

CENTRO UNIVERSITÁRIO 7 DE SETEMBRO
Especialização em Ciência de Dados com Big Data, BI e Data Analytics
Disciplina: Análise Estatística de Dados
Professor: Prof. Dr. Marcio Mota

Atividade Final

Equipe: Grupos divididos em equipes entre 3 e 4 alunos.

Data Entrega: 1 de junho de 2020.

A Atividade Final deverá ser entregue até às 23h59min do dia 1 de junho de 2020 por email em formato .PDF e .XLS(X) APENAS ao e-mail marcio@marciomota.com. Cada aluno será responsável a formar e/ou entrar em equipes. O Professor não realizará interferência na formação das equipes. Cada dia de atraso na entrega do trabalho escrito implicará em uma penalidade de 10% da nota do trabalho. Casos de plágio serão considerados ZERO na nota final. Não haverá outra forma de avaliação a ser substituída pela Atividade Final.

Questão 1 (3 pontos):

Uma empresa seguradora deseja prever com maior segurança as possibilidades de indenização das apólices de seguro de automóveis quando assume um novo contrato. Assim, através do levantamento de registros anteriores, obteve os seguintes dados:

Observação	Grupo	Tempo de habilitação	Solteiro (1) Casado (2) Filhos? (3)	Número de multas
1	1	20	3	1
2	1	21	3	0
3	1	25	3	2
4	1	25	2	3
5	1	18	2	2
6	1	23	1	2
7	2	9	3	6
8	2	12	2	4
9	2	15	1	3
10	2	14	2	2
11	2	15	1	5
12	2	10	3	5
13	2	8	2	4
14	3	7	2	13
15	3	11	1	15
16	3	10	2	9
17	3	7	2	6
18	3	9	1	10
19	3	1	3	8
20	3	3	1	5

Esses casos foram divididos em três grupos: (1) Baixo Risco, (2) Médio Risco e (3) Alto Risco. As variáveis consideradas representativas foram “tempo de habilitação”, “número de multas” sofridas desde que obteve sua habilitação e estado civil/existência ou não de filhos, assumindo os valores de 1 para solteiro, 2 para casado e 3 com filhos.

Pede-se:

1) Usando o método da distância de Mahalanobis, estabeleça o grau de risco de cada novo potencial cliente cujos dados são:

Novas observações				
Observação	Grupo	Tempo de habilitação	Solteiro (1) Casado (2) Filhos? (3)	Número de multas
1	?	8	1	3
2	?	2	1	6
3	?	8	2	2
4	?	20	2	4
5	?	7	3	8
6	?	15	3	1
7	?	6	2	10
8	?	3	3	5

2) Indique a probabilidade em que cada um dos novos clientes pode pertencer ao Grupo Correspondente e cria uma Regra de Negócio diante dessa problemática. Aponte ainda qual dessas três variáveis é a que mais discrimina a classificação dos Grupos.

Questão 2 (3 pontos):

O controlador da XPTO deseja determinar a influência das variáveis mão de obra (MO) e energia elétrica (EE) nos custos totais de fabricação (CTF) de seus produtos. Para isso, fez um levantamento dos valores destas variáveis nos últimos 24 meses e os resultados são mostrados na tabela, onde XXX representa os três últimos algarismos do seu número de matrícula na UNI7 (escolha a matrícula de um dos alunos da equipe, se for o caso).

Meses	CTF	MO	EE
1	5XXX8	2378	980
2	XXX34	2295	945
3	62XXX	2450	930
4	6XXX7	2487	995
5	71XXX	2390	985
6	46XXX	2550	1010
7	XXX67	2440	998
8	5XXX9	2590	1025
9	7XXX3	2610	1100
10	54XXX	2575	1045
11	XXX98	2490	1038
12	XXX75	2580	1095
13	6XXX3	2395	1150
14	65XXX	2640	1030
15	4XXX9	2595	1085
16	XXX68	2720	1175
17	7XXX0	2690	1190
18	6XXX8	2565	1165
19	69XXX	2585	1200
20	XXX49	2615	1195
21	60XXX	2590	1210
22	7XXX1	2630	1189
23	58XXX	2680	1205
24	63XXX	2700	1200

Faça as análises de regressão simples e múltipla e determine o modelo de previsão para os CTF em função de MO e EE. Calcule os CTF para os seguintes valores:

Meses	MO	EE
25	2695	1120
26	2584	1185
27	2710	1178
28	2705	1205
29	2715	1195

Questão 3 (4 pontos):

Desenvolva uma aplicação simulada em que a Regressão Logística seja a técnica de Análise Multivariada de Dados mais adequada para encontrar a Probabilidade do Evento $p(\text{evento})$ ocorrer. Indique a variável dependente e, pelo menos, três variáveis independentes estatisticamente significantes que possam apontar o(a) aumento/diminuição de chance de identificar a probabilidade do $p(\text{evento})$. Indique todos os testes adequados, apresente a Tabela de Classificação com VP, FP, FN e VN e as medidas de desempenho. Crie uma regra de classificação e apresente a solução da Regra de Negócio da simulação.