



# Statistical Methods for Data Mining

MMA AND MECD, 2<sup>nd</sup> SEMESTER, 2019/2020

## Project 1

Handed out on March 23, 2020.

To be **handed back** by **April 30**, 2020.

As a result of an experimental a dataset has to be analysed. The researcher has rudimentary knowledge about Data Mining and wants your help to work on his/her problem. You can grab the dataset (*vide* Tables 1 and 2) and a brief description of the observed variables at: <http://web.tecnico.ulisboa.pt/ist13493/MEDM2020/Project1/>

1. Perform a exploratory data analysis and discuss what you have learned from this analysis.
2. Solve your classification problem. Consider several classification methods and discuss how can they contribute for the solution of your problem. Include in your discussion topics such as the options that you have made in building each classifier; interpretation of results; validation of the methods used and possible assumptions; advantages and disadvantages of each alternative; etc. Have in mind that some of the explanatory variables may be irrelevant to the classification problem and that you may need to do some preprocessing methodologies of your data set e.g. dimensionality reduction techniques.
3. Imagine you are going to meet the researcher who contacted you. Report to him/her what you have learned about the problem. Discuss limitations of the analysis you have done and provide suggestions for future work.

The datasets are distributed as follows:

Table 1: Assignment of each dataset to each group of students.

Number	Name	Group	Folder
81603	Inês Sofia Nunes Duarte	1	Group1Data
83397	Filipa Leitão Bento	1	
97011	João Patriarca Pais Marques	1	
97154	Francisco Mendes Malheiro Rodrigues	1	
97256	Miguel Machado Moita de Deus	1	
83628	Bruno Filipe Antunes Costa	2	Group2Data
83638	Maria Seabra Dourado Eusébio	2	
83640	Mariana Esteves Oliva	2	
83756	Bruno Miguel Dos Reis Melgão	2	
86963	Catarina Alexandra Conde Ribeiro Oliveira	2	
87005	Gonçalo Leote Cardoso Mestre	3	Group3Data
87101	Pedro Miguel Martins Chanca	3	
87118	Rui Miguel Ferreira Lopes de Paiva Vasconcelos	3	
97136	Maria Soares de Almeida	3	
97160	João Bernardo Garcia de Moraes	3	
83646	Miguel De Lacerda e Costa Serra do Nascimento	4	Group4Data
83657	Tiago André Anciães dos Santos	4	
84142	Miguel Ângelo Cardoso Engrossa	4	
86592	Maria Beatriz Silva Santiago	4	
87936	Rodrigo Mendes Ribeiro	4	
86581	Carolina Isabel Matias Vicente	5	Group5Data
86594	Mariana Sacramento Espada Venâncio Carrasco	5	
97636	Catharina Bratterud Lilleengen	5	
97808	Michala Zwicki Blicher	5	
97816	Marguerite Saucé	5	
34290	Nuno Miguel Fernandes Clemente da Encarnação Rosinha	6	Group6Data
76561	Tiago Vaz Mendes	6	
95485	Asish Ghosh	6	
97181	António Rafael Mendes Rosa	6	
97008	Eduardo Miguel Ribeiro Bártolo	6	

Table 2: Assignment of each dataset to each group of students.

Number	Name	Group	Folder
78882	Samuel Pinto Ramos de Pina	7	Group7Data
96986	Regina de Brito Duarte	7	
97165	Bárbara Palmeiro Tavares	7	
97166	Diogo André Guarda Teixeira	7	
97205	Ricardo Carvalho Santos	7	
86798	Miguel Lourenço Figueira Francisco Farinha	8	Group8Data
95193	Luis Oliva Fontecha	8	
97569	Mahima Raut	8	
97781	Sahil Rajesh Kumar	8	
39953	Pedro Luís Ruivo Dias	9	Group9Data
79410	Sara Ventura de Melo Furtado da Cruz	9	
81234	Vasco Paradanta Gomes Ribeiro da Silva	9	
81269	João Miguel Fernandes Torres	9	
94775	Pedro Moura Soares de Almeida	10	Group10Data
95357	Renan Yudi Hamada Nunes	10	
97108	Sebastião Maria De Aguiar Martins e Maia Cerqueira	10	
97133	Eduardo Manuel Canejo Machado	10	
86579	Bruno Miguel Repolho Pires	11	Group11Data
86585	David Tadeu Lopes Moreira	11	
77039	Margarida Gerald Oliveira Moreira de Campos	12	Group12Data
83627	André Filipe Andrade Baptista Beirão Nunes	12	
83647	Miguel Figueiredo Vieira	12	

**About the report:**

- The report should not exceed 20 pages.
- Do not forget topics such as objectives; estimation methods; decisions; conclusions of the study; and references.
- The R code (or other code), the data cleaned or transformed (where applicable) as well as the report, must be send by email to:  
`conceicao.amado@tecnico.ulisboa.pt`