rapidminer

Text, Web, and Social Media Mining

What is this training about?

This Text, Web, and Social Media Mining course is a 2 days course and is an introduction into knowledge discovery using unstructured data like text documents, web and social media contents. It focuses on the necessary preprocessing steps and the most successful methods for automatic text classification including: Naive Bayes, Support Vector Machines (SVM), and clustering. Hands-on exercises will be carried out using RapidMiner Studio, hence there will be an introduction section on the first day to help participants familiarize with the environment.

Upon completion of this course, participants will have a solid understanding of typical text mining workflows and be able to identify techniques for processing unstructured data, apply different statistical text-processing methods, and perform content classification & clustering.

Practical exercises during the course prepare students to take the knowledge gained and apply to their own text mining challenges. Examples include: adaptive personal news filtering, patent clustering, sentiment analysis of text documents like news, web pages, blogs, e-mail, or PDF documents. Since the class labs are hands-on and performed on the participants' personal laptops, students will take actual classwork home with them, which will provide a jumpstart to the real world.

Course Objectives

After the training, students will have the ability to:
- Identify techniques for processing unstructured data
- Transform textual data into a structured format
- Apply different statistical text-processing methods
- Perform text classification and text clustering
- Work on popular tasks like sentiment analysis or opinion mining

About Go Training

Go Training applies effective pedagogical methodologies that demonstrate case studies and hands-on practical skills, in addition to explaining clearly how things work in principle. Every course that we conduct is delivered by a subject matter expert who holds the academic qualification and working experience in that specialization. On the days when they are not teaching, our trainers work on consultancy projects and technical deliveries. Their work has received numerous recognition and awards in the industry. Our team of trainers has been invited as keynote speakers at numerous international conferences, and as principal consultants for various industries.
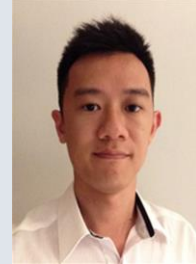
Date: 5 – 6 December 2016
           (Monday - Tuesday)
Time: 0900 - 1700
Venue: Kuala Lumpur, Malaysia.

HRDF Claimable

Course Outline

- **Introduction to RapidMiner Studio**

- **Loading of Texts**
  - Loading from Flat Files
  - Loading from Data Sets
  - Loading from Databases
  - Loading from Web Sources (e.g. Web Pages, Twitter)

- **Concepts**
  - Documents
  - Tokens

- **Visualization**
  - Visualizing Documents and Tokens
  - High Dimensional Visualizations for Transformed Documents

- **Handling Unstructured Data**
  - Preprocessing of Textual Data
  - Tokenizing
  - Stemming
  - Filtering of Tokens
  - Term Frequencies
  - Document Frequencies
  - TF-IDF

- **Advanced Modeling**
  - Methods for High Dimensional Data
  - Support Vector Machines
  - Text Classification
  - Text Clustering

- **Web Mining**
  - Fetching data from Twitter
  - Crawling the Web
  - Extracting Information from Web Sites
  - Transforming Web Sites to Documents

About the Instructor

Jaden graduated from Wichita State University (USA) in 2010, with a Bachelor's Degree in Aerospace Engineering. His primary design work in college was the design-build-fly of a radio-controlled aircraft. Other notable efforts include the investigation of the aerodynamic behaviors of an automobile and the Global Design Challenge, a collaborative airframe design competition sponsored by Spirit Aerosystems. The close-to-industry project management and collaboration settings in these projects played a part in preparing Jaden well for a jump-start in his career.

Jaden began his career as an Application Engineer in TechSource Systems, the SEA sole distributor of MATLAB. As a consultant, he focused on technical computing areas such as data analytics, artificial intelligence, image processing, application deployment, etc. During this time, Jaden has supported many engineering and research professionals and assisted them in creating more efficient workflows and obtaining better results.

In 2013, Jaden transitioned to the MATLAB technical sales role, taking charge of commercial accounts in Malaysia and Indonesia. He works with decision-making executives and has established a thorough understanding on business requirements that necessitates adoption of new technologies. In 2014, Jaden has co-founded Quandatics, which focuses on developing and providing advanced analytics services and solutions. Jaden is now heading the business development and technical marketing teams in Quandatics and his clientele-base includes MNCs, GLCs, government agencies, and higher-education institutions.

Go Training
wholly owned by iRadar Sdn Bhd
HRDF Approved Training Provider (Category A)

No. 36, Jalan IMJ 1, Taman Industri Malim
Jaya, 75250 Melaka, Malaysia.
t   +606 336 6016
f   +606 252 3059
w www.gotraining.com.my
[f] fb.com/gotraining.com.my
[in] linkedin.com/company/gotraining

To register, please contact:
m +6010 663 1852
e   yiwei@gotraining.com.my