63.10% |
| Total | | 393 | 393 |
| Work | Housewife | 34 | 8.65% |
| | Freelancer | 9 | 2.29% |
| | Public Sector | 28 | 7.12% |
| | Private Sector | 205 | 52.16% |
| | Study (High School/D3/S1/S2) | 37 | 9.41% |
| | Businessman | 80 | 20.36% |
| Total | | 393 | 393 |
| Monthly Expenses | > 10,000,000 | 38 | 9.67% |
| | 2,000,001 - 4,000,000 | 119 | 30.28% |
| | 4,000,001 - 6,000,000 | 109 | 27.74% |
| | 6,000,001 - 8,000,000 | 74 | 18.83% |
| | 8,000,000 - 10,000,000 | 53 | 13.49% |
| Total | | 393 | 393 |
| Monthly Income | <Rp. 10,000,000 | 118 | 30.03% |
| | > Rp. 40,000,000 | 33 | 8.40% |
| | Rp. 10,000,000 - Rp. 2,000,0000 | 130 | 33.08% |
| | Rp. 20,000,000 - Rp. 3,000,0000 | 72 | 18.32% |
| | Rp. 30,000,000 - Rp. 4,000,0000 | 40 | 10.18% |
| Total | | 393 | 393 |

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

**Table 2.** Demographic Profile Respondent Upper Middle Class

| | Category | Frequency | Percentage (%) |
|---|---|---|---|
| Gender | Man | 16 | 34.04% |
| | Woman | 31 | 65.96% |
| Total | | 47 | 47 |
| Age Group | 17-24 years | 4 | 8.51% |
| | 25-34 years | 12 | 25.53% |
| | 35-44 years | 11 | 23.40% |
| | 45-54 years | 16 | 34.04% |
| | 55-60 years | 4 | 8.51% |
| Total | | 47 | 47 |
| Marital status | Divorced | 1 | 2.13% |
| | Bachelor | 7 | 14.89% |
| | Marry | 39 | 82.98% |
| Total | | 47 | 47 |
| Work | Housewife | 10 | 21.28% |
| | Freelancer | 2 | 4.26% |
| | Public Sector | 5 | 10.64% |
| | Private Sector | 16 | 34.04% |
| | Businessman | 14 | 29.79% |
| Total | | 47 | 47 |
| Monthly Expenses | > 10,000,000 | 47 | 100.00% |
| Total | | 47 | 47 |
| Monthly Income | > Rp. 40,000,000 | 11 | 23.40% |
| | Rp. 10,000,000 - Rp. 20,000,000 | 15 | 31.91% |
| | Rp. 20,000,000 - Rp. 30,000,000 | 1 | 2.13% |
| | Rp. 30,000,000 - Rp. 40,000,000 | 20 | 42.55% |
| Total | | 47 | 47 |

Source: Primary Data Processed

## Hyperparameter Tuning

The parameters tested include the number of trees (n_estimators), the maximum depth of the tree (max_depth), the minimum number of samples to split nodes (min_samples_split), and the minimum number of samples on each leaf (min_samples_leaf). The tuning process is done by performing 5-fold cross-validation to get a better performance estimation.

**Table 3.** Tuning Parameters with Tested Values

| Parameter | Tested Value |
|---|---|
| n_estimators | 100, 200, 300 |
| max_depth | 5, 10, 15, None |
| min_samples_split | 2, 5, 10 |
| min_samples_leaf | 1, 2, 4 |

Source: Proceed by Python

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

### Model Evaluation

Table 4 shows the comparison of midlle class dan upper class test data. To evaluate the performance of the trained Random Forest model, several evaluation matrices were used, namely accuracy, precision, recall, and F1-score.

