

Data
mining

Algorithms

**MACHINE
LEARNING**

Artificial
intelligence

Machine Learning & Advanced Analytics

Learn the tools and techniques to gain new insights

Two day seminar by
Mike Ferguson

AdeptEvents

- Tools and techniques for discovery, analysis and visualisation of multi-structured data
- Develop analytical models using supervised and unsupervised machine learning
- Develop machine learning models at scale with Apache Spark and Hadoop
- Text and sentiment analysis, Clickstream analysis, Graph analysis
- Analyse fast data in real-time using streaming analytics
- Use Self-Service BI reports to leverage machine learning and advanced analytics quickly and easily

VENUE
Amersfoort/Utrecht Area

TIME
9:30 – 17:00 hours

REGISTRATION
www.adeptevents.nl





Machine Learning & Advanced Analytics

Learn the tools and techniques to gain new insights

Today, with most people connected to the Internet, the power of the customer is almost limitless. The Internet has given them freedom to choose in a way that business could never have imagined. They can browse your competitors' web sites with ease. They can compare prices, they can view sentiment about your business, and they can switch loyalty in a single click any time anywhere all from a mobile device.

In addition, the emergence of social media sites means that customers also have a voice. They can express opinion and sentiment about products and brands on Twitter Facebook, and review web sites and create social networks by attracting followers, and following others. For many CEOs, customer retention, loyalty, service and growth are top of their agenda. In addition improving operational effectiveness is also high on their priority list. The only way they can achieve this is to acquire more data. CMOs also want access to new data to enrich what they already know about customers. New data is needed to provide insight on customer on-line behaviour for better segmentation and to understand the value of a customers' social network and not just the customer. In addition, COOs want more data to become more effective in operations. Instrumentation is therefore being added so that operations can capture new data. With so much demand we are now in an era where data has never before been so important to business in helping to create competitive advantage.

This new 2-day seminar looks at the need to capture new data sources to add to what we already know and use machine learning to automatically discover, profile and catalog what is in these data sources. It then looks at how machine learning and advanced analytical techniques such as text analyses, sentiment analysis, graph and streaming analytics can be used at scale on Big data to provide new insight that helps foster growth, reduce costs and improve effectiveness for competitive advantage.



Learning objectives

Attendees to this seminar will learn:

- How data and analytical characteristics can dictate the approach taken and tools needed to conduct exploratory analytics
- How to develop analytical models using supervised and unsupervised machine learning
- How to develop machine learning models at scale on Apache Spark and Hadoop
- Tools for building machine learning models



- Tools and techniques for discovery, analysis and visualisation of multi-structured data
- Text and sentiment analysis
- Scaling text analysis to run on Hadoop, MapReduce and Spark
- Clickstream analysis
- Graph analysis – 4 graph analytical techniques to identify shortest path, analyse connectivity, identify communities, determine influencers and important people in social networks etc.
- Scale graph analysis on Apache Spark GraphX
- Analyse fast data in real-time using streaming analytics
- Leverage machine learning and advanced analytics quickly and easily from self-service BI reports and dashboards for access over the web and on mobile device.

Target audience

Business Analysts, data scientists, BI Managers, data warehousing professionals, enterprise architects, data architects, CIO's, IT Managers



MIKE FERGUSON

Mike Ferguson is Managing Director of Intelligent Business Strategies Limited. As an analyst and consultant he specialises in business intelligence / analytics, data management, big data and enterprise business integration. With over 34 years of IT experience, Mike has consulted for dozens of companies on business intelligence strategy, technology selection, enterprise architecture, and data management. He has spoken at events all over the world and written numerous articles. Formerly he was a principal and co-founder of Codd and Date Europe Limited – the inventors of the Relational Model, a Chief Architect at Teradata on the Teradata DBMS and European Managing Director of

Database Associates. He teaches popular master classes in Big Data, New Technologies for Data Warehousing and BI, Operational BI, Enterprise Data Governance, Master Data Management, Data Integration and Enterprise Architecture.

He teaches popular master classes in Big Data, Predictive and Advanced Analytics, Fast Data and Real-time Analytics, Enterprise Data Governance, Master Data Management, Data Virtualisation, Building an Enterprise Data Lake and Enterprise Architecture. Mike was one of the speakers at a previous edition of our yearly conference, the Data Warehousing & Business Intelligence Summit.



Course description

MODULE 1: AN INTRODUCTION TO DATA EXPLORATION, DISCOVERY AND VISUALISATION

This session introduces the relatively area of data discovery and visualisation and looks at why businesses now need.

