Conforme a matriz de correlação abaixo, recomenda-se excluir as seguintes variáveis:

* regiao\_ibge
* qtd\_comodos\_dormitorio\_fam
* cod\_banheiro\_domic\_fam

**Correlações acima de 0.8:**

Colunas: uf\_ibge e regiao\_ibge - Correlação: 0.97

Colunas: cod\_local\_domic\_fam e cod\_especie\_domic\_fam - Correlação: 0.73

Colunas: qtd\_comodos\_domic\_fam e qtd\_comodos\_dormitorio\_fam - Correlação: 0.81

Colunas: cod\_agua\_canalizada\_fam e cod\_abaste\_agua\_domic\_fam - Correlação: 0.74

Colunas: cod\_agua\_canalizada\_fam e cod\_banheiro\_domic\_fam - Correlação: 0.83

Gráfico

Descrição gerada automaticamente

**Resultados**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | DecisionTree | | RandomForest | | XGBoost | | CatBoost | |
| SB | CB | SB | CB | SB | CB | SB | CB |
| Variaveis sem tratamento | | | | | | | | |
| F1 Score | 0.4816 | 0.5999 | 0.5267 | 0.6620 | 0.5378 | 0.6514 | 0.5383 | 0.6509 |
| Acurácia | 0.6394 | 0.6008 | 0.6437 | 0.6636 | 0.6623 | 0.6544 | 0.6653 | 0.6536 |
| Variaveis normalizadas | | | | | | | | |
| F1 Score | 0.4816 | 0.5987 | 0.5296 | 0.6895 | 0.5378 | 0.6645 | 0.5396 | 0.6629 |
| Acurácia | 0.6394 | 0.6075 | 0.6452 | **0.6921** | 0.6623 | **0.6703** | 0.6658 | 0.6686 |
| Variaveis dumies | | | | | | | | |
| F1 Score | 0.4660 | 0.5941 | 0.5158 | 0.6926 | 0.5376 | 0.6576 | 0.5357 | 0.6603 |
| Acurácia | 0.6359 | 0.6090 | 0.6360 | **0.6941** | 0.6628 | 0.6629 | 0.6640 | 0.6663 |

**Com Balanceamento**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Variaveis sem tratamento | Variaveis normalizadas | Variaveis dumies |
| Decision Tree | qtde\_pessoas  dias\_cadastramento  cod\_material\_piso\_fam  dias\_atualizacao  uf\_ibge  qtd\_comodos\_domic\_fam  cod\_iluminacao\_domic\_fam  ind\_parc\_mds\_fam  cod\_calcamento\_domic\_fam  classf | qtde\_pessoas  dias\_cadastramento  cod\_material\_piso\_fam  uf\_ibge  dias\_atualizacao  qtd\_comodos\_domic\_fam  ind\_parc\_mds\_fam  cod\_iluminacao\_domic\_fam  cod\_local\_domic\_fam  classf | qtde\_pessoas  cod\_material\_piso\_fam\_5  dias\_cadastramento  qtd\_comodos\_domic\_fam  dias\_atualizacao  uf\_ibge\_35  uf\_ibge\_21  cod\_material\_piso\_fam\_4  uf\_ibge\_41  classf\_3 |
| Random Forest | qtde\_pessoas  dias\_cadastramento  uf\_ibge  qtd\_comodos\_domic\_fam  dias\_atualizacao  cod\_material\_piso\_fam  classf  cod\_calcamento\_domic\_fam  cod\_escoa\_sanitario\_domic\_fam  cod\_material\_domic\_fam | qtde\_pessoas  dias\_cadastramento  uf\_ibge  qtd\_comodos\_domic\_fam  dias\_atualizacao  cod\_material\_piso\_fam  classf  cod\_calcamento\_domic\_fam  cod\_escoa\_sanitario\_domic\_fam  cod\_material\_domic\_fam | qtde\_pessoas  dias\_cadastramento  qtd\_comodos\_domic\_fam  dias\_atualizacao  cod\_material\_piso\_fam\_5  cod\_material\_piso\_fam\_2  cod\_escoa\_sanitario\_domic\_fam\_1  cod\_calcamento\_domic\_fam\_1  uf\_ibge\_35  cod\_iluminacao\_domic\_fam\_1 |
| XGBoost | qtde\_pessoas  dias\_cadastramento  dias\_atualizacao  uf\_ibge  qtd\_comodos\_domic\_fam  cod\_material\_piso\_fam  cod\_iluminacao\_domic\_fam  cod\_local\_domic\_fam  ind\_parc\_mds\_fam  cod\_escoa\_sanitario\_domic\_fam | qtde\_pessoas  uf\_ibge  dias\_cadastramento  dias\_atualizacao  qtd\_comodos\_domic\_fam  cod\_material\_piso\_fam  classf  cod\_iluminacao\_domic\_fam  ind\_parc\_mds\_fam  cod\_local\_domic\_fam | qtde\_pessoas  dias\_cadastramento  dias\_atualizacao  qtd\_comodos\_domic\_fam  cod\_material\_piso\_fam\_5  cod\_iluminacao\_domic\_fam\_1  classf\_1  uf\_ibge\_41  uf\_ibge\_29  uf\_ibge\_35 |
| CatBoost | qtde\_pessoas  uf\_ibge  dias\_atualizacao  dias\_cadastramento  qtd\_comodos\_domic\_fam  cod\_material\_piso\_fam  ind\_parc\_mds\_fam  cod\_iluminacao\_domic\_fam  cod\_local\_domic\_fam  classf | qtde\_pessoas  uf\_ibge  qtd\_comodos\_domic\_fam  dias\_cadastramento  dias\_atualizacao  cod\_material\_piso\_fam  classf  ind\_parc\_mds\_fam  cod\_iluminacao\_domic\_fam  cod\_local\_domic\_fam | qtde\_pessoas  qtd\_comodos\_domic\_fam  dias\_cadastramento  dias\_atualizacao  cod\_material\_piso\_fam\_5  uf\_ibge\_41  uf\_ibge\_29  uf\_ibge\_35  uf\_ibge\_51  uf\_ibge\_43 |

Melhores colunas:

* qtde\_pessoas
* uf\_ibge
* qtd\_comodos\_domic\_fam
* dias\_cadastramento
* dias\_atualizacao
* cod\_material\_piso\_fam
* classf
* ind\_parc\_mds\_fam
* cod\_iluminacao\_domic\_fam
* cod\_local\_domic\_fam