**Table 4.** Comparison of Middle and Upper Middle Class Test Data

| Middle | | Upper Class | |
|---|---|---|---|
| Metrics | Nilai | Metrics | Nilai |
| Accuracy | 0.44 | Accuracy | 0.7 |
| Precision | 0.42 | Precision | 0.715 |
| Recall | 0.44 | Recall | 0.7 |
| F1-score | 0.41 | F1-score | 0.7 |

Source: Proceed by Python

### Confusion Matrix

In table 5, From this matrix, it can be seen that the model tends to classify the Intermediate data quite well, especially for the negative class. However, the model still misclassifies Intermediate data that is actually positive as negative quite often. The model shows less good performance in classifying Intermediate to Upper data. The model often misclassifies Middle to Upper data that is actually positive as Middle. This indicates that the model still has difficulty distinguishing between these two classes.

**Table 5.** Confusion Matrix

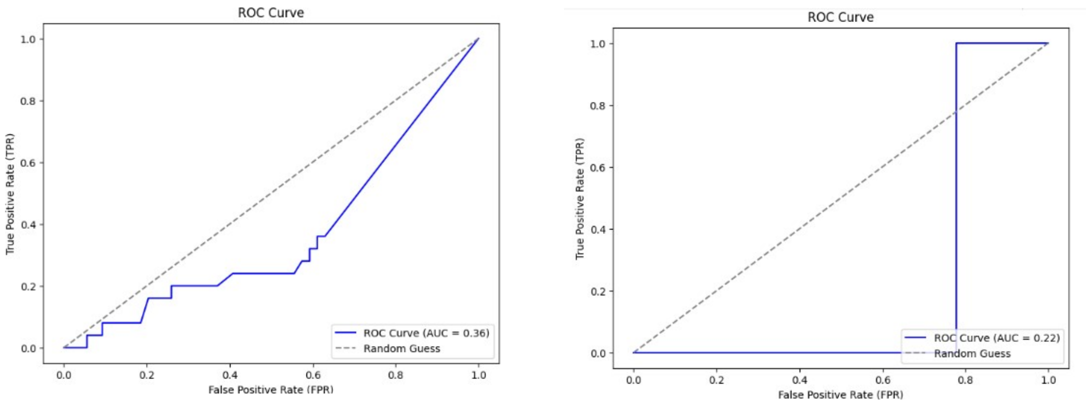
| | Middle | | Upper Middle | |
|---|---|---|---|---|
| | Positive Predictions | Negative Prediction | Positive Predictions | Negative Prediction |
| Current Positive | 0 | 3 | 3 | 0 |
| Current Negative | 0 | 15 | 0 | 1 |

Source: Procced by Python

### ROC-AUC Curve

In order to get a more comprehensive picture of the model's ability to distinguish between positive and negative classes, an analysis was conducted using the ROC curve. The ROC curve illustrates the trade-off between True Positive Rate (TPR) and False Positive Rate (FPR) at various threshold values. The area under the ROC curve (AUC) indicates the model's solving ability between the two classes.

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

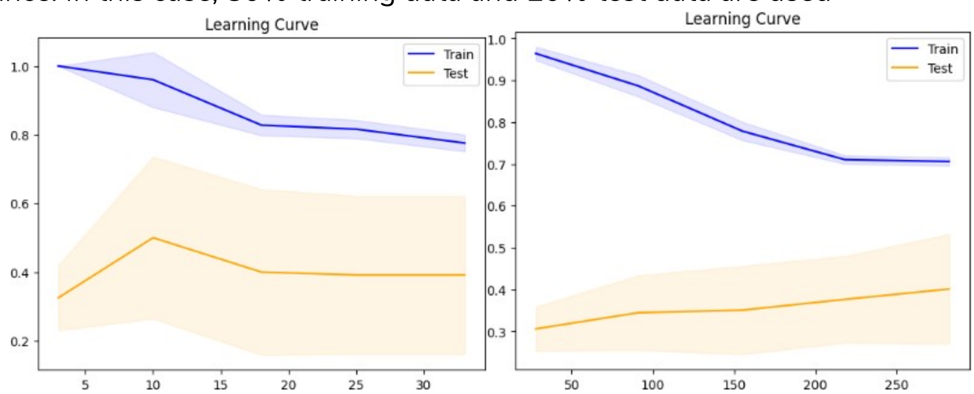



Source: Procced by Python

Figure 2. ROC-AUC Curve of Middle Class and Upper Middle Class

## R-Learning Curve

The learning curve is used to identify whether the model is experiencing overfitting or underfitting, as well as to determine whether adding training data can improve model performance. In this case, 80% training data and 20% test data are used




