



Aplicação de Aprendizado de Máquina para o Comportamento de Areias Puras

Lucas Marques Pires da Silva

Mestrando, COPPE/UFRJ, Rio de Janeiro, Brasil, lucas.silva@coc.ufrj.br

Gustavo Vaz de Mello Guimarães, D.Sc.

Docente, UFRJ, Macaé, Brasil, gvmg@poli.ufrj.br

RESUMO: O conceito de aprendizado de máquina vem sendo empregado largamente em diversas áreas da indústria, mas na geotecnia esse conceito ainda não é tão explorado. Essas técnicas, que inclusive são creditadas de prever o comportamento humano, tem o potencial de prever o comportamento do solo. Porém, cabe ressaltar que para o aprendizado de máquina funcionar bem, ele deve ser “ensinado” a partir de um banco de dados confiável. A partir de um banco de dados de ensaios de cisalhamento direto para as areias puras das praias de Ipanema (Rio de Janeiro), Itaipuaçu (Niterói) e Cavaleiros (Macaé), o aprendizado de máquina é capaz de prever se o comportamento de uma areia pura, predominantemente quartzosa, é contrátil ou dilatante. Este banco de dados foi gerado a partir de resultados de ensaios de cisalhamento direto. As amostras ensaiadas possuíam diferentes densidades relativas e foram submetidas a distintas tensões confinantes, assim foi possível identificar se o solo ensaiado possuía comportamento dilatante ou contrátil. Um código de aprendizado de máquina foi escrito com as seguintes metodologias: *Dummy Classifier*, *LinearSVC Classifier*, *MultinomialNB Classifier*, *AdaBoost Classifier*, *OneVsRest Classifier*, *OneVsOne Classifier* e *Decision Tree*. O algoritmo apresentou resultados excelentes mostrando que é uma prova de tecnologia e que para um banco de dados mais abrangente mais perguntas poderão ser respondidas.

PALAVRAS-CHAVE: Aprendizado de máquina; Areias puras; Comportamento de areias.

ABSTRACT: The concept of machine learning has been widely used in several areas of industry, but in geotechnics, this concept is still not so explored. These techniques, widely used to predict human behavior, can also predict soil behavior. However, it is noteworthy that for machine learning to work well, it must be “taught” from a reliable database. Based on a database of direct shear tests for the pure sands of the beaches of Ipanema (Rio de Janeiro), Itaipuaçu (Niterói) and Cavaleiros (Macaé), machine learning is able to predict whether the behavior of a sand pure, predominantly quartzose, is contractile or expansive. This database was generated from the direct shear test. Samples had different relative density and submitted to many confining stress. Therefore, it was possible to identify the samples behavior like expansive or contractile. Machine learning code was written with the following methodologies: *Dummy Classifier*, *LinearSVC Classifier*, *MultinomialNB Classifier*, *AdaBoost Classifier*, *OneVsRest Classifier*, *OneVsOne Classifier* and *Decision Tree*. The algorithm showed excellent agreement with the hypotheses and for a better database more questions could be answered.

KEYWORDS: Machine learning; Pure sands; Sand’s behavior

1 Introdução

No mínimo, desde a Grécia antiga, há mais de três mil anos, a humanidade vem tentando entender o mundo ao seu redor. A cada descoberta realizada, naturalmente, cresce o nível de complexidade a ser desbravado. Desde Riemann (1851), problemas com mais de três variáveis vêm se tornando comuns.

Devido a limitação do entendimento humano à três dimensões, problemas multi-dimensionais são de difíceis soluções. Com efeito, correlacionar um valor com duas ou mais variáveis é um trabalho árduo para o ser humano.

Para solucionar tal problema, Rosenblatt (1957) desenvolveu uma metodologia de ajuste de funções lineares de multi-variáveis que consegue resolver casos binários. Essa metodologia ficou conhecida como “*perceptron*” que é um neurônio artificial computacional.

