MACHINE LEARNING USING JULIA

Machine learning is one of the fastest-growing and most exciting fields out there. In this course, you'll develop a clear understanding of the motivation for machine learning models, and design intelligent systems that learn from datasets.
We will introduce the basic concepts of machine learning models. You will learn to solve new classes of problems that were once thought prohibitively challenging, and come to better appreciate the complex nature of human intelligence as you solve these same problems effortlessly using machine learning algorithms. We will be using Julia as a programming language to work on machine learning projects.

Julia, an open source project, made available under the MIT license is fast, dynamic typed language. More on https://julialang.org/. Julia Pro by Julia computing also distributes Julia along with Juno IDE. The Julia script can be written on Juno IDE (https://junolab.org/).

The Juno IDE is in maintenance mode now and another IDE named Visual Studio Code (VSCode) will see all the new feature development for use with Julia. It is preferable to use VSCode text editor with Julia. VSCode can be downloaded from https://code.visualstudio.com/download

However, we will be using jupyter notebook for working on Julia script.

WHO SHOULD ATTEND

Irrespective of type of industry (retail, e-commerce, manufacturing, real estate & construction, telecom, hospitality, banking, healthcare, IT, supply chain &logistic, etc.); data forms the crux of decision making. This course is designed for graduates and post graduates who will venture into the corporate set up and will be assisting the management in various decision making

process. This course is equally suited to hone up analytical skills and business acumen of mid-level and senior level corporate professionals trying to understand the nuances of data science and help them the machine learning techniques an efficient way to generate insights for customers which in turn optimizes the bottom line of organizations.

HARDWARE, SOFTWARE REQUIREMENT AND INSTALLATION:

Follow details in this link: https://statsmalai.com/ and https://statsmalai.com/julia/

PRE-REQUISITE & COURSE DELIVERABLE

1. Participants should have basic programming skills. Participants are expected to spend time with the code set as a home assignment to leverage the classroom training hours to the fullest.
2. High speed internet connection.
3. Deliverable: Julia code and dataset. Soft copy of the content being covered (PDF file)

COURSE OUTLINE

Day 1: Understanding Anaconda Framework platform and other useful packages in Julia

Session 1–Introduction to Business Analytics

- What is Business Analytics – Tools, Techniques, Context
- Why is it needed and how industries are adopting it
- Different components of analytics – Descriptive, Predictive and Prescriptive
- Different types of machine learning algorithms–Supervised and Unsupervised learning

Session 2 & 3–Introduction to Anaconda and Julia

- Overview of Jupyter Notebook
- Julia – Variables, objects, loops, conditions, function.
- Julia Data structures – lists, tuples, dictionaries, sets
- Overview of Data Analysis Stack – DataFrames.jl, CSV.jlSeaborn, Plots.jl, GLM.jl, GLMNET.jl, MultivariateStats.jl, ScikitLearn.jl

Day 2: Understanding regression and its implementation using Julia

Session 1–Data Exploration and visualization

- Loading data from Files
- Data manipulation - Filtering, Grouping, Ordering of data
- Dealing with missing Data
- Drawing Histograms, Bar charts, Scatter Plot, Box Plots
- Understanding Basic Statistics, Distributions, Correlations

Session 2 & 3–Lab 1: Linear Regression

- Understanding Regression and Examples
- Understanding loss function and gradient descent
- Building Linear Regression Model
- Creating Training, validation and Test Data Sets, Cross validations

Day 3: Understanding logistic regression and its implementation using Julia

Session 1–Lab 1: Linear Regression

- Understanding Evaluation Metrics: RMSE, R-square
- Case study using regression techniques and hands-on using Julia code for regression

Session 2&3–Lab 2: Logistic Regression

- Understanding Classification and Examples
- Introduction to Logistic Regression, strategy to find the optimal cut-off
- Loss function and regularization
- Understanding Evaluation Metrics: Confusion Matrix, Precision, Recall, Accuracy etc.

