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 2

2.0. Review of Related Literature, Studies, Systems and Technologies

This chapter will provide an overview of related literature, studies, systems, and technologies that are relevant to this project's Application of Machine Learning to Strategic Location Analysis of the Construction Projects of Harvard Multiland Homes Real Estate Corporation Leading to Optimal Site Selection.

2.1.0. Review of Related Literature and Studies

2.1.1. Related Literature on the Correlation of Attributes

This first part of this section focuses on different literature and studies that support high correlations between some of the main attributes (e.g., geographic, socioeconomic, population, project costs, housing value, human resource, and environmental factors) stated in the previous chapter, which are to be used for analyzing the optimality of a location's site in helping executive decision-making for future projects of Harvard Multiland Homes Real Estate Corporation.

The Impact of Human Capital on Economic Growth

This paper by Pelinescu (2015) discusses the exploration of the role of human capital in economic growth across EU countries. The paper's findings confirm that there is a statistically significant and positive relationship between GDP per capita and specific human capital indicators such as the number of employees with secondary education. It emphasizes that education, skills, and innovation are core elements of human capital and are essential for achieving sustainable economic development. It is

also showed through using a panel regression model that countries with stronger human capital foundations tend to exhibit greater potential for economic growth.

The study confirmed that overall, through innovation and education-related employment, human capital plays a key role in influencing GDP per capita, with further support from the R-squared value of the model which amounts to 0.99687, which indicates a very strong explanatory power.

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.

The Human Resources as an Important Factor of Regional Development

This study by Jašková and Havierniková (2020) involves the correlation between the following attributes: geographic location and human resources. It analyzes different indicators under the following: Demographic Development, Economic Level and Employment, Social and Living Standards, and Health, Education, and Crime. These are utilized for analysis in the importance of human resources as a factor of regional development in Slovakia, specifically using advanced statistical techniques and tools such as Principal Component Analysis (PCA), cluster analysis, and composite indicators in ranking each region's human resources development potential.

Results show a clear correlation between geographic location and the availability and quality of human resources, with factorial distinctions of the analysis on population density, education levels, higher concentration of universities, employment rates, healthcare facilities and existence of businesses between regions with urban centers, and geographic positions of capital, eastern regions, and others in the discovery of overall development potentials across its country of focus.

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.

Adapting to Flood Risk: Evidence from a Panel of Global Cities

This research by Gandhi, S., et al. (2022) involves the correlation between the following attributes: flooding risk (environmental risk), population, and income. It indicates that cities experiencing frequent floods often face reduced population growth and economic activity. Factors such as income levels and the presence of protective infrastructure, like dams, play significant roles in mitigating these adverse effects. For instance, wealthier cities tend to suffer fewer deaths from flood events, and cities protected by dams experience faster population growth.

The Effect of Flood Risk on Residential Land Prices

This paper examines residential land prices in regions with varying flood exposure levels. It shows that properties within flood-prone zones are consistently undervalued compared to their non-risk counterparts, with estimated price discounts ranging between 4% and 12%. The authors attribute this trend to buyer perceptions, insurance obligations, and regulatory constraints. For construction firms, this data is crucial for understanding how location affects project profitability, and it supports strategic planning around whether to invest in mitigation or avoid risky zones entirely. It also emphasizes the need for reliable risk mapping and socio-economic data when making site selection decisions.

Analyzing the Impact of Location Factors on Building Construction Cost in Sri Lanka

This research by Parameswaran, T., Jayawickrama, T., & Melagoda, D. (2019) involves the correlation between the following attributes: construction costs (project costs), and geographic location. It focuses on the increasing importance of location factors in determining building construction costs globally. Over the past few decades, the influence of location on construction costs has become more significant, prompting the need for accurate estimation methods that can reflect the impact of geographic variables. The goal is to provide clients with reliable cost estimates, reducing the risk of unexpected fluctuations in expenses and ensuring that designs remain within the client's budget.

The location of a project is a critical factor in its success or failure, particularly for commercial ventures that depend on specific location conditions (Rymarzak and Seiminska, 2012). The cost of construction materials, labor, and other resources can differ not only over time but also by geographic area (Grogan and Ichniowski, 2009). As such, understanding how location influences construction expenses is essential when making cost projections (Akanni et al., 2015).

Getting Environmental Information from Construction Cost Databases: Applications in Brazilian Courses and Environmental Assessment

This paper by Borja, L., César, S., Cunha, R., & Kiperstok, A. (2019) involves the correlation between the following attributes: environmental

information (environmental data), and construction cost (project cost). It explores how incorporating environmental indicators into construction cost databases can enhance the communication of sustainable practices in the industry. It reviews the use of cost databases in environmental studies, identifies the most commonly cited databases in Brazilian civil engineering courses, and evaluates these databases using a matrix that compares cost data with environmental information.

The study found that databases like CYPE, TCPO, and ORSE performed better in integrating environmental factors than SINAPI, BDCCM, and BCCA. Despite their overall low control over environmental aspects such as water and energy consumption, machine circulation, and pollution generation, the research indicates that incorporating environmental indicators into these tools can significantly improve their ability to promote sustainable practices to a broad user base.

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.

Property Prices and Disaster (In)Justice: Fifty Years of Demographic Change in Hawai'i County's Lava Hazard Zones

Mauro (2022) analyzed the relationship between environmental risk and social vulnerability in Hawai'i County using U.S. census data from 1970 to 2020, overlaid with lava hazard zone maps. The study found that higher-risk zones had consistently higher proportions of low-income and unemployed populations, and lower median property values. Pearson correlation analyses from 1990 to 2020 showed a negative relationship between property values and multiple social vulnerability indicators.