Source: Procced by Python

Figure 3 Learning Curve Model for Middle Class and Upper Middle Class

## Feature Importance

Feature importance analysis is performed to understand which features contribute most to the model predictions. The goal of this analysis is to find which features have the greatest influence on the model output.

Table 6. Features Importance in Random Forest Models

| Middle | | Upper Middle | |
|---|---|---|---|
| Features | Importance | Features | Importance |
| Expenditure Per Month | 0.317146 | Age Group | 0.332814 |
| Age Group | 0.266974 | Job_Status | 0.264237 |
| Job_Status | 0.147961 | Work | 0.241226 |
| Work | 0.125229 | Gender | 0.116831 |
| Gender | 0.107753 | Marital status | 0.044892 |

Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

| Middle | | Upper Middle | |
|---|---|---|---|
| Marital status | 0.034935 | Expenditure Per Month | 0.000000 |

Source: Proceed by Python

### Validation Model

To estimate model performance more accurately and reduce the risk of overfitting, cross validation was carried out using the K-fold Cross Validation method with K = 5 dividing the dataset into five equal parts. Each part is used in turn as test data, and the remaining parts are used as training data. Training and testing were carried out five times.

**Table 7** K-fold Cross Validation Results

| | Middle | Upper Class |
|---|---|---|
| Mean Score | 0.37275985663082434 | 0.46785714285714286 |
| Score Variance | 0.0029993112595560452 | 0.025816326530612242 |

Source: Proceed by Python

### Deployment

At the Deployment stage, the machine learning model that has been built is used to predict the monthly income of potential buyers in the middle and upper middle class segments. The next step is to integrate this model into the company's CRM system, so that predictions can be used in real-time to support data-based sales strategies.

**Table 8** Summary of Statistics Hypothesis Testing Results

| | | |
|---|---|---|
| Income Forecast Per Month | Middle | Rp 20.000.000 - Rp 30.000.000 |
| | Upper Middle | Rp 30.000.000 - Rp 40.000.000 |

Source: Proceed by Python

### Results and Discussion

The Results and Discussion chapter begins with a descriptive analysis to provide an overview of the characteristics of middle-class and upper-middle-class respondents. The data used came from a field survey involving 393 middle-class and 47 upper-middle-class respondents. The analysis includes frequency and percentage distributions for variables such as gender, age group, marital status, occupation, expenditure per month, and income per month. All these data were processed using Microsoft Excel (Data source: Respondent Results processed through Microsoft Excel). The results of this descriptive analysis provide an initial overview of the respondents' profiles and the differences in characteristics between the two groups, showing the basis for further analysis using the Random Forest method.

Prior to modeling with Random Forest, a crucial feature engineering stage was conducted. This process includes data transformation to handle categorical variables. Categorical variables (Gender, Age Group, Marital Status, Occupation) were converted to numeric using the Label Encoding method, and then data scaling was performed using StandardScaler. In addition, the addition of a new feature, namely "Status_Employment", which is a combination of marital status and employment, is expected to improve the accuracy of the model. The feature selection process was carried out to select the most relevant and influential variables for income prediction. All of these processes were carried out with the help of relevant Python libraries (Data source: Python analysis results). After feature engineering, the Random Forest model is built and trained using the processed data. The hyperparameter tuning process is performed to optimize the performance of the model, finding the best combination of parameters to achieve optimal prediction accuracy (Data source: Python analysis results).

