UNIVERSITY OF SANTO TOMAS
Institute of Information and Computing Sciences
Department of Information Systems

Machine Learning-Enhanced Site Selection Optimization for Strategic Location Analysis of Harvard Multiland Homes Real Estate Corporation Construction Projects

A Capstone Project

presented to the

Department of Information Systems

Institute of Information and Computing Sciences

University of Santo Tomas

in partial fulfillment

of the requirements for the degree of

Bachelor of Science in Information Systems

Chua, Kyle Steven T.

Concepcion, Margareth Samantha G.

Jacobo, Mikaella P.

Lazaro, Adrian DM.

Catubag, Joseph Richard G.

May 2025

TABLE OF CONTENTS

Chapter 3 **1**

3.0. Methodology 1

3.1. Data Gathering (with ETL) 1

3.2. Presentation of Constellation Schema 3

3.3.0. Mock-Ups *(images too large to upload in gforms size limit)*

3.4.0. Business Analytics Model and Testing 12

3.5. Business Analytics Tools, Techniques, and Specific Applications 36

3.6. Risk Assessment and Mitigation 39

Appendices **42**

Appendix A: THS1 Forms 42

THS1-Form1: Title Proposal Form 42

THS1-Form2: Thesis Group Advisorship Agreement Form 47

THS1-Form3A: Endorsement for Project Proposal 48

THS1-Form4A: Panel Member's Availability Confirmation 49

Appendix B: Consolidated Comments of Panel Members 50

LIST OF FIGURES

Figure 3.4.1. CRISP-DM Diagram 12

Figure 3.4.2. Conceptual Data Pipeline 14

LIST OF TABLES

Table 3.2.1. Dataset Description 4

Table 3.5.1. Risk Assessment Matrix 40

Table 3.5.2. Risk Assessment and Mitigation 41

Chapter 3

3.0. Methodology

This chapter will provide the project's methodologies that will be used for implementation. It will run through the overall procedure of how the project will come to be with the relevant technologies and resources that will be utilized by the project proponents in the application of a Machine Learning-Enhanced Site Selection Optimization for Strategic Location Analysis of Harvard Multiland Homes Real Estate Corporation Construction Projects.

3.1. Data Gathering (with ETL)

The project proponents will use the ETL (Data Extraction, Transformation, and Loading) process to prepare the data required for the system development.

a. Data Gathering

The company-provided data will be gathered and collected through communications with the client or with their representative for their preferred mode of communication. Furthermore, other relevant data that are not available from the client's database will be gathered by exploring credible public online sources like the Philippine Statistics Authority's (PSA) OpenSTAT online platform, the Humanitarian Data Exchange (HDX), geoportal PH, and the Philippine Atmospheric Geophysical and Astronomical Services Administration (PAGASA).

b. Extraction

The raw data will be extracted from the project client's database, which consists of their project costs, and from external databases, which consist of publicly available online data on geographic locations, environmental risks, human resource, population, housing value, and socioeconomic status.

c. Transformation

After the extraction, the data will undergo transformation to enhance its integrity. More detailed data preprocessing steps will include imputation of missing values, normalization of continuous variables, and encoding of categorical features.

i. Null Data

To handle null data, the researcher will consider replacing missing data with the mean value of that particular column. However, if the generated values will compromise data integrity, the researcher will opt to remove the record with null values.

ii. Duplicate Data

To handle duplicate data, the project proponents will assess the importance and identify if both are significant. Otherwise, if one of the duplicate data is found to be redundant, it will be removed to maintain data accuracy.

iii. Standardizing Data Format

To address inconsistent data formats, the data will be standardized to maintain a consistent structure in the data and compatibility across different systems.

d. Loading

Following the transformation process, the transformed data will be stored in a landing table, which will serve as the foundation for succeeding steps like data mapping and data warehousing up until the machine learning processes for analytics.

3.2. Presentation of Constellation Schema

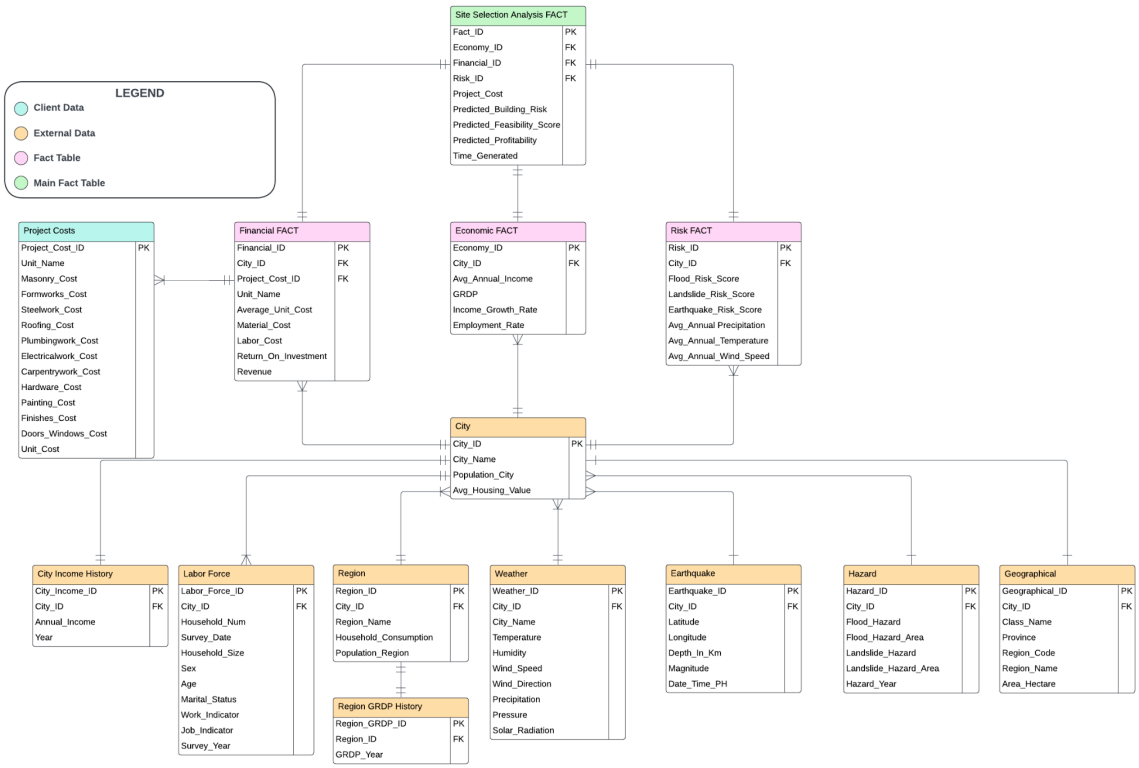


Figure 3.2.1. Constellation Schema

Figure 3.2.1 depicts the constellation schema, which provides the entities, attributes and key identifiers, and the entity relationships of the project's datasets. This will be the guiding blueprint in structuring the project's data warehouse from the gathered databases. For more clarity, table 3.2.1 in the next pages provides the descriptions of each column of each entity dataset.

