

DATA SCIENCE

MRD offers a complete package of learning, practicing and executing data science and mathematical modules into day-today activities.

Transform the way you discover, analyse and present data.

And become a more valued asset by learning the latest advanced analytics techniques for solving critical business challenges across every domain.

Program Modules:

Module	Topic
CI/CD training	Learn basic principles of CI/CD
	Get familiar with Github/GitFlow
	Learn basic knowledge of error handling/unit test
	Get to know how CI works in a company
	Learn how to use unit testing in R package: test that
	Develop a code in CI manner
Data analysis and visualization	Writing unit test for the code in R
	Learn how to program in R & Python
	Learn how to use R package: Shiny
	Learn basic knowledge of SQL
	Learn how to clean and format data for analysis
	Learn how to write data content quality report
	Learn common data analysis concepts: distributions, correlation, regression, and time series analysis, in a real case demonstration
	Learn how to build analysis and visualization tool by using Shiny
	Get familiar with two key tools: FLP1 and PriceCypher
	Practice the skill of data exploration and insights discovery on a real client dataset
	Build a Shiny tools to present bussiness insighs
Feature Selection	Learn the operations of feature selection; remove feature, re-code feature, combine features and add new feature
	Learn the operations of feature transformation: scaling, filtering, and aggregation
	Understand different methodologies of feature importance ranking: PCA, ANOVA, correlation test, Chi-Squared test, etc
	Understand the reason behind feature selection and data transformation
	Learn how to select smallest set of features that best captures data for business objectives in a real case demonstration
	Be able to measure the difference between what a model needs and how business works
	Understand the effect of removing/selecting features in the model and its impact
	Practice the skill of feature discovery on a real open source dataset
Ask the right questions, make the right judgement	
Model development	Learn different encoding methodologies: label encoding, frequency encoding, one hot encoding, mean encoding
	Understand machine learning models: linear regression, logistic regression, SVM, KNN, Random forest, gradient boosting

	Understand cross validation technique: Holdout, K-Fold, Stratified K-Fold, Leave-P-Out
	Learn how to do hyperparameters tuning: grid search, random search, and informed search
	Learn how to build a model that meet the business objective in a real case demonstration
	Practice the skill of modelling on a real client dataset
Model validation	Understand what overfitting is and underfitting
	Understand validation process and importance
	Understand different evaluation metrics: RMSE, MAE, MAPE, R2, Bias
	Learn how to evaluate a model that meet the business objective in a real case demonstration
	Practice the skill of model validation on an open source data set
Model deployment	Know how to create a REST API and understand the core principles.
	Learn basic knowledge of Docker containers
	Understand how different environments work (dev/uat/prod)
	Learn how to build a REST API by using R package: plumber
	Learn how to build a Docker container
	Learn how to deploy a container in Azure
	Learn how the azure pipeline works
	Build a full model deployment in Azure with live API connections
	Learn how API Connections work
	Understand database connections and its impact on model speed.
	Optimize a containerized solution