These findings highlight the value of integrating environmental risk, socioeconomic data, and land value into spatial analysis. This supports the application of interactive heatmap dashboards to visualize risk and vulnerability factors for more informed and equitable location-based decision-making.

2.1.2. Related Literature on the Minimum Credibility of Historical Data

This second part of this section focuses on articles and studies that support and suggest the credibility of historical data in maximizing their analytical use for

different technological systems of companies and businesses, especially in forecasting.

How much data is required for forecasting?

This article by Anamind (2020) says that, even though a longer data history helps the system better differentiate between external factors like promotions, natural disasters, or pandemics, which can cause anomalies that may confuse the system and lead to inaccurate forecasts, a forecasting software can typically identify seasonality, trend, and level components automatically with just two years of data (minimum).

Historical data for accurate sales forecasting

This article by verteego (2020) says that to create an accurate forecast, at least two years of data history are sufficient and essential, because contact center arrival patterns are seasonal, with variations occurring monthly, yearly, and weekly. Two years of data help separate trends and seasonality, while accounting for external factors like marketing campaigns or special events that may cause demand spikes in one year but not in another. The third year helps distinguish between regular patterns and anomalies, ensuring a more reliable forecast. Thus, three years of data are crucial for accurate and consistent forecasting.

2.2.0. Review of Related Systems and Technologies

This section of the chapter will provide an overview of related existing systems and technologies that are relevant to this project's Application of Machine Learning to Strategic

Location Analysis of the Construction Projects of Harvard Multiland Homes Real Estate Corporation Leading to Optimal Site Selection.

2.2.1. Related Systems and Technologies

This first part of this section focuses on different technological and system-focused studies that have utilized different analytical models involving site selection and other similar concepts, which the project proponents may use as reference in the future methodologies of analyzing the optimality of a location's site to help with executive decision-making for future projects of Harvard Multiland Homes Real Estate Corporation.

Research on Site Selection Planning of Urban Parks Based on POI and Machine Learning—Taking Guangzhou City as an Example

This technological research study by Tang x., et al. (2024) explores the beneficial enhancements that data-driven approaches can provide for urban park planning in Guangzhou, China. Using big data from Points of Interest (POIs), where the researchers have collected 663,749 POI entries across 15 categories, and user activity footprints, where the researchers have collected 171,958 of, the researchers processed and visualized these data using ArcGIS 10.2 to assess park distribution, density, and service pressure in achieving their goal to identify optimal locations for future urban parks. The ID3 decision tree machine learning algorithm was the main analytical methodology used to train a predictive model classifying city grid cells as suitable or unsuitable for park development based on surrounding POIs.

The results of the study found geographical imbalances in existing parks in Guangzhou, China, highlighting clustering in central districts and a scarcity in suburban and peripheral areas. It overall showcases the advantageous effects of integrating machine learning with urban planning tools for a more data-informed decisions in both planning and possible developments of infrastructures in specific locations.

Site Selection of Digital Signage in Beijing: A Combination of Machine Learning and an Empirical Approach

This technological research study by Wang, Y., et al. (2020) delves on the accuracy and scientific foundation improvements of digital signage placements within Beijing, China's Sixth Ring Road. It presents a hybrid methodology between traditional empirical models and advanced machine learning techniques, integrating them to enhance site selection decision-making. Data involved in this study that are also processed using ArcGIS includes population census, housing prices, social media check-ins (from Sina Weibo), transportation centrality, and POI (Point of Interest) data. The study's methods combines the utilization of an improved Huff model for spatial accessibility analysis with machine learning algorithms like Backpropagation Neural Network (BPNN), Random Forest (RF), and Support Vector Regression (SVR) to calculate deployment potential. K-means clustering and Principal Component Analysis (PCA) were also used for dimensionality reduction and categorization of grid data into high, medium, and low deployment potential zones.

The results demonstrated that combining empirical and machine learning approaches produced more accurate and interpretable outcomes

than traditional models alone, verified through ROC curves and cross-validation. The results and findings shows the applicability of an integrated model not just to digital signage site selection, but also broadly to strategic site selection plannings in real estate and construction projects.

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.

The suitability assessment for land territorial spatial planning based on ANN-CA model and the Internet of Things

This technological research study by Nie, Z. (2024) is a presentation of a comprehensive framework for urban development suitability evaluation. It utilized a hybrid of the following technologies: Artificial Neural Networks (ANN), Cellular Automata (CA), and an integration of Internet of Things (IoT) data, where data such as land use, elevation, slope, water resources, and proximity to infrastructure are integrated into a multi-indicator suitability assessment system.

The study's methodology takes on a detailed ANN-CA model framework, which includes five stages: data preprocessing, Markov prediction, ANN training, model correction, and prediction. By using predictive modeling together with real-time IoT monitoring and smart urban planning, this research study provides a scalable solution for intelligent land use and construction site selection.

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.

Optimal site selection for solar photovoltaic (PV) power plants using GIS and AHP: A case study of Malatya Province, Turkey

This technological research study by Colak, H. E., Memisoglu, T., & Gercek, Y. (2019) explores the use of Geographic Information Systems (GIS) integrated with the Analytic Hierarchy Process (AHP) and how these can help guide the site selection process for solar power plants effectively, specifically at the Malayta Province of Turkey. It identified spatial and environmental criteria for analysis for its methodology which included the following: solar radiation, slope, land cover, proximity to transmission lines, transformer centers, roads, fault lines, and residential zones. GIS spatial analysis was then used to make a cost surface map for further descriptive visualization of optimal installation sites for solar PV systems.

The AHP model that was utilized involved pairwise comparisons of criteria to calculate consistency ratios and determine weight values. The study's results gave a conclusion that supports the reliability and

practicality of the integration of AHP-GIS for optimizing renewable energy infrastructure planning.