The performance of the Random Forest model was evaluated using several metrics, including accuracy, precision, recall, and F1-score. The evaluation results showed significant performance differences between the two groups of respondents. The model showed higher accuracy in predicting the income of upper-middle class respondents (~70%) compared to middle class respondents (~44%). (Data source: Python analysis results). Furthermore, the confusion matrix was visualized to analyze the type of misclassification that occurred. ROC-AUC and learning curve analysis were used to detect potential overfitting or underfitting. Feature importance analysis of the Random Forest model reveals the variables that have the most influence on income prediction in each group. In the middle class group, Expenditure Per Month and Age Group have the highest weight. Whereas in the upper middle class group, Age Group, Employment Status, and Occupation contributed the most. (Data source: Python analysis results).

To ensure the generalizability of the model and reduce the risk of overfitting, model validation was conducted using the k-fold cross-validation technique with k=5. The cross-validation results show a higher average score for the upper-middle class (~47%) than the middle class (~37%), indicating better stability of model performance for the upper-middle class. (Data source: Python analysis results). The deployment phase involves applying the trained model to predict the monthly income of potential buyers at PT XYZ. These predictions are integrated into the company's CRM system to support real-time decision-making in sales strategy.

In conclusion, this study shows that the Random Forest model, within the CRISP-DM framework, can identify factors that influence monthly revenue in both market segments. Although the model performed better in the middle and upper segments, there is still room for improvement in accuracy and expansion of analysis in future research.

## CONCLUSION

The conclusion of the study shows that the Random Forest model, integrated with the CRISP-DM framework, is able to identify significant factors affecting monthly income in middle-class (Spending Per Month and Age Group) and upper-middle-class (Age Group, Employment Status, and Occupation) prospective residential buyers at PT XYZ. Although the model shows better results in the upper middle class (accuracy 0.64-0.68) than the middle class (accuracy 0.46-0.48), the overall accuracy still needs to be improved. Suggestions for future research include adding more variables, exploring other methods (such as Gradient Boosting or Neural Networks), and adding a temporal dimension to the data. Implementation suggestions include the development of inclusive financial policies, data-based financial education, and the application of data analytics technology in financial companies.

## REFERENCES

Adetunji, A. B., Akande, O. N., Ajala, F. A., Oyewo, O., Akande, Y. F., & Oluwadara, G. (2021). House Price Prediction using Random Forest Machine Learning Technique. Procedia Computer Science, 199, 806–813. https://doi.org/10.1016/j.procs.2022.01.100

Aggarwal, C. C. (2020). Data Mining: The Textbook. Springer. Aggarwal, C. C. (2020). Machine Learning for Text. Springer.

Autor, D., & Dorn, D. (2021). The Changing Structure of Work and Its Impact on Income Inequality. Economic Studies Quarterly, 92(2), 147-172.

Autor, D., & Salomons, A. (2019). New Frontiers: The Evolving Impact of Automation on Jobs and Wages. Journal of Economic Perspectives, 33(2), 3-30.

BBC News Indonesia. (2024). Nasib jadi kelas menengah di Indonesia – Banting tulang, makan tabungan, dan dilupakan - BBC News Indonesia. BBC News Indonesia. https://www.bbc.com/indonesia/articles/cy4l3z2e8xro

Blau, F. D., & Kahn, L. M. (2020). The Gender Wage Gap: Extent, Trends, and Explanations. Journal of Economic Literature, 58(3), 789-865.

Blundell, R., Pistaferri, L., & Preston, I. (2019). Consumption Inequality and Income Dynamics. American Economic Review, 109(9), 3027-3072.

Blustein, D. L., Kenny, M. E., Di Fabio, A., & Guichard, J. (2019). Expanding the Impact of the Psychology of Working: Engaging Psychology in the Struggle for Decent Work and Human Rights. Journal of Career Assessment, 27(1), 3-28.

BM Lukita Grahadyarini. (2019). Transformasi Bisnis Lippo Karawaci - Kompas. https://www.kompas.id/baca/utama/2019/03/13/transformasi-bisnis-lippo- karawaci

Browning, M., & Crossley, T. F. (2021). Household Spending and Consumption: A Behavioral Perspective. Journal of Economic Behavior & Organization, 185, 523-540.

