



Department of Computer Engineering

Faculty of Engineering

Kasetsart University

[After dead-line not accepted in all cases]

1) Attributes of the car colour, type, origin and the target variable stolen can be either yes or no.

1. Red, Sport, Domestic, Yes
2. Red, Sport, Domestic, No
3. Red, Sport, Domestic, Yes
4. Yellow, Sport, Domestic, No
5. Yellow, Sport, Imported, Yes
6. Yellow, SUV, Imported, No
7. Yellow, SUV, Imported, Yes
8. Yellow, SUV, Domestic, No
9. Red, SUV, Imported, No
10. Red, Sport, Imported, Yes

Show your steps (Python or RapidMiner) in predicting the class for the car with attribute values (Red, SUV, Domestic).

2) Comparing KNN and Logistic Regression for Predicting Survival on the Titanic

Objective:

1. Download the "Titanic Survivors" dataset from Kaggle.
2. Write a Python program to compare KNN and Logistic Regression models for predicting passenger survival on the Titanic.
3. Explain the techniques used and the chosen parameters for each model.
4. Analyze and compare the performance of both models using appropriate metrics.

Instructions:

1. Load and Prepare Data:

- Use the pandas library to load the data.
- Explore the data and handle missing values.
- Convert categorical variables to numerical ones.
- Split the data into training and testing sets.

2. Build KNN Model:

- Define the K value (number of neighbors).
- Train the KNN model on the training set.
- Predict survival on the testing set.

3. Build Logistic Regression Model:

- Define the parameter value
- Train the Logistic Regression model on the training set.
- Predict survival on the testing set.

4. Compare Models:

- Calculate performance metrics such as accuracy, precision, recall, and F1-score.
- Compare the results of KNN and Logistic Regression models.
- Analyze the reasons for any differences (if present).

5. Explain Results:

- Summarize the findings of both models.
- Discuss the advantages and disadvantages of each model.
- Recommend the most suitable model for predicting survival on the Titanic.

Note:

- This exercise can be customized further, such as adding other models for comparison or performing deeper analysis of the results.
- A good response should provide clear explanations and analyze the results logically.