Site selection of wind farms using GIS and multi-criteria decision making method in Wafangdian, China

This technological research study by Xu, Y., et al. (2020) focuses on optimizing the site selection for onshore wind farms in Wafangdian, China. Its methodology integrates the following metods: Geographic Information System (GIS) and advanced Multi-Criteria Decision-Making (MCDM); and it proposes a methodology of combining Interval Analytic Hierarchy Process (IAHP) and Stochastic VIKOR in order to handle uncertainties in both criteria weighting and suitability scoring. The criteria for evaluation included data such as wind speed, slope, distance to electricity grid, distance to main roads, protected bird areas, and distance to urban areas.

The study's conclusion of its methods' results states that the IAHP–stochastic VIKOR–GIS integrated approach provides a scientifically viable framework for site selection planning for other renewable energy infrastructures.

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.

Towards Big Data-driven Construction Industry

This technological research article by Li, F., et al. (2023) examines the transformation of the construction industry driven by Industry 4.0 technologies such as the Internet of Things (IoT), cloud computing, and robotics. It emphasizes the importance of utilizing big data from various sources to enhance the construction process across its entire lifecycle. By leveraging big data, construction efficiency can be improved, material waste reduced, and planning and decision-making processes enhanced. Additionally, big data applications can make construction sites safer. The article reviews the benefits of big data in construction, assesses the current

state of the industry, and discusses unresolved challenges. Finally, it offers insights into the future potential of big data in construction.

Big data can be applied in site selection for construction projects. By utilizing various data sources, including geographic information systems (GIS), satellite imagery, and demographic data, big data can help identify optimal locations for new construction projects. The researchers project leverages geospatial data as part of a big data approach to identify optimal construction sites.

Application of Machine Learning for Optimal Wind Farm Location

This technological research study by Shahzad, U. (2021) focuses on determining the optimal location for wind farms within a power transmission network to improve sustainability and reduce environmental impacts. The study aims to minimize the system's operating cost while considering potential outages, specifically (N-1) line and (N-1) bus outages, which refer to the failure of a single transmission line or bus in the network. A machine learning algorithm is employed to predict the best locations for wind farms in a computationally efficient way. The research provides a unique solution for optimal wind farm placement, along with a priority order list, which can help in integrating multiple wind farms into the power transmission system efficiently.

The approach of using machine learning algorithms to predict optimal locations in the wind farm placement study is very similar to the researchers use of ML for construction site selection. Just as the wind farm

study leverages ML to determine the best locations while considering system constraints and minimizing costs, our project uses machine learning to identify the most strategic construction sites based on various geospatial and environmental factors.

Hospital Site Suitability Assessment Using Three Machine Learning Approaches: Evidence from the Gaza Strip in Palestine

This technological research study by Almansi, K. Y., et al. (2021) addresses the challenges Palestinian healthcare institutions face in providing effective services, particularly in crisis situations. In the Gaza Strip, issues such as inadequate spatial distribution, poor emergency response times, and limited infrastructure hinder healthcare accessibility. The research focuses on identifying optimal hospital locations in Gaza by creating suitability maps using environmental, topographic, and geodemographic factors. Machine learning techniques, such as correlation-based feature selection (CFS), support vector machine (SVM), multilayer perceptron (MLP), and linear regression (LR), are employed to identify key parameters that influence hospital site suitability. The models are validated using cross-validation and receiver operating characteristic (ROC) curves, showing high accuracy rates for the MLP, SVM, and LR models

The results highlight that machine learning approaches are effective and reliable for determining suitable hospital locations, offering a promising method for improving healthcare delivery planning and implementation in the region. The researchers project employs similar ML models to determine the best locations for construction projects.

Furthermore, the use of suitability maps in the healthcare study aligns with our approach to visualizing and analyzing data in spatial heatmaps, helping decision-makers easily identify areas that meet specific criteria for development.

Designing Interactive Web-GIS Dashboards for Data-Driven Emergency Management in Smart Cities

Han, T. and Choi, G. (2024) developed an interactive Web-GIS dashboard designed for emergency response in smart cities, integrating real-time data such as CCTV footage, IoT sensors, and public datasets into a user-centered, map-based interface. Their system supports layered visualizations (e.g., weather, infrastructure, emergency locations), enabling responders to monitor and coordinate actions efficiently. Built through Agile UX practices and evaluated with real users, the dashboard emphasizes intuitive navigation, dynamic data interaction, and rapid decision-making.

This study serves as a practical reference for the capstone project by illustrating effective integration of diverse datasets into a Web-GIS interface. Emulating Han and Choi's layered data visualization and user-centric design approach can enhance the project's interactive heatmap dashboard, facilitating informed decision-making for Harvard Multiland Homes executives during site evaluations. **(Tool: GIS (Python Packages or GIS Tools))**

Data Visualization in Big Data Analysis: Applications and Future Trends

Ouyang (2024) conducted a comprehensive technological research study on the role of data visualization in big data analysis, emphasizing its significance in enhancing analytical efficiency and insight extraction. The research systematically reviews the theoretical foundations and technological evolution of data visualization, addressing challenges such as massive data processing, real-time visualization demands, and multi-dimensional data representation. Through extensive literature analysis, the study explores innovative applications across various domains, including business intelligence, scientific research, and public decision-making. Key future trends identified involve interactive, real-time, and immersive visualization technologies, alongside the integration of artificial intelligence to enhance visualization capabilities. The paper also highlights the importance of fostering interdisciplinary collaboration and data democratization through improved visualization literacy, standardized frameworks, and open-source tool sharing.