Table 3.2.1. Dataset Description

CLIENT DATA	
Project Costs	
COLUMNS	**DESCRIPTION**
Project_Cost_ID	Unique identifier for project costs record
Unit_Name	Name of the construction unit
Masonry_Cost	Cost associated with masonry work
Formworks_Cost	Cost for formwork in construction
Steelwork_Cost	Cost of steel components used
Roofing_Cost	Cost for roofing materials
Plumbingwork_Cost	Cost related to plumbing tasks
Electricalwork_Cost	Cost for electrical installations
Carpentrywork_Cost	Cost associated with carpentry tasks

Hardware_Cost	Cost of hardware components
Painting_Cost	Cost for painting services
Finishes_Cost	Cost for finishing touches
Doors_Windows_Cost	Cost of doors and windows
Unit_Cost	Total cost per construction unit
EXTERNAL DATA	
City	
COLUMNS	**DESCRIPTION**
City_ID	Unique identifier for each city
City_Name	Name of the city
Population_City	Total city population
Avg_Housing_Value	The average selling price of residential properties.
City Income History	
COLUMNS	**DESCRIPTION**
City_Income_ID	Unique identifier for city income history
Annual_Income	Yearly income at historical time points

Labor Force	
COLUMNS	**DESCRIPTION**
Labor_Force_ID	Unique identifier for labor force records
Household_Num	Household identification number
Survey_Date	Date the survey was conducted
Household_Size	Number of people in the household
Sex	Gender of the surveyed person
Age	Age of the respondent
Marital_Status	Marital status of the individual
Work_Indicator	Indicates employment status
Job_Indicator	Type of job
Survey_Year	Year of the survey
Region	
COLUMNS	**DESCRIPTION**
Region_ID	Unique identifier for region records

Region_Name	Name of region
Household_Consumption	Consumption statistics per household
Population_Region	Population within the region
Region GRDP History	
COLUMNS	**DESCRIPTION**
Region_GRDP_ID	Unique identifier for GRDP history
GRDP_Year	Year for the GRDP record
Weather	
COLUMNS	**DESCRIPTION**
Weather_ID	Unique identifier for weather records.
Temperature	Measured temperature
Humidity	Humidity level
Wind_Speed	Speed of wind
Wind_Direction	Direction of wind flow
Precipitation	Amount of precipitation
Pressure	Atmospheric pressure
Solar_Radiation	Solar energy received

Earthquake	
COLUMNS	**DESCRIPTION**
Earthquake_ID	Unique identifier for earthquake records
Latitude	Exact location of the earthquake's epicenter
Longitude	
Depth_In_Km	Depth of area affected by earthquake
Magnitude	The size of the earthquake
Date_Time_PH	Date and Time earthquake was recorded
Hazard	
COLUMNS	**DESCRIPTION**
Hazard_ID	Unique identifier for hazard records
Flood_Hazard	Presence or degree of flood hazard
Flood_Hazard_Area	Area affected by flood hazard
Landslide_Hazard	Presence or degree of landslide hazard
Landslide_Hazard_Area	Area affected by landslide hazard
Hazard_Year	Year hazard was recorded

Geographical	
COLUMNS	**DESCRIPTION**
Geographical_ID	Unique identifier for geographical records
Class_Name	Land use or land cover classification of the area e.g., Built-up, Open Forest, Annual Crop, etc.)
Province	Province where the city is located.
Region_Code	Regional code designation
Area_Hectare	Area in hectares.
FACT TABLES	
Economic	
COLUMNS	**DESCRIPTION**
Economy_ID	Unique identifier for each economy record.
Avg_Annual_Income	Average annual income in the city.
GRDP	(Gross Regional Domestic Product) The total economic output produced within a city or region.
Income_Growth_Rate	Annual income growth rate
Employment_Rate	Percentage of the working-age population that is currently employed.

Financial	
COLUMNS	**DESCRIPTION**
Financial_ID	Unique Identifier for financial records
Average_Unit_Cost	Average cost per unit
Material_Cost	Total cost of physical construction materials used for a specific unit
Labor_Cost	Total cost on human labor required for constructing a unit
Return_On_Investment	Estimated ROI from the project
Revenue	Aggregation between Economic Factors and Cost Factors
Risk	
COLUMNS	**DESCRIPTION**
Risk_ID	Unique identifier for each environmental risk record.
Flood_Risk_Score	A numerical index representing the likelihood and severity of flood events in a city.
Landslide_Risk_Score	A numerical index representing the likelihood and severity of landslides in a city.
Avg_Annual_Precipitation	Yearly average rainfall
Avg_Annual_Temperature	Yearly average temperature

Avg_Annual_Wind_Speed	Yearly average wind speed
MAIN FACT	
Site Selection Analysis	
COLUMNS	**DESCRIPTION**
Fact_ID	Unique identifier for site selection records
Project_Cost	Total cost of the project
Predicted_Building_Risk	Estimated construction risk
Predicted_Feasibility_Score	A computed score that estimates the practicality and likelihood of successfully completing a construction project
Predicted_Profitability	Estimated profitability of the site
Time_Generated	For data versioning.

3.4.0. Business Analytics Model and Testing

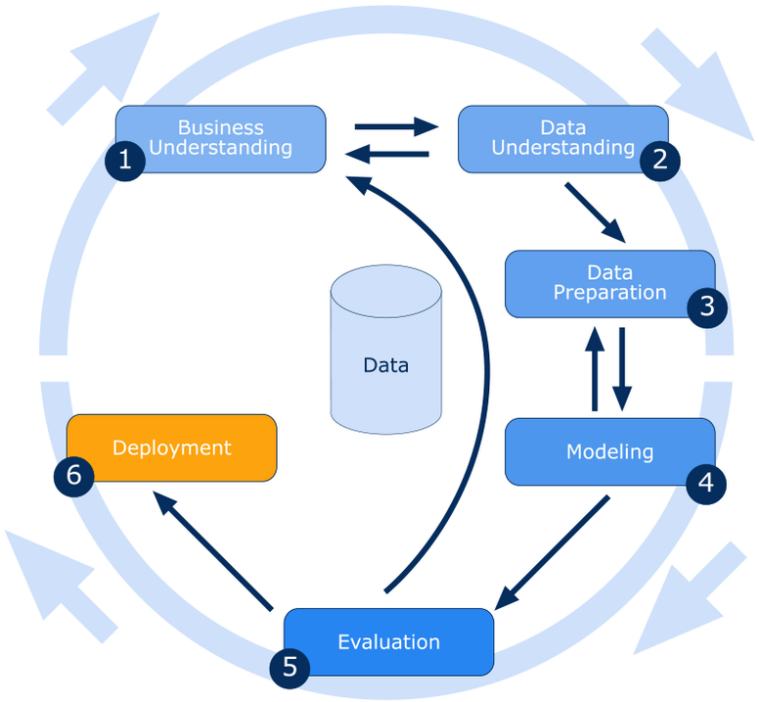


Figure 3.4.1. CRISP-DM Diagram

The CRISP-DM or Cross-Industry Standard Process for Data Mining Diagram above will serve as the overall roadmap for the proponents to follow from implementing the project up to its methodologies for machine learning and analytics shown in the conceptual framework in chapter 1. Below are the descriptions of each step involved in this methodological roadmap:

1. Business Understanding

This phase involves gaining a clear and comprehensive understanding of the project objectives, goals, and requirements from the business perspective of the project's client. It focuses on identifying the key business problems to be solved, aligning data mining goals with business priorities, and establishing success criteria to ensure the project delivers real value.

2. Data Understanding

This phase focuses on collecting, exploring, and evaluating the initial data. It involves identifying data quality issues, discovering patterns, and gaining insights that help shape the direction of the project.

3. Data Preparation

This phase involves selecting, cleaning, and transforming raw data into a suitable format for modeling. This includes handling missing values, removing duplicates, creating new variables, and integrating data from multiple sources.

