

Uso de Inteligência artificial para determinação da resistência à compressão no concreto simples

AMANDA ISABELA DE CAMPOS (DRE: 120074842)

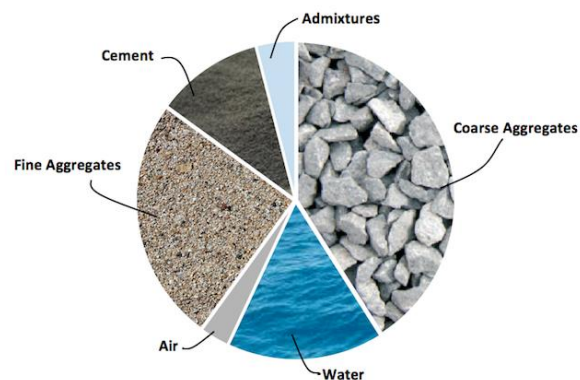
DISCIPLINA: INTRODUÇÃO À CIÊNCIA DOS DADOS (COC800)

PROF.: ROGÉRIO ESPÍNDOLA



I. Introdução

- Aplicar algoritmos de IA em um conjunto de dados conhecido como *Concrete Compressive Strength Data Set* [1], dessa forma é possível prever a resistência compressão (f_{ck});
- Atributos de entrada: idade do concreto e ingredientes que compõem a mistura;
- O intuito dessa análise de dados é permitir que empresas e/ou laboratórios possam ter maior controle dos concretos produzidos no pátio de obras.

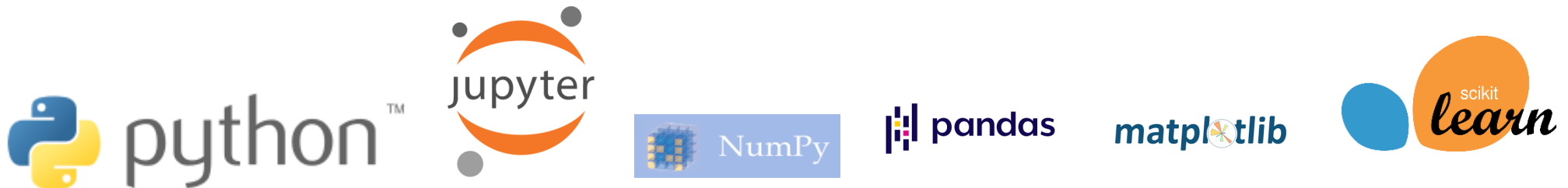


I. Introdução

- O trabalho de Yeh [2] analisou uma parcela desse conjunto de dados utilizando Redes Neurais;
 - A proposta deste trabalho engloba fazer análises com diversos algoritmos de I.A. ampliando a base de algoritmos utilizadas para esse tipo de problema que é tradicional na área de Engenharia Civil.
 - Conjunto de dados: (UCI):
 - composto por 9 atributos para 1030 registros
 - com uma variável de saída (f_{ck}).
- Variáveis de Entrada:
 - 1 - Cimento -- quantitativo -- kg em uma mistura m3
 - 2 - Escória de alto-forno -- quantitativa -- kg em uma mistura m3
 - 3 - Cinzas volantes -- quantitativo -- kg em uma mistura m3
 - 4 - Água -- quantitativo -- kg em uma mistura m3
 - 5 - Superplastificante -- quantitativo -- kg em um mistura m3
 - 6 - Agregado Graúdo -- quantitativo -- kg em uma mistura m3
 - 7 - Agregados Finos -- quantitativo -- kg em uma mistura m3
 - 8 - Idade -- quantitativa -- Dia (1~365)

2. Materiais e Métodos

- Linguagem de programação Python: Numpy, Pandas, Matplotlib, Seaborn e ScikitLearn.
- Técnica de validação cruzada ($k = 10$)
- Modelos de Regressão Linear, Regressão LASSO, Regressão de Ridge, K vizinhos mais próximos (KNN), AdaBoost, Árvores de Decisão, Florestas aleatórias e Gradiente Boosting.