- New data sources – Structured versus multi-structured data
- What are the different analytical workloads that dictate the need for data discovery and visualisation?
- The data discovery and visualisation process
- What is exploratory analysis?
- What is Data Discovery and Visualisation?
- Why do businesses need this new capability? – Example use cases
- Skills required for Data Discovery and Visualisation
- Types of Data Discovery and Visualisation tools?

MODULE 2: GETTING STARTED WITH PREDICTIVE ANALYTICS AND MACHINE LEARNING

As we move into the era of smart business, looking back in time is not enough to make good decisions. Companies have to also model the future to forecast and predict so that they can anticipate problems and act in a timely manner to compete. Predictive analytics is a therefore a key part of any BI initiative and should be integrated into analysis, reporting and dashboards. This session introduces predictive analytics and how shows how it can be used in analysis and in business optimisation.

- What is predictive analytics?
- Technologies and methodologies developing predictive analytical models
- Using supervised learning to develop predictive models for automatic classification
- Popular predictive algorithms, e.g. Linear regression, decision trees, random forest, neural networks
- Implementing in-Hadoop, in-memory analytics using Apache Spark and SAS Viya & LASR server
 - Data Science Workbooks using Jupyter Apache Zeppelin and Databricks Cloud
 - Accessing data in HDFS using SQL to build models
 - Accessing in Hadoop and Spark machine learning algorithms from data mining tools

- Deploying predictive analytical models in analytical databases and in Hadoop
- Integrating predictive analytics with event stream processing for automated analysis of fast data in every-day business operations
- Clustering data using unsupervised learning algorithms

MODULE 3: ADVANCED ANALYTICS FOR MULTI-STRUCTURED DATA

This session looks at emerging analytical technologies for multi-structured data and explores how you can use them to improve business insight. Not all analytical projects are implemented using relational database technology, especially when it comes to very large data volumes with unstructured content, semi-structured JSON or XML data, sensor data, and clickstreams. This session looks at the emergence of advanced analytics using Big Data NoSQL Platforms like Hadoop. It looks at the approaches to analysing complex unstructured and social content and the challenges of creating valuable business insight from multiple sources of unstructured content.

- Techniques for producing insight from unstructured content
- Tools and techniques for analysing text
- Understanding the 'voice of the customer' using sentiment analytics on email and social media data
- Clickstream analysis
- Graph analysis
 - Path analytics
 - Connectivity analysis
 - Community analysis
 - Centrality analysis
 - Finding Influencers in social networks
 - Calculating follower susceptibility to be influenced
- Streaming analytics
 - What is data-in-motion
 - Use cases for streaming data
 - Time series analysis and streaming data
 - Tools for managing streaming ingest, e.g. StreamSets, Hortonworks Data Flow
 - Data Science tools



- g. Cloudera Data Science Workbench, IBM Data Science Experience
 - Deep Learning
- Google Tensorflow, deepsense.io, Microsoft Cognitive Toolkit (CNTK)
 - Artificial Intelligence
- Open source streaming engines – Apache Apex, Apache Storm, Apache Spark, Apache Flink, Google Data Flow
- Commercial streaming analytics products
- Developing streaming analytics applications with no programming
- Modernising your architecture to accommodate streaming data
- Future proofing your architecture
 - IBM Watson

MODULE 4: SEARCH, BI & BIG DATA

This session will examine the growing role of search in an analytical environment both as an information consumer tool for self-service BI and as a way of analysing both structured and unstructured data. Search has been incorporated into BI tools for some time, but with the emergence of Big Data as a platform for analysing unstructured information, it is taking on a major new role. Search is a simple mechanism that is familiar to most people, and opening up the interactive use of BI via search can have enormous business benefits. Search can be used to grow the use of BI to a much wider group of users and also provide a way to extract additional insight from unstructured content.

Topics that will be covered include:

- Why Search and BI?
- The growing importance of analysing unstructured content
- The implications of Big Data on search and BI
- Creating search indexes on multi-structured data
- Building dashboards and reports on top of search engine indexed content
- Using search to analyse multi-structured data
- The integration of search with traditional BI platforms
- Using Search to find BI content and metrics
- Guided analysis using multi-faceted search
- The search based analytical tools marketplace: Apache Solr (Lucene), Attivio, Cloudera Search, Connexica, HP IDOL, IBI WebFocus Magnify, IBM Watson Explorer, Microsoft, Quid, SAP Lumira, Splunk, Thoughtspot

MODULE 5: DEPLOYING AND USING SELF-SERVICE DATA DISCOVERY AND VISUALISATION TOOLS

Self-service data discovery and visualisation tools are frequently sold into business departments so that local business analysts can start building their own BI applications without having to wait for IT. These tools offer the attraction of agile development and much faster time to value. When business areas buy them it often means that development starts without any IT guidance and quickly spreads to other parts of the business with little thought for integration or re-use. The result is that inconsistency and chaos can quickly set in. This session looks at best practices in deploying these tools and how to maximise business benefit through re-use and integration with predictive and advanced analytics deployed in-database, in-Hadoop, in-Spark and in-streaming analytics platforms to leverage analytics at scale. It also looks at newly emerging OLAP on Hadoop to enable scalable multi-dimensional analysis.