COURSE SCHEDULE

Day 1: Understanding Anaconda Framework platform and other useful packages in Julia

This day will be about basic concepts in Julia and statistics

Topic	Session	From	To
Introduction to business analytics	1	5:00 PM*	6:15 PM*
Introduction to Anaconda framework and Julia	2	6:25 PM*	7:40 PM*
Introduction to Anaconda framework and Julia…cont.	3.1	7:45 PM*	8:15 PM*

Day 2: Understanding regression and its implementation using Julia

This day will be about underlying concepts of regression

Topic	Session	From	To
Introduction to Anaconda framework and Julia…cont.	3.2	4:00 PM*	4:45 PM*
Data Exploration and visualization	4	5:00 PM*	6:15 PM*
Lab 1: Linear Regression…cont.	5	6:25 PM*	7:40 PM*
Lab 1: Linear Regression…cont.	6.1	7:45 PM*	8:15 PM*

Day 3: Understanding logistic regression and its implementation using Julia

This day will be about underlying concepts of logistic regression

Topic	Session	From	To
Lab 1: Linear Regression…cont.	6.2	9 AM	9:45 PM
Lab 1: Linear Regression…cont.	7	10 AM	11:15 AM
Lab 2: Logistic Regression	8	11:30 AM	12:45 PM
Lab 2: Logistic Regression…cont.	9	01:15 PM	2:30 PM

* **Tentatively Schedule. To be confirmed.**

INSTRUCTOR PROFILE

Rahul Kumar An engineering graduate from National Institute of Technology, Jaipur and an alumnus of IIM, Bangalore, as well. He has a professional career spanning 16 years in corporate and academia and is still passionate about exploring data using a variety of tools and techniques. He has worked with the likes of Satyam computers, Nokia Siemens and Deloitte Consulting but the urge to explore the uncharted territory led to few start-ups. Few failed and some saw the light of the day!!! Currently, I work as a freelancer as well as a founding member of AwesomeStats Consulting; a company, run by few passionate folks, which is primarily focused towards training and consulting in the field of data science.

For over seven years, he is also associated as a consultant at Data Centre and Analytics Lab, IIMB. He has executed several analytics projects for large corporates. Few of his work in the field of analytics includes predicting credit scoring for co-operative banks of Karnataka; predicting renege/attrition issue for fashion retail company; NPS for a reputed medical equipment manufacturer; predicting design issues for a leading US auto manufacturer; sales and warranty forecasting for a leading auto manufacturer; anomaly detection for a paper-based consumer products company based out of US.

He has imparted 400+ sessions in R/Python/Julia in short and long duration programs at IIM Bangalore. As a freelancer, he has undertaken equivalent number of sessions for working professionals in various corporates. Few of the prominent corporate clients, I have worked with are General Electric, Cisco, Deloitte Consulting, United Health Group, HSBC, Flipkart, Fidelity Investments, General Motors, JP Morgan, TVS Motors, Raukten, Hudson Bay etc.
As a part of the Data Centre and Analytics Lab, IIMB and Analytics society of India, he also imparted several sessions in faculty developments programs (FDP) and conducted workshops in IIM Bangalore, PSG Tech Coimbatore, PSGIM Coimbatore, CIT Coimbatore, LBSIM New Delhi, Christ University Bangalore, IMT Hyderabad, JIM Trichy, SRM Chennai, SDM Mysore. Some publications and have also presented papers in several national and international conferences.

- Rahul K., Seth N., Dinesh Kumar U. (2018) Spotting Earnings Manipulation: Using Machine Learning for Financial Fraud Detection. In: Bramer M., Petridis M. (eds) Artificial Intelligence XXXV. SGAI 2018. Lecture Notes in Computer Science, vol 11311. Springer, Cham
- Invited talk on "Using Machine Learning Algorithms to Detect Earnings Manipulations" at 5th International Conference on Business Analytics and Intelligence, IIM Bangalore 11th-13th December 2017.

- Paper on "Predicting Net Promoter Score (NPS) to Improve Patient Experience at Manipal Hospitals" published at Harvard Business Publishing, September 2017.
- Paper on "Behavioral Modeling to Predict Renege" published at Harvard Business Review, January 2016.
- Paper Presentation at CMMI conference organized by CMMI Institute, 10-11 Dec 2014 at Shenzhen, China.
- Paper Presentation at SEPG Europe conference organized by SEI | Carnegie Mellon University, 5-7 June 2012 at Madrid, Spain.

He has also undergone workshop on the usage of statistical models and techniques from ISI Bangalore. His other certifications include DB2 certification from IBM and ISO 9001:2008 lead auditor certification by DNV India.