4. Modeling

In this phase, various modeling techniques will be selected and applied to the prepared data. The project proponents will experiment with different algorithms and refine their parameters to optimize the effectiveness and accuracy of the data analyses. Often, modeling reveals data issues, so iteration with the data preparation phase may occur with this project.

5. Evaluation

After building the models, they will be evaluated to ensure they meet the business objectives of the project's client by interpreting the results, validating the models' performance, and deciding whether the model is good enough to move to deployment or if further refinement is needed.

6. Deployment

In the deployment phase, the final models will be integrated into the business process by embedding them into the actual system dashboard, or developing decision

support tools. Proper monitoring and and testing plans for maintenance are also established here.

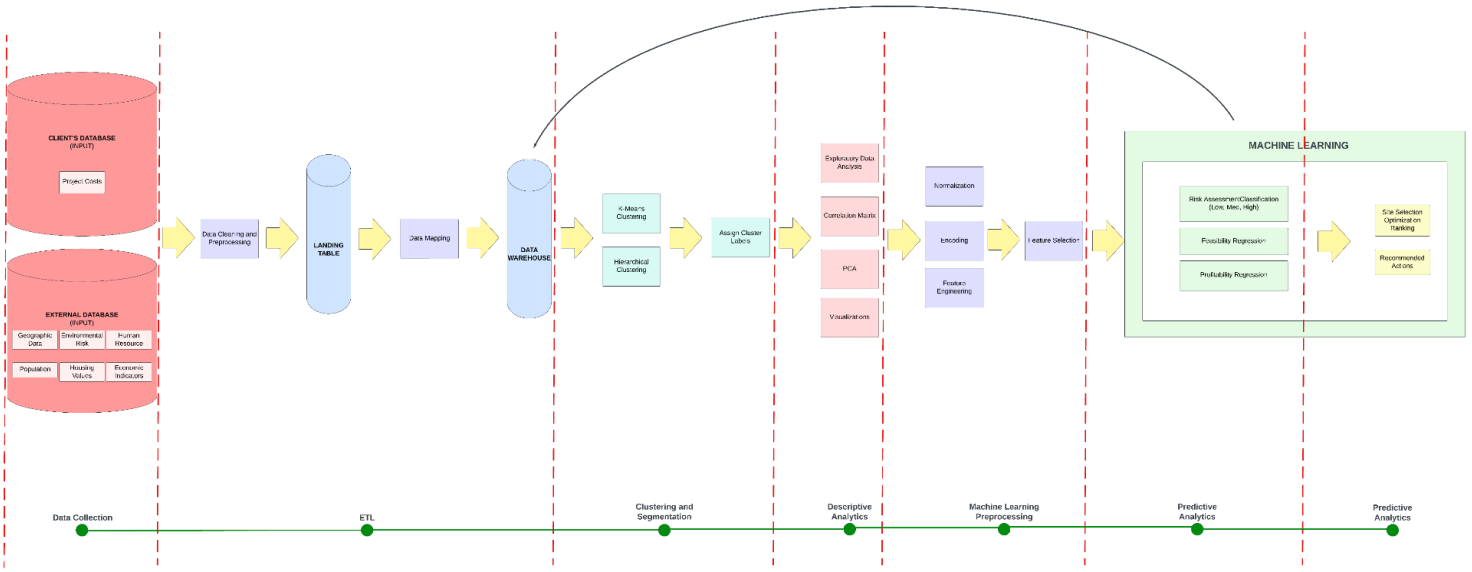


Figure 3.4.2. Conceptual Data Pipeline

Basing on the project's Conceptual Framework, Figure 3.3.2 illustrates the project's Conceptual Data Pipeline, which details the flow of data from collection to consumption. This pipeline serves as a roadmap for managing all collected data inputs, guiding the application of machine learning and other analytical methods to produce predictive insights. These insights will support site selection by leveraging relevant factors and employing suitable models for each type of analytics, as outlined in the following sections for the project proponents' consideration.

Data Collection and ETL

The first phase of the pipeline involves collecting data from internal and external sources. Internal data is derived from the client's database, particularly focusing on project

costs. External data includes geographic data, population statistics, housing values, environmental risk factors, economic indicators, and human resource data.

Once collected, the data undergoes ETL (Extract, Transform, Load) processes. This phase involves data cleaning to handle missing values, normalization to standardize feature scales, encoding of categorical variables, and composite feature engineering. After cleaning, the data is mapped to consistent formats and stored in a landing table before being integrated into the centralized data warehouse. This warehouse serves as the central repository from which subsequent analytical processes draw their inputs.

Clustering and Segmentation

After data integration, clustering methods are applied to group sites based on similarities in characteristics. Techniques such as K-Means and Hierarchical Clustering are employed to segment data, helping identify homogeneous groups among potential construction sites. Assigning cluster labels enables the system to organize the data into meaningful categories, which enhances the precision of subsequent analysis.

Descriptive Analytics

During the descriptive analytics stage, Principal Component Analysis (PCA) is applied to reduce dimensionality while retaining the most significant variance among features. This step helps the team identify key factors that influence site selection. Following PCA, a correlation matrix is generated to visualize relationships between variables such as project costs, population density, housing values, and environmental

risk. This analysis is crucial for identifying potential multicollinearity, which could impact model performance. Additionally, visualizations are created to present risk levels, economic indicators, and housing values across clustered groups, aiding in comprehensive data interpretation.

Machine Learning Preprocessing

Before model training, additional preprocessing steps are conducted to enhance model performance. This includes data normalization to ensure uniform feature scaling and advanced feature engineering to capture latent patterns within the data. Techniques such as one-hot encoding for categorical data and scaling for numerical variables are employed. Feature selection methods are also applied to retain only the most impactful variables for predictive modeling.

Predictive Analytics

The project aims to develop three distinct machine learning models to support site selection:

1. **Risk Assessment Model (Classification)**

   This model predicts the risk level associated with building at a particular site, categorized as Low, Medium, or High. Algorithms such as Random Forest and XGBoost are evaluated for their ability to classify sites based on factors including flood risk, landslide risk, income levels, and population density.

2. **Feasibility Prediction Model (Regression)**

   This model forecasts the feasibility of site development, determining whether the location is suitable for timely construction and sales. Regressors like Random Forest Regression and Gradient Boosting (e.g., XGBoost) are used, with inputs including GRDP, housing values, and project costs.

3. **Profitability Prediction Model (Regression)**

   This model estimates the potential return on investment (ROI) for each site. Depending on the performance during model evaluation, either Random Forest Regression or regularized linear regression techniques (e.g., Ridge or Lasso) will be utilized. These models will help predict the financial viability of selected sites.

These predictive models will potentially forecast site feasibility and profitability, guiding the decision-making process for the project's stakeholders.

Prescriptive Analytics

After generating predictive insights, the system produces a prescriptive output by ranking the optimal sites for construction projects of Harvard Multiland Homes Real Estate Corporation. This ranking is presented through an interactive heatmap dashboard, allowing users to visualize feasible and profitable locations. The ranking model employs Multi-Criteria Decision Making (MCDM) techniques, specifically the Interval Analytic Hierarchy Process (IAHP), to account for uncertainties in criteria weighting and suitability scoring. The system can also suggest alternative sites for consideration based on varying factorial conditions, assisting executive decision-makers in making informed choices.