CariProperti. (2023). Cari Properti Terbaru dan Terlengkap di Karawaci [Info + Harga]. https://cariproperti.com/tangerang/karawaci

Clarke, N., & Mahadi, N. (2020). Mutual Recognition Respect Between Leaders and Followers: Its Relationship to Employee Outcomes. Journal of Business Ethics, 164(3), 485-503.

CNN Indonesia. (2024). Mengintip Peran Kelas Menengah ke Ekonomi RI, Besarkah_. https://www.cnnindonesia.com/ekonomi/20240808164953- 532-1130857/mengintip-peran-kelas-menengah-ke-ekonomi-ri-besarkah

Deaton, A., & Muellbauer, J. (2020). Consumption and Income Patterns in a Changing Economy. Economic Perspectives Journal, 38(4), 345-368.

DQ Lab. (2023). Manfaat Machine Learning Dalam Real Estate dan Properti. Https://Dqlab.Id/Manfaat-Machine-Learning-Dalam-Real-Estate-and-Properti. https://dqlab.id/manfaat-machine-learning-dalam-real-estate-and- properti

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Dynan, K. E., Skinner, J., & Zeldes, S. P. (2020). Do the Rich Save More? A Reassessment of Savings Patterns Across Income Levels. Journal of Economic Perspectives, 34(3), 185-204.

Eickmeyer, K. J., & Manning, W. D. (2021). Marriage, Cohabitation, and Financial Decision-Making. Journal of Family and Economic Issues, 42(1), 34-49.

ekon.go.id. (2024). Hadirkan Para Menko Periode Sebelumnya, Menko Airlangga_ Kita Dorong Bersama Potensi Kelas Menengah - Kementerian Koordinator Bidang Perekonomian Republik Indonesia. https://www.ekon.go.id/publikasi/detail/5942/hadirkan-para-menko- periode-sebelumnya-menko-airlangga-kita-dorong-bersama-potensi-kelas- menengah

Frech, A., & Williams, K. (2019). Marital Status and Well-Being: A Comprehensive Review. Journal of Marriage and Family, 81(3), 825-843.

Gallie, D., Felstead, A., & Green, F. (2021). The Quality of Work and Wellbeing: A Longitudinal Analysis. Journal of Happiness Studies, 22(5), 2277-2299.

Galuh Mafela Mutiara Sujak. (2023). Bisakah Menggunakan Machine learning Untuk Menentukan Harga Properti? In Kementrian Keuangan Republik Indonesia (pp. 1–1). https://www.djkn.kemenkeu.go.id/kpknl-surabaya/baca-artikel/16335/Bisakah-Menggunakan-Machine-Learning- Untuk-Menentukan-Harga-Properti.html

Google Inc. (2017). Machine Learning -- Explore -- Google Trends. https://trends.google.com/trends/explore?q=Machine%5C Learning

Han, J., Pei, J., & Kamber, M. (2019). Data Mining: Concepts and Techniques. Elsevier.

Harjadi, C. N. (2024). Pasar Global Kecerdasan Buatan Diperkirakan Melesat pada 2024 - GoodStats. https://Goodstats.id/article/pasar-global- kecerdasan-buatan-diperkirakan-melesat-pada-2024-1pFZM

Hastie, T., Tibshirani, R., & Friedman, J. (2021). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Hurd, M. D., & Rohwedder, S. (2018). Spending and Well-Being in Retirement. Journal of Public Economics, 168, 109-121.