Ouyang's exploration of advanced visualization techniques and AI integration informs the capstone project's approach to developing a GIS-based heatmap dashboard. By incorporating interactive and real-time visualization technologies, along with machine learning algorithms for predictive analytics, the project aims to provide Harvard Multiland Homes executives with a sophisticated tool for evaluating potential construction sites. **(Tool: Python Data Visualization Packages or Power BI)**

A Comparative Study for Landslide Susceptibility Assessment Using Hybrid Machine Learning Models

Deng et al. (2022) conducted a technological research study to evaluate the effectiveness of hybrid machine learning models—Random Forest (RF), Forest by Penalizing Attributes (FPA), and Rotation Forest (ROF)—in assessing landslide susceptibility within Yaozhou District, Tongchuan City, China. The research utilized twelve factors, including elevation, slope angle, and lithology, to generate landslide susceptibility maps based on two evaluation units: slope units divided by curvature and 30-meter resolution grid units. The findings indicated that the RS-RF model applied to slope units exhibited superior spatial prediction capabilities. Additionally, the study demonstrated that the curvature-based division of slope units is particularly suitable for Loess tableland regions, offering valuable insights for selecting evaluation units in landslide susceptibility assessments.

The methodologies employed by Deng et al. offer valuable insights for the capstone project, particularly in integrating environmental factors into machine learning models for site selection analysis. Applying similar techniques can enhance the accuracy and reliability of site evaluations, providing Harvard Multiland Homes with a robust decision-support tool for identifying optimal construction sites.

Autonomous GIS: the next-generation AI-powered GIS

Li and Ning (2023) introduced the concept of Autonomous Geographic Information Systems (GIS), integrating Large Language

Models (LLMs) like GPT-4 to automate spatial data collection, analysis, and visualization. They proposed five key capabilities for such systems: self-generating, self-organizing, self-verifying, self-executing, and self-growing. Their prototype, LLM-Geo, demonstrated these capabilities by autonomously processing spatial tasks and producing accurate outputs, including maps and graphs, without human intervention. This advancement aims to make spatial analysis more efficient and accessible.

Incorporating LLMs into GIS, as demonstrated by Li and Ning, aligns with the capstone project's goal of automating site selection processes. Adopting autonomous GIS principles can lead to the development of a system capable of self-directed data analysis and visualization, thereby enhancing decision-making efficiency for Harvard Multiland Homes in identifying optimal construction sites.

Data Science for Geographic Information Systems

Oliveira et al. (2024) explore the integration of data science techniques into Geographic Information Systems (GIS), highlighting the evolution of GIS into comprehensive spatial analysis platforms. The study emphasizes the adoption of machine learning and big data methodologies, enabling GIS to process complex and voluminous datasets beyond traditional capabilities. A case study in disaster management is presented, utilizing aerial data from Tróia, Portugal, to demonstrate the extraction of actionable insights from raw data. The authors conclude by discussing future research directions for the continued integration of data science and GIS.

The integration of data science techniques into GIS, as discussed by Oliveira et al., offers valuable insights for enhancing the analytical capabilities of the capstone project. By incorporating machine learning algorithms and big data methodologies, the project can effectively process and analyze complex datasets related to potential construction sites. This approach aims to provide Harvard Multiland Homes with a sophisticated tool for data-driven decision-making in site selection.

idwMapper: An interactive and data-driven web mapping framework for visualizing and sensing high-dimensional geospatial (big) data

Sarigai et al. (2024) introduced idwMapper, an interactive web mapping framework designed to visualize and analyze high-dimensional geospatial big data. The framework leverages coordinated-view visualization and cross-filtering techniques to enable users to explore complex datasets through multiple linked visual representations. To demonstrate its versatility, the authors implemented three web applications across different domains: iLit4GEE-AI for literature analysis, iWURanking for university rankings, and iTRELISmap for scholar mapping. These applications underscore idwMapper's capacity to facilitate intuitive and scalable geospatial data analysis.

The idwMapper framework offers innovative approaches that can be adapted to the capstone project's development of an interactive heatmap dashboard. Integrating coordinated-view visualization and cross-

filtering techniques can enhance the dashboard's functionality, enabling users to effectively explore and analyze complex geospatial datasets. Implementing such methodologies can improve the decision-making process for site selection by providing comprehensive and user-friendly visualizations of potential construction locations.

Design and Implementation of an Interactive Visualization Dashboard for Monitoring the Flood Vulnerability and Mapping

This technological research study by Azizah and Wijayanto (2023) develops a web-based interactive visualization dashboard to estimate flood vulnerability using satellite imagery. The dashboard integrates data extracted from satellite imagery in CSV format, which is then stored in GeoJSON format and visualized using a 2D approach with the qgis2web Python plugin and JavaScript leaflet. The interactive dashboard includes features like zoom, search, and a "show me where I am" function, allowing users to monitor flood-prone areas and support decision-making for flood mitigation.

This approach aligns with the capstone project, which also focuses on developing an interactive dashboard to support decision-making in site selection for construction projects. Both projects use data-driven visualizations to analyze and present complex factors, with the capstone project employing an interactive heatmap dashboard to assess site

suitability for construction. This dashboard will similarly aid in strategic decision-making, much like the flood vulnerability dashboard helps in flood mitigation planning.

Optimizing Residential Construction Site Selection in Mountainous Regions Using Geospatial Data and explainable AI

This technological research study by Alqahtani et al. (2024) applies Fuzzy Analytic Hierarchy Process (Fuzzy-AHP) and Deep Neural Networks (DNN) with eXplainable Artificial Intelligence (XAI) to identify optimal construction sites in Abha, Saudi Arabia, based on sixteen environmental and socio-economic parameters. Its data-driven approach aligns with the capstone project entitled *“Machine Learning-Enhanced Site Selection Optimization for Strategic Location Analysis of Harvard Multiland Homes Real Estate Corporation Construction Projects”*, which similarly uses machine learning to support informed and strategic site selection.

