

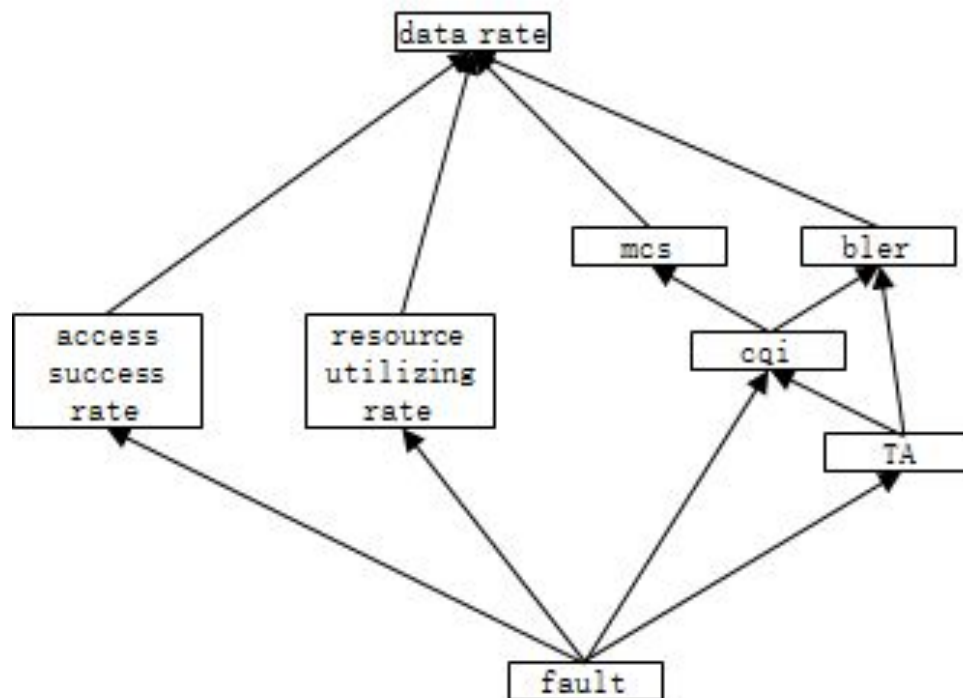
Projeto Princípios e Técnicas da Análise Estatística Experimental:

Detecção de falha em antenas 5G

Ana Caroline Silveira da Penha
Estela de Andrade Joffily
Johnattan Douglas Ferreira Viana
Tertius Sudário Gomes Ferraz

Fault Dataset (ITU Challenge)

- **7258** files (.CSV);
- Análise por hora de atenas de redes
- Impacto de falhas em KPIs (Key Performance Indicator)



Fault Dataset (Atributos)

- NE ID;
- Access Success Rate;
- **Resource Utilizing Rate;**
- TA (Time Advanced);
- **BLER (Block Error Rate);**
- CQI (Channel Quality Indicator);
- **MCS (Modulation and Coding Scheme);**
- Data rate of the hour;
- Fault duration (in seconds) during the hour;
- “Distance” (Relation) of the NE to the fault.

