# Course 3- Data Science Methodology

## Week 3

## Quiz 1: From Deployment to Feedback

**1.**

Question 1

Feedback is not required once the model is deployed because the Model Evaluation stage would have assessed the model and made sure that it performed well.

1 / 1 point

True

False

Correct

Correct.

**2.**

Question 2

A data scientist determines that building a recommender system is the solution for a particular business problem at hand. What stage of the data science methodology does this represent?

1 / 1 point

Analytic Approach.

Model Evaluation.

Modeling.

Deployment.

Correct

Correct. The selection of a model to use should happen in the Analytic Approach stage.

**3.**

Question 3

A data scientist, John, was asked to help reduce readmission rates at a local hospital. After some time, John provided a model that predicted which patients were more likely to be readmitted to the hospital and declared that his work was done. Which of the following best describes this scenario?

1 / 1 point

John still needed to collect more data.

John's mistake is that he lied in the Analytic Approach step of the data science methodology.

Even though John only submitted one solution, it might be a good one. However, John needed feedback on his model from the hospital to confirm that his model was able to address the problem appropriately and sufficiently.

John only provided one model as a solution and he should have provided multiple models.
Correct
Correct.

## 4.

Question 4
Which of the following represent the two important characteristics of the data science methodology?
1 / 1 point

It is a highly iterative process and immediately ends when the model is deployed.

It has no endpoint because data collection occurs before identifying the data requirements.

It is a highly iterative process and it never ends.

It immediately ends when the model is deployed because no feedback is required.
Correct
Correct.

## 5.

Question 5
For predictive models, a test set, which is similar to – but independent of – the training set, is used to determine how well the model predicts outcomes. This is an example of what step in the methodology?
1 / 1 point

Model Evaluation.

Analytic Approach.

Deployment.

Data Requirements.
Correct
Correct.

## 6.

Question 6
What are three important reasons that data scientists should maintain continuous communication with business sponsors throughout a project?
1 / 1 point

So that business sponsors can ensure the work remains on track to generate the intended solution.
Correct
Correct.

Actually, data scientists do not need to maintain a continuous communication with business sponsors and stakeholders.

So that business sponsors can review intermediate findings.
Correct
Correct.

So that business sponsors can provide domain expertise.
Correct
Correct.

# Final Assignment

## Predicting Credit Card Default: A Data-driven Approach for Risk Assessment

Submitted on June 11, 2023

Shareable Link

**PROMPT**

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

The topic that I have chosen to apply data science methodology to is Credit Cards. My problem statement will be Identifying Credit card frauds.

**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?".

Problem: Credit Card Default Prediction

In the context of credit cards, the problem we aim to solve is to predict the likelihood of a customer defaulting on their credit card payment. Defaulting refers to the failure of a cardholder to make the required minimum payment within a specified time frame. By accurately identifying customers at a higher risk of defaulting, financial institutions can take proactive measures such as offering financial counseling, adjusting credit limits, or implementing collection strategies to mitigate potential losses.

Question: Can we predict the likelihood of a credit card customer defaulting on their payment using data on their transaction history, demographics, and credit-related attributes?

**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: To answer the question of predicting credit card default likelihood, we can adopt a supervised machine learning approach. Specifically, we can use classification algorithms to build a predictive model that classifies customers into default and non-default categories based on their transaction history, demographics, and credit-related attributes. Techniques such as logistic regression, decision trees, random forests, or gradient boosting can be explored to develop the predictive model.

2. Data Requirements: To successfully build the predictive model, we would require a dataset that includes historical credit card transaction data along with associated customer information. The dataset should contain features such as transaction amounts, transaction dates, customer demographics (age, gender, location), credit utilization, credit limit, payment history, and any other relevant attributes that could be indicative of default likelihood.

3. Data Collection: Data collection would involve acquiring the required credit card transaction data and customer information from the financial institution or relevant sources. This process may involve obtaining permission and adhering to data privacy regulations. It is essential to ensure that the data collected is representative of the population and includes a sufficient number of default and non-default instances.

4. Data Understanding and Preparation: In this stage, we would analyze and explore the collected data to gain insights into its characteristics. This involves performing descriptive statistics, data visualization, and feature engineering to transform and prepare the data for modeling. Additionally, steps such as handling missing values, addressing outliers, and encoding categorical variables may be necessary to ensure the quality and suitability of the data for modeling.

5. Modeling and Evaluation: In this stage, we would train and evaluate the predictive models using the prepared dataset. We would split the data into training and testing sets to assess the performance of the models. Multiple classification algorithms can be implemented, and their performances can be compared using appropriate evaluation metrics such as accuracy, precision, recall, and F1-score. Hyperparameter tuning and cross-validation techniques can be applied to optimize the model's performance. The final model would be selected based on its ability to accurately predict credit card default likelihood.

By completing these stages, we would be able to develop a predictive model that can answer the question of predicting the likelihood of credit card customers defaulting on their payment using the available data.

# Final Exam

## 1.

Question 1
Which of the following is the first state of the data science methodology?
1 / 1 point

Modeling

Data Collection

Data Understanding

Business Understanding
Correct

## 2.

Question 2
Business Understanding is the least important stage in the data science methodology because none of the other stages depend on it.
1 / 1 point

True.

False
Correct

## 3.

Question 3
According to the videos, you can think of the _____ and _____ stages as a cooking task, where the problem at hand is a recipe, and the data to answer the question is the ingredients.
1 / 1 point

Business Requirements; Presentation Requirements

Data Analysis; Presentation Requirements

Analytics; Business Requirements

Data Requirements; Data Collection
Correct

## 4.

Question 4
In the Data Collection stage, techniques such as descriptive statistics and visualization can be applied to the data set, to assess the content, quality, and initial insights about the data
1 / 1 point

True

False
Correct

**5.**

Question 5
A _____ is used for predictive modeling.
1 / 1 point

Technique set

Modeling set

Analysis set

Training set
Correct

**6.**

Question 6
A false-positive is what type of error?
1 / 1 point

Type II error

Type L error

Type I error

Type III error
Correct

**7.**

Question 7
The Data Understanding stage encompasses what?
1 / 1 point

All activities related to constructing the dataset.

Sorting the data.

Transforming data

Removing redundant data.
Correct

**8.**

Question 8
Select the correct statement about the Data Preparation stage.
1 / 1 point

The Data Preparation stage involves addressing missing values.

The Data Preparation stage involves correcting invalid values and addressing outliers.

The Data Preparation stage involves removing duplicate data.

The Data Preparation stage involves properly formatting the data.

All of the above statements are correct.
Correct

## 9.

Question 9
The final stages of the data science methodology are an iterative cycle between what stages?
1 / 1 point

Modeling, Deployment, Data Understanding, Data Preparation

Data Preparation, Evaluation, Feedback, Deployment

Modeling, Evaluation, Deployment, and Feedback.

Data Understanding, Data Preparation, Evaluation, and Feedback.
Correct

## 10.

Question 10
Deploying a model into production represents the beginning of an iterative process from _____, then Model Refinement, and to Redeployment.
1 / 1 point

Scalability

Data Storage

Feedback

None of the above
Correct

## 11.

Question 11
The data science methodology is a specific strategy that guides processes and activities relating to data science only for text analytics.
1 / 1 point

True

False
Correct

## 12.

Question 12
As a data scientist what type of analysis is used with descriptive statistics and data visualization techniques?
1 / 1 point

- Exploratory
- Deep learning
- Extraction
- Classification

Correct