The study's use of AI to analyze factors such as slope, land use, and flood risk, and its output of suitability maps and “no-build zones,” directly supports the capstone project’s goal of developing an interactive heatmap dashboard. This dashboard will visualize the influence of various factors on site suitability, aiding Harvard Multiland Homes Real Estate Corporation in making data-informed construction decisions while balancing safety, environmental, and development needs.

A Machine Learning Approach to Predict Site Selection from the Perspective of Vitality Improvement

This technological research study by Zhao et al. (2024) on the selection of construction sites for Cultural and Museum Public Buildings (CMPBs) integrates Artificial Neural Networks (ANNs) and Genetic Algorithms (GAs) to improve site selection and planning. Using Shanghai as a case study, the researchers analyzed 344 sites with 39 infrastructure data sets and vitality values, predicting vitality improvements and assessing model applicability across different scales. This approach enhances early planning, design, and operational management of CMPBs.

The study's predictive modeling and data analysis align with the capstone project. Both projects use machine learning to analyze site suitability factors, with the capstone project focusing on developing an interactive heatmap dashboard to support data-informed construction decisions, similar to how the model aids CMPB planning and development.

Machine Learning Algorithms for Construction Projects Delay Risk Prediction

This technological research study by Gondia et al. (2020) focuses on leveraging machine learning models to predict project delays in the construction sector, which is known for its complexity and interdependent risk sources. The researchers identified key delay risk factors and compiled a multivariate data set of previous project performance. They applied decision tree and naïve Bayesian classification algorithms, with the latter providing better predictive performance. The study's goal was to enable evidence-based decision-making for proactive project risk management,

using machine learning to analyze the dynamic interdependencies of delay risk sources.

This approach aligns with the capstone project. Both projects utilize machine learning to analyze complex data and support decision-making. The capstone project focuses on developing an interactive heatmap dashboard to visualize the influence of various site suitability factors, similarly aiding proactive decision-making, much like how the machine learning models in the study assist in predicting and managing project delays.

A Comprehensive Business Location Choice Model Leveraging Machine Learning in Systematic Choice Set

This technological research study by Mahmud and Habib (2024) develops a two-stage location choice framework for business establishments within Halifax Regional Municipality, utilizing machine learning to generate a systematic choice set. By combining unsupervised machine learning techniques with the mixed multinomial logit model, the study enhances the precision and robustness of location choice models. The findings reveal patterns such as wholesalers prioritizing highway proximity and transportation businesses valuing accessibility, offering valuable insights for commercial vehicle modeling and sustainable urban development.

Both the study and the capstone project use machine learning to analyze location suitability, with the focus on visualizing various factors influencing site selection. The capstone project will develop an interactive

heatmap dashboard to support decision-making, similar to how machine learning helps uncover latent patterns in business location choices, enabling data-driven insights.

Sustainability Model to Select Optimal Site Location for Temporary Housing Units: Combining GIS and the MIVES–Knapsack Model

This technological research study by Hosseini et al. (2022) presents a novel multi-criteria decision-making approach for selecting optimal site locations for temporary housing (TH) in post-disaster recovery programs. By combining a knapsack algorithm with the integrated value model for sustainability assessment (MIVES), the model objectively quantifies sustainability indicators (economic, environmental, and social aspects). It integrates data from a geographic information system (GIS) to evaluate multiple alternatives, aiming to maximize stakeholder satisfaction and enhance the sustainability of site selections.

This approach aligns with the capstone project, which also focuses on strategic site selection using machine learning. Both projects analyze multiple factors that influence location suitability, with the capstone project using an interactive heatmap dashboard to visualize the influence of various factors. This dashboard will support data-driven decisions, similar to how the model in the study helps optimize temporary housing site selection by considering spatial and sustainability-related aims.

Risk Identification, Assessments, and Prediction for Mega Construction Projects: A Risk Prediction Paradigm Based on Cross Analytical-Machine Learning Model

This technological research study by Chattapadhyay et al. (2021) develops a risk prediction system for construction megaprojects using a cross-analytical machine learning model. It identifies 63 risk factors related to cost, time, quality, and scope, collected from industry experts. The system employs K-means clustering, enhanced by genetic algorithms (GA-K-means), to identify high-risk factors and their associated sub-components. This model helps stakeholders pinpoint critical risks, enabling proactive management strategies that can enhance overall project performance.

This approach aligns with the capstone project, which also applies machine learning to assess location suitability for construction projects. Both projects focus on analyzing multiple factors influencing decision-making, with the capstone project utilizing an interactive heatmap dashboard to visualize site suitability factors. This dashboard will support decision-making by offering a data-driven approach, similar to how the risk prediction model helps manage potential risks in construction projects.

2.2.2. Related Technological Tools to be Used

This second part of this section focuses on the different technological tools to be used for aiding and supporting the project proponents in accomplishing this project's target technological and document-focused goals.

The impact of ChatGPT on higher education.

This technological research study by Dempere, J., et al. (2023) explores the rapid advancement of AI and the implications of its capabilities in the field of academics by utilizing the following databases to search for literature on AI chatbots' impact on Higher Education Institutions (HEIs): PubMed, IEEE Xplore, and Google Scholar.

Results of the study found diversity in different perspectives on ChatGPT's potential in education, particularly for research support, automated grading, and enhanced human-computer interaction. Its capacious and highly accessible functionalities are able to draw significantly useful insights for effectively creating necessary content in the field of research, which can remarkably make a difference for researchers on the condition of appropriately following and conforming ethical steps in writing.