Action Library and Model Retraining

The Action Library acts as the centralized repository for processed outputs, scoring data, and ranked site recommendations. This library feeds directly into the decision-making tools and dashboards. Additionally, a feedback loop is established, where data from the Action Library is periodically fed back into the data warehouse. This triggers retraining of the models when significant new data is available, maintaining the model's accuracy and relevance as the environment and data evolve.

Model Testing and Evaluation

For model testing and evaluation, classification models will be assessed using performance metrics such as accuracy, F1-score, and AUC-ROC, with k-fold cross-validation to ensure generalizability. Regression models will be evaluated using RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and R^2 (coefficient of determination), along with residual analysis to inspect assumptions and outlier influence.

a. Accuracy

Accuracy is a fundamental metric used to evaluate classification models. It measures the proportion of correctly predicted observations over the total number of predictions made. While accuracy provides a general idea of how well the model performs, it can be misleading in the presence of imbalanced classes, where one category significantly outnumbers another, since a model may appear accurate by simply predicting the dominant class. In this project, accuracy will be used alongside other metrics to ensure balanced evaluation of risk level predictions.

b. F1-Score

The F1-score offers a more balanced performance assessment for classification tasks, especially when dealing with class imbalance. It is the harmonic mean of precision and recall, providing a single score that accounts for both false positives and false negatives. This metric is particularly useful when the cost of misclassifying a high-risk site as low-risk (or vice versa) is significant. In this context, F1-score ensures that the classification model not only achieves high correctness but also maintains meaningful sensitivity to critical risk classifications.

c. AUC-ROC

The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is a classification performance metric that evaluates the model's ability to distinguish between classes across various threshold settings. A higher AUC indicates a better model at separating classes, such as high-risk versus low-risk locations. This metric is valuable in site risk assessment, as it reflects the likelihood that the model ranks a randomly chosen positive instance higher than a randomly chosen negative one.

d. K-Fold Cross-Validation

K-fold cross-validation is a resampling method used to assess model generalizability. The dataset is divided into k equal parts or folds; the model is trained on k-1 folds and validated on the remaining fold. This process repeats k

times, with each fold serving as the validation set once. The results are then averaged to provide a robust estimate of model performance. Applying this method helps mitigate overfitting and ensures that model performance is not overly dependent on a specific subset of the data.

e. Root Mean Squared Error (RMSE)

RMSE is a widely used metric in regression tasks, measuring the square root of the average squared differences between predicted and actual values. It emphasizes larger errors more than smaller ones, making it sensitive to outliers. In this study, RMSE will be used to evaluate the feasibility and profitability regression models, with lower RMSE values indicating better predictive accuracy and consistency.

f. Mean Absolute Error (MAE)

MAE calculates the average absolute difference between predicted values and actual observations. Unlike RMSE, it treats all errors equally without giving additional weight to larger deviations. This makes MAE a more interpretable and stable measure when the model is exposed to minor prediction inaccuracies. It will be applied to assess the average prediction error in estimating feasibility scores.

g. R-Squared (R^2)

R^2, or the coefficient of determination, evaluates how well the independent variables explain the variation in the dependent variable. A higher R^2 indicates that the model captures more of the data's underlying patterns. In this project, R^2 will provide insight into how effectively the selected features contribute to predicting feasibility and profitability, helping assess the explanatory strength of the regression models.

h. Residual Analysis

Residual analysis involves examining the differences between actual and predicted values to validate regression model assumptions. This method helps detect issues such as non-linearity, heteroscedasticity, and the presence of outliers. By analyzing residual patterns, the project team can determine whether additional data preprocessing or model refinement is required, thereby improving the robustness of the predictive models.

Test Cases

This next section outlines the test cases designed to verify the system's functionality and performance. The test cases cover key features to ensure the application works as expected in different scenarios. The goal is to ensure the system meets the required specifications and provides a smooth user experience.

I. User Authentication

TEST CASE						
Test Case ID: UA_0001				**Test Designed by: Mikaella P. Jacobo**		
Test Priority (Low/Medium/High): High				**Test Designed date: 05/07/25**		
Module Name: Authentication				**Test Executed by:**		
Test Title: Login using valid credential				**Test Execution date:**		
Description: Test the Site Selection System Login Page for valid users						
Pre-conditions: User has account in the system						
Dependencies: Authentication Database connected properly.						
Step	**Test Steps**	**Test Data**	**Expected Result**	**Actual Result**	**Status (Pass/ Fail)**	**Notes**
1	Navigate the login page		Authentication succeeds			
2	Enter valid email	Email: correctEmail				
3	Enter valid password	Password: correctPassword				
4	Click the “Login” button					
Post-conditions: User is authenticated.						

TEST CASE						
Test Case ID: UA_0002				**Test Designed by: Mikaella P. Jacobo**		
Test Priority (Low/Medium/High): High				**Test Designed date: 05/07/25**		
Module Name: Authentication				**Test Executed by:**		
Test Title: Login using invalid credential				**Test Execution date:**		
Description: Ensures that the system denies access for invalid credentials						
Pre-conditions: User has account in the system						
Dependencies: Authentication Database connected properly.						
Step	**Test Steps**	**Test Data**	**Expected Result**	**Actual Result**	**Status (Pass/Fail)**	**Notes**
1	Navigate the login page		Authentication fails			
2	Enter invalid email	Email: correctEmail				
3	Enter invalid password	Password: correctPassword				
4	Click the “Login” button					
Post-conditions: User is rejected.						

II. Importing Data

TEST CASE						
Test Case ID: ID_0001				**Test Designed by: Mikaella P. Jacobo**		
Test Priority (Low/Medium/High): High				**Test Designed date: 05/07/25**		
Module Name: Authentication				**Test Executed by:**		
Test Title: Import Process for City Data				**Test Execution date:**		
Description: The user can import city data into the system						
Pre-conditions: User has account in the system						
Dependencies: UA_0001, HM_0001						
Step	**Test Steps**	**Test Data**	**Expected Result**	**Actual Result**	**Status (Pass/Fail)**	**Notes**
1	Click the “Import Data” button		A modal will appear with an upload button			
2	Click the “Upload button”		The data will be uploaded into the system			
Post-conditions: The uploaded dataset will be included in the process of data warehousing up to all of the analytical processes						

III. Interactive Heat Map Dashboard

TEST CASE						
Test Case ID: HM_0001				**Test Designed by: Mikaella P. Jacobo**		
Test Priority (Low/Medium/High): High				**Test Designed date: 05/07/25**		
Module Name: Dashboard				**Test Executed by:**		
Test Title: Regional Map Display				**Test Execution date:**		
Description: Ensures that the map will display specific regions.						
Pre-conditions: User is logged in.						
Dependencies: UA_0001, ID_0001						
Step	**Test Steps**	**Test Data**	**Expected Result**	**Actual Result**	**Status (Pass/Fail)**	**Notes**
1	Access the login page	Email: correctEmail Password: correctPassword	The user will be redirected to the map display.			
2	Import data		The map will only display the following regions: Central Luzon, NCR, and Calabarzon			
Post-conditions: The map will display specific regions.						

TEST CASE						
Test Case ID: HM_0002				**Test Designed by: Mikaella P. Jacobo**		
Test Priority (Low/Medium/High): High				**Test Designed date: 05/07/25**		
Module Name: Dashboard				**Test Executed by:**		
Test Title: Access Map Toolbar				**Test Execution date:**		
Description: Ensures that the button performs its intended function correctly						
Pre-conditions: User is logged in.						
Dependencies: UA_0001, ID_0001						
Step	**Test Steps**	**Test Data**	**Expected Result**	**Actual Result**	**Status (Pass/Fail)**	**Notes**
1	Click the “?” button		An Info box should be displayed, providing instructions on how to navigate the dashboard			
2	Click the “+” button		The user should be able to zoom in on the map to view the city.			
3	Click the “-” button		The user should be able to zoom out on the map			
Post-conditions: The map toolbar functions correctly.						