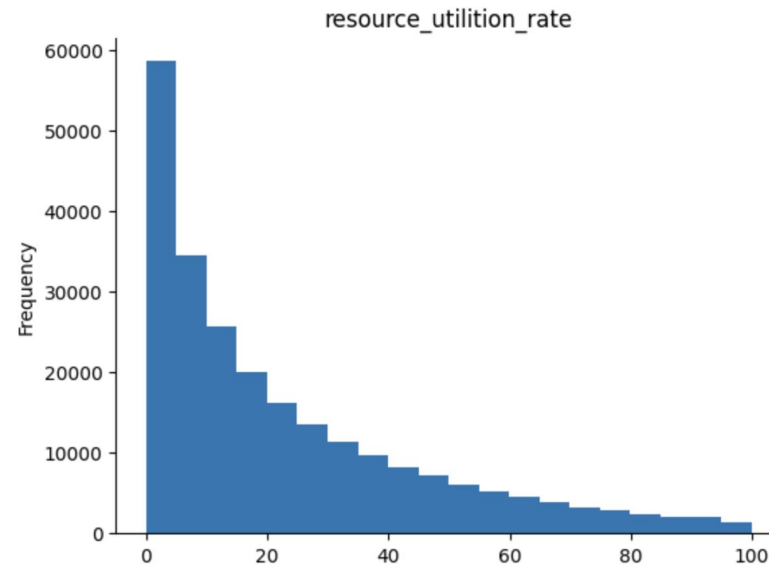
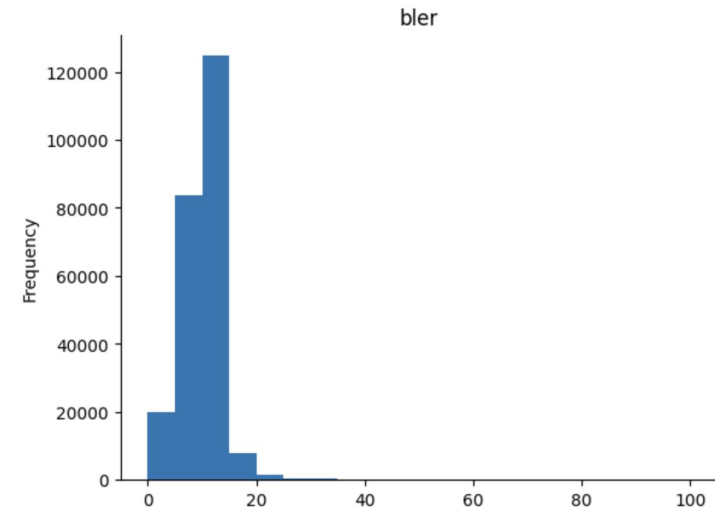
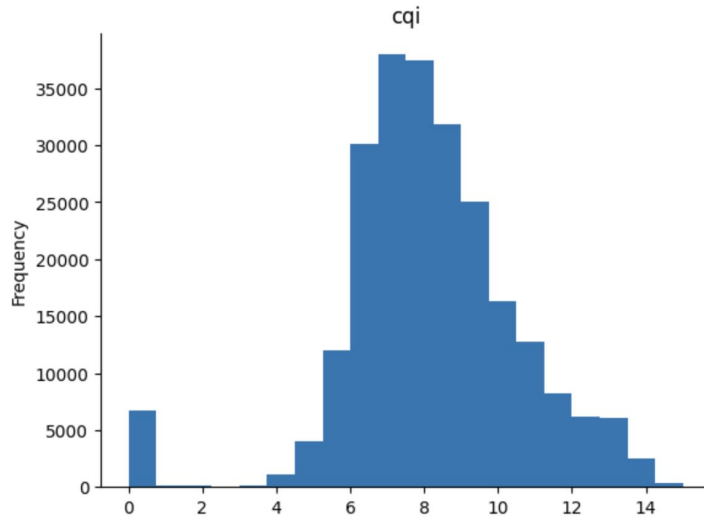
Objetivos

- Encontrar um **padrão de falhas** em uma antena 5G, através das métricas apresentadas.
- Para isso, inicializamos fazendo uma análise exploratória dos dados e levantando resultados estatísticos para um melhor resultado.
- Foram identificados testes que compararam métricas e analisaram suas correlações, com o objetivo de identificar um padrão que caracteriza uma "Falha".

