Data Science with Python Programming

- Presentation By Uplatz
- Contact us: https://training.uplatz.com
- Email: info@uplatz.com
- Phone: +44 7836 212635

Data Science Methodology (Part-2)

Uplatz

Learning outcomes:

Module 1: From Problem to Approach

- Business Understanding
- Analytic Approach

Module 2: From Requirements to Collection

- Data Requirements
- Data Collection

Module 3: From Understanding to Preparation

- Data Understanding
- Data Preparation

Module 4: From Modeling to Evaluation

- Modeling
- Evaluation

Module 5: From Deployment to Feedback

- Deployment
- Feedback

Learning outcomes:

Module 4: From Modeling to Evaluation

- Modeling
- Evaluation

Module 5: From Deployment to Feedback

- Deployment
- Feedback

Summary

Module 4: From Modeling to Evaluation

Modeling:

Welcome to Data Science Methodology From Modeling to Evaluation Modeling - Concepts!

Modelling is the stage in the data science methodology where the data scientist has the chance to sample the sauce and determine if it's bang on or in need of more seasoning!

This portion of the course is geared toward answering two key questions:

First, what is the purpose of data modeling, and

second, what are some characteristics of this process?

Data Modelling focuses on developing models that are either descriptive or predictive. An example of a descriptive model might examine things like: if a person did this, then they're likely to prefer that. A predictive model tries to yield yes/no, or stop/go type outcomes.

Module 4: From Modeling to Evaluation

Modeling:

These models are based on the analytic approach that was taken, either statistically driven or machine learning driven.
The data scientist will use a training set for predictive modelling. A training set is a set of historical data in which the outcomes are already known. The training set acts like a gauge to determine if the model needs to be calibrated.
In this stage, the data scientist will play around with different algorithms to ensure that the variables in play are actually required. The success of data compilation, preparation and modelling, depends on the understanding of the problem at hand, and the appropriate analytical approach being taken. The data supports the answering of the question, and like the quality of the ingredients in cooking, sets the stage for the outcome.

Module 4: From Modeling to Evaluation

Modeling:

Constant refinement, adjustments and tweaking are necessary within each step to ensure the outcome is one that is solid. In **John Rollins' descriptive Data Science Methodology,** the framework is geared to do 3 things: First, understand the question at hand. Second, select an analytic approach or method to solve the problem, and third, obtain, understand, prepare, and model the data.

The end goal is to move the data scientist to a point where a data model can be built to answer the question.

With dinner just about to be served and a hungry guest at the table, the key question is: Have I made enough to eat?

Well, let's hope so. In this stage of the methodology, model evaluation, deployment, and feedback loops ensure that the answer is near and relevant.

Module 4: From Modeling to Evaluation

Evaluation:

A model evaluation goes hand-in-hand with model building as such, the modeling and evaluation stages are done iteratively.
Model evaluation is performed during model development and before the model is deployed.
Evaluation allows the quality of the model to be assessed but it's also an opportunity to see if it meets the initial request.
Evaluation answers the question: Does the model used really answer the initial question or does it need to be adjusted?
Model evaluation can have two main phases.
The first is the diagnostic measures phase, which is used to ensure the model is working as intended.
If the model is a predictive model, a decision tree can be used to evaluate if the answer the model can output, is aligned to the initial design. It can be used to see where there are areas that require adjustments.

Module 4: From Modeling to Evaluation

Evaluation:

If the model is a descriptive model, one in which relationships are being assessed, then a testing set with known outcomes can be applied, and the model can be refined as needed. The second phase of evaluation that may be used is statistical significance testing.

This type of evaluation can be applied to the model to ensure that the data is being properly handled and interpreted within the model. This is designed to avoid unnecessary second guessing when the answer is revealed. Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future.

Module 4: From Modeling to Evaluation

Evaluation:

Methods for evaluating a model's performance are divided into 2 categories: namely, **holdout** and **Cross-validation**. Both methods use a test set (i.e. data not seen by the model) to evaluate model performance. It's not recommended to use the data we used to build the model to evaluate it. This is because our model will simply remember the whole training set, and will therefore always predict the correct label for any point in the training set. This is known as **overfitting**.

Module 4: From Modeling to Evaluation

Evaluation:

The purpose of **holdout** evaluation is to test a model on different data than it was trained on. This provides an unbiased estimate of learning performance.

Cross-validation is a technique that involves partitioning the original observation dataset into a training set, used to train the model, and an independent set used to evaluate the analysis.

Module 5: From Deployment to Feedback

Deployment :

Welcome to Data Science Methodology From Deployment to Feedback - Deployment!

While a data science model will provide an answer, the key to making the answer relevant and useful to address the initial question, involves getting the stakeholders familiar with the tool produced.

In a business scenario, stakeholders have different specialties that will help make this happen, such as the solution owner, marketing, application developers, and IT administration.

Once the model is evaluated and the data scientist is confident it will work, it is deployed and put to the ultimate test. Depending on the purpose of the model, it may be rolled out to a limited group of users or in a test environment, to build up confidence in applying the outcome for use across the board.

Module 5: From Deployment to Feedback

Deployment :

In preparation for solution deployment, the next step was to assimilate the knowledge for the business group who would be designing and managing the intervention program to reduce readmission risk.

In this scenario, the business people translated the model results so that the clinical staff could understand how to identify high-risk patients and design suitable intervention actions.

Deployment of an ML-model simply means the integration of the model into an existing production environment which can take in an input and return an output that can be used in making practical business decisions.

Module 5: From Deployment to Feedback

Deployment :

The concept of deployment in data science refers to the application of a model for prediction using a new data. Building a model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data science process.

Module 5: From Deployment to Feedback

Feedback :

Once in play, feedback from the users will help to refine the model and assess it for performance and impact.

The value of the model will be dependent on successfully incorporating feedback and making adjustments for as long as the solution is required. Throughout the Data Science Methodology, each step sets the stage for the next.

Making the methodology cyclical, ensures refinement at each stage in the game. The feedback process is rooted in the notion that, the more you know, the more that you'll want to know. That's the way **John Rollins** sees it and hopefully you do too. Once the model is evaluated and the data scientist is confident it'll work, it is deployed

and put to the ultimate test: actual, real-time use in the field.

Summary

I hope you've learned how to think like a data scientist, including taking the steps involved in tackling a data science problem and applying them to interesting, real-world examples.

These steps have included:

forming a concrete business or research problem, collecting and analysing data, building a model, and understanding the feedback after model deployment.

I hope you've also learned methodical ways of moving from problem to approach, including the importance of understanding the question, working with the data, specifically, determining the data requirements, collecting the appropriate data, the business goals and objectives, and picking the most effective analytic approach to answer the question and solve the problem.

Summary

You've also learned how to model the data by using the appropriate analytic approach, based on the data requirements and the problem that you were trying to solve. Once the approach was selected, you learned the steps involved in evaluating and deploying the model, getting feedback on it, and using that feedback constructively so as to improve the model.
Remember that the stages of this methodology are iterative!
This means that the model can always be improved for as long as the solution is needed, regardless of whether the improvements come from constructive feedback, or from examining newly available data sources.
Your success within the data science field depends on your ability to apply the right tools, at the right time, in the right order, to the address the right problem.
And that is the way John Rollins sees it!

Thank you