QUESTION SHEET
DAC 2021

DATA ANALYSIS
COMPETITION
2021

Data Analysis Competition (DAC) 2021

1. The Question sheet is accessible on 11th - 19th July 2021. Report has to be submitted maximum by 19th July 2021 at 23.59 Western Indonesian Time (GMT+7)

2. Once the report is submitted, it cannot be replaced or revised for any reasons.

3. All round will be conducted online.

4. The Semi-final round will be held on 28th August 2021.

5. The final round will be held on 4th September 2021.

GUIDELINE ELIMINATION ROUND

1. Participants, who are registered as teams consist of 2 active students from the same university, will compete in the elimination round.

2. Participants pay for the account activation fee to activate their registered accounts.

3. Participants with the activated accounts may access the question via PRS website during the elimination round.

4. Participants are allowed using any programming language.

5. The time period to accomplish and collect the answers is 9 days, from 11th July 2021 until 19th July 2021.

6. Participants are not allowed to submit the answers exceeds the allocated time. The subject of answer sheet is : Registration Number_Team's name_University.

7. Participants who qualify for the semifinal round are 30 teams. The semifinalists are selected with rank based selection.

GENERAL INSTRUCTIONS ELIMINATION ROUND

1. Write your results of the analysis on the answer sheet we provide. In addition, we require you to :

    a. Write your analysis completely and clearly.

    b. Write the report with maximum 10 pages in English. Use paper size A4, font Times New Roman size 12 with line spacing 1.15 and margin 3 3 3 3.

    c. Participant are recommended to write the code in the attachment page right after the article. The attachment page are not included in the previous 10 pages and has no page limit

2. Save the answer sheet with the format file:

    • For the article is pdf and docx.

    • For the prediction is csv.

    with the format name: Registration Number _Team's name_University

3. All answers must be compressed in a RAR format with the format name: Registration Number_Team's name_University.rar. Example: ID-19-01-0248_PRS2021_ITS.rar (All participants may check the registration number at the pre-registration webpage on PRS Website)

4. Submit your team's answer maximum on 19th July 2021 at 23.59 Western Indonesian Time (GMT+7) on the PRS website. If participants send in format which does not match, then the participant's answer sheet will not be corrected.

5. Participants are allowed using any programming language but not allowed using excel only

6. All kinds of cheating, including copy answers from other teams will result in disqualification.

7. The judge's decision cannot be contested.

8. Any participant who breaks the general instructions will be disqualified.

Himpunan Mahasiswa Statistika ITS (HIMASTA-ITS)

Gedung H Lantai III, Jl. Arief Rahman Hakim
Kampus ITS Sukolilo Surabaya
Email: eventhimastaits@gmail.com
Contact Person: Catur (085155430660)
Wanda (081327522030)

Data Analysis Competition (DAC) 2021

Nowadays, the world's startup industry has grown rapidly. Indonesia itself has 230 start-ups, four of which have become unicorns. Startups have revolutionized various industrial sectors, one of which is the property industry especially for residence. It happened because, according to the Vice President of Non-Subsidized Mortgage and Consumer Lending Division of PT Bank Tabungan Negara (Persero) Tbk Suryanti Agustinar, the housing sector experienced growth of 2% in the third quarter of 2020 and in 2021 it is projected that economic growth will reach 5% in the housing sector. This is also supported by the statement of the General Chairperson of the Indonesian Real Estate Brokers Association (AREBI) Lukas Bong who said that the property trend in 2021 is more towards landed houses compared to apartments. This revolution has many benefits for residence agents and customers. With app developer startups now planning a tough holiday. All can be done in just a hand and a matter of minutes. However, does the user easily search for the house in residence that matches with their preference? Of course it is difficult. Therefore, the application development team created a residence recommendation system according to user preferences. The data is collected to obtain information that helps the application development team see customer status.

Table 1. Variable default on the dataset

Name of Variable	Description
time_date	Timestamp
site	ID of the site
continent_id	ID of continent
buyer_country	The ID of the country the customer is located
buyer_region	The ID of the region the customer is located
buyer_city	The ID of the city the customer is located
distance	Physical distance between a regency and a customer at the time of search. A null means the distance could not be calculated

buyer_id	ID of user
mobile	1 when a user connected from a mobile device, 0 otherwise
package	1 if the click/buying was generated as a part of a package (i.e. combined with a furniture), 0 otherwise
channel_id	ID of a marketing channel
buying_date	Buying date
dealing_date	Dealing with seller date
adults	The number of adults specified in the room
children	The number of (extra occupancy) children specified in the room
room	The number of rooms specified in the search
destination_id	ID of the destination where the regency search was performed
destination_type	Type of destination
regency_continent	Regency continent
regency_country	Regency country
regency_market	Regency market
dealing	1 if dealing, 0 if a click
cnt	Number of similar events in the context of the same user session
regency_cluster	ID of regency cluster

The residence agency has provided logs of customer behavior. These include what customers searched for, how they interacted with search results (click/buy), the price interval they want, the facilities they want, etc.

We are here interested in studying the characteristics of customers who decide their preference for a house in a residence. We provide you 10000 customers with different characteristics. We have cluster residences based on price, customer star ratings, geographical locations relative to city center, etc (maybe it can be explained in more detail). As a consultant, you are asked to determine which residence's cluster is chosen by customers based on several clusters that you specify according to your creativity. If you think you find inconsistent information regarding the data in relation to the description, you are required to make a decision on your own and give your justification why you take such an action. To report the results, you are required to write in the article format (see General Instruction) of a maximum of 10 pages and the attachment of your code.

The paper should include four chapters as follows:

a. Chapter I: Determine the variable you choose for the prediction and give the justification.

b. Chapter II: Determine the methodology you use for the analysis and explain your reasons why you choose such a method (mention and explain the theory briefly). Start from preprocessing, modelling until validation.

c. Chapter III: Show the result of data exploration and explain the results of your analysis.

d. Chapter IV: Conclude your results and give suggestions based on your results.

Himpunan Mahasiswa Statistika ITS (HIMASTA-ITS)

Gedung H Lantai III, Jl. Arief Rahman Hakim
Kampus ITS Sukolilo Surabaya
Email: eventhimastaits@gmail.com
Contact Person: Catur (085155430660)
Wanda (081327522030)