Data Science Methodology

Peer-graded Assignment

Lukas Schumi

Final Assignment to finish the course
Data Science Methodology

Vienna, Austria
April 2020

Contents

1 Instructions **2**

2 Topic **3**

2.1 Business Understanding

What is problem you trying to solve? 3

2.2 Analytic Approach

How can you use the data to answer the question? 3

2.3 Data Requirements

What data do you need to answer the question? 3

2.4 Data Collection

Where is the data coming from (identify all sources) and how will you get it? 3

2.5 Data Understanding

Is the data that you collected representative of the problem to be solved? 4

2.6 Data Preparation

What additional work is required to manipulate and work with the data? 4

2.7 Model Training

In What way can the data be visualized to get the answer that is required? 4

2.8 Model Evaluation

Does the model used really answer the initial question or does it need to be adjusted? 5

1 Instructions

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.

You are required to:

- Describe the problem, related to the topic you selected.
- 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?".

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.

- Analytic Approach
- Data Requirements
- Data Collection
- Data Understanding and Preparation
- Modeling and Evaluation

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

2 Topic

The topic I chose is a small pharmaceutical Business in Vienna which tries to sell more products.

2.1 Business Understanding
What is problem you trying to solve?

The problem at hand is, that sales are stagnating. To solve the issue and increase revenue a new model of advertising will be introduced. According to salesforce.com the ROI of newsletter used in businesses in 2018 was up to 3800 percent, also newsletter are a great way to analyze customer behavior and engagement.

The past of said pharmaceutical Business suggests that direct marketing to customers are not that efefctive.

Who is Your Audience? For the reason, that customers can buy most of the products only in pharmacies, the approach is taken to advertise directly to pharmacies. Those in return will be keen to sell their stock to customers.

So a **newsletter to pharmacies** will be introduced.

2.2 Analytic Approach
How can you use the data to answer the question?

The focus will lie on the question: **do pharmacies buy more when they get a newsletter?** They will recieve Emails for a while and then the data will be evaluated to see if the revenue increased.

2.3 Data Requirements
What data do you need to answer the question?

1.) Open and click rate as well as deliverability of the mails
2.) Visits to Homepage directly from mail
3.) Salesnumbers through qualified sales representatives
4.) Direct sales

2.4 Data Collection
Where is the data coming from (identify all sources) and how will you get it?

The collected data will be grouped in two categories: concerning sales and concerning mails.
The sales numbers are easily available in the internal system. Also feedback from the sales representatives will be taken into account.

An external mailservice will be used to send and track the newsletter. Data like open rate will be available through the service. Analytical tools of the homepage will track how many visits will be coming directly from a link in the newsletter.
Taking into account, that a lot of orders are taken through phone, the fact that someone is calling because of the newsletter will also be noted, as well as special offers only for subscribers to distinguish the sales from regular ones.

2.5 Data Understanding

Is the data that you collected representative of the problem to be solved?

Because there are not that many pharmacies in Austria, there will be no problem collecting the data from all of them.

2.6 Data Preparation

What additional work is required to manipulate and work with the data?

The sales numbers as well as the data from the mail service will be nicely structured already. They only need to be combined in a database which can be easily done by IDing the pharmacies.
The data collected from phone calls have to taken in a structered form to be useful. The feedback from sale representatives has to be considered individually.

2.7 Model Training

In What way can the data be visualized to get the answer that is required?

The model can be achieved easily through a simple decission tree considering the following questions:

Did the pharmacies even get the newsletter? If yes: did they open it? If yes: did they buy more in general? If yes: did they buy more of the advertised products?

To get a better understanding visual tools will be used.
A regression curve of sales over time can show if there is a general increase in revenue and barplots of sales of products bevor and after they were advertised in the newsletter can give an idea if the measurements worked.

2.8 Model Evaluation

Does the model used really answer the initial question or does it need to be adjusted?

Every node of the decission tree has to be evaluated seperately.
If the pharmacies never got the emails or they received bugged a new service tool has to be used. If the mails are not getting opend a lot, a way to advertise better must be found. Either trough more attractive headlines in the mails or directly through the representatives. If all those factors can be excluded, one can conclude that marketing might not be the problem of stagnating revenue but maybe the price or quality of the products.