TEST CASE						
Test Case ID: HM_0003				**Test Designed by: Mikaella P. Jacobo**		
Test Priority (Low/Medium/High): High				**Test Designed date: 05/07/25**		
Module Name: Dashboard				**Test Executed by:**		
Test Title: Selected Region Display				**Test Execution date:**		
Description: Ensures that the user can view the city ranking and recommendation						
Pre-conditions: User is logged in.						
Dependencies: UA_0001, HM_0001						
Step	**Test Steps**	**Test Data**	**Expected Result**	**Actual Result**	**Status (Pass/Fail)**	**Notes**
1	Click a specific region		The ranking and recommendation will appear in the sidebar			
Post-conditions: The user will be able to view the ranking and recommendation.						

TEST CASE						
Test Case ID: HM_0004				**Test Designed by: Mikaella P. Jacobo**		
Test Priority (Low/Medium/High): High				**Test Designed date: 05/07/25**		
Module Name: Dashboard				**Test Executed by:**		
Test Title: City-Level Map View				**Test Execution date:**		
Description: The city view will be visible when the map is zoomed in closely.						
Pre-conditions: User is logged in.						
Dependencies: UA_0001, HM_0001						
Step	**Test Steps**	**Test Data**	**Expected Result**	**Actual Result**	**Status (Pass/Fail)**	**Notes**
1	Zoom in the regional map		The city view should be displayed when the map is zoomed in			
Post-conditions: The user can view the city-level map						

TEST CASE						
Test Case ID: HM_0005				**Test Designed by: Mikaella P. Jacobo**		
Test Priority (Low/Medium/High): High				**Test Designed date: 05/07/25**		
Module Name: Dashboard				**Test Executed by:**		
Test Title: Selected City Display				**Test Execution date:**		
Description: Ensures that the user can view recommendation						
Pre-conditions: User is logged in.						
Dependencies: UA_0001, HM_0001, HM_0004						
Step	**Test Steps**	**Test Data**	**Expected Result**	**Actual Result**	**Status (Pass/Fail)**	**Notes**
1	Click a specific region then zoom in		The city on that specific region should be visible			
2	Click a specific city		The recommendation will appear in the sidebar			
Post-conditions: The user will be able to view recommendations.						

TEST CASE						
Test Case ID: HM_0006				**Test Designed by: Mikaella P. Jacobo**		
Test Priority (Low/Medium/High): High				**Test Designed date: 05/07/25**		
Module Name: Dashboard				**Test Executed by:**		
Test Title: Provincial-Level Map View				**Test Execution date:**		
Description: The provincial view will be visible when the map is zoomed in closely. (only applicable to central luzon and CALABARZON)						
Pre-conditions: User is logged in.						
Dependencies: UA_0001, HM_0001, HM_0004						
Step	**Test Steps**	**Test Data**	**Expected Result**	**Actual Result**	**Status (Pass/Fail)**	**Notes**
1	Zoom in the city-level map		The provincial view should be displayed when the map is zoomed in			
Post-conditions: The user can view the provincial-level map						

TEST CASE						
Test Case ID: HM_0007				**Test Designed by: Mikaella P. Jacobo**		
Test Priority (Low/Medium/High): High				**Test Designed date: 05/07/25**		
Module Name: Dashboard				**Test Executed by:**		
Test Title: Selected Provincial Display				**Test Execution date:**		
Description: Ensures that the user can view recommendations. (only applicable to central luzon and CALABARZON)						
Pre-conditions: User is logged in.						
Dependencies: UA_0001, HM_0001, HM_0004, HM_0006						
Step	**Test Steps**	**Test Data**	**Expected Result**	**Actual Result**	**Status (Pass/Fail)**	**Notes**
1	Click a specific city then zoom in		The province on that specific city should be visible			
2	Click a specific province		The recommendation will appear in the sidebar			
Post-conditions: The user will be able to view recommendations.						

TEST CASE						
Test Case ID: HM_0008				**Test Designed by: Mikaella P. Jacobo**		
Test Priority (Low/Medium/High): High				**Test Designed date: 05/07/25**		
Module Name: Dashboard				**Test Executed by:**		
Test Title: View Report Overview				**Test Execution date:**		
Description: The user will be able to view the report overview						
Pre-conditions: The user selected a specific region, or province						
Dependencies: UA_0001, HM_0001, HM_0003, HM_0005, HM_0007						
Step	**Test Steps**	**Test Data**	**Expected Result**	**Actual Result**	**Status (Pass/Fail)**	**Notes**
1	Click the “Overview” tab		The Total Population, Total Income, Total GRDP, Project Cost, Building Feasibility Score, Building Risk, Profitability, and Work Force Indicator will be displayed			
Post-conditions: The report overview is displayed.						

TEST CASE						
Test Case ID: HM_0009				**Test Designed by: Mikaella P. Jacobo**		
Test Priority (Low/Medium/High): High				**Test Designed date: 05/07/25**		
Module Name: Dashboard				**Test Executed by:**		
Test Title: View Report Risk Indicator				**Test Execution date:**		
Description: The user will be able to view the report risk indicator						
Pre-conditions: The user selected a specific region, city, or province						
Dependencies: UA_0001, HM_0001, HM_0003, HM_0005, HM_0007						
Step	**Test Steps**	**Test Data**	**Expected Result**	**Actual Result**	**Status (Pass/Fail)**	**Notes**
1	Click the “Risk Indicator” tab		The risk meter, flood score , and landslide score will be displayed.			
Post-conditions: Report risk indicator data are displayed						

IV. Search Functionalities

TEST CASE						
Test Case ID: S_0001				**Test Designed by: Mikaella P. Jacobo**		
Test Priority (Low/Medium/High): High				**Test Designed date: 05/7/25**		
Module Name: Searching				**Test Executed by:**		
Test Title: Test Search for Regions and Cities				**Test Execution date:**		
Description: Validates the search feature for finding specific regions or cities.						
Pre-conditions: Regions and cities exist in the system.						
Dependencies: UA_0001, HM_0001, HM_0002, HM_0003						
Step	**Test Steps**	**Test Data**	**Expected Result**	**Actual Result**	**Status (Pass/Fail)**	**Notes**
1	Enter search query	Query: "Central Luzon" Or “Tarlac City”	All possible matching result will be displayed in a drop-down list beneath the search bar			
2	Select the region or city that matches your search		A sidebar will appear containing recommendations when the query corresponds to a specific region/city on the map.			
Post-conditions: Search results display specific region/city based on query.						

V. Logout

TEST CASE						
Test Case ID: L_0001				**Test Designed by: Mikaella P. Jacobo**		
Test Priority (Low/Medium/High): Low				**Test Designed date: 05/07/25**		
Module Name: Logout				**Test Executed by:**		
Test Title: Site Selection System Log-out				**Test Execution date:**		
Description: Test the logging out of the Site Selection System						
Pre-conditions: User is logged in.						
Dependencies: UA_0001						
Step	**Test Steps**	**Test Data**	**Expected Result**	**Actual Result**	**Status (Pass/Fail)**	**Notes**
1	Click the "Logout" button		Prompt message will appear to confirm logout.			
2	When the user click the "No" button		The user will be returned to the page they were viewing prior to clicking the logout button.			
3	Click "Yes" button		The user is redirected back to the log-in screen			
Post-conditions: The user is logged out of the Site Selection System and can log-in again with valid credentials.						