Com base no “*perceptron*” foram desenvolvidos vários algoritmos de ajustes de funções de multi-variáveis. Uma característica desse conjunto de metodologias é que, exceto no caso linear, a ordem de entrada



dos dados no algoritmo a fim de ajustar a função leva a funções diferentes, assim a forma de entrada dos dados no método precisa ser feita de forma arbitrária. Outra particularidade é que o banco de dados precisa ser dividido de forma aleatória em duas partes, a primeira parte será empregada no ajuste de função e a segunda parte no teste de acurácia da função.

Cada dado entra no ajuste da função de forma recursiva a medida que os dados vão sendo considerados (“loop”). Essa natureza do método pode ser interpretada como um aprendizado. Como esses cálculos são feitos geralmente em computadores, dá-se o nome de aprendizado de máquina (“machine learning”).

O aprendizado de máquina é utilizado diariamente em diversas atividades desde a “internet” (compras “on-line”) até o meio acadêmico (astronomia), ou seja, do comportamento humano a propriedades intrínsecas da natureza. Percebe-se que uma gama muito variada de eventos consegue ser interpretada pelo aprendizado de máquina, não há razão de se pensar que a geotecnia seja uma exceção.

Como aprendizado de máquina não é difundido no contexto da engenharia geotécnica, o presente trabalho é apenas uma prova de conceito onde a partir de um banco de dados de ensaios de cisalhamento direto em areias tenta-se prever o comportamento do material (dilatante ou contrátil). Para implementar o aprendizado de máquina foi utilizada a linguagem de programação “Python” através da biblioteca “scikit-learn”. Destaca-se que esta linguagem é bastante conhecida, principalmente, pela facilidade de implementação de aprendizado de máquina, inteligência artificial e redes neurais.

2 Banco de Dados

Foi constituído um banco de dados de ensaios de cisalhamento direto para as areias puras das praias de Itaipuaçu (Niterói/RJ), Ipanema (Rio de Janeiro/RJ) e Cavaleiros (Macaé/RJ), estudadas por Nunes (2014), Simões (2015) e Pinheiro (2018) respectivamente. Ressalta-se que o experimento é válido durante a fase de cisalhamento do ensaio de cisalhamento direto.

Além de parâmetros de resistência e deformabilidade, o ensaio de cisalhamento direto mostra o comportamento contrátil ou dilatante da areia cisalhada conforme a figura 1.

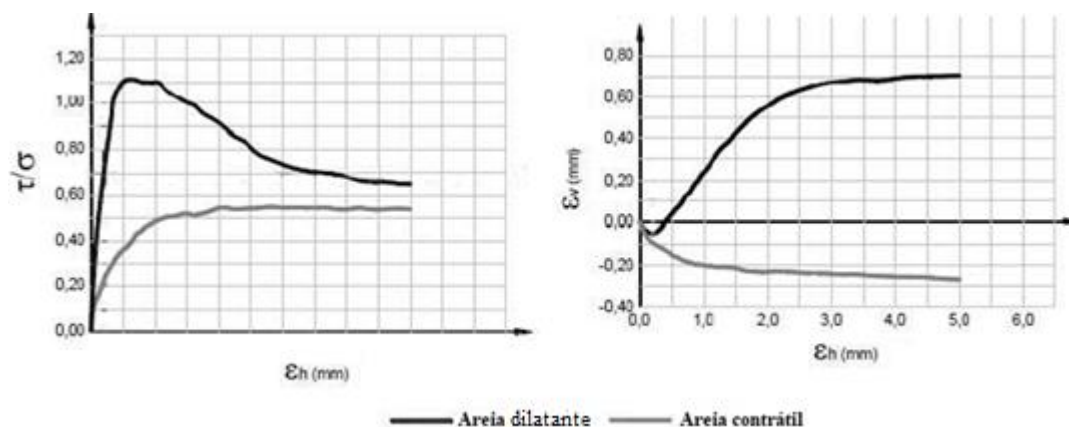


Figura 1. Comportamento de areias sobre ensaio de cisalhamento direto.

