Predictive Analytics & Data Mining

Description

With data science and the combination of Big data and Machine Learning Techniques, suddenly the kind of questions that most professional ask since many years get a crystal-clear answer: Which of my customers are most likely to churn? How to minimize credit risks? How to attract best talents? How to predict next security breach? Who are my best prospects? How can we increase Lifetime Value of customers?

All Strategic decisions are now based on data.

In this context any business professional (Marketing, Sales, Finance, Human Resource) needs to understand and master data and AI, those technologies will not be any more “black boxes” that only specialists like data scientists can handle.

The first and main objective of this course is to give you all the keys for becoming autonomous about data in order to solve a business challenge or problem and take a decision. with big amount of data.

The second objective is to give you the knowledge to be integrated in large data projects and work with specialists like data scientists, IT specialist, legal specialist, understand their constraints and vocabulary.

Prerequisite:

This course will be a little bit technical you will practice predictive analysis on real datasets using R& R Studio and Dataiku DSS, so you have to install those tools prior to the course on your desktop.

You also need to have some basics in maths and statistics

Learning Objectives and Outcomes

- Know how and where to Find appropriate data in order to solve a business complex topic
- Preparing data for analysis by using best-in-class tools
- Analyzing those data to get business insights or take decisions
- Visualizing your results and present management synthesis of your work
- Understanding the main models and algorithm and select one to make predictive analysis
- Running a predictive analysis using Machine Learning techniques
- Using your own judgment to evaluate quality of predictions

Predictive Analytics & Data Mining

Course Schedule and Contents

Session#1 3hrs	▪ Introduction, main concepts and vocabulary ▪ Practicing R & R studio ▪ Datavisualisation, graphics, descriptive analysis ▪ Data preparation, enriching data ▪ The package ggplot ▪ Enriching data: joining dataset ▪ First analysis: social survey dataset ▪ Basic stats: correlation, influence of variables ▪ Case study 1 # diamonds dataset
Session#2 3hrs	▪ Using DSS to conduct a data analysis case study sale ▪ Predictive analysis with machine learning algorithms ▪ Case study 1: predict health expenses ▪ Using linear regression on numeric variables ▪ Case study 2: predict quality of wines using decision trees ▪ Improving model with random Forrest
Session#3 3hrs	▪ Case Study 3: predictive analysis on Wisconsin breast cancer dataset ▪ Using K means algorithm for the prediction ▪ Case study 4: prediction of spam vs ham sms, Natural Language Processing algorithm ▪ Using Naïve Bayes algorithm for classification ▪ Producing cloud words ▪ Case study 4: Using classification knn algorithm on teenagers dataset
Session#4 3hrs	▪ Using DSS to produce a predictive analysis ▪ Final Case study: choice of subject , datasets ▪ Analysis on subject chosen

Grading

Course case study:	40%
Final case study	40%
Participation in class:	20%

Policies

- I expect you to turn-in your reports on time to receive proper credit/grade.
- Any work submitted must be your own.
- I expect everyone to contribute equally to group assignments
- Attendance in every class is expected and class participation and discussion is strongly encouraged.
- Late work will not be accepted unless prior arrangements have been made directly with me.

Good Luck!