Digital Skills Training Programs at Knowledge Technology

MASTERING MACHINE LEARNING (AI)

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

- This 10-day course provides you advanced techniques in machine learning. You will delve into statistical learning theory and supervised learning; design efficient algorithms; learn about creating Recommendation Engines; use multi-class classification and deep learning; and more.

- By the end of this course, you will be able to perform machine learning in the cloud using AWS in various scenarios with different datasets.

Course Title	Mastering Machine Learning
Duration	10 Days
Trainer	Assoc. Prof. Dr. Rayner Alfred
Max Participants	25

SYNOPSIS

This course will teach you advanced techniques in machine learning. You will delve into statistical learning theory and supervised learning; design efficient algorithms; learn about creating Recommendation Engines; use multi-class classification and deep learning; and more. You will explore, in depth, topics such as data mining, classification, clustering, regression, predictive modeling, anomaly detection, boosted trees with XGBOOST, and more. More than just knowing the outcome, you'll understand how these concepts work and what they do. With a slow learning curve on topics such as neural networks, you will explore deep learning, and more. By the end of this course, you will be able to perform machine learning in the cloud using AWS in various scenarios with different datasets.

LEARNING OUTCOMES

What participants will gain at the end of the course

➢ Gain deep insights into the application of machine learning tools in the industry

➢ Manipulate data to prepare it for analysis

➢ Master the skill of recognizing techniques for effective visualization of data

➢ Understand why and how to create test and training data sets for analysis

➢ Master fundamental learning methods such as linear and logistic regression

➢ Comprehend advanced learning methods such as support vector machines

➢ Learn how to use R tool in a cloud service such as Amazon

JUSTIFICATION TO LEARN MACHINE LEARNING

Machine Learning is no longer just a niche subfield of computer science, but technology giants have been using it for years – Machine learning algorithms power Walmart product recommendations, surge pricing at Uber, fraud detection at top financial institutions, content that Twitter, LinkedIn, Facebook and Instagram display on social media feeds or Google Maps. Machine learning products are being used daily, perhaps without realizing it. The future of machine learning is already here, it's just that machine learning career is exploding now because of smart algorithms being used everywhere from email to mobile apps to marketing campaigns. If you are in search of the most in-demand and most-exciting career domains, gearing up yourself with machine learning skills is a good move now.

Machine learning is the shining star of the moment. With every industry looking to apply AI in their domain, studying machine learning opens world of opportunities to develop cutting edge machine learning applications in various verticals – such as cyber security, image recognition, medicine, or face recognition. With several machine learning companies on the verge of hiring skilled ML engineers, it is becoming the brain behind business intelligence.

TOPICS LIST

[1] Data Mining Process
[2] Linear Regression - The Blocking and Tackling of Machine
[3] Logistic Regression and Discriminant Analysis
[4] Advanced Feature Selection in Linear Models
[5] More Classification Techniques - K-NN and Support Vector Machines
[6] Classification and Regression Trees
[7] Neural Networks and Deep Learning
[8] Cluster Analysis
[9] Principal Components Analysis
[10] Market Basket Analysis, Recommendation Engines, and Sequential Analysis
[11] R on the Cloud

COURSE SYLLABUS (10 DAYS)

DAY	TOPICS COVERED	TIME
One	**MODULE 1: DATA MINING PROCESS** ➢ The process and Business understanding ➢ Identifying the business objective ➢ Assessing the situation, Determining the analytical goals ➢ Producing a project plan, Data understanding and preparation ➢ Modelling and Evaluation, Deployment and Algorithm flowchart **MODULE 2: LINEAR REGRESSION - THE BLOCKING AND TACKLING OF MACHINE LEARNING** ➢ Univariate linear regression - Business understanding ➢ Multivariate linear regression - Business understanding ➢ Other linear model considerations ➢ Qualitative features and Interaction terms Participants will learn the process of data mining and a simple, yet extremely effective technique called linear regression that has been used for a long time to describe data.	8am – 5pm
Two	**MODULE 3: LOGISTIC REGRESSION AND DISCRIMINANT ANALYSIS** ➢ Classification methods and linear regression, ➢ Logistic regression ➢ Discriminant analysis overview and application	8am – 5pm