2.1 Nunes (2014)

Nunes (2014) realizou 30 ensaios de cisalhamento direto. As tensões confinantes utilizadas foram de 25 kPa, 50 kPa, 75 kPa, 100 kPa, 150 kPa, 300 kPa, 500 kPa, 750 kPa, 1000 kPa e 1500 kPa para as compacidades variando de 0% até 100%. Vale ressaltar que o ensaio de 1500 kPa de tensão confinante com compacidade relativa elevada não foi realizado.

2.2 Simões (2015)

Simões (2015) realizou 30 ensaios de cisalhamento direto. As tensões confinantes utilizadas foram de 25 kPa, 50 kPa, 75 kPa, 100 kPa, 150 kPa, 300 kPa, 500 kPa, 750 kPa, 1000 kPa e 1250 kPa para as compacidades variando de 30% até 101%.



2.3 Pinheiro (2018)

Pinheiro (2018) realizou 24 ensaios de cisalhamento direto. As tensões confinantes utilizadas foram de 12,5 kPa, 25 kPa, 50 kPa, 100 kPa, 200 kPa, 400 kPa, 800 kPa e 1600 kPa para as compacidades variando de 30% até 97%.

3 Métodos

3.1 Generalidades

Os procedimentos adotados neste trabalho estão associados a três etapas: (i) pré-processamento; (ii) processamento e (iii) pós-processamento.

(i) Pré-processamento

Nessa etapa foi realizada a coleta dos dados para formar o banco de dados que no presente trabalho possui 83 elementos. Assim, classifica-se em dilatante e contrátil, vale ressaltar que para uma análise adequada utilizou-se na terceira coluna o valor 0 para indicar o caráter contrátil e o valor 1 para o caráter dilatante. Nesse ponto ocorre a randomização dos dados e o sorteio dos dados para o conjunto treino e para o conjunto teste, no presente trabalho a proporção adotada foi de 88% e 12% respectivamente.

De acordo com o estado da arte de aprendizado de máquina, dever-se-ia realizar a normalização (“*encoding*”) dos dados, porém essa etapa não foi realizada no presente trabalho. E na hora de inserir os dados no algoritmo, os mesmos foram embaralhados.

(ii) Processamento

O processamento dos dados é feito a partir de modelos de “*machine learning*” da biblioteca do “*scikit-learn*” que são: “*Dummy Classifier*”, “*LinearSVC Classifier*”, “*MultinomialNB Classifier*”, “*Adaboost Classifier*”, “*OneVsRest Classifier*”, “*OneVsOne Classifier*” e “*Decision Tree*”.

Cabe ressaltar que cada algoritmo mencionado não é nada a mais que um ajuste de uma função (linear no caso “*LinearSVC Classifier*”), portanto cada ajuste possui um nível de complexidade, associando o tempo a acurácia do método, conforme a figura 2.

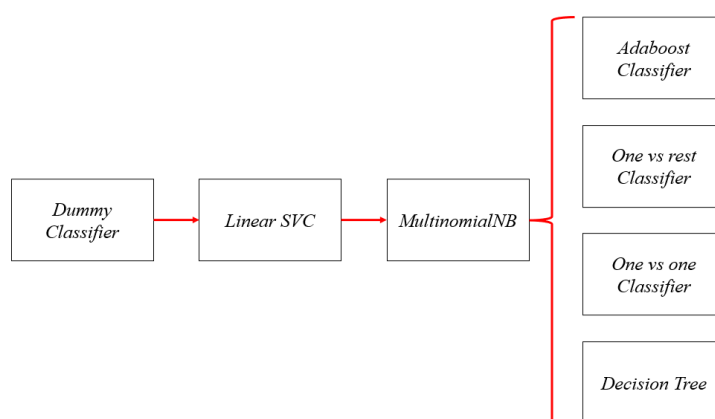


Figura 2. Nível de complexidade dos algoritmos

(iii) Pós-processamento

Com o modelo de aprendizado de máquina calibrado, inicia-se a fase teste para entender se o aprendizado foi efetivo ou não, assim utiliza-se o grupo separado para o teste a fim de comparar as previsões feitas pelo algoritmo com os dados reais.