2.1 Modelos de regressão

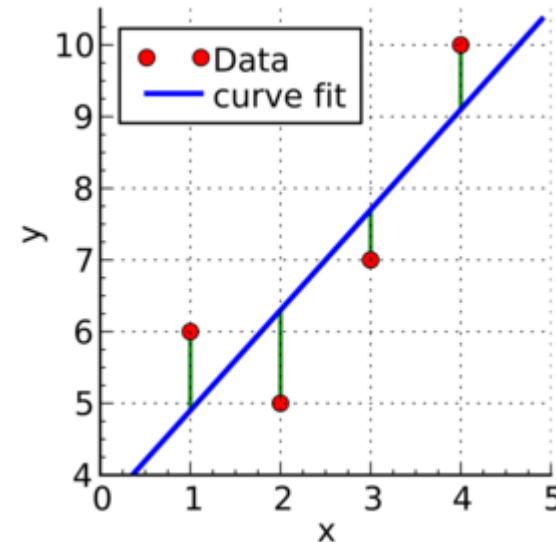
- Regressão Linear

$$\hat{y}(w, x) = w_0 + w_1x_1 + \dots + w_px_p$$

$$\min_w ||Xw - y||_2^2$$

- Regressão Lasso

$$\min_w \frac{1}{2n_{\text{samples}}} ||Xw - y||_2^2 + \alpha ||w||_1$$

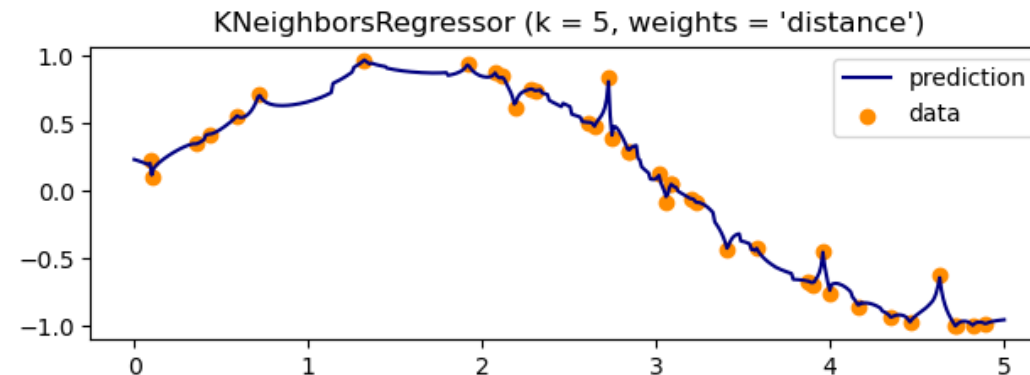


2.1 Modelos de regressão

- Regressão de Ridge

$$\min_w ||Xw - y||_2^2 + \alpha ||w||_2^2$$

- Regressão KNN (K vizinhos mais próximos)



2.1 Modelos de regressão

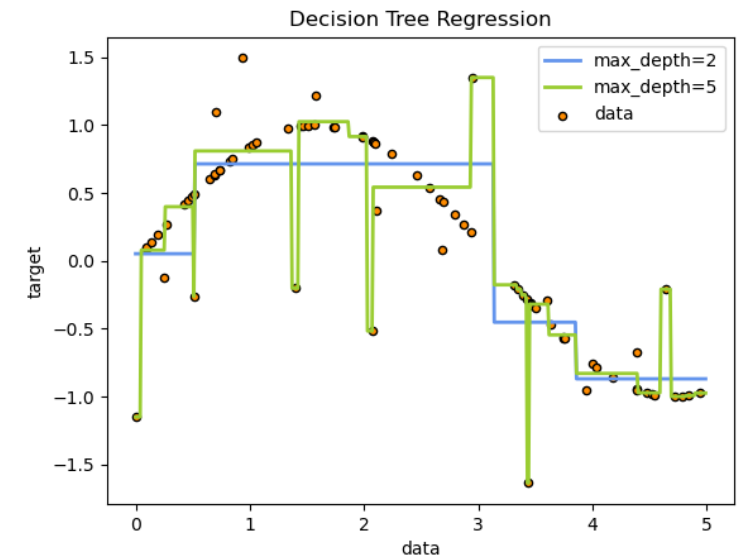
■ Regressão AdaBoost

- Combina hipóteses fracas dentro de uma única hipótese de predição que, de acordo com o esperado, será mais precisa que qualquer outra gerada anteriormente;
- Comparado à estratégias evolucionárias.

$$L_t(x_i) = 1 - \exp\left(-\frac{|f_t(x_i) - y(x_i)|}{\max_{i=1\dots m} |f_t(x_i) - y(x_i)|}\right)$$