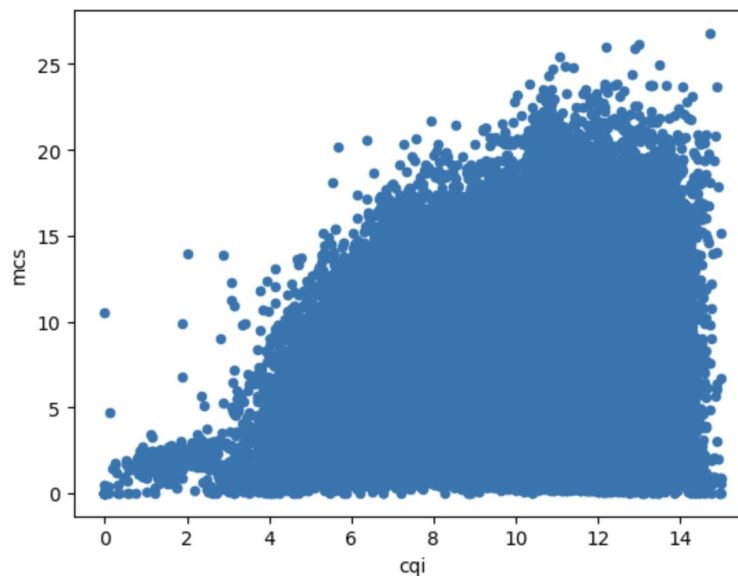
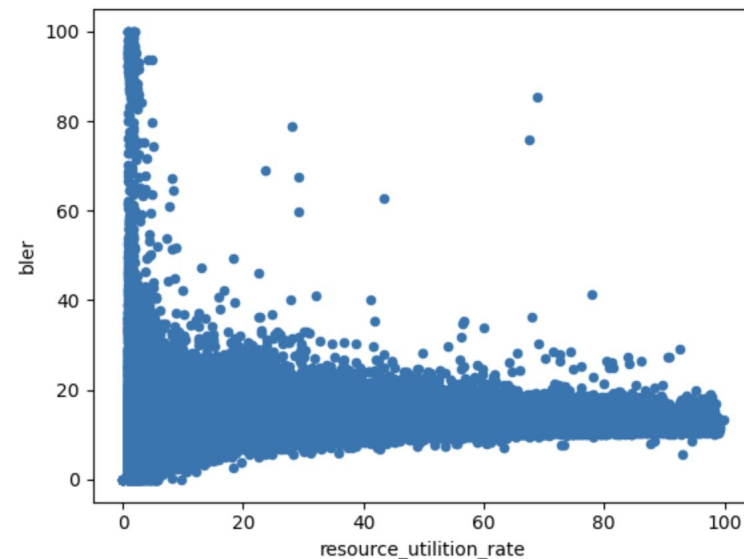
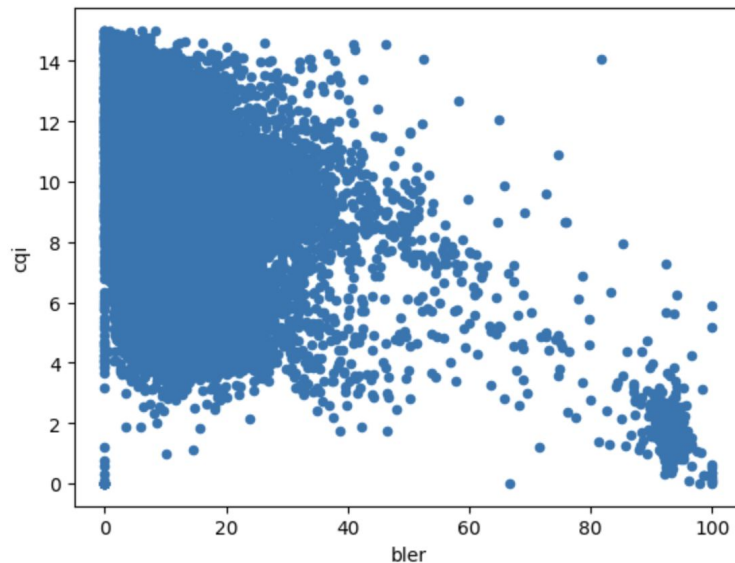
Descrição dos dados