3.2 *Dummy Classifier*

O “*Dummy Classifier*” é o método mais básico e é utilizado como comparação para os outros métodos. Basicamente, esse método consiste em analisar o banco de dados e verificar qual das classes é a predominante e, assim, estimar que todo o banco de dados se comporta dessa forma.

Em um caso que um banco de dados é composto por 100 elementos, sendo 80 da classe A e 20 da classe B, o “*Dummy Classifier*” estimará que todos os 100 elementos se comportarão conforme a classe A gerando uma taxa de acerto de 80%. Com efeito, os métodos mais avançados deverão acertar mais do que a taxa de acerto do “*Dummy Classifier*”, ou seja, nesse caso 80%.

3.3 *LinearSVC Classifier*

O método “*LinearSVC Classifier*” é um classificador linear onde cada parâmetro tem um peso diferente, mas todos os parâmetros são diretamente proporcionais a classificação.

3.4 *MultinomialNB Classifier*

Esse é um classificador baseado na teoria de *Bayes* (“*Naive Bayes classifier*”) para modelos multinomiais. Segundo Scikit-Learn (2020) esse classificador é adequado para a categorização em classes discretas.

3.5 *AdaBoost Classifier*

Segundo Scikit-Learn (2020) o “*AdaBoost Classifier*” é um metaestimador que começa ajustando um classificador no conjunto de dados original e, em seguida, ajusta cópias adicionais do classificador no mesmo conjunto de dados, mas onde os pesos das variáveis classificadas incorretamente são ajustados de modo que os classificadores subsequentes se concentrem mais nos casos difíceis.

3.6 *OneVsRest Classifier*

Segundo Scikit-Learn (2020) essa estratégia consiste em ajustar um classificador por classe. Para cada classificador, a classe é ajustada em relação a todas as outras classes. Além de sua eficiência computacional uma vantagem desta abordagem é sua facilidade de interpretação. Como cada classe é representada por um e, apenas um classificador, é possível obter conhecimento sobre a classe inspecionando seu classificador correspondente. Esta é a estratégia mais comumente usada para classificação multiclasse.

3.7 *OneVsOne Classifier*

Segundo Scikit-Learn (2020) essa estratégia consiste em ajustar um classificador por par de classes. No momento da previsão, a classe que recebeu mais votos é selecionada. Este método é geralmente mais lento do que “*OneVsRest Classifier*”, devido à sua complexidade.

3.8 *Decision-Tree*

“*Decision-Tree*” é um método que possui como estratégia a divisão dos dados em busca de encontrar uma árvore de decisão para saber como o objeto é classificado.

4 Desempenho das Análises

Existem vários índices que mensuram a qualidade da análise realizada pelo algoritmo na fase teste. Para facilitar as análises utiliza-se o conceito da matriz confusão. Assim, definiu-se o verdadeiro positivo quando o algoritmo acerta que o solo contraiu; o falso negativo quando o algoritmo prediz que o solo dilatou, mas na



realidade o solo contraiu; o verdadeiro negativo quando o algoritmo acerta que o solo dilata; e o falso positivo quando o algoritmo prediz que o solo dilatou, mas na realidade o solo contraiu. A figura 3 apresenta o padrão da matriz confusão. As equações 1, 2, 3, 4, 5 e 6 definem, respectivamente: a acurácia, erro, razão de verdadeiros positivos, razão de falsos positivos, precisão e “re-call”.

		Clase prevista	
		Positivo	Negativo
Clase real	Positivo	Verdadeiros positivos (VP)	Falsos negativos (FN)
	Negativo	Falsos positivos (FP)	Verdadeiros negativos (VN)

Figura 3. Matriz confusão.

$$ACC = \frac{VP+VN}{VP+VN+FP+FN} \quad (1)$$

$$ERR = 1 - ACC \quad (2)$$

$$TPR = \frac{VP}{FN+VP} \quad (3)$$

$$FPR = \frac{FP}{FP+VN} \quad (4)$$

$$PRE = \frac{VP}{VP+FP} \quad (5)$$

$$REC = \frac{VP}{VP+FN} \quad (6)$$

Cada um desses índices possui uma interpretação. A acurácia mede quantas previsões foram corretas no total de previsões, já o erro mede quantas previsões foram erradas no total de previsões. A razão de verdadeiros positivos mede o percentual de valores positivos reais que são previstos como verdadeiros positivos e a razão de falsos positivos mede o percentual de valores negativos reais que são previstos como falsos positivos. A precisão mede o percentual de previsões positivas que correspondem a verdadeiros positivos já o “re-call” mede o percentual de valores positivos reais que são previstos como verdadeiros positivos.