Indopremier. (2024). Sektor Properti Bertumbuh, Lippo Karawaci (LPKR) Pacu Penjualan Proyek Andalan. https://www.indopremier.com/ipotnews/newsDetail.php?jdl=Sektor_Properti_Bertumbuh Lippo_Karawaci_(LPKR)_Pacu_Penjualan_Proyek_Andalan&news_id=454971&group_news=RESEARCHNEWS&news_date=&taging_subtype=LPKR&name=&search=y_general&q=LippoKarawaci&halaman=1

Infobanknews. (2024). Ini Bukti Nyata Pentingnya Peran Kelas Menengah untuk Ekonomi Indonesia _ Infobanknews. https://infobanknews.com/ini-bukti- nyata-pentingnya-peran-kelas-menengah-untuk-ekonomi-indonesia/

Jaggia, S., Kelly, A., Lertwachara, K., & Chen, L. (2020). Applying the CRISP- DM Framework for Teaching Business Analytics. Decision Sciences Journal of Innovative Education, 18(4), 612–634. https://doi.org/10.1111/dsji.12222

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning with Applications in R. Springer.

Jappelli, T., & Pistaferri, L. (2020). The Economics of Consumption: Theory and Evidence. Oxford University Press.

Kahneman, D., & Deaton, A. (2020). High Income Improves Evaluation of Life but Not Emotional Well-Being Beyond a Threshold. Proceedings of the National Academy of Sciences, 117(4), 2200-2205.

Kotu, V., & Deshpande, B. (2019). Data Science: Concepts and Practice. Morgan Kaufmann.

Krueger, D., & Perri, F. (2019). Macroeconomic Shocks and Household Spending Behavior. Journal of Economic Dynamics and Control, 104, 1-20.

Kuhn, M., & Johnson, K. (2019). Feature Engineering and Selection: A Practical Approach for Predictive Models. CRC Press.

Kumar, R., Singh, P., & Sharma, A. (2021). Real Estate Price Prediction Using Random Forest Model. Journal of Machine learning Applications, 12(4), 215-229.

Lee, R., & Mason, A. (2018). Population Aging and the Generational Economy: A Global Perspective. Journal of Economic Perspectives, 32(3), 129-152.

Liputan6. (2024). Banyak Orang Ingin Beli Rumah di Tangerang, LPKR Tawarkan Proyek Lippo Village hingga Park Serpong - Bisnis Liputan6. https://www.liputan6.com/bisnis/read/5660883/banyak-orang-ingin-beli- rumah-di-tangerang-lpkr-tawarkan-proyek-lippo-village-hingga-park- serpong?page=2

Liu, H., Waite, L. J., & Shen, S. (2022). Marital Status, Social Support, and Mental Health: A Longitudinal Study. International Journal of Public Health, 67, 56-68.

Lusardi, A., & Mitchell, O. S. (2020). Financial Literacy and Planning Across the Age Spectrum. Journal of Financial Economics, 144(2), 300-325.

Marban, O., Mariscal, G., & Segovia, J. (2018). A Data Mining & Knowledge Discovery Process Model. Journal of Intelligent Information Systems, 40(2), 223–240.

Muhammad Rizki Vauzi. (2022). Lippo Karawaci (LPKR) Kejar Target Pra Penjualan, Intip Strateginya Yuk! https://www.emitennews.com/news/lippo-karawaci-lpkr-kejar-target-pra- penjualan-intip-strateginya-yuk

Muhammad, F. I. (2024). Analisis Penentuan Solusi Desain Properti Berdasarkan Pendapatan Generasi Milenial Di Yogyakarta Menggunakan N-Vivo. STORAGE: Jurnal Ilmiah Teknik Dan Ilmu ..., 3(3), 145–155.https://journal.literasisains.id/index.php/storage/article/view/3920%0Ahttps://journal.literasisains.id/index.php/storage/article/download/3920/1716

Nelson, M. R., Brunel, F. F., & Supphellen, M. (2019). Gender Differences in Consumer Behavior: Exploring Sustainable Preferences. Journal of Consumer Research, 45(6), 1152-1175.

Nelson, R. E., & Thompson, B. M. (2019). Age and Social Dynamics: Implications for Research and Policy. Population Studies Journal, 73(2), 245-268.

News, A. (2024). Menjaga kelas menengah untuk ekonomi yang stabil - ANTARA News. https://www.antaranews.com/berita/4301775/menjaga-kelas- menengah-untuk-ekonomi-yang-stabil

Perelli-Harris, B., & Styrc, M. (2019). Marriage and Mental Health: Trends Across Demographic Groups. Demographic Research, 40(8), 231-258.