	resource_utilition_rate	TA	bler	cqi	mcs	data_rate	fault_duration	relation	isAnomaly
count	239069.000000	239069.000000	239069.000000	239069.000000	239069.000000	239069.000000	239069.000000	239069.000000	239069.000000
mean	23.120502	1.884526	10.101276	8.179196	8.443441	18.240187	305.706369	0.085621	0.094952
std	22.807407	1.162158	4.837337	2.407428	3.380533	17.198204	991.788277	0.296449	0.293149
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-7.341916	0.000000
25%	5.128000	1.200028	8.370138	6.859920	6.655544	6.465350	0.000000	0.000000	0.000000
50%	15.129787	1.606796	10.418485	8.041363	8.553159	12.853785	0.000000	0.000000	0.000000
75%	34.487000	2.194757	11.777301	9.497022	10.415853	24.641657	0.000000	0.000000	0.000000
max	99.897872	7.000000	100.000000	15.000000	26.794554	366.040967	3600.000000	1.000000	1.000000

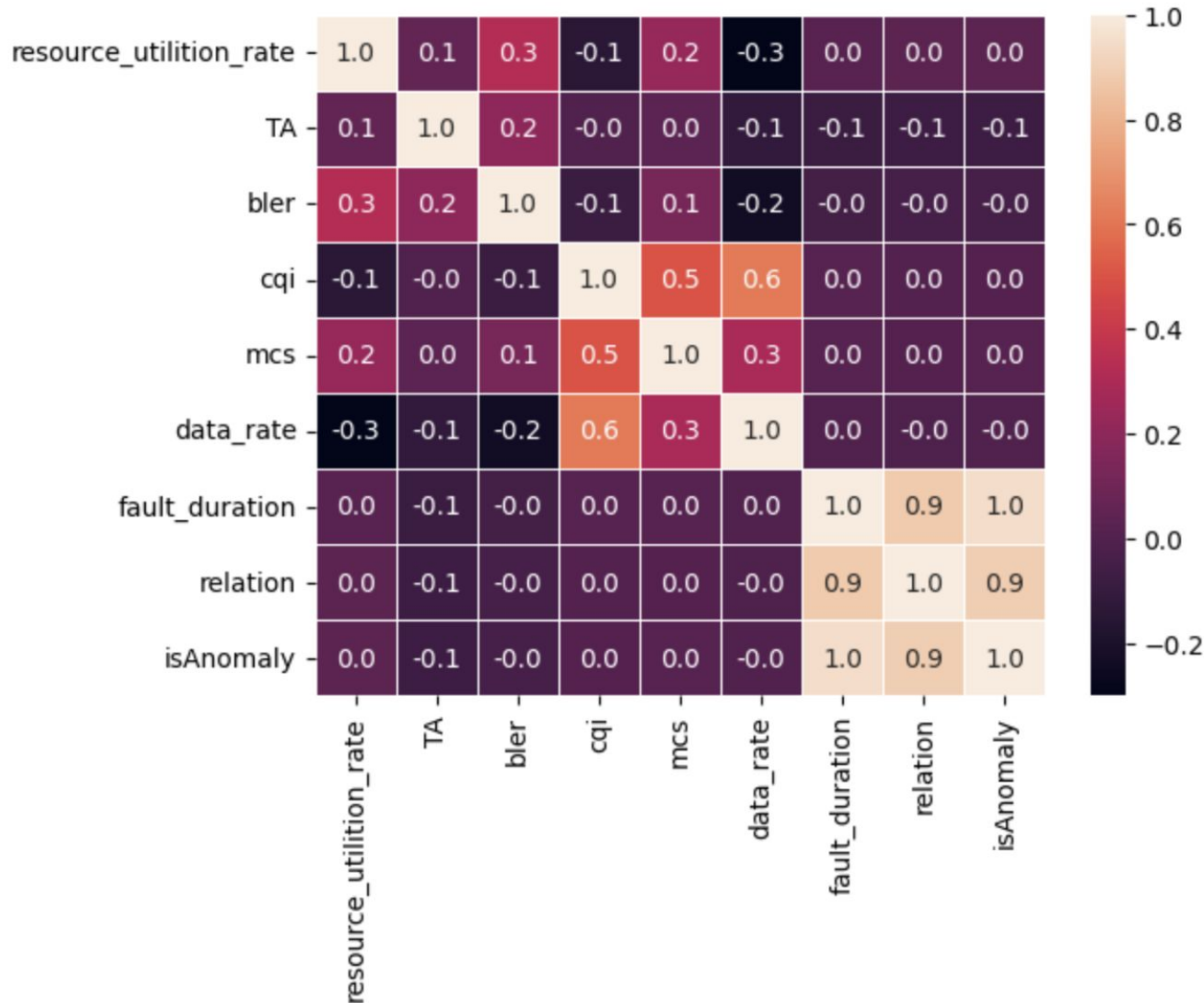
Distribuição dos dados



Gráficos de Dispersão



Matriz de Correlação (Pearson)



Matriz de Correlação

- Pearson considera a correlação linear dos dados (Spearman).
- Através do Scatter Plotter, não identificamos uma linearidade entre eles.
- Com os modelos estatísticos escolhidos, conseguimos analisar uma relação mais precisa.

Testes Não Paramétricos

- Teste de "spearman"
- Teste de "Kendall"
- Teste "Anova"
- Teste de "Bartlett"

Teste de Bartlett

- Avaliar a igualdade de variâncias entre diferentes grupos