5 Resultados

Foram sorteados 10 resultados para o teste, sendo que 6 de amostras que contraiam e 4 de amostras que dilataram.

A figura 4 apresenta a matriz confusão do “Dummy Classifier” e a tabela 1 o resultado dos indicadores.

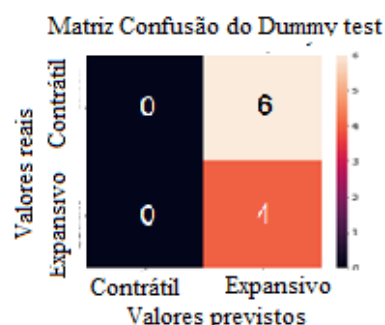


Tabela 1. Indicadores do dummy test	
Estimador	Valor
Acurácia	40%
Erro	60%
Razão de verdadeiros positivos	0%
Razão de falsos positivos	0%
Precisão	-
Re-call	0%

Figura 4. Matriz confusão do “Dummy classifier”.

A figura 5 apresenta a matriz confusão do “LinearSVC Classifier” e a tabela 2 o resultado dos indicadores. Cabe ressaltar que esse método apresentou problemas de convergência.

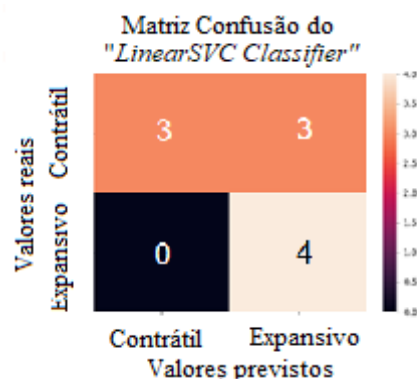


Tabela 2. Indicadores do linear classifier	
Estimador	Valor
Acurácia	70%
Erro	30%
Razão de verdadeiros positivos	50%
Razão de falsos positivos	0%
Precisão	100%
Re-call	50%

Figura 5. Matriz confusão do “LinearSVC Classifier”.



A figura 6 apresenta a matriz confusão do classificador “*MultinomialNB Classifier*” com base no teorema de Bayes (“*Bayes classifier*”) e a tabela 3 o resultado dos indicadores.

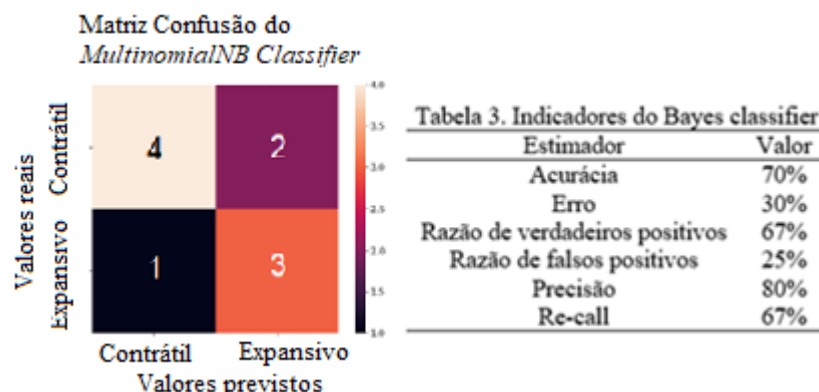


Figura 6. Matriz confusão do “*MultinomialNB Classifier*”.

A figura 7 apresenta a matriz confusão do “*Adaboost Classifier*” e a tabela 4 o resultado dos indicadores.

