

Digital Skills
Training
Programs at
Knowledge
Technology

INTRODUCTION
TO BIG DATA
ANALYTICS
AND
VISUALIZATION

- The Big Data Analytics and Visualization course is a rigorous 10-Days certificate course designed for working professionals to develop practical knowledge and skills and accelerate entry into data science careers.
- Covers three five areas
 - Data Analysis with R
 - R Graphs
 - Learning Data Mining with R
 - Mastering R for Quantitative Finance
 - Machine Learning with R



KNOWLEDGE TECHNOLOGY
R E S E A R C H U N I T

Course Title	Introduction to Big Data Analytics and Visualization
Duration	10 Days
Trainer	Assoc. Prof. Dr. Rayner Alfred
Cost	Email ralfred121@gmail.com or call 013-881-9966 for quotations
Max Participants	25

SYNOPSIS

This course comprises of five main modules. It begins by looking at the **Data Analysis with R** module. This module will help you navigate the R environment. You'll gain a thorough understanding of statistical reasoning and sampling. Finally, you'll be able to put best practices into effect to make your job easier and facilitate reproducibility.

The second place to explore is **R Graphs**. This module will help you leverage powerful default R graphics and utilize advanced graphics systems such as lattice and ggplot2, the grammar of graphics. Through inspecting large datasets using table plot and stunning three-dimensional visualizations, you will know how to produce, customize, and publish advanced visualizations using this popular, and powerful, framework.

With the third module, **Learning Data Mining with R**, you will learn how to manipulate data with R using code snippets and be introduced to mining frequent patterns, association, and correlations while working with R programs. Discover how to write code for various predication models, stream data, and time-series data. You will also be introduced to solutions written in R based on RHadoop projects. You will finish this module feeling confident in your ability to know which data mining algorithm to apply in any situation.

The **Mastering R for Quantitative Finance** module pragmatically introduces both the quantitative finance concepts and their modeling in R, enabling you to build a tailor-made trading system on your own. By the end of the module, you will be well versed with various financial techniques using R and will be able to place good bets while making financial decisions.

Finally, we'll look at the **Machine Learning with R module**. With this module, you'll discover all the analytical tools you need to gain insights from complex data and learn how to choose the correct algorithm for your specific needs. Through full engagement with the sort of real-world problems data-wrangers face, you'll learn to apply machine learning methods to deal with common tasks, including classification, prediction, forecasting, market analysis, and clustering.

LEARNING OUTCOMES

What participants will gain at the end of the course

- Gain a thorough understanding of statistical reasoning and sampling, learn Bayesian methods for estimating parameters, and identify and manage problematic data points.
- Learn to create diverse types of bar charts, produce and customize density plots and histograms with lattice and ggplot2, and construct various types of three-dimensional, plots using three-dimensional visualizations.
- Learn to manipulate data with R using code snippets, develop best practices in the fields of graph mining and network analysis, and know spatial data mining, text mining, and web data mining.
- Learn to analyse high frequency financial data, build, calibrate, test, and implement theoretical models; create a winning arbitrage, speculation, or hedging strategy customized to your risk preferences.
- Learn to build common machine learning algorithms with real-world data science applications; classify your data with Bayesian and nearest neighbour methods; and evaluate and improve the performance of machine learning models.

JUSTIFICATION TO LEARN BIG DATA ANALYTICS AND VISUALIZATION

Big data analytics examines large amounts of data to uncover hidden patterns, correlations and other insights. With today's technology, it's possible to analyze your data and get answers from it almost immediately – an effort that's slower and less efficient with more traditional business intelligence solutions.

The concept of big data has been around for years; most organizations now understand that if they capture all the data that streams into their businesses, they can apply analytics and get significant value from it. But even in the 1950s, decades before anyone uttered the term "big data," businesses were using basic analytics (essentially numbers in a spreadsheet that were manually examined) to uncover insights and trends.

The new benefits that big data analytics brings to the table, however, are speed and efficiency. Whereas a few years ago a business would have gathered information, run analytics and unearthed information that could be used for future decisions, today that business can identify insights for immediate decisions. The ability to work faster – and stay agile – gives organizations a competitive edge they didn't have before.

Big data analytics helps organizations harness their data and use it to identify new opportunities. That, in turn, leads to smarter business moves, more efficient operations, higher profits and happier customers.

TOPICS LIST

- [1] Data Analysis with R
- [2] R Graphs
- [3] Learning Data Mining with R
- [4] Mastering R for Quantitative Finance
- [5] Machine Learning with R

COURSE SYLLABUS (10 DAYS)

