Peer-graded Assignment: Final Assignment

1. [**Instructions**](https://in.coursera.org/learn/data-science-methodology/peer/LRAgH/final-assignment)
2. [My submission](https://in.coursera.org/learn/data-science-methodology/peer/LRAgH/final-assignment/submit)
3. [Discussions](https://in.coursera.org/learn/data-science-methodology/peer/LRAgH/final-assignment/discussions)

In this Assignment, you will demonstrate your understanding of the data science methodology by applying it to a given problem. Pick one of the following topics to apply the data science methodology to:

1. Emails
2. Hospitals
3. Credit Cards

You will have to play the role of the client as well as the data scientist to come up with a problem that is more specific but related to these topics.

Peer-graded Assignment: Final Assignment

1. [Instructions](https://in.coursera.org/learn/data-science-methodology/peer/LRAgH/final-assignment)
2. [**My submission**](https://in.coursera.org/learn/data-science-methodology/peer/LRAgH/final-assignment/submit)
3. [Discussions](https://in.coursera.org/learn/data-science-methodology/peer/LRAgH/final-assignment/discussions)

Credit card fraud detection

Submitted on February 21, 2022

[Shareable Link](https://in.coursera.org/learn/data-science-methodology/peer/LRAgH/final-assignment/review/uRfy7JKGEeyUBxJw3wg2Sw)

**PROMPT**

Which topic did you choose to apply the data science methodology to? **(2 marks)**

I have chosen the topic - Application of Data Science to detect fraud transactions using "**Credit Cards**" for this task as I am currently working in BFSI sector.

**RUBRIC**

Did the student pick one of the three topics proposed in the assignment overview?

|  |  |  |
| --- | --- | --- |
|  | 0 points  No |  |
|  | **2 points**  **Yes** | DU |

**PROMPT**

Next, you will play the role of the client and the data scientist.

Using the topic that you selected, complete the Business Understanding stage by coming up with a problem that you would like to solve and phrasing it in the form of a question that you will use data to answer. **(3 marks)**

You are required to:

1. Describe the problem, related to the topic you selected.
2. Phrase the problem as a question to be answered using data.

For example, using the food recipes use case discussed in the labs, the question that we defined was, "Can we automatically determine the cuisine of a given dish based on its ingredients?".

As fraud transactions are always a risk with credit cards for both customers and banks, so it is very crucial to detect and stop any fraud transactions that is being done using credit cards in approval phase of the transaction.

Our problem as a question would be - "Can we automatically determine if the transaction made using the credit card is fraud or not"

**RUBRIC**

The student is required to come up a problem related to the topic they selected and the problem must be phrased as a question that can be answered using data. Use your best judgement to rate the student's completion of the Business Understanding stage.

|  |  |  |
| --- | --- | --- |
|  | 1 point  Poor. Some description is provided about the problem, but the question to be answered is missing. |  |
|  | 2 points  Good. The problem to be solved is described and a question is submitted but the question does not match the problem described. |  |
|  | **3 points**  **Excellent. The student gave sufficient description of the problem, and the question to be answered reflects the problem described.** | DU |

**PROMPT**

Briefly explain how you would complete each of the following stages for the problem that you described in the Business Understanding stage, so that you are ultimately able to answer the question that you came up with. **(5 marks)**:

1. Analytic Approach
2. Data Requirements
3. Data Collection
4. Data Understanding and Preparation
5. Modeling and Evaluation

You can always refer to the labs as a reference with describing how you would complete each stage for your problem.

1. Analytic Approach - As the problem requires the answer in terms of yes or no, so we will be using classification algorithm 2. Data Requirements - To create classification model for this problem, we will need the every credit card transaction record including transaction date, time, place, amount, merchant, customer details like salary, age, gender, etc., which will help to detect if the transaction seems to be abnormal for the customer. 3. Data Collection - Firstly we need to identify the source and destination of the required data. We will be collecting all the required data from different tables under bank databases into single place/table. 4. Data Understanding and Preparation - In this step we need to evaluate the different variables of our data in order to understand it better. For example, we would calculate univariate statistics, such as mean or median and the correlation between variables. So we need to evaluate the quality of the data. In the data preparation phase we have to prepare the data in an specific way depending on the model. 5. Modeling and Evaluation - Then we create a classification model, evaluate the outcome and perform the corresponding changes until we have a suitable model.

**RUBRIC**

The student is required to explain how they would complete each stage for the problem that they described in the Business Understanding stage. Use your best judgement to rate the student's description of each stage.

|  |  |  |
| --- | --- | --- |
|  | 1 point  Poor. Many stages are missing and insufficient description is provided. |  |
|  | 3 points  Good. At least three stages are described and the description is clear and applies to the question defined in the Business Understanding stage. However, some stages are missing. |  |
|  | **5 points**  **Excellent. All stages are described appropriately and the description is clear and applies to the question that they defined in the Business Understanding stage.** |  |