INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination : 2017–18

Course : Post Graduate Diploma in Business Analytics (First Year)

Subject : Computing for Data Sciences : BAISI–4 for PGDBA–I

Date : 20 September 2017 Maximum Marks : 40 + 10 (bonus) Duration : 4 Hours

This is an open-resources examination. You may use any online or offline academic material as reference during the test, and you are free to borrow codes from any non-interactive source on the Internet. However, posting any query on discussion forums, emails, blogs, and interactive online portals during the test will be considered unethical, and will draw penalty. You must refer to each and every online or offline source you use during the test at the end of your report. There is no need to explicitly cite the R (or Python) packages and libraries you use during the test, as they will anyway be a part of your code(s). Within-group discussion is permitted, but between-group discussion during the test will draw penalty.

Consultancy Project for RandMart™

Data : You are given a dataset — `randData.csv` — associated with this problem, along with an auxiliary description file — `randVars.csv`. The dataset `randData.csv` contains the records of 2823 departmental stores of a specific chain of stores — RandMart™ — in the USA, and the variables/features of this dataset are explained in the auxiliary description file `randVars.csv`.

Objective : RandMart™ wants to open up a *new* store in the USA, and they have hired your group as expert consultants to help them choose the best location for the new store. Analyze the given dataset `randData.csv` given by RandMart™, and advise them to the best of your expertise, the most *desirable* location(s) for the company to open up its new store in the USA.

Submission : Submit a *consultancy report* targeted at RandMart™, describing your approach and methodology for the analysis of the dataset, modelling of the problem, and prediction of the desired location(s). Along with the report, you should submit all relevant code(s) as well, so that the results you obtained from the dataset `randData.csv` are reproducible and verifiable.

Your report should not be more than 8–10 pages in length, and should touch upon the following points — problem statement, objective, description of data, exploratory data analysis, problem formulation (if required), choice of model, modelling the problem, comparison between models (if required), model tuning and validation, final model to fit on the dataset, prediction using the final model, intuitive justification for the prediction, your final suggestion(s) for the desired location(s) for the *new* store of RandMart™, remarks about the data, conclusion, and references.

Submit **'groupXX_report.pdf'** and **'groupXX_codes.zip'** via email to `sg.sourav@gmail.com`. Subject line of the email should be – "CDS 2017 : Mid-Sem Examination : Solution by Group XX".

Best of luck! ☺