- The Data Discovery and Visualisation tools marketplace – Arcadia Data, Information Builders WebFOCUS, Microsoft PowerBI, MicroStrategy Visual Intelligence, Qlik Sense, Tableau, SAS Visual Analytics, SAP Lumira, Zoomdata, etc.
- Key features of self-service data discovery and visualisation tools
- Requirements and best practices for successful self-service BI
- Self-service BI tool access to Big Data via SQL on Hadoop
- Simplifying Self-service BI tool data access to multiple data stores via data virtualisation – logical data warehouse
- Accessing in-database, in-Hadoop and in-Spark predictive analytics from self-service BI tools and spread sheets
- Accessing streaming data and real-time analytics from self-service BI tools and spreadsheets
- Integration with advanced analytics in the cloud and on-premises
- Scalable OLAP – Multi-dimensional analysis on Hadoop
- OLAP on Hadoop using AtScale, Kyvos Insights and Apache Kylin

INFORMATION

DATE AND TIME

The workshop will take place once or twice a year with the exact date and time available on our website. The programme starts at 9:30 am and ends at 5:00 pm on both days. Registration commences at 8.30 am and we recommend that you arrive early.

VENUE

Adept Events works with several venues in and near Amersfoort and Utrecht. We strive to provide you with the location details as soon as possible. The exact venue will be on our website and in the confirmation e-mail that you will receive one week prior to the event. Always check our website prior to your departure to ensure you have the exact location and directions.

HOW TO REGISTER

Please register online at www.adeptevents.nl. For registering by print, please scan the completed registration form and send this or your Purchase Order to customerservice@adeptevents.nl. We will confirm your registration and invoice your company by e-mail therefore please do not omit your e-mail address when registering.

REGISTRATION FEE

Taking part in this two-day workshop will only cost 1305 Euro when registering 30 days beforehand and 1450 Euro per person afterwards (excl. 21% Dutch VAT). This also covers documentation, lunch, tea/coffee.



Members of the DAMA are eligible for 10 percent discount on the registration fee.

In completing your registration form you declare that you agree with our Terms and Conditions.

TEAM DISCOUNTS

Discounts are available for group bookings of two or more delegates representing the same organization made at the same time. Ten percent off when registering 2 – 3 delegates and fifteen percent off for all delegates when registering four or more delegates (all delegates must be listed on the same invoice).

This cannot be used in conjunction with other discounts. All prices are VAT excluded.

PAYMENT

Full payment is due prior to the workshop. An invoice will be sent to you containing our full bank details including BIC and IBAN. Your payment should always include the invoice number

as well as the name of your company and the delegate name. For Credit Card payment please contact our office by e-mail mentioning your phone number so that we can obtain your credit card information.

CANCELLATION POLICY

Cancellations must be received in writing at least three weeks before the commencement of the workshop and will be subject to a € 75,- administration fee. It is regretted that cancellations received within three weeks of the workshop date will be liable for the full workshop fee. Substitutions can be made at any time and at no extra charge.

CANCELLATION LIABILITY

In the unlikely event of cancellation of the workshop for any reason, Adept Events' liability is limited to the return of the registration fee only. Adept Events will not reimburse delegates for any travel or hotel cancellation fees or penalties. It may be necessary, for reasons beyond the control of Adept Events, to change the content, timings, speakers, date and venue of the workshop.

MORE INFORMATION



+31(0)172 742680



<http://www.adeptevents.nl/paa>



seminars@adeptevents.nl



@AdeptEventsNL / <https://twitter.com/AdeptEventsNL>



<http://www.linkedin.com/company/adept-events>



<https://google.com/+AdeptEventsNL>



<https://www.facebook.com/AdeptEventsNL>



Visit our Business Intelligence and Datawarehousing website www.biplatform.nl and download the App



Visit our website on Software Engineering, www.release.nl and download the App

In-House Training

Would you like to run this course in-company for a group of persons? We can provide a quote for running an in-house course, if you offer the following details. Estimated number of delegates, location (town, country), number of days required (if different from the public course) and the preferred date/period (month). Please find more info on the **In-house page on our website**.