■ Regressão em árvore de decisão

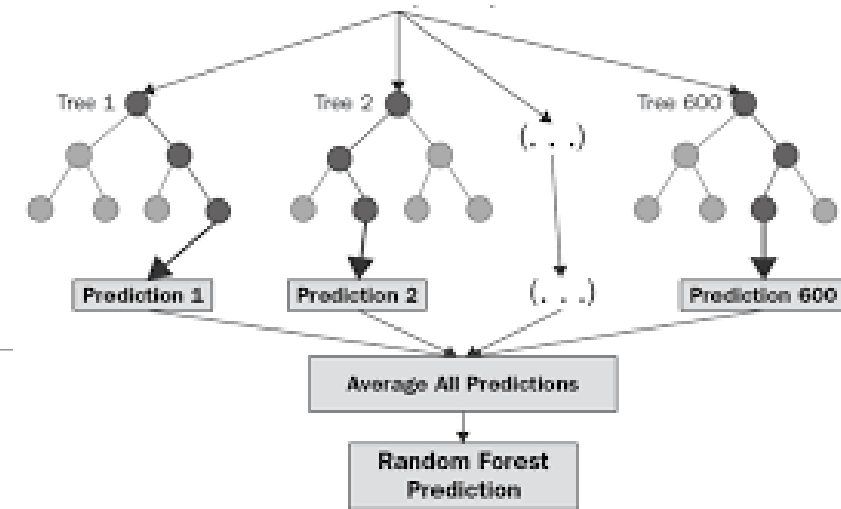
- Hierarquia de nós conectados por ramos. O nó interno é a unidade de tomada de decisão que avalia através de teste lógico qual será o próximo nó descendente ou filho.
- Nó folha está associado a um valor.



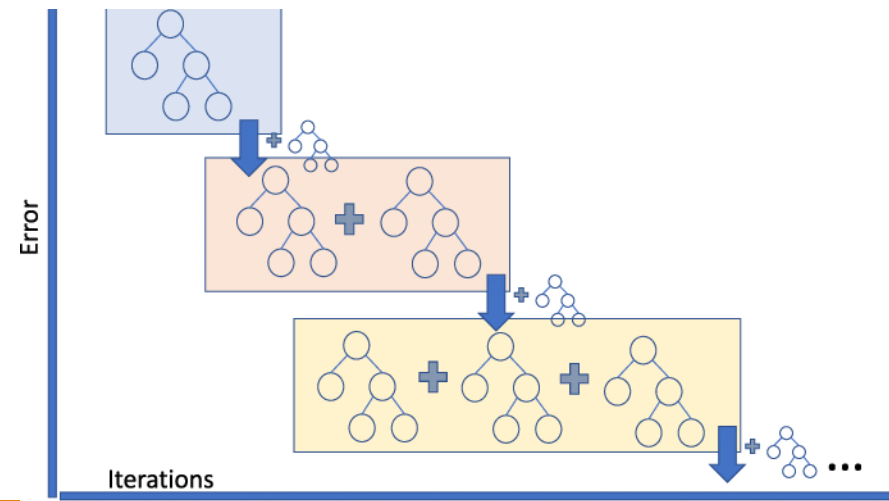
2.1 Modelos de regressão

■ Regressão de floresta aleatória

- É criada uma “floresta” aleatória com uma combinação (*ensemble*) de árvores de decisão;
- A ideia do método é que a combinação dos modelos de aprendizado aumenta o resultado geral.



■ Regressão Gradient Boosting



2.2 Estatísticas de Avaliação

- Raiz quadrada do erro médio quadrático REMQ (em inglês *RMSE*)

$$REMQ = \sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{y}(t) - y(t))^2}$$

- Coeficiente de determinação R^2

$$R^2 = \frac{\sum_{t=1 \dots N} (\hat{y}(t) - y(t))^2}{\sum_{t=1 \dots N} (y(t) - \bar{y})^2}$$