3.5. Business Analytics Tools, Techniques, and Specific Applications

From conversations and meetings with our client, below are some specific business questions or problems that they wish to be answered and solved with the implementation of this project for their company's executives' utilization for decision-making:

1. What available locations are the most optimal based on multiple strategic factors, and how can Harvard Multiland Homes Real Estate Corporation make smarter and data-driven decisions when choosing locations from those available locations for new construction projects to avoid unsuccessful ones?
2. How can multiple decision factors from applicable online sources and from Harvard Multiland Homes Real Estate Corporation's own data resources be combined systematically to predict potential risks and profitability using their relevance and associations?
3. What features significantly help in effectively and descriptively visualizing important information for smarter decision-making in site selection?

These questions above should serve as helpful guides for the project proponents to answer in alignment with the set general and specific objectives leading to the functional and serviceable output of an Interactive Heatmap Dashboard.

Furthermore, the tools that shall be utilized for practical use in achieving the goals of implementing this project's objectives include, but will not be limited to, the following:

a. **Python**

The project proponents will primarily make use of the programming language Python for most of its technical aspects including data preprocessing, data visualization, and data modeling, and with libraries such as scikit-learn,

xgboost, pandas, and geopandas for machine learning and geospatial data analysis. IDEs such as Spyder and Visual Studio Code shall serve as the platforms to make use of with Python according to the developers' preferences.

b. **Power BI**

For more advanced data transformation and visualization features, Power BI shall be the platform to export preprocessed datasets by using Python, and it can also be further utilized or even integrated to the actual Interactive Heatmap Dashboard for viewing relevant datasets in the project's descriptive, predictive, and prescriptive modeling of the available data. It shall be employed to create the interactive dashboard displaying risk classifications, feasibility scores, and profitability estimates.

Moreover, below are some other methods that the project proponents may consider, such as their budget, for effortless and less demanding utilizations of methods that lead to achieving the goals of implementing this project's objectives:

Amazon Web Services (AWS)

Amazon Web Services (AWS) is a comprehensive cloud computing platform that offers a wide range of scalable and cost-efficient services for data processing, storage, analytics, and machine learning. In the context of this project, AWS may serve as the backbone infrastructure supporting the entire analytics workflow. The following services are specifically the services to be considered for this project's implementation:

a. AWS Glue

This Amazon Service can automate the Extract, Transform, Load (ETL) process by pulling data from both internal and external databases into a centralized data repository.

b. AMAZON S3

This Amazon Service can serve as the data lake (“landing table”) for storing extracted, generated, and transformed datasets.

c. AMAZON Redshift or AMAZON Relational Database Service (PostgreSQL)

These Amazon Services may be used for structured data warehousing and querying from the data lake in AMAZON S3. Attribute selection may also be done here from the structured data warehouse.

d. AMAZON SageMaker

This Amazon Service may be utilized for its machine learning and analytics capabilities as it enables building, training, and deploying machine learning models for the project’s prediction goals on risk assessment, profitability prediction, and development feasibility.

e. AMAZON Athena

This Amazon Service may be used in querying the data stored in the action library from the deployed machine learning models by AMAZON SageMaker for essential and necessary data exploration.

Furthermore, these tools shall further prove to be necessary and applicable in applying the following techniques for the project:

a. Data Visualization

Application of Descriptive Analytics and the execution of clustering the relevant factors or attributes from the datasets, which will provide more sensible insights through the interactive heatmap dashboard.

b. **Predictive Modeling**

Application of Predictive Analytics and modeling to reveal optimal sites, predicting profitability by using Regression Analysis and Random Forest Regressor, and open up possible outcomes in choosing a site such as high risk low reward, low risk high reward, and etc. by using classification algorithms and methods such as Random Forest Classifier.

These visualizations and models garnered will further be put to use in the project's application of **Prescriptive Analytics** in directly showing the data-driven ranking of optimal sites in advising the next possible steps for the executives to decide on with the provided actionable insights for selecting the most suitable construction sites based on their risk profiles, development potential, and expected financial returns.

3.6. Risk Assessment and Mitigation

This next section in the next few pages shows Table 3.5.1: Presents the risk assessment scores and corresponding descriptions for each identified risk. The project proponents have identified potential risks that may impact the project's quality, security, and overall success. Recognizing and addressing these risks is essential to ensure the project's successful implementation and long-term viability.

Table 3.5.1. Risk Assessment Matrix

LEGEND					
Probability	**Severity**				
	Insignificant (1)	Minor (2)	Significant (3)	Major (4)	Severe (5)
Almost Certain (5)	Medium	Medium	High	High	High
Likely (4)	Low	Medium	Medium	High	High
Moderate (3)	Low	Medium	Medium	Medium	High
Unlikely (2)	Low	Low	Medium	Medium	Medium
Rare (1)	Low	Low	Low	Low	Medium

Probability (Likelihood of Occurrence)

This refers to the estimated likelihood that a particular event or risk will occur during the course of the project.

- **Almost Certain** – The event is expected to occur under normal circumstances.
- **Likely** – The event will probably occur in most situations.
- **Moderate** – The event may occur at some point.
- **Unlikely** – The event is possible but not anticipated in the near future.
- **Rare** - The event is highly unlikely to occur but cannot be completely ruled out.

Severity (Impact on Project Outcomes)

This describes the potential consequences of a risk event, should it occur.

- **Severe** – A serious problem that stops major parts of the project or causes a major system failure.
- **Major** – A big issue that causes delays, added costs, or the need to redo important work.

- **Significant** – A problem that affects the project noticeably but can be handled without major delays.
- **Minor** – A small issue that causes little disruption and is easy to fix.
- **Insignificant** – A very minor issue with no real effect on the project.

Table 3.5.2. Risk Assessment and Mitigation

ID	IDENTIFIED RISK	RISK SCORE (Probability x Severity)	MITIGATION ACTION
1	Model Inaccuracy and Bias	Medium (12)	Validate models using multiple metrics (e.g MAE, Silhouette Score, and etc.) and implement bias detection tools.
2	Insufficient or Poor Quality Data	Medium (6)	Conduct early data audits, and implement preprocessing checks.
3	Security and Data Privacy Concern	Medium (9)	Apply encryption techniques, and ensure compliance with local data protection laws.
4	Data Loss	High (15)	Implement regular automated backup processes.
5	Internet Issues	Low (4)	Ensure availability of alternative backup connection options.
6	Researcher's Technical Expertise	High (15)	Conduct training and development programs with Business Analytics/ML experts.

Appendices

Appendix A: THS1 Forms

THS1-Form1: Title Proposal Form

UNIVERSITY OF SANTO TOMAS
Institute of Information and Computing Sciences
Department of Information Systems
2nd Semester AY 2024-2025

CAPSTONE PROJECT PROPOSAL TITLE FORM

Name of the Proponents: Section: 3ISA Date: May 28, 2025
1. Chua, Kyle Steven, T.
2. Concepcion, Margareth Samantha, G.
3. Lazaro, Adrian DM.
4. Jacobo, Mikaella, P.