Perez, C. C. (2020). Invisible Women: Exposing Data Bias in a World Designed for Men. Chatto & Windus.

Property Lounge. (2024). Bagaimana Menggunakan Analytics untuk Mengidentifikasi Tren Pasar Properti_ Panduan Lengkap - Property Lounge. https://www.propertylounge.id/bagaimana-menggunakan- analytics-untuk-mengidentifikasi-tren-pasar-properti-panduan-lengkap/

Provost, F., & Fawcett, T. (2020). Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. O'Reilly Media.

Saez, E., & Zucman, G. (2019). The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay. W.W. Norton & Company.

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. Procedia Computer Science, 181(2019), 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Singh, A., Gupta, R., & Verma, K. (2021). Age-Related Health Risks and Public Health Interventions. International Journal of Health Policy and Management, 10(4), 567-576.

Smith, J. W., & Clurman, A. (2019). Generational Behaviors and Consumer Trends: Understanding Millennials and Beyond. Journal of Consumer Studies, 45(5), 1234-1256.

Smith, J., & Wang, L. (2021). Gender Differences in Health Risks: A Comparative Analysis. International Journal of Public Health, 66, 45-59.

Smith, T., Johnson, P., & Taylor, L. (2019). Market Segmentation Analysis Using Random Forest. Marketing Science Journal, 18(3), 145-160.

Sun, X., Zhao, L., & Chen, W. (2022). Big Data Analysis Using Random Forest Algorithm: Challenges and Opportunities. Journal of Data Science, 20(1), 67-84.

Tripathi, S., Muhr, D., Brunner, M., Jodlbauer, H., Dehmer, M., & Emmert-Streib,
F. (2021). Ensuring the Robustness and Reliability of Data-Driven Knowledge Discovery Models in Production and Manufacturing. Frontiers inArtificial Intelligence, 4(June), 1–20. https://doi.org/10.3389/frai.2021.576892

Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. IEEE Access, 7(c), 60134–60149. https://doi.org/10.1109/ACCESS.2019.2914999

Universitas Airlangga. (2024). Kelas Menengah Menurun, Ancam Pertumbuhan Ekonomi Indonesia_ - Universitas Airlangga Official Website. Universitas Airlangga. https://unair.ac.id/kelas-menengah-menurun-ancam- pertumbuhan-ekonomi-indonesia/

Valecha, H., Varma, A., Khare, I., Sachdeva, A., & Goyal, M. (2018). Prediction of Consumer Behaviour using Random Forest Algorithm. 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering, UPCON 2018. https://doi.org/10.1109/UPCON.2018.8597070

Wang, J., Li, Q., & Zhang, Y. (2020). Application of Random Forest in Healthcare Prediction. International Journal of Health Informatics, 9(2), 98-112.

Warr, P., & Nielsen, K. (2018). Well-Being and Work Performance: The Role of Job Stress and Engagement. Journal of Managerial Psychology, 33(2), 123- 139.

Wiemer, H., Drowatzky, L., & Ihlenfeldt, S. (2019). Data mining methodology for engineering applications (DMME)-A holistic extension to the CRISP-DM model. Applied Sciences (Switzerland), 9(12), 403–408. https://doi.org/10.3390/app9122407

Wirth, R., & Hipp, J. (2000). CRISP-DM : Towards a Standard Process Model for Data Mining. Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining.

Wood, W., & Eagly, A. H. (2018). Gender and Social Role Theory: A Developmental Perspective. Journal of Personality and Social Psychology, 115(5), 889-902.

Zhang, Z., Wang, S., & Sun, J. (2020). Random Forests: Principles and Applications in Data Science. Journal of Big Data, 7(1), 1-12.

Zissimopoulos, J., & Karney, B. (2020). Economic Outcomes of Marital Transitions. Journal of Economic Behavior & Organization, 177, 524-537.