3. Pré-processamento

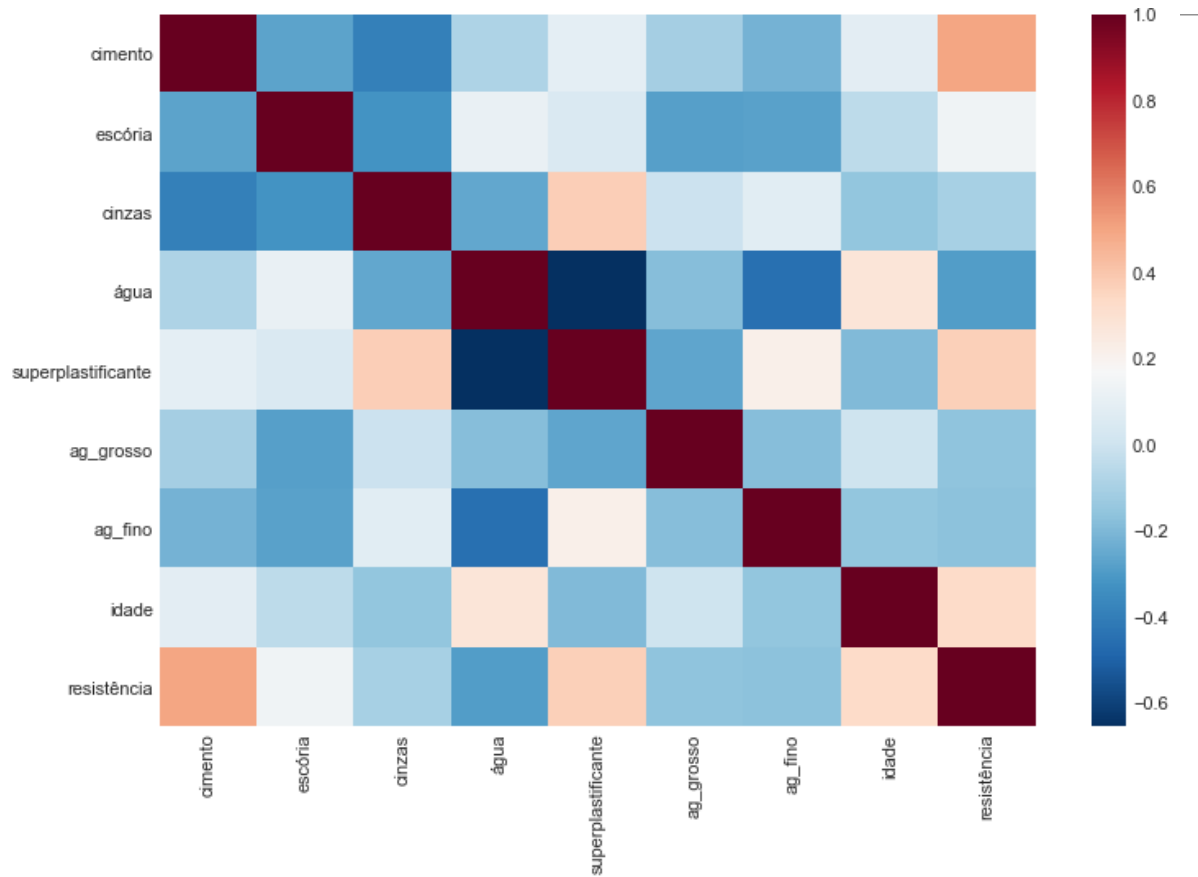


Figura 2. Matriz de correlação dos atributos com mapa de cores

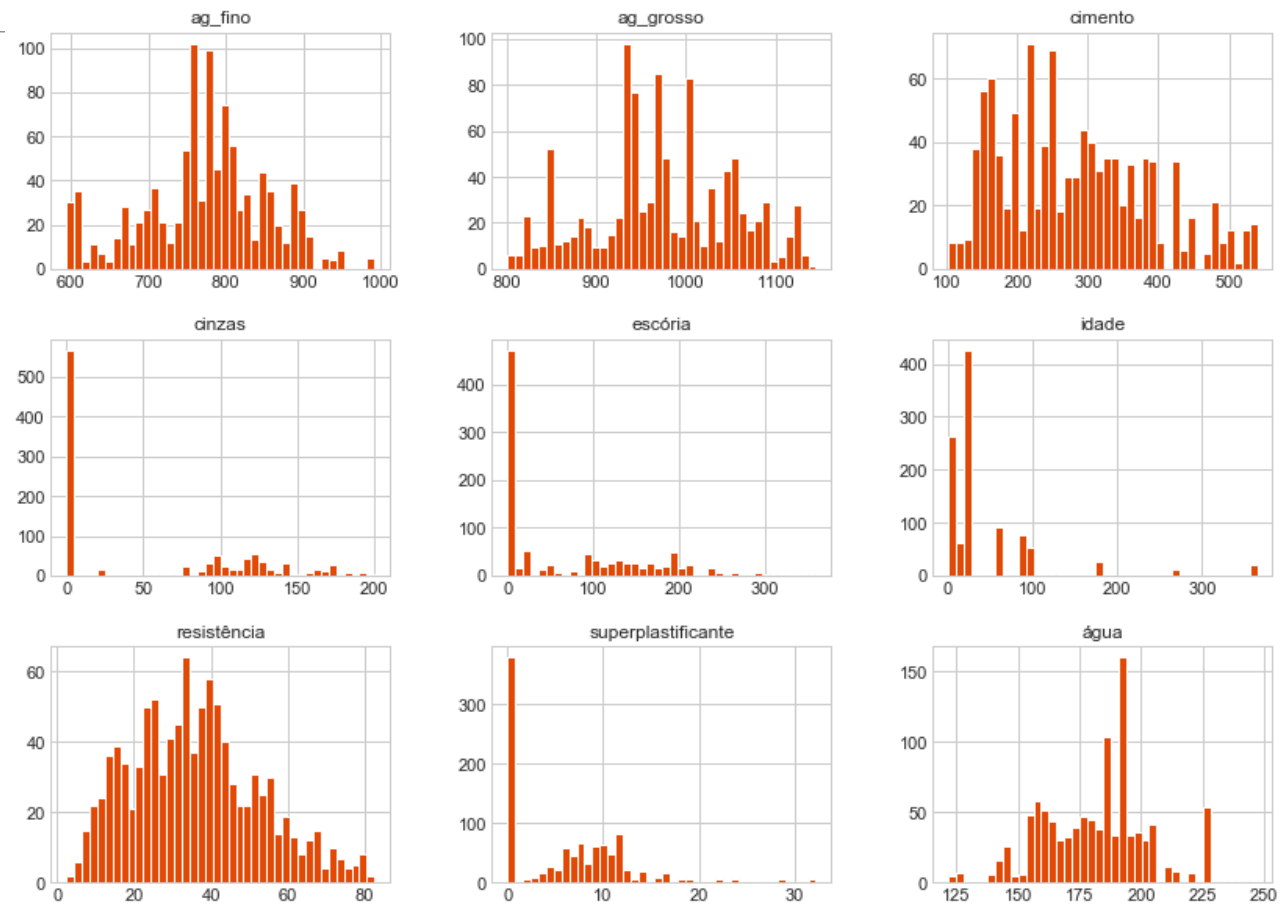


Figura 3. Histogramas dos atributos

3. Pré-processamento

- Não existem valores ausentes