1.0. **Proposed Capstone Title:** Machine Learning-Enhanced Site Selection Optimization for Strategic Location Analysis of Harvard Multiland Homes Real Estate Corporation Construction Projects

2.0. **Area of Investigation:**

Site selection is a crucial process for a real estate corporation to further commit on planning construction projects because many factors are necessary to be considered such as location and profitability. These can significantly help a company in identifying optimal and strategic ways of achieving further success in their builds. Overlooking these aspects could potentially redirect a company's projects further away from their goals to success.

3.0. **Importance / Significance of the Study:**
4.0. This project is dedicated to utilizing business analytics to process and interpret essential statistical data, facilitating strategic-level decision-making and location analysis for optimal site selection.

5.0. **Target Beneficiaries:**

Harvard Multiland Homes Real Estate Corporation, a registered developing company with the Securities and Exchange Commission with the primary purpose of "to own, use, improve, develop, subdivide, sell, exchange, lease and hold for investment or otherwise, real estate of all kinds including building houses, apartments, and other Structures". Our esteemed client from Harvard Multiland Homes Real Estate Corporation, **Caroline Chuateco**, holds the distinguished positions of Vice President, Treasurer, and Marketing Head, playing a pivotal role in the strategic and financial management of the organization.

UST:A022-02-F011 REV02 2/16/16

6.0. **Related Studies, Literature, Systems, and Technologies**

6.0.1. Related Studies & Literature

The Role of Population in Economic Growth

This study by Peterson (2017) involves the correlation between the following attributes: population and economic growth (socioeconomy). It utilizes historical data in highlighting population growth's influence in expanding the labor force and stimulating economic activity from a global perspective. It also emphasizes that population growth's impact on an economy's performance is influenced by other factors such as technological advancement, capital accumulation, and policy frameworks.

The study's analysis between population growth and economic growth results that both have a situational correlational relationship, since its correlation's positivity or negativity can also depend on specific contextual factors, like a society's strategic investments in education, healthcare, and infrastructure in enhancing a growing population's benefits.

Flooding Risk and Housing Values: An Economic Assessment of Environmental Hazard

This research by Daniel, V., Florax, R. J., & Rietveld, P. (2007) involves the correlation between the following attributes: flooding risk (environmental risk) and housing value. It examines the impact of climate change, fluctuating river cycles, and evolving water resource management practices on the spatial distribution of flood risk. It highlights how these factors have altered flood patterns and how they influence property values in flood-prone areas. The study reveals that properties located within the 100-year floodplain experience a reduction in value, with prices being 0.3% to 0.8% lower than those of properties outside of flood risk zones.

risk in housing purchases

Song Shi and Michael Naylor (2023) examined how an earthquake impacts household perceptions of seismic risk and real estate prices, using data from the 2010/2011 Canterbury earthquake in New Zealand. Their findings revealed that households initially underestimated seismic risks and overreacted after the quake, underscoring the importance of quake-related information in real estate pricing and risk management.

This study is closely related to the project as both involve data analysis for strategic site selection in real estate. While the study emphasizes the impact of seismic risk perceptions on decision-making, the project similarly uses data to analyze various factors influencing site selection. Both approaches aim to improve decision-making by providing insights that help assess risks and identify optimal locations for development.

6.0.2. Related Systems and Technologies

Planning and layout of tourism and leisure facilities based on POI big data and machine learning

This technological research study by Wu, S., et al. (2025) focuses on investigating the spatial arrangements of tourism infrastructure Beijing, China (focusing on six core district's of the city's main urvan area) with a proposal for a more scientific and efficient strategy for site selection using a data-driven approach. It utilizes Point of Interest (POI) data and population grid data in analyzing facility distribution.

The study's methodology is immersed around a supervised machine learning technique called the CART (Classification and Regression Trees) decision tree algorithm in handling mixed data types and high-dimensional datasets effectively, training it to evaluate a grid's suitability for leisure facility development. This was done by analyzing the presence urban service facilities such as hotels, public transportation, shopping areas, etc. as independent variables. The results of this study provides a scalable framework for strategic site selection using machine learning.

A framework for GIS-based site selection and technical potential evaluation of PV solar farm using Fuzzy-Boolean logic and AHP multi-criteria decision-making approach

This technological research study by Noorollahi, Y., et al. (2022) devices methodologies for site selection and potential evaluation of PV solar farm in Khuzestan, Iran that include a utilization of a multi-criteria decision-making approach with categorized data such as the following: climatic data (e.g., solar irradiance, sunshine hours), economic data (e.g., distance to roads, substations, urban/rural areas), orographic data (e.g., slope, elevation), and environmental data (e.g., land use). The study also further integrates Analytical Hierarchy Process (AHP), Fuzzy Logic, Boolean Logic, and GIS spatial analysis in determining the most suitable locations for solar farm development. A suitability map was also produced through a Weighted Linear Combination (WLC) overlay method with filtering features to exclude restricted zones such as fault lines, protected areas, and gas pipelines.

The study's results and findings state that its applied showcase a reliable approach to spatial decision-making. Furthermore, the flexibility of the model allows it to be adapted for other regions and renewable energy technologies.

The Impact of Machine Learning on Prescriptive Analytics for Optimized Business Decision-Making

This technological research study by Ara, A., et al. (2024) explores how integrating Machine Learning (ML) with Prescriptive Analytics enhances business decision-making by improving accuracy, efficiency, and forecasting. It highlights case studies across various industries, showcasing the competitive advantages

of adopting ML-driven tools. The paper also addresses challenges like data management, system integration, and skill gaps, offering best practices to overcome these obstacles. Ultimately, it emphasizes the importance of ongoing innovation in ML and prescriptive analytics for businesses to remain competitive in a data-driven world.

The study's focus on competitive advantages from Machine Learning mirrors the researchers aim to provide Harvard Multiland Homes a strategic edge by providing data-driven insights that streamline the site selection process. Moreover, the paper's discussion on overcoming challenges like data management and system integration is highly relevant as we work with large geospatial datasets and integrate them into ML models for predictive analysis, ensuring the project's success in a data-driven real estate market.

7.0. **Bibliography**

Ara, A., Maraj, A.,Rahman, A., Bari, H. (2024). The Impact of Machine Learning on Prescriptive Analytics for Optimized Business Decision-Making. International Journal of Management Information Systems and Data Science 1(1): 7-18.
https://www.researchgate.net/publication/379893157_THE_IMPACT_OF_MACHINE_LEARNING_ON_PRESCRIPTIVE_ANALYTICS_FOR_OPTIMIZED_BUSINESS_DECISION-MAKING

Daniel, V., Florax, R. J., & Rietveld, P. (2007). FLOODING RISK AND HOUSING VALUES: AN ECONOMIC ASSESSMENT OF ENVIRONMENTAL HAZARD. AgEcon Search.
https://doi.org/10.22004/ag.econ.7333

Noorollahi, Y., Ghenaatpisheh Senani, A., Fadaei, A., Simaee, M., & Moltames, R. (2022). A framework for GIS-based site selection and technical potential evaluation of PV solar farm using Fuzzy-Boolean logic and AHP multi-criteria decision-making approach. Renewable Energy, 186, 89–104.
https://doi.org/10.1016/j.renene.2021.12.124

Peterson, W. F. (2017). The Role of Population in Economic Growth. SAGE Open, 7(4).
https://doi.org/10.1177/2158244017736094

Shi, S., & Naylor, M. (2023). risk in housing purchases. Journal of Housing and the Built Environment.
https://doi.org/10.1007/s10901-023-10012-6

Wu, S., Wang, J., Jia, Y., Yang, J., & Li, J. (2025). Planning and layout of tourism and leisure facilities based on POI big data and machine learning. PLoS ONE, 20(3), e0298056–e0298056.
https://doi.org/10.1371/journal.pone.0298056

UNIVERSITY OF SANTO TOMAS
Institute of Information and Computing Sciences
Department of Information Systems
2nd Semester AY 2024-2025

Reviewed by/date	Comments/Suggestions/Remarks (please use additional sheets if necessary):
Asst. Prof. William A. Cortez	
Asst. Prof. Imelda E. Marollano	
Asst. Prof. Khrisnamonte M. Balmeo	

Panel Member's Recommendations: Please check which of the following actions you recommend:
( ) 1. Accept this proposal without any significant modifications suggested.
(X) 2. Accept this proposal but proponents must follow th prescribed modifications.

