Institute of Technology of Cambodia

Department: Applied Mathematics and Statistics

Subject: Program for Data science

Lecturer: CHAN Sophal

Project Guideline

Group 3: **Land Price Analysis**

Members:

1. KOH Tito   e20200353
2. CHEA Makara   e20201131
3. AOV Keatmeng   e20201812
4. BUT Cheableng   e20200861
5. HOK Kimleang   e20200637

Academic year
2022-2023

Project Guideline

Define problem and goal	• State the problem that we are going to solve with this project. • Goal: To find the best model for Data Analysis in Land Price Analysis.
Data Collection	• Collecting Data: Web scraping from the websites that contain the information of the land in Cambodia. • Understand the important features for the training model.
Exploratory Data Analysis (EDA) ➢ Data cleaning and Preprocessing	• Handle missing values. • Check and remove Data duplication. • Handle the outliers. • Preprocess the data by standardizing, normalizing, or transforming variables as necessary. • Perform feature engineering techniques.
➢ Data Visualization	• Explore and visualize the data to understand deeply about its characteristics. • Summarize the main statistical measures and distributions of the variables. • Identify patterns, correlations, or relationships between variables using statistical and visual techniques.
Modeling and Analysis	• Select models based on the nature of the problem and data. • Split the data into training and testing sets for model training and evaluation. • Train Data for each model that is chosen. • Evaluate the model performance using relevant metrics and techniques using MSE, RMSE, R squared, Accuracy score…etc.
Conclusion	• Select the best model that is useful for the Dataset. • Interpret the results and draw meaningful conclusions.