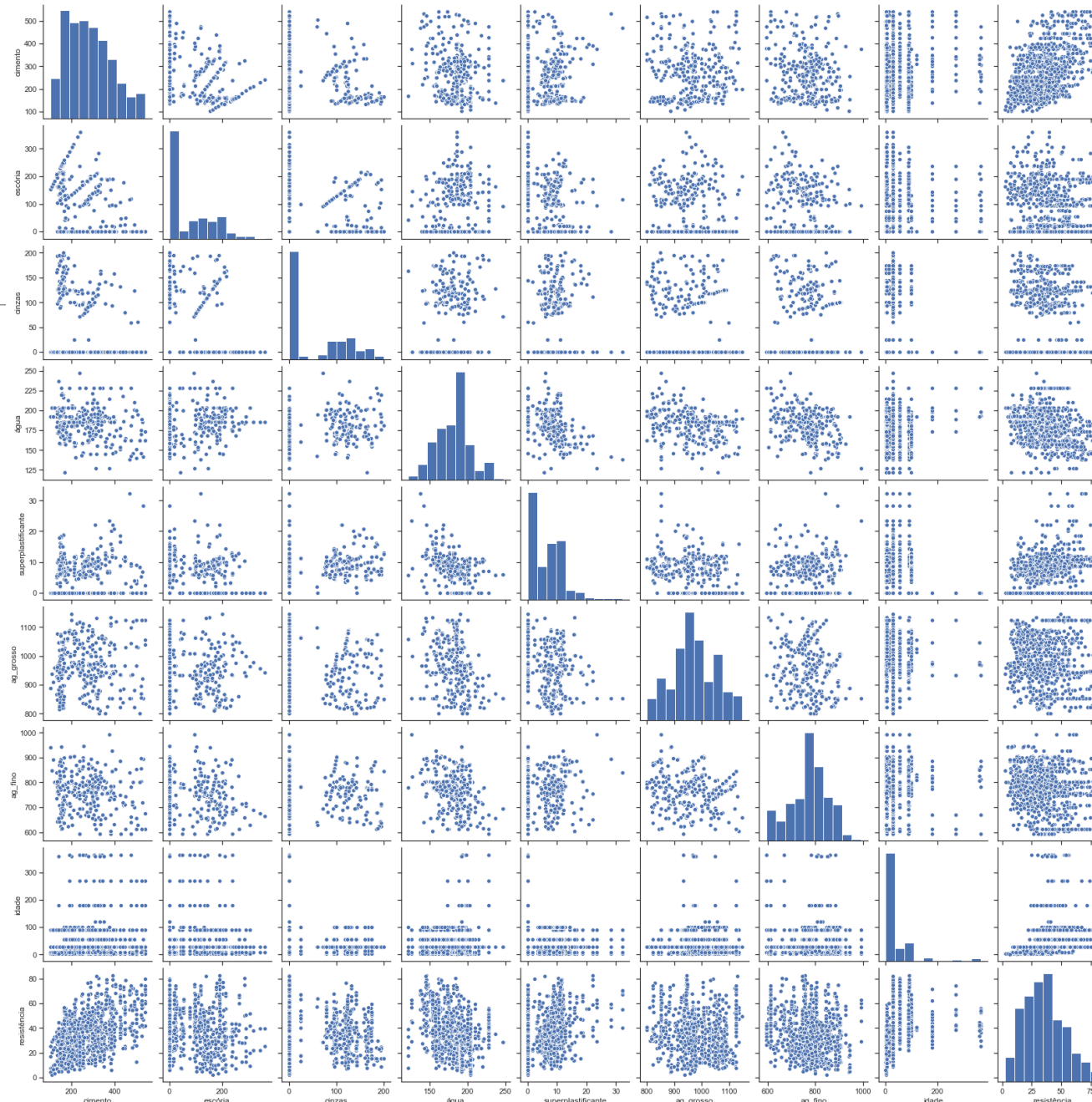


Figura 5. Gráfico de Projeção

4. Resultados

Tabela 1. Resultados dos diferentes modelos de inteligência artificial

	Treino e Teste*		Validação Cruzada (cv =10)	
	REQM	R2	REQM	R2
Regressão Linear	10.9439	0.6013	10.253646	0.620875