		➢ Multivariate Adaptive Regression Splines (MARS) ➢ Model selection	
		Participants will examine how they can develop algorithms to predict qualitative outcomes. Such outcome variables could be binary (male versus female, purchase versus does not purchase, tumor is benign versus malignant) or multinomial categories (education level or eye color). Regardless of whether the outcome of interest is binary or multinomial, the task of the analyst is to predict the probability of an observation belonging to a category of the outcome variable. In other words, they develop an algorithm to classify the observations.	
Three	**MODULE 4: ADVANCED FEATURE SELECTION IN LINEAR MODELS** ➢ Regularization in a Nutshell, ➢ Business case and understanding ➢ Data understanding and preparation ➢ Modelling and evaluation, Model selection Participants will look at the concept of regularization where the coefficients are constrained or shrunk towards zero. There is several methods and permutations to these methods of regularization, but we will focus on Ridge regression, Least Absolute Shrinkage and Selection Operator (LASSO), and finally, elastic net, which combines the benefit of both techniques into one.	8am – 5pm
Four	**MODULE 5: MORE CLASSIFICATION TECHNIQUES - K-NN AND SUPPORT VECTOR MACHINES** ➢ K-nearest neighbours, Support vector machines ➢ Business case, Feature selection for SVMs In this module, participants will delve into two nonlinear techniques: K-Nearest Neighbours (KNN) and Support Vector Machines (SVM). These techniques are more sophisticated than what they have learnt earlier because the assumptions on linearity can be relaxed, which means a linear combination of the features to define the decision boundary is not needed.	8am – 5pm
Five	**MODULE 6: CLASSIFICATION AND REGRESSION TREES** ➢ An overview of the techniques ➢ Random forest regression ➢ Random forest classification ➢ Feature Selection with random forests In the previous modules, participants examined the techniques used to predict either a quantity or a label classification. In this module, participants will apply them to both types of problems. They will also approach the business problem differently than in the previous modules.	8am – 5pm
Six	**MODULE 7: NEURAL NETWORKS AND DEEP LEARNING**	8am – 5pm

	➢ Introduction to neural networks ➢ Deep learning, a not-so-deep overview ➢ Business understanding, Data understanding and preparation ➢ Modelling and evaluation, An example of deep learning In this module, participants will learn how the methods work, their benefits, and inherent drawbacks so that they can become conversationally competent about them. They will work through a practical business application of a neural network. Finally, they will apply the deep learning methodology in a cloud-based application.	
Seven	**CHAPTER 8: CLUSTER ANALYSIS** ➢ Hierarchical clustering, K-means clustering ➢ Gower and partitioning around medoids ➢ Random forest ➢ Business understanding ➢ Data understanding and preparation ➢ Modelling and evaluation Participants will learn a popular and powerful technique known as cluster analysis. With cluster analysis, the goal is to group the observations into several groups (k-groups), where the members in a group are as similar as possible while the members between groups are as different as possible.	8am – 5pm
Eight	**MODULE 9: PRINCIPAL COMPONENTS ANALYSIS** ➢ An overview of the principal components Analysis ➢ Business understanding ➢ Data understanding and preparation ➢ Modelling and evaluation In this module, participants will see how to reduce the dimensionality and improve the understanding of their data by grouping the correlated variables with Principal Components Analysis (PCA). Then, they will use the principal components in supervised learning.	8am – 5pm
Nine	**MODULE 10: MARKET BASKET ANALYSIS, RECOMMENDATION ENGINES, AND SEQUENTIAL ANALYSIS** ➢ An overview of a market basket analysis ➢ Business understanding ➢ Data understanding and preparation, ➢ Modelling and evaluation ➢ An overview of a recommendation engine ➢ Sequential data analysis, Sequential analysis applied In this module, participants will learn association rules and recommendation engines. Association rule analysis is commonly referred to as market basket analysis as one is trying to understand what items are purchased together. With recommendation engines, the goal is to provide a customer with	8am – 5pm

		other items that they will enjoy based on how they have rated previously viewed or purchased items.	
Ten		**MODULE 11: R ON THE CLOUD** ➢ Creating an Amazon Web Services account ➢ Launch a virtual machine ➢ Start RStudio In this module, participants will learn using cloud-based computing with R in which it can facilitate the seamless work across multiple locations and provide you with tremendous computing power, which can be quickly scaled up or down as needed. This can be a significant cost saving.	8am – 5pm

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.