React

Deshpande, C. (2024) defines React as a JavaScript library for building user interfaces. Although it's not a programming language, it has become one of the most popular tools in web development since its release in May 2013, widely adopted for creating dynamic and responsive frontend applications. This framework will be used by the researchers to build the dashboard and web application for the study, providing a responsive and interactive user interface. React will ensure that the web application is both highly interactive and responsive, enabling efficient exploration of different

construction sites based on various factors, ultimately supporting data-driven decision-making for site selection at Harvard Multiland Homes.

Geospatial and Heatmap Data Visualization Using Python

This technological research article by Ault, S. V., et al. (2025) is about the use of Python for geospatial data visualization and analysis. It explains how geospatial data, which describes geographic locations and attributes, can be manipulated and visualized using Python libraries. The article specifically focuses on Pandas, a library for data manipulation, and Geopandas, which extends Pandas to handle geospatial data by providing a GeoDataFrame for spatial operations. It also discusses the concept of spatial heatmaps, a visualization technique used to represent the density or intensity of data points within a geographical area, helping to highlight areas of high and low concentration.

Geospatial data analysis plays a crucial role for construction companies seeking optimal site selection. Python libraries like Pandas and Geopandas are key tools in this process. Furthermore, heatmaps offer a powerful way to visualize data density, helping identify areas with high or low concentrations of important factors like infrastructure and land use. By combining these geospatial analysis and visualization techniques, construction companies can make more informed, data-driven decisions for selecting ideal sites for development.

2.3. Synthesis

Several studies has found strong relationships between geographic, socioeconomic, environmental, and human resource variables in influencing construction-related decisions, with further recommendations of including other factors for further association of appropriate attributes with each other. It was also explored through other studies that the reliability of using two to three years of historical data can be used for accurate forecasting, which supports predictive analytics in system development. A significant portion of related technologies centered on integrating machine learning models such as Random Forest, Decision Trees, and Neural Networks into site suitability evaluations across various industries. Lastly, the technological tools and frameworks, like React, Python geospatial libraries, and data visualization dashboards describe practical implementations relevant to building interactive decision-support systems.

These relevant discussions further contribute in supporting the proposed study entitled “Machine Learning-Enhanced Site Selection Optimization for Strategic Location Analysis of Harvard Multiland Homes Real Estate Corporation Construction Projects”. The correlation-based literature states the selection of the most impactful variables for location analysis, which are available gathered data, and will be gathered data, by this project’s proponents. Studies on machine learning applications provide methodological references that can be approached and adapted to construction site selection, while GIS-based dashboards offer inventiveness for presenting outputs in an accessible and executive-friendly format. The emphasis on historical data further validates the use of previous project records and external datasets in training predictive models. Together, these works strengthen the study's foundational logic by reinforcing the relevance of data-driven, AI-enhanced approaches in real estate and construction planning.

All stated reviewed studies and technologies can significantly contribute in shaping the future of this study's development process. The methodologies adapted from related studies, such as supervised learning, suitability mapping, and spatial heatmap visualization, inform the structure and functionality of the system to be developed. Moreover, real-world use cases from various industries using different factors such as renewable energy, healthcare, and tourism validated the practicality and scalability of machine learning models in spatial decision-making.

To conclude, the reviewed literature and technologies gives assistance to the feasibility and necessity of developing a machine learning-powered decision-support system for optimal site selection. By aligning the project's components with proven practices and insights from multiple fields, the study ensures the reliability of the methodologies that can be used and the practical applicability.

Bibliography

Almansi, K. Y., Shariff, A. R. M., Abdullah, A. F., & Ismail, S. N. S. (2021). Hospital Site Suitability Assessment Using Three Machine Learning Approaches: Evidence from the Gaza Strip in Palestine. Applied Sciences, 11(22), 11054. https://doi.org/10.3390/app112211054

Alqahtani, D., Mallick, J., Alqahtani, A. M., & Talukdar, S. (2024). Optimizing Residential Construction Site Selection in Mountainous Regions Using Geospatial Data and eXplainable AI. Sustainability, 16(10), 4235. https://doi.org/10.3390/su16104235

Anamid. (2020). How much data is required for forecasting? – Anamind Advanced Analytical Planning. https://www.anamind.com/anamind-blog/how-much-data-is-required-for-forecasting/

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

Ault, S. V., Liao, S. N., & Musolino, L. (2025). 9.4 Geospatial and heat map data Visualization using Python - Principles of Data Science | OpenStax. https://openstax.org/books/principles-data-science/pages/9-4-geospatial-and-heat map-data-visualization-using-python

Azizah, W. R., & Wijayanto, A. W. (2023). Design and Implementation of an Interactive Visualization Dashboard for Monitoring the Flood Vulnerability and Mapping. Proceedings of the International Conference on Data Science and Official Statistics, 2023(1), 218–232. https://doi.org/10.34123/icdsos.v2023i1.362

Borja, L., César, S., Cunha, R., & Kiperstok, A. (2019). Getting Environmental Information from Construction Cost Databases: Applications in Brazilian Courses and Environmental Assessment. Sustainability, 11(1), 187. https://doi.org/10.3390/su11010187

Chattapadhyay, D. B., Putta, J., & Rao P, R. M. (2021). Risk Identification, Assessments, and Prediction for Mega Construction Projects: A Risk Prediction Paradigm Based on Cross Analytical-Machine Learning Model. Buildings, 11(4), 172. MDPI. https://doi.org/10.3390/buildings11040172

Colak, H. E., Memisoglu, T., & Gercek, Y. (2019). Optimal site selection for solar photovoltaic (PV) power plants using GIS and AHP: A case study of Malatya Province, Turkey. Renewable Energy, 149, 565–576. https://doi.org/10.1016/j.renene.2019.12.078