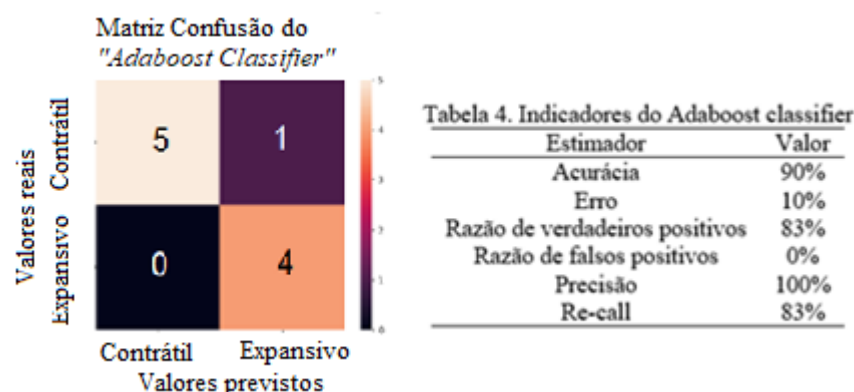


Figura 7. Matriz confusão do “*Adaboost Classifier*”.

A figura 8 apresenta a matriz confusão do “*OneVsRest Classifier*” e a tabela 5 o resultado dos indicadores.

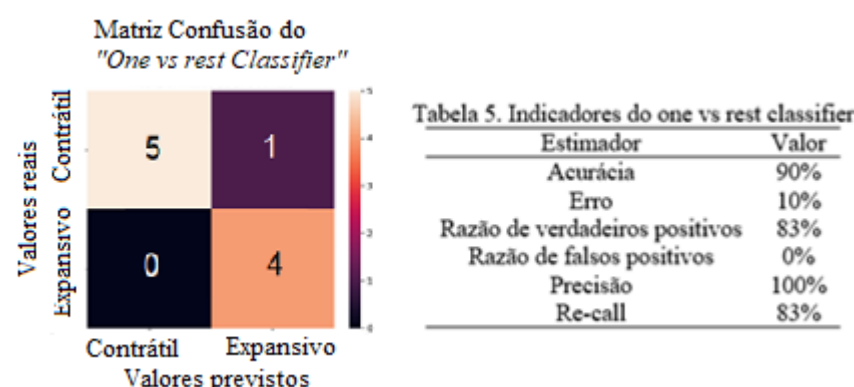


Figura 8. Matriz confusão do “*OneVsRest Classifier*”.

A figura 9 apresenta a matriz confusão do “*OneVsOne Classifier*” e a tabela 6 o resultado dos indicadores.

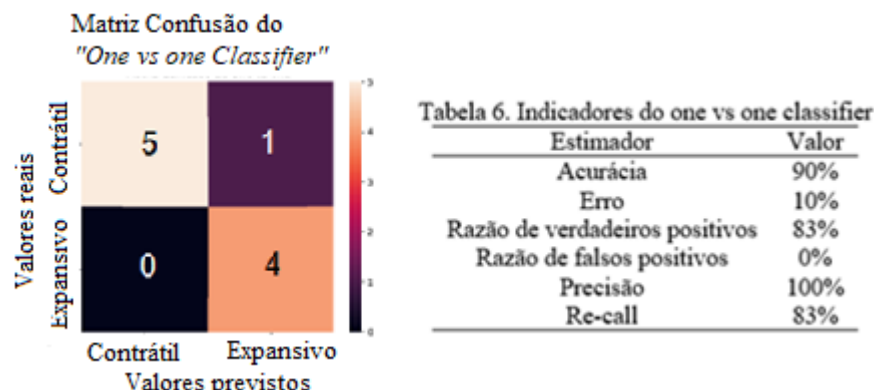


Figura 9. Matriz confusão do "OneVsOne Classifier".

A figura 10 apresenta a matriz confusão do "Decision-Tree" e a tabela 7 o resultado dos indicadores.