DAY	TOPICS COVERED	TIME
One and Two	<p>MODULE 1: DATA ANALYSIS WITH R</p> <ul style="list-style-type: none"> ➤ Basic of R and statistical reasoning ➤ Describing relationships ➤ Testing Hypotheses ➤ Bayesian methods <p>Participants will gain a thorough understanding of statistical reasoning and sampling, learn Bayesian methods for estimating parameters, and identify and manage problematic data points.</p>	8am – 5pm
Three and Four	<p>MODULE 2: R GRAPHS</p> <ul style="list-style-type: none"> ➤ Using basic graph functions ➤ Creating scatter plots ➤ Creating different charts ➤ Creating maps <p>Participants will learn to create diverse types of bar charts, produce and customize density plots and histograms with lattice and ggplot2, and construct various types of three-dimensional plots using three-dimensional visualizations.</p>	8am – 5pm
Five and Six	<p>MODULE 3: LEARNING DATA MINING WITH R</p> <ul style="list-style-type: none"> ➤ Market basket analysis ➤ Classification algorithms ➤ Cluster analysis ➤ Graph mining and network analysis ➤ Mining text and web data <p>Participants will learn to manipulate data with R using code snippets, develop best practices in the fields of graph mining and network analysis, and know spatial data mining, text mining, and web data mining.</p>	8am – 5pm
Seven and Eight	<p>MODULE 4: MASTERING R FOR QUANTITATIVE FINANCE</p> <ul style="list-style-type: none"> ➤ Factor models ➤ Big data analysis ➤ Interest rate derivations and models ➤ Technical analysis ➤ Neural networks ➤ Log-optimal portfolios ➤ Systemic risks <p>In this module, participants will learn to analyse high frequency financial data, build, calibrate, test, and implement theoretical models; create a winning arbitrage, speculation, or hedging strategy customized to your risk preferences.</p>	8am – 5pm
Nine and Ten	<p>MODULE 5: MACHINE LEARNING WITH R</p> <ul style="list-style-type: none"> ➤ Classification using Nearest Neighbours ➤ Classification using Naive Bayes, ➤ Forecasting numeric data - regression methods ➤ Neural networks and support vector machines ➤ Evaluating and improving model performance <p>In this module, participants will learn to build common machine learning algorithms with real-world data science applications; classify your data with Bayesian and nearest neighbour methods; and evaluate and improve the performance of machine learning models.</p>	8am – 5pm

TRAINER'S BIOGRAPHIES



RAYNER ALFRED

ASSOCIATE PROFESSOR OF COMPUTER SCIENCE

Certified IBM DB2 Academic Associate, Certified Tester Foundation Level (CTFL)

AREAS OF SPECILIZATION: Advanced Machine Intelligence, Data Analytics, Data Mining, Information Retrieval, Artificial Intelligence, Machine Learning, Knowledge Discovery

ADDRESS: Knowledge Technology Research Group, Faculty of Computing and Informatics, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah.

CONTACT: Mobile: 6013-881-9966, eMail: ralfred@ums.edu.my

Rayner Alfred is an Associate Professor of Computer Science at the Faculty of Computing and Informatics, Universiti Malaysia Sabah in Malaysia that focuses on Data Science and Software Engineering programmes. He leads and defines projects around knowledge discovery, information retrieval and machine learning that focuses on building smarter mechanism that enables knowledge discovery in structured and unstructured data. His work addresses the challenges related to big data problem: How can we create and apply smarter collaborative knowledge discovery and machine learning technologies that bridge the structured and unstructured data mining and cope with the big data problem.

Rayner completed his PhD in 2008 looking at intelligent techniques using machine learning to model and optimize the dynamic and distributed processes of knowledge discovery for structured and unstructured data. He holds a PhD degree in Computer Science from York University (United Kingdom), a master's degree in computer science from Western Michigan University, Kalamazoo (USA) and a Computer Science degree from Polytechnic University of Brooklyn, New York (USA) where he was the recipient of the *Myron M. Rosenthal Academic Achievement Award* for the outstanding academic achievement in Computer Science in 1994. He has authored and co-authored more than 100 journals/book chapters and conference papers, editorials, and served on the program and organizing committees of numerous national and international conferences and workshops.

Rayner is currently a member of IEEE, a Certified Software Tester (CTFL) from the International Software Testing Qualifications Board (*ISTQB*), and a certified IBM DB2 Academic Associate (IBM DB2 AA). He leads the Advanced Machine Intelligence (AMI) research group in UMS and he has led several projects related to knowledge discovery and machine learning on Big Data. Rayner is also the recipient of the Research Fellow at Japan Advanced Institute of Science and Technology (JAIST), Japan. He is also the recipient of multiple GOLD awards at national and international research exhibitions in Data Mining and Machine Learning based solutions (Face Recognition and Knowledge Discovery), that include International Trade Fair Ideas in Nuremberg, Germany (iNEA2018) International Invention Innovation Competition in Toronto, Canada (iCAN 2018), Seoul International Invention Exhibition in Seoul, Korea (SIIF 2010). He has secured RM6,931.433.00 worth of project grants. Some of his project researches include biometric authentication using face recognition, building security based on plate number recognition using deep learning, sentiment analysis for Malay and English in measuring public opinion, news-news correlation trending, machine learning algorithm-based solution for predicting diseases in health care, smart monitoring using an ensemble based face recognition system and smart information management and retrieval to name a few. Some of the completed projects include Semantic Multi-Agent For Knowledge Sharing, developing an Evolutionary-Based Ensemble Classifier Framework for Learning Big Relational Data, developing a genetic-based hierarchical agglomerative clustering technique for parallel clustering of bilingual corpora based on reduced terms, enhancing document Clustering By Integrating Semantic Background Knowledge and Syntactic

Features Into the BOW Representation and the fundamental Study on an Evolutionary Based Features Construction Methods for Data Summarization Approach to Predict Survival Factors of Coral Reefs in Malaysia, to name a few and also infrared face recognition based on ensemble approach.