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

Dempere, J., Kennedy Modugu, Allam Hesham, & Ramasamy, L. K. (2023). The impact of ChatGPT on higher education. Frontiers in Education, 8.

https://doi.org/10.3389/feduc.2023.1206936

Deng, N., Li, Y., Ma, J., Shahabi, H., Hashim, M., De Oliveira, G., & Chaeikar, S. S. (2022). A comparative study for landslide susceptibility assessment using machine learning algorithms based on grid unit and slope unit. Frontiers in Environmental Science, 10.

https://doi.org/10.3389/fenvs.2022.1009433

Deshpande, C. (2024). The best guide to know what is react.

https://www.simplilearn.com/tutorials/reactjs-tutorial/what-is-reactjs

Gandhi, S., Kahn, M., Kochhar, R., Lall, S., & Tandel, V. (2022). Adapting to Flood Risk: Evidence from a Panel of Global Cities.

https://doi.org/10.3386/w30137

Gondia, A., Siam, A., El-Dakhakhni, W., & Nassar, A. H. (2020). Machine Learning Algorithms for Construction Projects Delay Risk Prediction. Journal of Construction Engineering and Management, 146(1), 04019085.

https://doi.org/10.1061/(asce)co.1943-7862.0001736

Han, T., & Choi, G. (2024). Designing Interactive Web-GIS Dashboards for Data-Driven Emergency Management in Smart Cities. Journal of Logistics, Informatics and Service Science, 11(2).

https://doi.org/10.33168/jliss.2024.0222

Hansun, S., Suryadibrata, A., & Sandi, D. R. (2022). Deep Learning Approach in Predicting Property and Real Estate Indices. International Journal of Advances in Soft Computing and Its Applications, 14(1), 61–71.

https://doi.org/10.15849/ijasca.220328.05

Hosseini, S. M. A., Ghalambordezfooly, R., & de la Fuente, A. (2022). Sustainability Model to Select Optimal Site Location for Temporary Housing Units: Combining GIS and the MIVES–Knapsack Model. Sustainability, 14(8), 4453.

https://doi.org/10.3390/su14084453

Jašková D., et al. (2020). The Human Resources as an Important Factor of Regional Development | International Journal of Business and Society. Unimas.my.

https://publisher.unimas.my/ojs/index.php/IJBS/article/view/3364

Li, Z., & Ning, H. (2023, May 10). Autonomous GIS: the next-generation AI-powered GIS. arXiv.org.

https://arxiv.org/abs/2305.06453

Li, F., Laili, Y., Chen, X., Lou, Y., Wang, C., Yang, H., Gao, X., & Han, H. (2023). Towards big data driven construction industry. Journal of Industrial Information Integration, 35, 100483.
https://doi.org/10.1016/j.jii.2023.100483

Liu, B., Li, Y., Zhang, Y., Wu, Z., Pan, Y., & Li, M. (2025). Multi-criteria decision-making method for emergency shelter site selection considering flood risk: A case study of Zhuhai, China. Cleaner Engineering and Technology, 100892–100892.
https://doi.org/10.1016/j.clet.2025.100892

Mahmud, N., & Habib, M. A. (2024). A Comprehensive Business Location Choice Model Leveraging Machine Learning in Systematic Choice Set. Transportation Research Record: Journal of the Transportation Research Board, 2678(12), 1856–1871.
https://doi.org/10.1177/03611981241253609

Mauro, A. (2022). Property Prices and Disaster (In)Justice: Fifty Years of Demographic Change in Hawai'i County's Lava Hazard Zones. Lub.lu.se.
https://lup.lub.lu.se/student-papers/search/publication/9084540

Nie, Z. (2024). The suitability assessment for land territorial spatial planning based on ANN-CA model and the Internet of Things. Heliyon, 10(10), e31237–e31237.
https://doi.org/10.1016/j.heliyon.2024.e31237

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

Oliveira, A., Fachada, N., & Matos-Carvalho, J. P. (2024). Data Science for Geographic Information Systems. Data Science for Geographic Information Systems, 1–7. https://doi.org/10.1109/yef-ece62614.2024.10624902

Ouyang, W. (2024). Data visualization in Big Data Analysis: applications and future trends. Journal of Computer and Communications, 12(11), 76–85. https://doi.org/10.4236/jcc.2024.1211005

Parameswaran, T., Jayawickrama, T., & Melagoda, D. (2019). Analyzing the Impact of Location Factors on Building Construction Cost in Sri Lanka. https://www.ieomsociety.org/ieom2019/papers/622.pdf

Pelinescu, E. (2015). The Impact of Human Capital on Economic Growth. Procedia Economics and Finance, 22, 184–190. https://doi.org/10.1016/s2212-5671(15)00258-0

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

Sarigai, S., Yang, L., Slack, K., Lane, M. D., Buenemann, M., Wu, Q., Woodhull, G., Driscol, J. (2024). idwMapper: An interactive and data-driven web mapping framework for visualizing and sensing high-dimensional geospatial (big) data. (2024). ResearchGate.

https://www.researchgate.net/publication/378690200_idwMapper_An_interactive_and_data-driven_web_mapping_framework_for_visualizing_and_sensing_high-dimensional_geospatial_big_data

Shahzad, U. (2021). Application of machine learning for optimal wind farm location. Shahzad | Journal of Electrical Engineering, Electronics, Control and Computer Science.

https://jeeeccs.net/index.php/journal/article/view/276

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

Tang, X., Zou, C., Shu, C., Zhang, M., & Feng, H. (2024). Research on Site Selection Planning of Urban Parks Based on POI and Machine Learning—Taking Guangzhou City as an Example. Land, 13(9), 1362–1362.