Regressão LASSO	11.7651	0.5392	10.830778	0.579454
Regressão de Ridge	10.9435	0.6013	10.252755	0.620908
K vizinhos mais próximos	9.1241	0.7228	9.028723	0.705960
AdaBoost	8.2736	0.7721	7.589431	0.786656
Árvores de Decisão	9.0447	0.7276	7.298293	0.813419
Florestas aleatórias	6.2575	0.8696	5.341337	0.896260
Gradiente Boosting	6.4358	0.8621	5.163711	0.903980
Gradiente Boosting (gridsearchcv)	4.6268	0.9165	5.2943	0.9074

4. Resultados

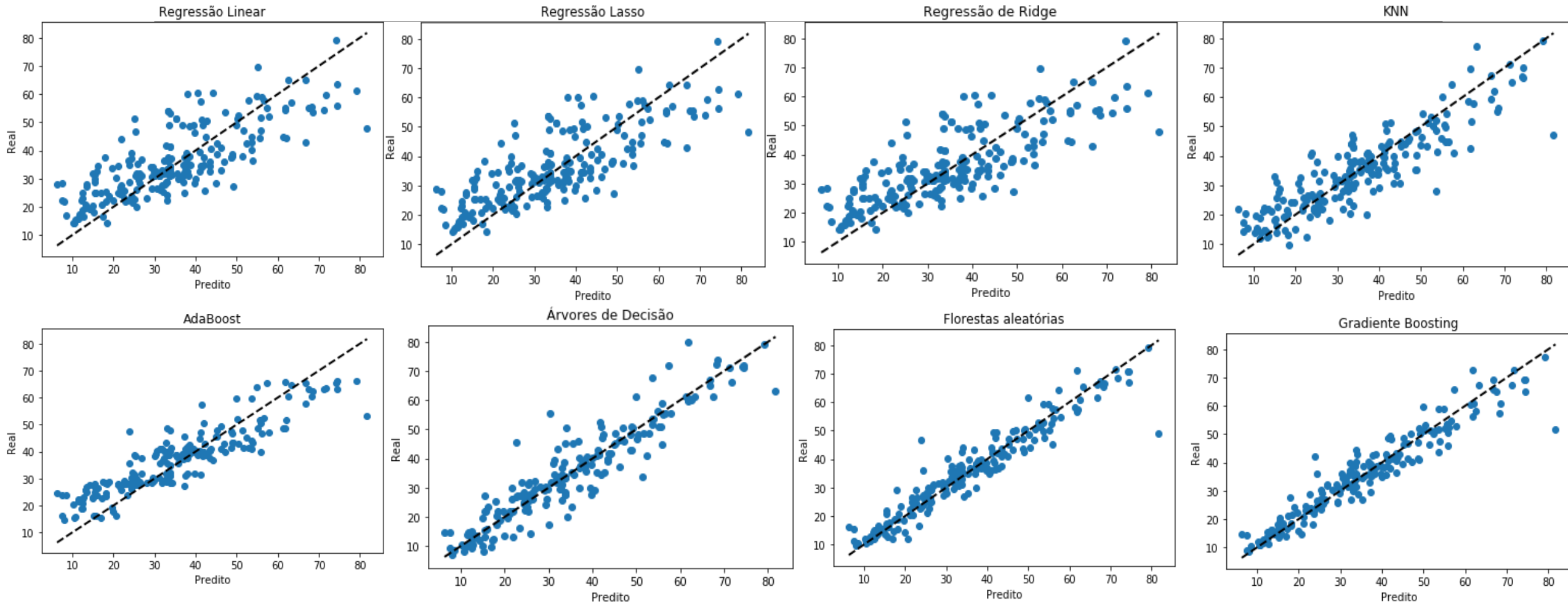
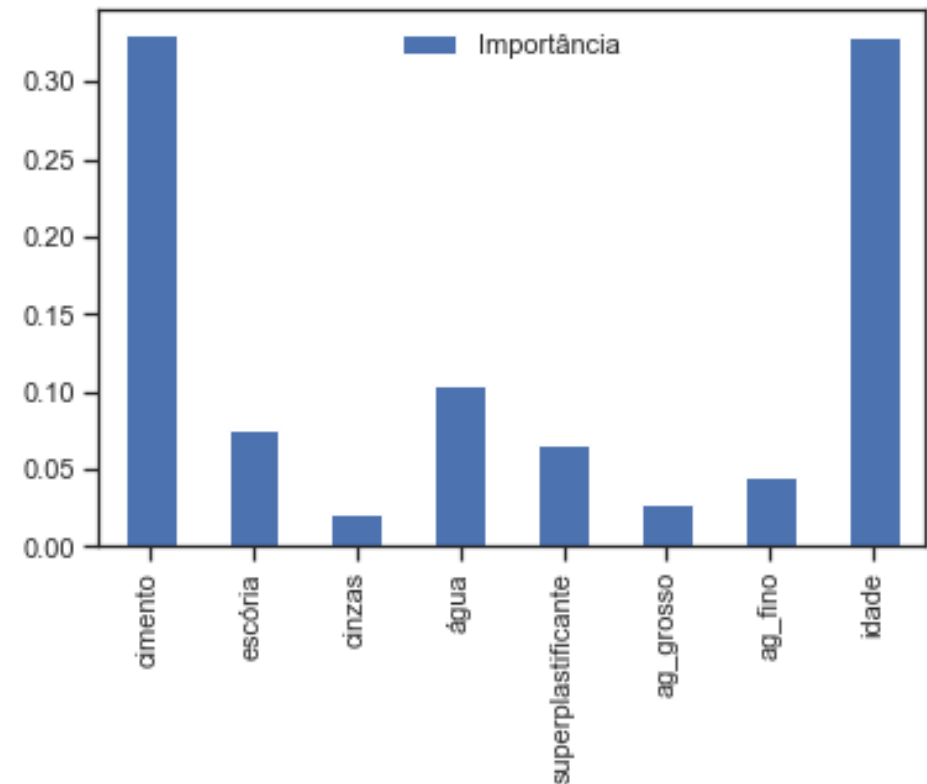