_____
_____

( ) 3. Do not accept this proposal and discontinue any further efforts on it.

THS1-Form2: Thesis Group Advisorship Agreement Form

UNIVERSITY OF SANTO TOMAS
Institute of Information and Computing Sciences
Department of Information Systems

2nd Semester AY 2024- 2025

THESIS GROUP ADVISORSHIP AGREEMENT

This is in acceptance of the Technical Adviser/Co-author rights for the Capstone Project Proposal **Machine Learning-Enhanced Site Selection Optimization for Strategic Location Analysis of Harvard Multiland Homes Real Estate Corporation Construction Projects**

As Technical Adviser, the tasks include the following:

- providing logistics analysis for the project
- overseeing the project's development phase
- providing input and critic on project documentation in terms of technical content and descriptions
- presenting the project alongside the student proponents in front of a panel

As a Co-author, the faculty member is a member of the proponents and must be involved in the project from conceptualization until the project is completed. He/she must also perform the tasks to be performed by the Technical Adviser.

The name of the Technical Adviser/Co-author will be included in all documentation as a Technical Adviser or Co-author depending on the chosen obligation. As such, they are obligated to make sure that student proponents defend their projects within the designated timeline.

They are also obligated to oversee the transition from THS1 to THS2 of the student proponents. Upon agreeing to be Technical Adviser in THS1, they too, also agree to be the group's Technical Adviser in THS2.

Student proponents will still need to do the project development within their own terms. A Technical Advisor/Co-author need not be part of the development aspect (i.e. writing program codes, implementing test case procedures, etc.); however, a technical advisor/co-author must be aware of the project's development process flow (i.e. what is the expected output of the given code, why that type of test case procedure was done, etc.).

Conforme:
Proponents: **Signature**

1. *Lazaro, Adrian* _____
2. *Chua, Kyle Steven* _____
3. *Concepcion, Margareth Samantha* _____
4. *Jacobo, Mikaella* _____

Joseph Richard G. Catubag, MBA
Technical Adviser/Co-author
Date: 05/28/2025

Asst. Prof. Jannette E. Sideño
Capstone Project Course Facilitator
Date:

THS1-Form3A: Endorsement for Project Proposal

UNIVERSITY OF SANTO TOMAS
Institute of Information and Computing Sciences
Department of Information Systems

2nd Semester AY 2024-2025

ENDORSEMENT FOR CAPSTONE PROJECT PROPOSAL

Project Title: Machine Learning-Enhanced Site Selection Optimization for Strategic Location Analysis of Harvard Multiland Homes Construction Projects

Proponents:
1. Chua, Kyle Steven T.
2. Concepcion, Margareth Samantha G.
3. Jacobo, Mikaella P.
4. Lazaro, Adrian DM.

Technical Adviser: ***Inst. Joseph Richard G. Catubag***

In partial fulfillment of the requirements for the degree of Bachelor of Science in Information Systems, the Capstone Project mentioned above, has been adequately prepared and submitted by the proponents and is hereby endorsed by the undersigned for Title oral examinations.

Inst. Joseph Richard G. Catubag
Technical Adviser/Co-author
Date: 5/8/2025

Asst. Prof. Janette E. Sideño
Capstone Project Facilitator
Date: 5-9-25

THS1-Form4A: Panel Members' Availability Confirmation

UNIVERSITY OF SANTO TOMAS
Institute of Information and Computing Sciences
Department of Information Systems

2nd Semester AY 2024-2025

PANEL MEMBERS' AVAILABILITY CONFIRMATION

Thesis Title: Machine Learning-Enhanced Site Selection Optimization for Strategic Location Analysis of Harvard Multiland Homes Construction Projects

Proponents:
1. ***Chua, Kyle Steven T.***
2. ***Concepcion, Margareth Samantha G.***
3. ***Jacobo, Mikaella P.***
4. ***Lazaro, Adrian DM.***

Scheduled Defense Date : ***May 16, 2025***

Time : ***11:30 am to 01:00 pm***

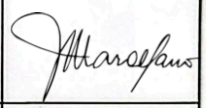
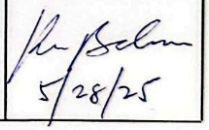
Room : ***1915***

Capstone Project Coordinator	Confirmation
Asst. Prof. Janette E. Sideño	5-8-2025

Panel Members	Confirmation
1. **Asst. Prof. William A. Cortez**	
2. Asst. Prof. Imelda E. Marollano	
3. Asst. Prof. Khrisnamonte M. Balmeo	

UNIVERSITY OF SANTO TOMAS
COLLEGE OF INFORMATION AND COMPUTING SCIENCES
Department of Information Systems

Consolidated Comments of Panel Members

Panel Member's Name (As indicated in the Schedule of Defense)	Suggestion during the Defense	Your Revision/Action Made	(Indicate the chapter and page)	Status (complied, not complied)	Signature of Panel Member with date
Asst. Prof. W. Cortez	State in your limitations how you will go about the missing data values beyond the years of available data you have gathered. Indicate that you will evaluate the effectiveness of the methods you will be using.	Proponents added a limitation indicating extrapolation to be implemented for missing necessary data. Proponents added a statement on evaluating and validating models performances before choosing what to use before project deployment.	Chapter 1: 1.4.2. Limitations Pg. 6 Chapter 3: 3.4.0. Business Analytics Model and Testing Pg. 69	complied	
Asst. Prof. I. Marollano	Indicate that you will evaluate the effectiveness of the methods you will be using.	Proponents added a statement on evaluating and validating models performances before choosing what to use before project deployment.	Chapter 3: 3.4.0. Business Analytics Model and Testing Pg. 69	compiled	
Asst. Prof.K. Balmeo	State in your limitations how you will go about the missing data values beyond the years of available data you have gathered.	Proponents added a limitation indicating extrapolation to be implemented for missing necessary data.	Chapter 1: 1.4.2. Limitations Pg. 6	complied	 5/28/25

UNIVERSITY OF SANTO TOMAS
COLLEGE OF INFORMATION AND COMPUTING SCIENCES
Department of Information Systems

Noted by:

Asst. Prof. Janette E. Sideño
Course Facilitator
Date: 5-28-25

Accepted by:

Asst. Prof. William A. Cortez
Panel Member 1

Asst. Prof. Imelda E. Marollano
Panel Member 2

Asst. Prof. Khrisnamonte M. Balmeo
Panel Member 3

2nd Floor, Blessed Pier Giorgio Frassati Building, UST España Boulevard,
Sampaloc, Manila Philippines 1015
Telephone No. 3406-1611 local 8518 - ...
http://www.ust.edu.ph Facebook : @USTIICS2014Official Twitter : @USTIICS2014