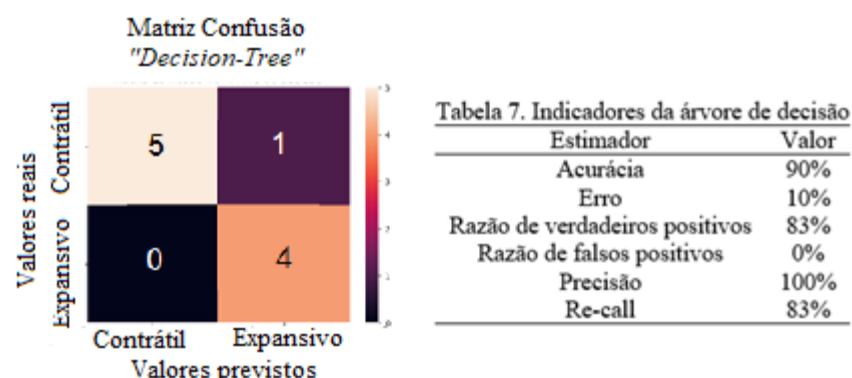


Figura 10. Matriz confusão do "Decision-Tree".

A figura 11 ilustra a árvore de decisão gerada pelo método "Decision-Tree". Esta árvore de decisão pode ser utilizada como um ábaco para tomadas de decisão de acordo com um algoritmo. No caso da figura 11 a regra adotada ao chegar em um determinado nó é: (I) seguir a seta para a esquerda se a afirmação do nó for verdadeira ou (II) seguir a seta para a direita se a afirmação do nó for falsa.

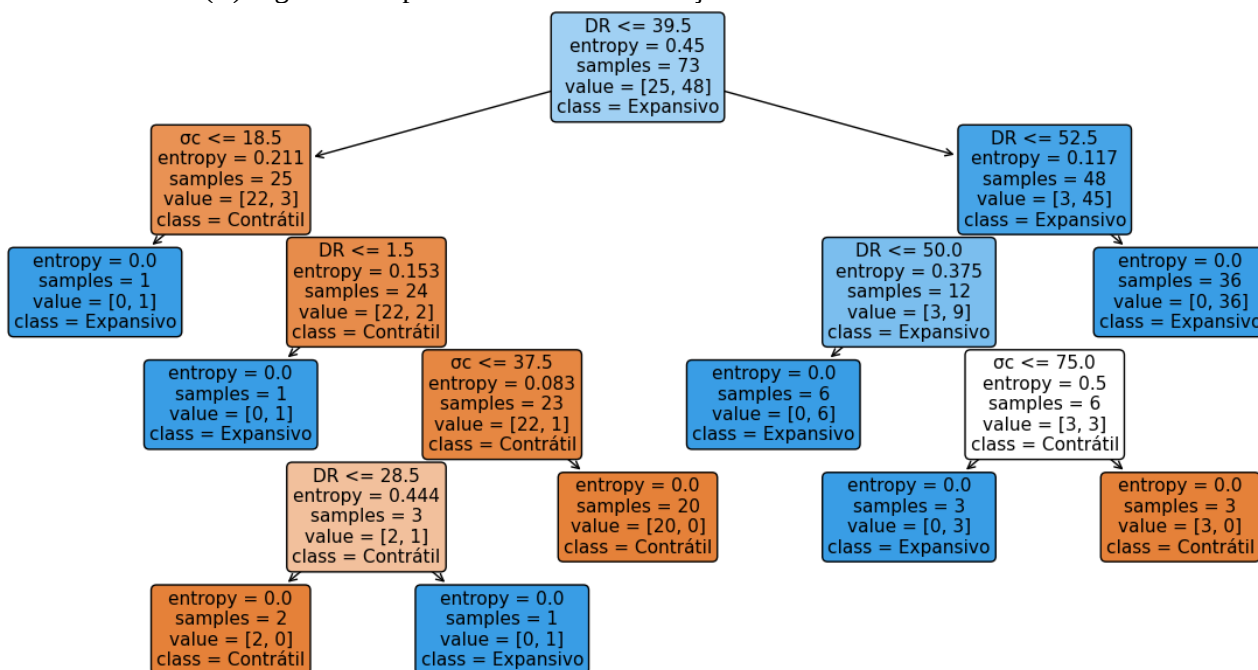


Figura 11. Árvore de decisão.