Figura 6. Correlação real x predito para os diferentes modelos

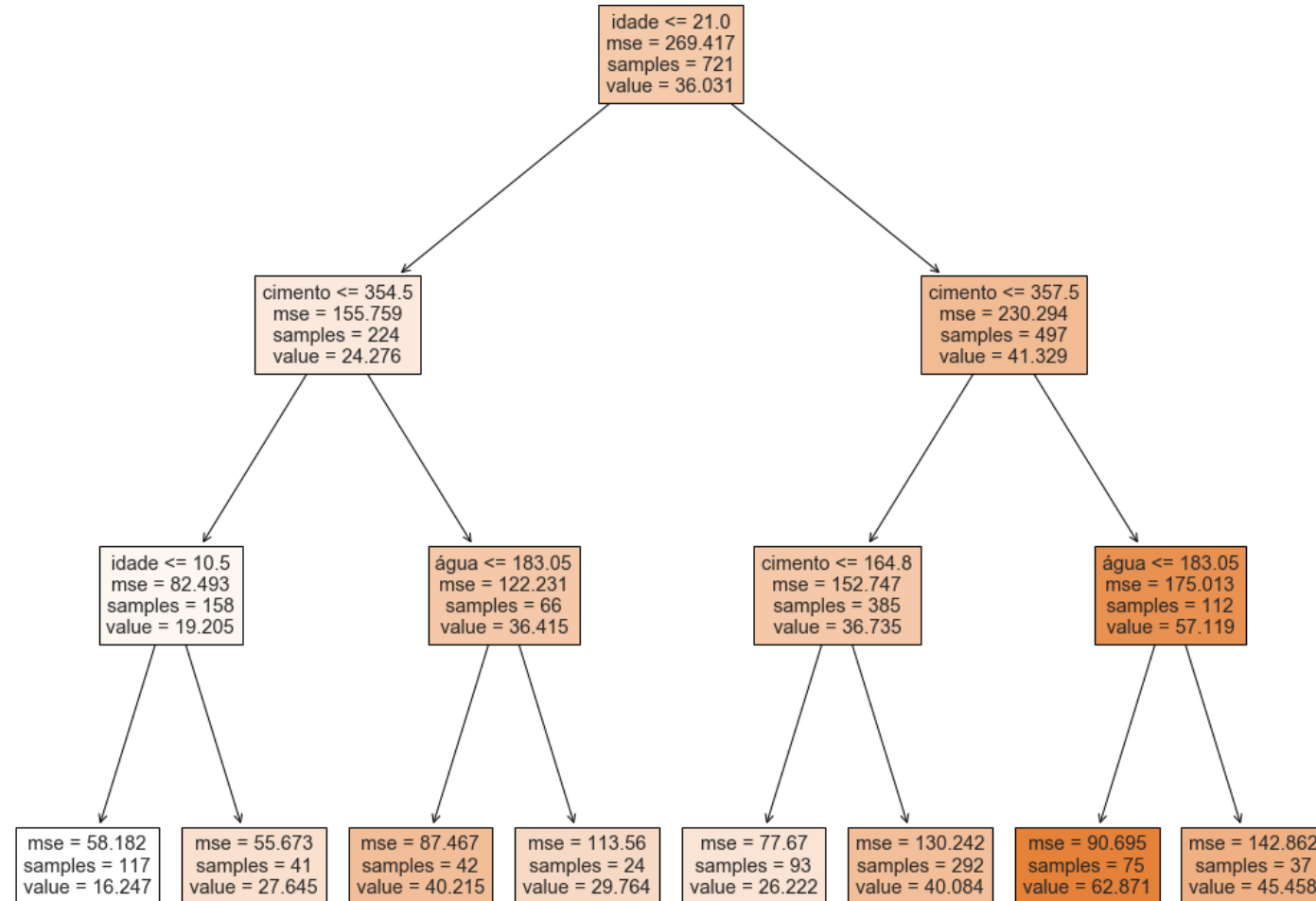
4.1 Importância das características

- Quanto mais características, mais provavelmente o modelo irá sofrer superajuste (*overfitting*);
- Para o modelo de floresta aleatória:

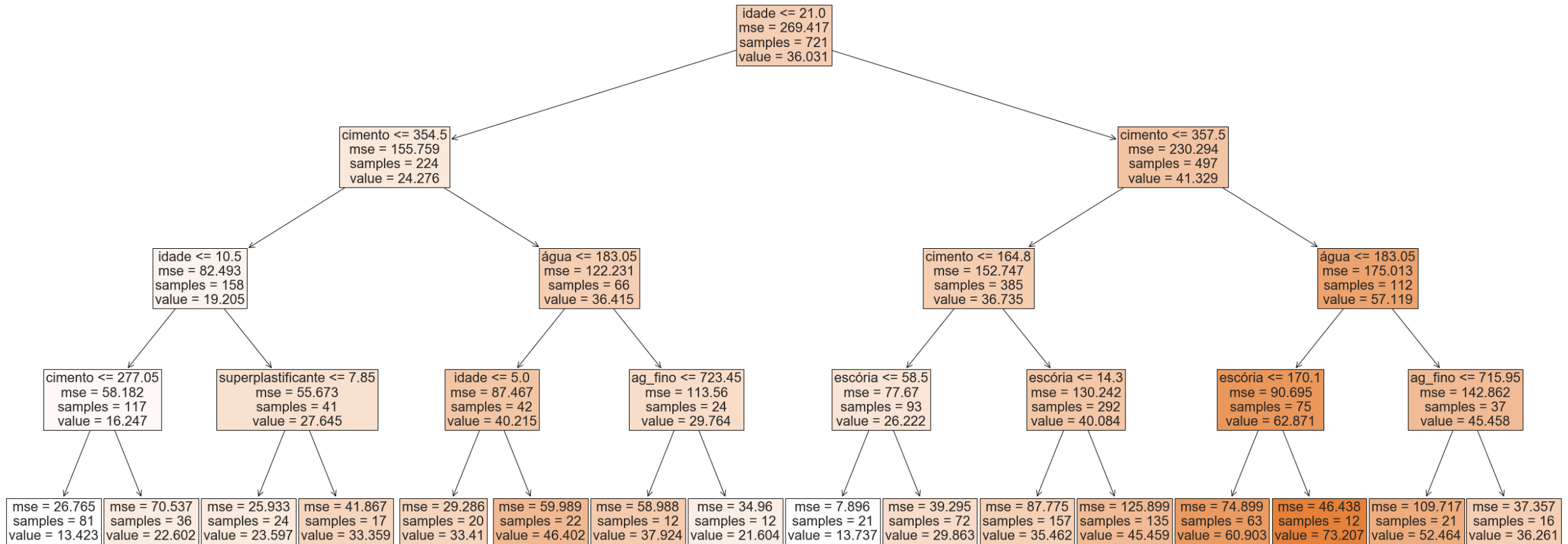
	Importância
cimento	0.330079
escória	0.075474
cinzas	0.021808
água	0.103928
superplastificante	0.066722
ag_grosso	0.027455
ag_fino	0.046049
idade	0.328486



Árvore de decisão do conjunto de dados



Árvore de decisão do conjunto de dados



4.2 Busca de parâmetros Gradient Boosting

- GridSearchCV

```
params = {'n_estimators': 500, 'max_depth': 5, 'min_samples_split': 2, 'learning_rate': 0.01, 'loss': 'ls'}
```

5. Conclusões

- Vantagens da aplicação da validação cruzada;
- Modelos mais simples e rápidos do que redes neurais e com resultados próximos;
- Melhor o controle de qualidade dos concretos produzidos no pátio de obras;
- Prever a resistência do concreto antes de executar ensaios exaustivos e demorados.