

Digital Skills  
Training  
Programs at  
Knowledge  
Technology

**BUSINESS  
INTELLIGENCE  
(BI)**

- This 5-days course provides data-driven and analytically focused approaches to help participants answer business questions in operations, marketing, and finance—a diverse perspective.
- Participants will also see how asking the right type of questions and developing the stories and visualizations helps them connect the dots between the data and the business.



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

<b>Course Title</b>	Introduction to Business Intelligence
<b>Duration</b>	5 Days
<b>Trainer</b>	Assoc. Prof. Dr. Rayner Alfred
<b>Max Participants</b>	25
<b>SYNOPSIS</b>	
<p>Introduction to Business Intelligence gives participants a way to explore the world of business intelligence through the eyes of an analyst working in a successful and growing startup company. Participants will learn R through use cases supporting different business functions. This course also provides data-driven and analytically focused approaches to help participants answer business questions in operations, marketing, and finance—a diverse perspective. Participants will also see how asking the right type of questions and developing the stories and visualizations helps them connect the dots between the data and the business.</p>	
<b>LEARNING OUTCOMES</b>	
<p>At the end of the course, participants will be able to</p> <ul style="list-style-type: none"> <li>➤ Perform data preparation that covers data extraction, data transformation and data loading, managing and understanding data and data cleaning</li> <li>➤ Perform data analysis that covers Linear Regression for Business, Cluster Analysis, and Time Series Analysis</li> <li>➤ Perform data visualization that cover the presentation of results and visualizing the Data's Story</li> </ul>	

## JUSTIFICATION TO LEARN BUSINESS INTELLIGENCE

Business intelligence allows businesses to make better decisions by accessing big data, and cloud computing has leveled the playing field. Business intelligence allows small, medium and large businesses to make better decisions by accessing big data. Even small companies that don't accumulate a large volume of data can derive considerable benefits from better analytics. Initially only large businesses could afford the cost of using Business Intelligence software because of the cost of the software itself and the infrastructure to run it on. However, recent technological innovations like Software as a Service (SaaS) on a cloud computing platform have leveled the playing field.

Business Intelligence (BI) is not a single piece of software or even a suite of software to crunch big data. Instead, it is an umbrella term that includes best business practices, software, infrastructure, and any other tools to optimize decision making and enhance performance. BI is about information analysis. A database pulls together information from the different parts of your business, then applications convert that raw data into reports, charts, and other analytical tools to provide insight. BI can be used in any number of different ways:

- It can be used to look at any aspect of your business.
- It can be used to see how well your sales efforts are going.
- It can be used to assess the efficiency of your manufacturing processes.
- It can be used to measure staff performance.

## TOPICS LIST

- [1] Extract Transform and Load (ETL)
- [2] Data Cleaning
- [3] Exploratory Data Analysis
- [4] Linear Regression for Business
- [5] Data Mining with Cluster Analysis
- [6] Time Series Analysis
- [7] Visualizing the Data's Story
- [8] Web Dashboards with Shiny
- [9] Business Intelligence Project

## COURSE SYLLABUS (10 DAYS)

DAY	TOPICS COVERED	TIME
One	<p><b>MODULE 1: EXTRACT, TRANSFORM AND LOAD (ETL)</b></p> <ul style="list-style-type: none"> <li>➤ Understanding big data in BI analytics</li> <li>➤ Extracting data from sources</li> <li>➤ Transforming data to fit analytic needs</li> <li>➤ Loading data into business systems for analysis</li> </ul> <p>Participants will learn how to extract, transform and load their data.</p>	8am – 5pm
	<p><b>MODULE 2: DATA CLEANING</b></p> <ul style="list-style-type: none"> <li>➤ Summarizing your data for inspection</li> <li>➤ Finding and fixing flawed data</li> <li>➤ Converting inputs to data types suitable for analysis</li> <li>➤ Adapting string variables to a standard</li> </ul> <p>Participants will learn how to clean their data for data analysis.</p>	8am – 5pm