- Aplicado antes de testes de análise de variância (ANOVA) para verificar se as variâncias dos grupos são estatisticamente iguais
- H_0 : As variâncias são iguais em todas as amostras
- H_1 : As variâncias são diferentes

Teste ANOVA

- Análise de Variância é usado para comparar as médias de três ou mais grupos
 - Ele determina se há diferenças estatisticamente significativas entre as médias dos grupos
 - Requer a normalidade dos dados
 - Assume que as variâncias dos grupos são estatisticamente iguais
-
- H_0 : As médias são iguais em todas as amostras.
 - H_1 : As médias são diferentes.

Teste de Bartlett e ANOVA

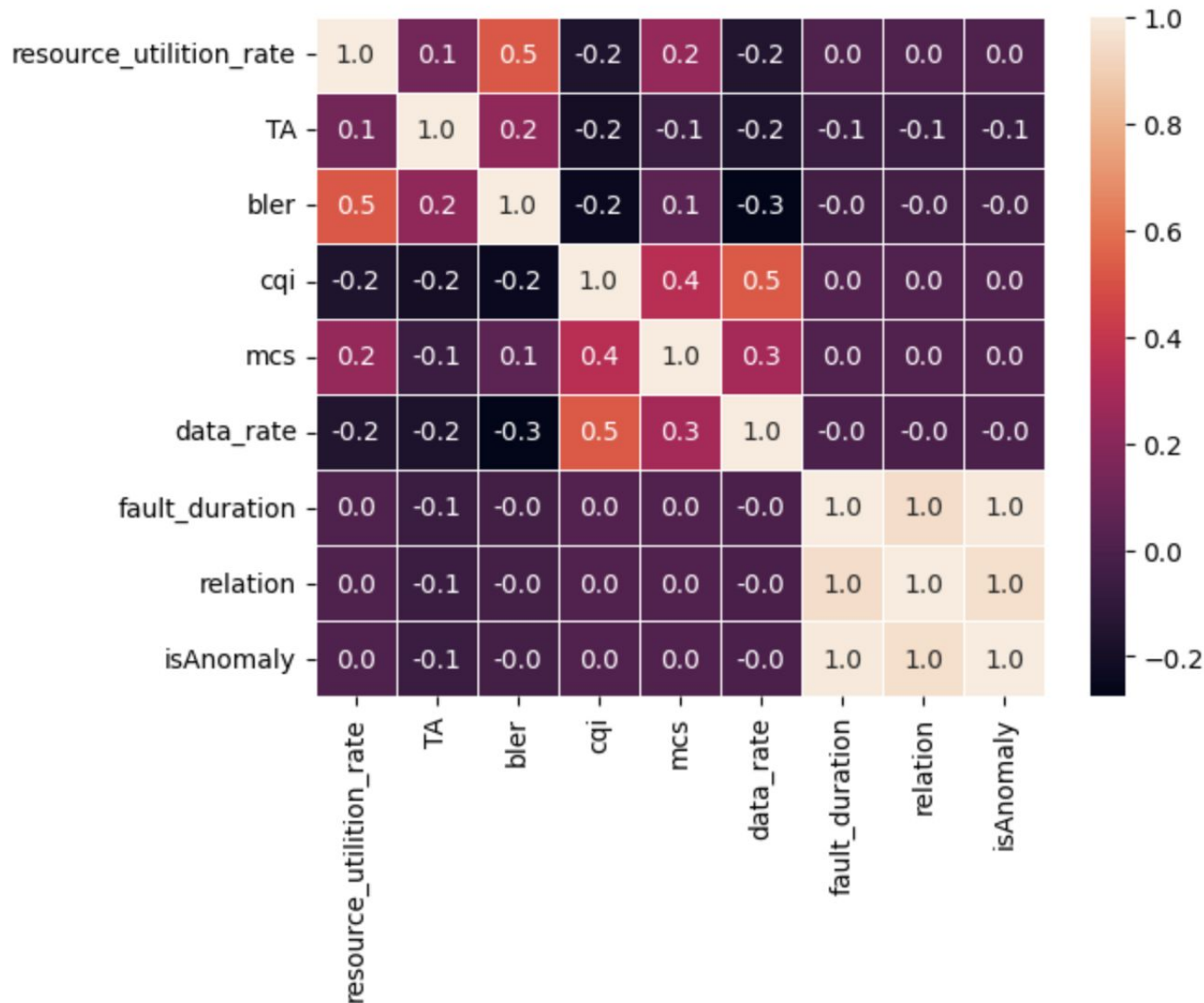
Teste de Bartlett:

Estatística de Bartlett: 4004.4448749261273

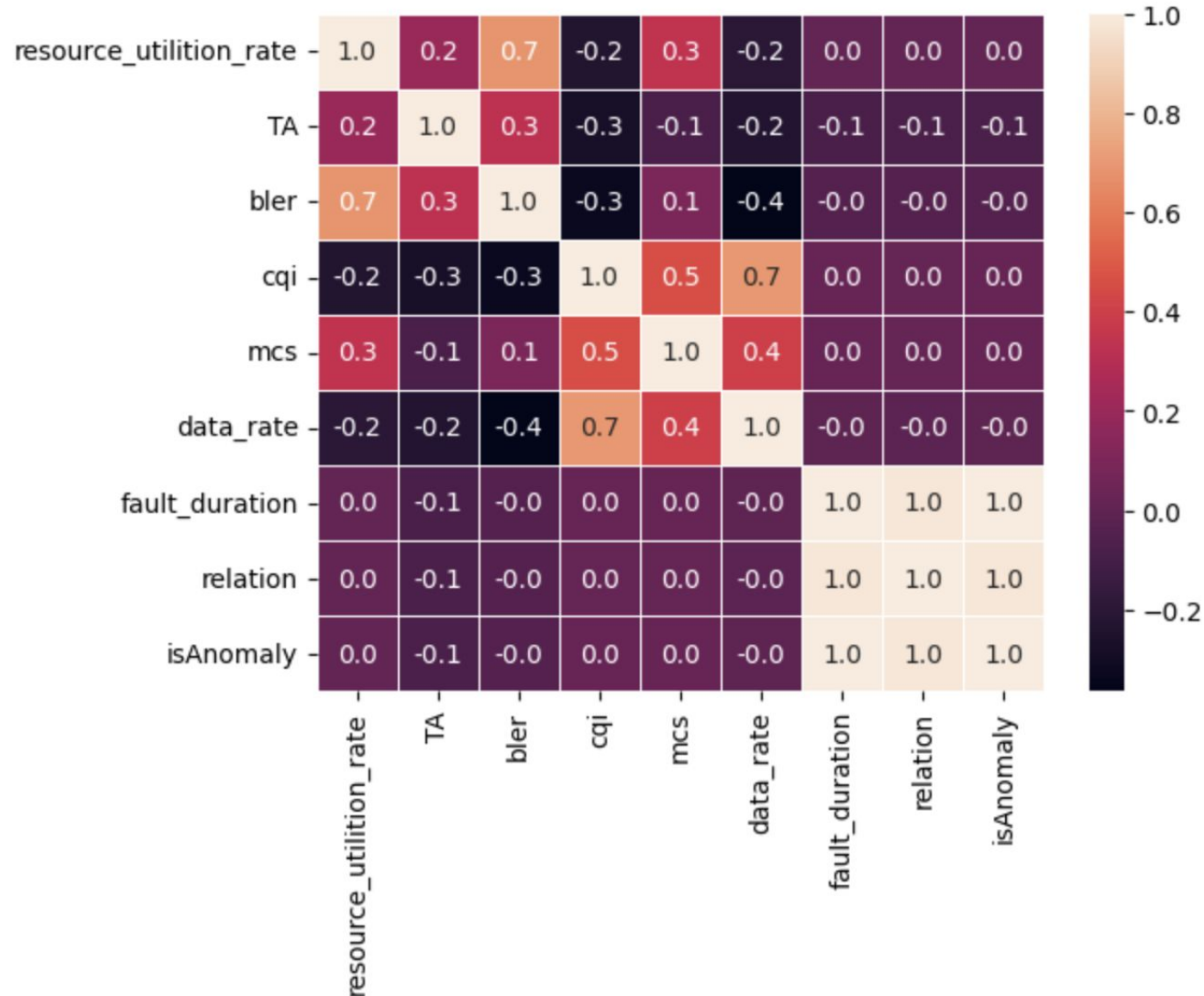
Valor p: 0.0

- Como o valor p deu ≤ 0.05 , o teste rejeitou a hipótese nula de que as variâncias das amostras são estatisticamente iguais.
- Como o teste ANOVA requer que as variâncias sejam estatisticamente iguais, não podemos aplicá-lo.

Matriz de Correlação (Kendall)



Matriz de Correlação (Spearman)



Testes Não Paramétricos

- Teste de Spearman (Entre RU e bler):

coeficiente de correlação de Spearman: 0.6882051668612339

Valor p: 0.0

Rejeitar H_0 – há evidências de diferença significativa.

- Teste de Kendall:

Correlação entre resource_utilition_rate e TA: 0.1328795142453435

Correlação entre resource_utilition_rate e bler: 0.525572663453551

Correlação entre resource_utilition_rate e cqi: -0.16474462098358736

Correlação entre resource_utilition_rate e mcs: 0.23699034281958678

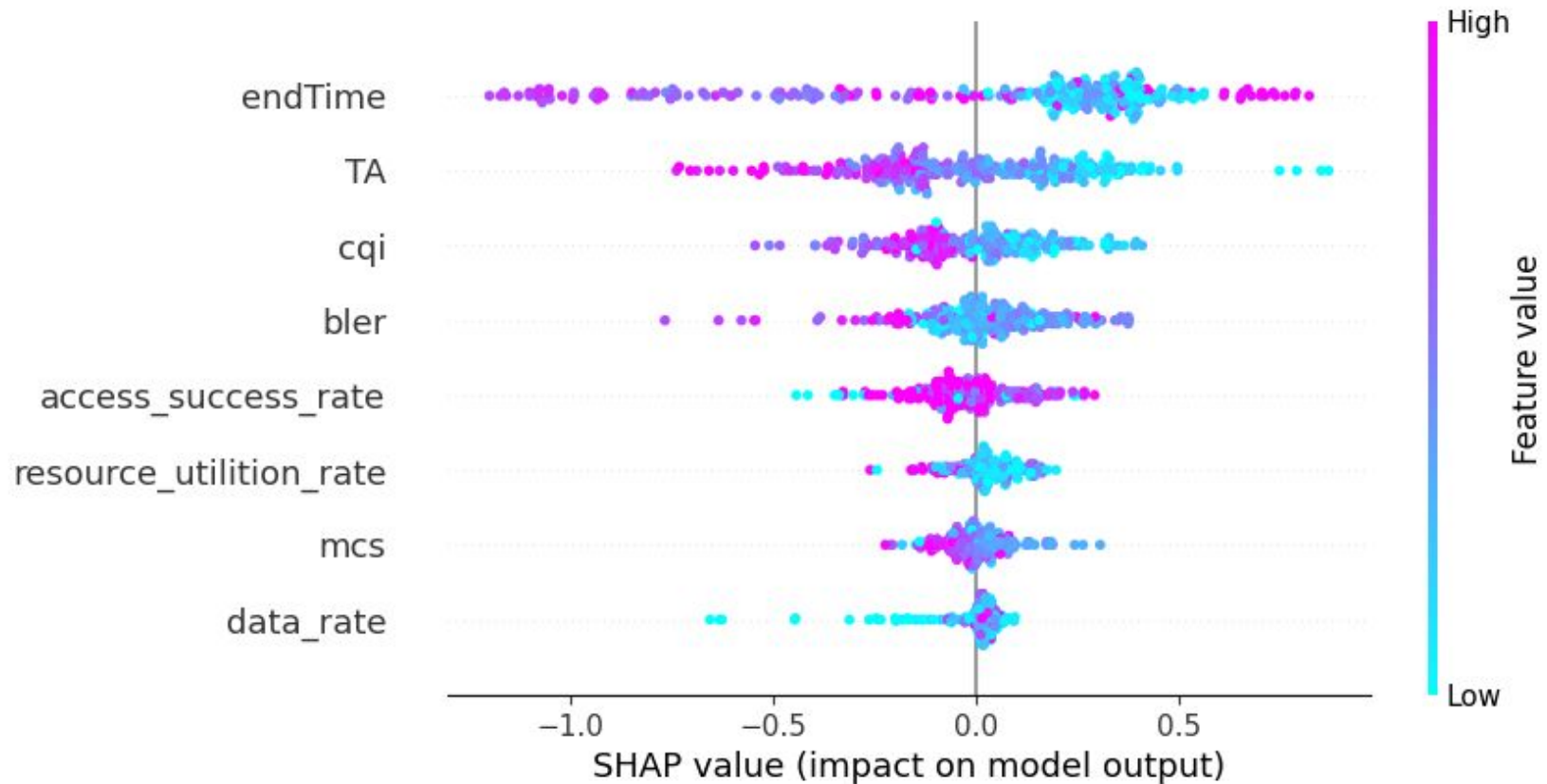
Correlação entre resource_utilition_rate e data_rate: -0.155137728190413

Correlação entre resource_utilition_rate e fault_duration: 0.009550192780816518

Correlação entre resource_utilition_rate e relation: 0.00901534065704249

Correlação entre resource_utilition_rate e isAnomaly: 0.010287335762897462

Shap Values



Modelos de Aprendizagem Levantados

- Árvore de Decisão, com a técnica SMOTE.
- *Feature importance* e *Feature engineering*.
- MLP's (Multilayer Perceptron), fazendo uso do framework Tensorflow + Keras para construção das redes neurais.
- Sklearn Regressor.
- XGBRegressor.

O que apresentou um melhor desempenho foi Tensorflow Keras 2, com score de 0.59.

Conclusões

- Os testes de Spearman e Kendall emergem como ferramentas mais apropriadas para avaliar a relação entre variáveis em conjuntos de dados não lineares, uma vez que o uso de Scatter Plots revelou a natureza não linear dos dados apresentados, enquanto a matriz de correlação de Pearson, por sua vez, é mais adequada para dados lineares.
- Essa abordagem proporciona uma análise mais precisa, destacando valores próximos que refletem a verdadeira associação entre as variáveis.