

Artificial Intelligence Mid Exam



From:

AL AZHAR RIZQI RIFA'I FIRDAUS

Class:

2 I

Absence:

01

Student Number Identity:

2241720263

Department:

Information Technology

Study Program:

Informatics Engineering

Case Study

You are a real estate businessman. You have several units of houses, land, and shophouses spread across Malang City, and property acquisition plans are seen as future prospects. As a businessman your goal is to get good sales, and good profits.

To achieve that goal, you have several strategies that are divided into the following categories:

1. Determine the best selling price for each unit
2. Determine the target marketing on social media
3. Choose the right location for the property unit to be acquired
4. Determine the specifications for the house
5. Choose the priority of the unit owned, between a house, land, or shophouse

Your job is to choose one of the five strategies above!

Question

1. From one of the strategies you choose, define a clear problem with machine learning as the solution approach (10points)

- Problem: The problem we need to address is to determine the optimal location for acquiring new property units (houses, land, or shophouses) in Malang City to maximize sales and profits. We want to identify the neighborhoods or areas with the highest potential for property appreciation and demand.

2. Develop a data ingestion strategy, determine what data will be processed, where it will come from (10 points)

Data Source: We will gather data from various sources, including:

1. Real estate listings and transaction data from local real estate agencies.
2. Demographic and economic data for different neighborhoods in Malang City.
3. Social media and online forums where people discuss their preferences and concerns about various neighborhoods.
4. Historical property price data for different locations.

3. Make a sample data, at least 20 items (20 points)

Number	Property Type	Location	Price (IDR)	Area (Meter)	Nearby Amenities	Demand Score (0-10)
1	House	Klojen	700,000,000	200	School, Hospital	9
2	House	Ijen	1,200,000,000	150	Hospital, Park	8
3	House	Araya	1,500,000,000	220	Market, School	10
4	House	Sukarno Hatta	900,000,000	230	University, Market	9
5	House	Dieng	1,100,000,000	190	Market, School	8
6	Land	Tidar	700,000,000	300	School, Hospital	9
7	Land	Tidar	1,200,000,000	400	Hospital, Park	8
8	Land	Klojen	1,500,000,000	500	Market, School	9
9	Land	Ijen	900,000,000	450	University, Market	10
10	Land	Sukarno Hatta	1,100,000,000	440	Market, School	8
11	Apartment	Araya	700,000,000	80	School, Hospital	9
12	Apartment	Tidar	1,200,000,000	90	Hospital, Park	9
13	Apartment	Dieng	1,500,000,000	100	Market, School	9
14	Apartment	Sukarno Hatta	900,000,000	110	University, Market	9
15	Apartment	Sawojajar	1,100,000,000	120	Market, School	8
16	Building	Dieng	700,000,000	150	School, Hospital	8
17	Building	Tidar	1,200,000,000	190	Hospital, Park	8
18	Building	Klojen	1,500,000,000	180	Market, School	10
19	Building	Ijen	900,000,000	160	University, Market	10
20	Building	Dieng	1,100,000,000	170	Market, School	10

4. What kind of machine learning model/method will be used? (10 points)

- We can use a regression model to solve this problem. Specifically, we can apply multiple linear regression to predict property prices based on various features, including location, area, nearby amenities, and demand score.

5. Explain whether the method is supervised or unsupervised? (10 points)

- The chosen method is supervised learning because we have labeled data. We can train the model using historical property data where we know the actual prices, and we want the model to predict property prices for new, unlabeled data.

6. Explain why do you choose this model? (20 points)

Reasons for choosing multiple linear regression:

- Interpretability: Linear regression allows us to understand the impact of each feature on property prices.
- Simplicity: It's a simple yet effective model for price prediction.
- Existing Data: We have a dataset with labeled examples (property prices), which is suitable for supervised learning.
- Potential for Improvement: We can extend this model to include more complex features or use other regression techniques if needed.

7. Explain the machine learning method that will be used! (20 points)

- Multiple Linear Regression: This method models the relationship between the dependent variable (property price) and multiple independent variables (location, area, nearby amenities, demand score) using a linear equation. The model aims to find the best-fitting line that minimizes the difference between the predicted property prices and the actual prices in the training data.

The linear equation for multiple linear regression is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$$

- \hat{Y} is the predicted property price.
- β_0 is the intercept.
- $\beta_1, \beta_2, \beta_3, \beta_4$ are the coefficients for the independent variables (location, area, nearby amenities, demand score).
- X_1, X_2, X_3, X_4 are the values of the independent variables.
- ϵ is the error term.

The goal is to find the values of the coefficients (β) that minimize the error (ϵ) and provide accurate price predictions for new properties. The model can be evaluated using metrics like Mean Squared Error (MSE) or R-squared to measure its performance.