https://doi.org/10.3390/land13091362

Verteego. (2020). Historical data for accurate sales forecasting. What data is needed to forecast sales?

https://www.verteego.com/en/what-data-is-needeed-to-forecast-sales

Wang, Y., Li, S., Zhang, X., Jiang, D., Hao, M., & Zhou, R. (2020). Site Selection of Digital Signage in Beijing: A Combination of Machine Learning and an Empirical Approach. ISPRS International Journal of Geo-Information, 9(4), 217. https://doi.org/10.3390/ijgi9040217

Wei, F., & Zhao, L. (2022). The effect of flood risk on residential land prices. Land, 11(10), 1612. https://doi.org/10.3390/land11101612

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

Xu, Y., Li, Y., Zheng, L., Cui, L., Li, S., Li, W., & Cai, Y. (2020). Site selection of wind farms using GIS and multi-criteria decision making method in Wafangdian, China. Energy, 207, 118222. https://doi.org/10.1016/j.energy.2020.118222

Zhao, B., Zheng, H., & Cheng, X. (2024). A Machine Learning Approach to Predict Site Selection from the Perspective of Vitality Improvement. Land, 13(12), 2113–2113. https://doi.org/10.3390/land13122113

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***

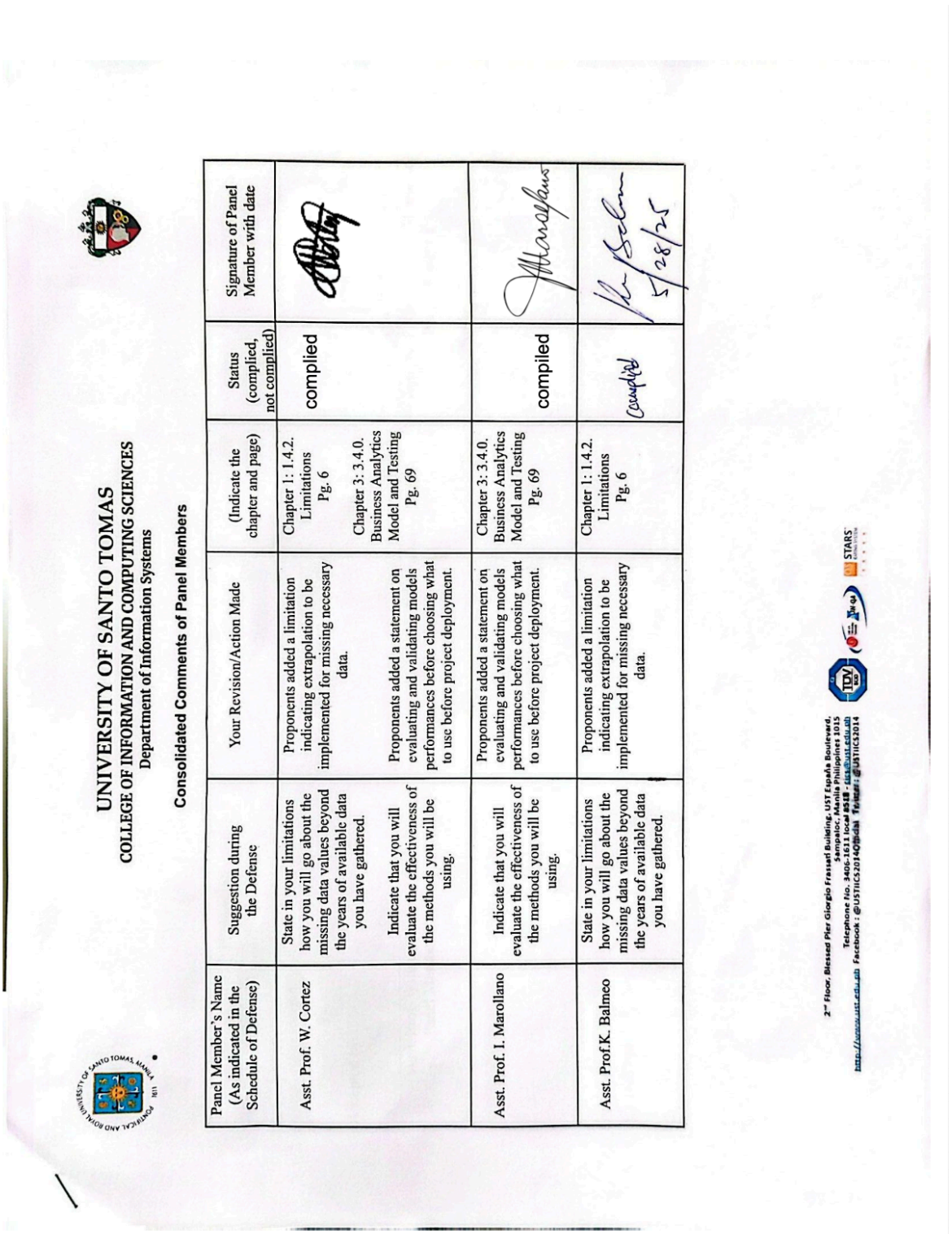
Time : ***11:30 am to 01:00 pm***

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	

Appendix B: Consolidated Comments of Panel Members



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	complied	
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

2nd Floor, Blessed Pier Giorgio Frassati Building, UST España Boulevard, Sampaloc, Manila Philippines 1015
Telephone No. 3406-1611 local 8518 - fcisms@ust.edu.ph
http://www.ust.edu.ph Facebook : @USTINCS2014Official Twitter: @USTINCS2014

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 - ccs@ust.edu.ph
http://www.ust.edu.ph Facebook : @USTIICS2014Official Twitter : @USTIICS2014