	<p><b>MODULE 3: EXPLORATORY DATA ANALYSIS</b></p> <ul style="list-style-type: none"> <li>➤ Understanding exploratory data analysis</li> <li>➤ Analysing a single data variable</li> <li>➤ Analysing two variables together</li> <li>➤ Exploring multiple variables simultaneously</li> </ul> <p>Participants will learn how to perform an exploratory data analysis.</p>	8am – 5pm
Two	<p><b>MODULE 4: LINEAR REGRESSION FOR BUSINESS</b></p> <ul style="list-style-type: none"> <li>➤ Understanding linear regression</li> <li>➤ Checking model assumptions</li> <li>➤ Using a simple linear regression</li> <li>➤ Interpreting model output</li> <li>➤ Predicting unknown outputs with an SLR</li> <li>➤ Working with big data using confidence intervals</li> <li>➤ Refining data for simple linear regression</li> <li>➤ Transforming data</li> <li>➤ Handling outliers and influential points</li> <li>➤ Introducing multiple linear regression</li> </ul> <p>In this module, participants will learn to analyse data using a linear regression method.</p>	8am – 5pm
Three	<p><b>MODULE 5: DATA MINING WITH CLUSTER ANALYSIS</b></p> <ul style="list-style-type: none"> <li>➤ Explaining clustering analysis</li> <li>➤ Partitioning using k-means clustering</li> <li>➤ Exploring the data</li> <li>➤ Running the kmeans() function</li> <li>➤ Interpreting the model output</li> <li>➤ Developing a business case</li> <li>➤ Clustering using hierarchical techniques</li> <li>➤ Cleaning and exploring data</li> <li>➤ Running the hclust() function</li> <li>➤ Visualizing the model output</li> <li>➤ Evaluating the models</li> <li>➤ Choosing a model</li> <li>➤ Preparing the results</li> </ul> <p>In this module, participants will learn how to perform data mining with cluster analysis.</p>	8am – 5pm
	<p><b>MODULE 6: TIME SERIES ANALYSIS</b></p> <ul style="list-style-type: none"> <li>➤ Analysing time series data with linear regression</li> <li>➤ Linearity, normality, and equal variance</li> <li>➤ Prediction and confidence intervals</li> <li>➤ Introducing key elements of time series analysis</li> <li>➤ The stationary assumption</li> <li>➤ Differencing techniques</li> <li>➤ Building ARIMA time series models</li> <li>➤ Selecting a model to make forecasts</li> <li>➤ Using advanced functionality for modelling</li> </ul> <p>Participants will learn how to perform the time series analysis.</p>	8am – 5pm

	<p><b>MODULE 7: VISUALIZING THE DATA'S STORIES</b></p> <ul style="list-style-type: none"> <li>➤ Visualizing data</li> <li>➤ Calling attention to information</li> <li>➤ Empowering user interpretation</li> <li>➤ Plotting with ggplot2</li> <li>➤ Geo-mapping using Leaflet</li> <li>➤ Learning geo-mapping</li> <li>➤ Extending geo-mapping functionality</li> <li>➤ Creating interactive graphics using rCharts</li> <li>➤ Framing the data story</li> <li>➤ Learning interactive graphing with JavaScript</li> </ul> <p>Participants will learn how to visualize the data using R</p>	8am – 5pm
Four	<p><b>MODULE 8: WEB DASHBOARDS WITH SHINY</b></p> <ul style="list-style-type: none"> <li>➤ Creating a basic Shiny app</li> <li>➤ The ui.R file</li> <li>➤ The server.R file</li> <li>➤ Creating a marketing-campaign Shiny app</li> <li>➤ Using more sophisticated Shiny folder and file structures</li> <li>➤ The www folder</li> <li>➤ The global.R file</li> <li>➤ Designing a user interface</li> <li>➤ The head tag</li> <li>➤ Adding a progress wheel</li> <li>➤ Using a grid layout</li> <li>➤ UI components of the marketing-campaign app</li> <li>➤ Designing the server-side logic</li> <li>➤ Variable scope</li> <li>➤ Server components of the marketing-campaign app</li> <li>➤ Deploying your Shiny app</li> <li>➤ Located on GitHub</li> <li>➤ Hosted on RStudio</li> <li>➤ Hosted on a private web server</li> </ul> <p>Participants will learn how to develop a web dashboard using Shiny.</p>	8am – 5pm
Five	<p><b>MODULE 9: BUSINESS INTELLIGENCE PROJECT</b></p> <ul style="list-style-type: none"> <li>➤ Business Intelligence Best Practices</li> <li>➤ Business Intelligence Project Management</li> <li>➤ Presentations</li> </ul>	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](mailto: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.