Deloitte.

Data Science Academy - Intermediate

Move your analytical skills to a higher level

Do you have an analytical mindset, are you working with data a lot at work, be it as part of some SQL database, in Excel or in Tableau, and would you like to move further and start working on more complex projects? You may have completed a number of on-line DataCamp or Udemy training sessions in which you have learnt many tools, such as R or Python, and you may have some experience with applying various ML algorithms to simple illustrative tasks. Nevertheless, approaching and delivering a solution for some real business issue is hard to imagine for you? If most of the above applies to you, Data Science Academy is right for you.

Within eight weeks, we will solve together a real business problem relating to credit risk and provide an end-to-end solution. You will learn to use SQL, Python, R and Tableau to solve all partial tasks typically solved by a data scientist in individual project phases – from understanding the business issue, understanding data, data preparation and modelling to the assessment of model quality and application to production.

As opposed to online courses, which concentrate on providing partial knowledge and skills, in DS Academy you will learn applying the knowledge in a single workflow to solve a real problem. Continuous **individual consulting with lecturers** is a great advantage of DS Academy. Thanks to the project and personal approach, you will be able to take part in data and analytical projects in your own firm immediately after the completion of the Academy, or even in the course of your studies.

The content of our training:

- At the beginning, we will define project goals and criteria for success/failure evaluation based on the business information.
- In the SQL database, we will prepare data and calculate predictors.
- We will show you how to use indexing, SQL coding standards and best practices to achieve the optimal inquiring in terms of processing requirements and to understand our SQL code within a year or two whenever we will have to get back to it.
- We will use the libraries of R (*dplyr, tidyr, ggplot2, DataExplorer*) and Python (*pandas, matplotlib, seaborn*) for exploratory data analysis.
- Using various libraries in R and Python (eg *stringr, forcats, lubridate, recipes, caret, smbinning, numpy, pandas, scikit-learn)* we will create new, potentially useful variables as part of feature engineering and perform the transformations needed on all data, such as the imputation of missing values, binning, one hot encoding, dimension reduction, normalisation, etc.
- Using the *scikit-learn* ML library in Python, we will try to identify groups of customers in our data with a similar profile that the marketing department will be able to address with a targeted financial services offering.
- Using the *caret* ML library in R, we will try to identify the key predictors of risk loans and forecast capability to repay a loan for individual loan applicants.
- In R and in Python, we will learn to use cross validation to fine-tune hyperparameters of the used models and resampling to treat situations when there is a major disproportion in the representatives of individual predicted categories.
- We will show you how to assess the quality of the data models we have created in terms of statistics and business.
- We will try and improve our performance metrics using an ensemble model combining the outputs of a large number of various types of models.

- We will put our models in production where they will be able to assist business in making better decisions and we will create an easily reproducible report on the project for our colleagues in *R Markdown*, or *Jupyter Notebook*.
- We will use Tableau to visualise the credit data that will help business people get a better insight into the status of some business relevant metrics and apply them in making your own exploratory analysis.

(Some of) your training takeaways:

- You will obtain a number of key skills and knowledge to be able to realise end-to-end data and analytical projects of medium complexity.
- In the SQL database, you will be able to prepare and pre-process data you will need in order to model the defined business issue.
- Using Python and R, you will learn to quickly orientate yourselves in data, evaluate its relevance for the defined business issue and make the data transformations needed.
- Using the same tools, you will be able to apply a cluster analysis and selected binar classifiers to the relevant business issues and you will be able to assess the quality of the resulting data models from the statistical and business perspectives.
- You will know the processes to fine-tune hyperparameters of data models and to treat a significant disproportion in predicted categories and you will learn to use the ensembling principle in data modelling.
- You will know how to deploy the created models in business practice and how to apply them to new data. You will also be able to create an easily reproducible report on your project.
- You will know the basic principles of creating useful, practically applicable and aesthetically attractive data visualisations in Tableau.

What you should know:

D-S Academy is primarily intended for those willing to move from the beginner to the intermediate level in the data science. To complete the course, you need the basic knowledge of SQL as the course only focuses on more advanced instructions. You can test your knowledge in a **simple test**. We will be happy to provide you with recommendation on how to complete your knowledge before the course starts.

Lecturers:

Luděk Stehlík, Advanced Analytics Deloitte

Expert on using data and data analysis in business process optimisation.

Luděk has more than 10 years of work experience in positions in which he uses his data science knowledge to help clients from various industries (retail, pharma, telco, manufacturing) achieve their business and strategic goals through more effective managing of their resources, from employees and customers to stock management. As part of his academic career, Luděk is engaged in the teaching and/or research in the area of cognitive psychology, statistical considerations and mathematic modelling of the cognitive functions related to language understanding and considerations.

Eliška Valterová, Advanced Analytics Deloitte

Leading expert on data visualisation, dashboard creation and business intelligence. She has more than six years of experience in e-commerce, marketing and finance projects.
