
Introduction to Machine Learning, AI & Data Science with Azure ML and SQL Server

Varighed: 2 Days Kursus Kode: M030

Beskrivelse:

This live classroom course is new for 2018! It focuses on the newest technologies of Microsoft Machine Learning Server and SQL Server 2017.

Målgruppe:

Analysts, budding data scientists, database and BI developers, programmers, power users, DBAs, predictive modellers, forecasters, consultants.

Agenda:

- **Why attend this class?**
 - Because of Rafal's 10+ years of real-world machine learning experience.
 - You will not only learn all the concepts and tools that you need to know from a great teacher who has trained almost 500 data scientists world-wide, a highly-respected presenter, capable of holding your attention, but, above all, from a practitioner of machine learning. Rafal Lukawiecki has been delivering ML, data mining, and data science projects for customers in retail, banking, entertainment, healthcare, manufacturing, education, and government sectors for over ten years.
-

Indhold:

To deliver the best possible training we follow the industry. The agenda and course content are subject to continuous improvement and revision without further notice.

Machine Learning Fundamentals.

We begin with a thorough introduction of all of the key concepts, terminology, components, and tools. Topics include:

- Machine learning vs. data mining vs. artificial intelligence.
- Tool landscape: open source R vs. Microsoft R, Python, SQL Server, ML Server, Azure ML.
- Teamwork.
- Algorithms.

There are hundreds of machine learning algorithms, yet they belong to just a dozen of groups, of which 5 are in very common use. We will introduce those algorithm classes, and we will discuss some of the most often used examples in each class, while explaining which technology tools (Azure ML, SQL, or R) provide their most convenient implementation. You will also learn how to find more algorithms on the Internet and how to figure out if they are any good for real use. Topics include:

- What do algorithms do?
- Algorithm classes in R, Python, ML Server, Azure ML, and SSAS Data Mining.
- Supervised vs. unsupervised learning.
- Classifiers.
- Clustering.
- Regressions.
- Similarity Matching.
- Recommenders.

Data

Machine learning requires you to prepare your data into a rather unique, flat, denormalised format. While features (inputs) are always necessary, and you may need to engineer thousands of them, we do not need labels (predictive outputs) in all cases. Topics include:

- Cases, observations, signatures.
- Inputs and outputs, features, labels, regressors, independent and dependent variables, factors.
- Data formats, discretization/quantizing vs. continuous.
- Indicator columns.
- Feature engineering.
- Azure ML data preparation and manipulation modules.
- Moving data around and its storage, SQL vs. NoSQL, files, data lakes, BLOBs, and Hadoop.

Process of Data Science

The process consists of problem formulation, data preparation, modelling, validation, and deployment—in an iterative fashion. You will briefly learn about the CRISP-DM industry-standard approach but the key subject of this module will teach you how to apply the scientific method of reasoning to solve real-world business problems with machine learning and statistics. Notably, you will learn how to start projects by expressing needs as hypotheses, and how to test them. Topics include:

- CRISP-DM.
- Stating business question in data science term.
- Hypothesis testing and experiments.
- Student's t-test.
- Pearson chi-squared test.
- Iterative hypothesis refinement.

Introduction to Model Building

At the heart of every project we build machine learning models! The process is simple and it follows a well-trodden path. In this module you will build your first decision tree and get it ready for validation in the next module. Topics include:

- Connecting to data.
- Splitting data to create a holdout.
- Training a decision tree.
- Scoring the holdout.
- Plotting accuracy.

Introduction to Model Validation

The most important aspect of any data science, artificial intelligence, and machine learning project is the iterative validation and improvement of the models. Without validation, your models cannot be reliably used. There are several tests of model validity, most importantly those that check accuracy and reliability. Topics include:

- Testing accuracy.
- False positives vs. false negatives.
- Classification (confusion) matrix.
- Precision and recall.
- Balancing precision with recall vs. business goals and constraints.
- Introduction to lift charts and ROC curves.
- Testing reliability.

Testing usefulness

Flere Informationer:

For yderligere informationer eller booking af kursus, kontakt os på tlf.nr.: 44 88 18 00

training@globalknowledge.dk

www.globalknowledge.com/da-dk/

Global Knowledge, Stamholmen 110, 2650 Hvidovre