6 Conclusão

No contexto do presente artigo, que é uma prova de tecnologia, o método de aprendizado de máquina se mostrou eficaz.

É importante destacar que quando um solo apresenta um comportamento dilatante ou contrátil, naturalmente, algumas hipóteses em termos de engenharia geotécnica podem ser destacadas. Com efeito, pode-se definir se uma determinada areia possui a capacidade de se liquefazer, bem como, qual tipo de ruptura uma fundação superficial pode assumir (e.g.: generalizada ou localizada/puncionamento).

A matriz confusão se demonstrou um conceito de simples aplicação e que auxiliou muito no cálculo dos índices que auxiliaram na avaliação do desempenho dos algoritmos utilizados.

O “*Dummy Classifier*” apresentou um resultado de 40% de acurácia, já o método “*LinearSVC Classifier*” teve um desempenho de 70% na acurácia, porém não apresentou convergência, isso deve-se à natureza da curva índice de vazios crítico por tensão confinante crítica. O “*MultinomialNB Classifier*” também apresenta uma acurácia de 70%, porém esse método apresenta uma convergência. Os métodos “*Adaboost Classifier*”, “*OneVsRest Classifier*”, “*OneVsOne Classifier*” e “*Decision-Tree*” apresentaram uma acurácia de 90%.

No presente trabalho, os métodos mais avançados não apresentaram um ganho de acurácia relevante, isso deve-se ao banco de dados pequeno, se o banco de dados for enriquecido com mais dados, métodos mais complexos podem exibir alguma vantagem.

Se os métodos mais avançados não apresentarem um ganho de acurácia relevante, devem ser utilizados com cautelas, pois o tempo de execução do programa é maior que os métodos mais simples.

Em particular, o método de árvore de decisão possui como “*output*” um ábaco que pode ajudar a Engenheiros que não são acostumados à programação a tomar decisões baseadas em aprendizado de máquina sem ter que programar.

O problema abordado no presente artigo é um dos problemas que o aprendizado de máquina pode resolver. Cabe salientar, que esses métodos, não necessariamente estão associados a grandezas de desempenho de fundações (e.g.: recalque) e, sim, que eles podem ser muito úteis para estimar parâmetros geotécnicos com uma acurácia maior que as correlações utilizadas normalmente na prática da geotecnia.

Por fim, cabe destacar que é fascinante pensar que profissionais recém-formados podem utilizar inteligências artificiais treinadas por profissionais mais experientes. Este procedimento acaba sendo uma forma natural de transferência de conhecimento através das gerações.

REFERÊNCIAS BIBLIOGRÁFICAS

- Nunes, V. P. (2014) *Ensaio de Caracterização Geotécnica da Areia da Praia de Itaipuaçu*, Trabalho de Conclusão de Curso, Rio de Janeiro: Escola Politécnica/UFRJ, Rio de Janeiro, pp. 148.
- Pinheiro, G. P. (2018) *Caracterização Geotécnica em Laboratório da Areia da Praia dos Cavaleiros – Macaé/RJ*, Trabalho de conclusão de Curso, Macaé: Engenharia Civil, Universidade Federal do Rio de Janeiro, pp. 124.
- Riemann, B. (1851) *Grundlagen für eine allgemeine Theorie der Funktionen einer veränderlichen complexen Größe*. Tese de doutorado, Universidade de Göttingen, Alemanha.
- Rosenblatt, F. (1957) *The Perceptron, A Perceiving and Recognizing Automaton*, Project Para Report No. 85-460-1, Cornell Aeronautical Laboratory (CAL), jan. 1957.
- Scikit-Learn (2020), *User Guide*, Disponível em: < https://scikit-learn.org/stable/user_guide.html > Acessado em: 12/12/2021.
- Simões, F. B. (2015) *Caracterização da Areia da Praia de Ipanema*, Trabalho de conclusão de Curso, Rio de Janeiro: Escola Politécnica, Universidade Federal do Rio de Janeiro, pp. 124.