



**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
FACULDADE DE ENGENHARIA CIVIL**



MAX CANTÃO DE LEÃO

**AVALIAÇÃO DE PARÂMETROS DE ENSAIOS DE COMPACTAÇÃO A
PARTIR DE PROPRIEDADES ÍNDICES DO SOLO USANDO TÉCNICAS DE
MACHINE LEARNING**

**BELÉM-PA
2026**

MAX CANTÃO DE LEÃO

**AVALIAÇÃO DE PARÂMETROS DE ENSAIOS DE COMPACTAÇÃO A
PARTIR DE PROPRIEDADES ÍNDICES DO SOLO USANDO TÉCNICAS DE
MACHINE LEARNING**

Trabalho de Conclusão de Curso
apresentado a Faculdade de Engenharia Civil
do Instituto de Tecnologia da Universidade
Federal do Pará, como parte dos requisitos para
obtenção do título de Bacharel em Engenharia
Civil

Orientador: Prof. D.Sc. Gérson Jacques
Miranda dos Anjos.

BELEM-PA
2026

MAX CANTÃO DE LEÃO

**AVALIAÇÃO DE PARÂMETROS DE ENSAIOS DE COMPACTAÇÃO A
PARTIR DE PROPRIEDADES ÍNDICES DO SOLO USANDO TÉCNICAS DE
MACHINE LEARNING**

Trabalho de Conclusão de Curso
apresentado a Faculdade de Engenharia Civil
do Instituto de Tecnologia da Universidade
Federal do Pará, como parte dos requisitos para
obtenção do título de Bacharel em Engenharia
Civil

Orientador: Prof. D.Sc. Gérson Jacques
Miranda dos Anjos.

Belém, ____ de _____ de _____.

Aprovado por:

Prof. D. Sc. Gérson Jacques Miranda dos Anjos

Orientador - Universidade Federal do Pará – UFPA

Professor Dr. Salim Habib Fraiha Neto

Membro da banca - Universidade Federal do Pará –
UFPA

Professor Dr. Rodrigo da Cruz de Araújo

Membro da banca - Universidade Federal do Pará –
UFPA

*Dedico este trabalho a minha mãe,
aos amigos e colegas pelo incentivo e apoio
constante.*

AGRADECIMENTOS

Agradeço primeiramente a Deus, por ter sido o meu alicerce e a fonte de toda a força necessária para superar os momentos de maior dificuldade, tanto ao longo desta graduação quanto nos desafios impostos pela vida.

À minha mãe, cuja presença transcende a distância física. Mesmo não estando mais entre nós, seu legado de amor e apoio permanece como o combustível que me impulsionou a chegar até aqui. Tudo o que conquistei é, também, fruto da base que ela me deu.

Ao meu orientador, Prof. D.Sc. Gérson Jacques Miranda dos Anjos, pela paciência, pelo apoio constante e por compartilhar seu conhecimento com tanta generosidade. Sua orientação foi fundamental para que esta jornada chegasse a um bom termo.

À Danielly, minha companheira de todas as horas. Agradeço por estar ao meu lado em cada passo dessa caminhada, oferecendo suporte, compreensão e por ser o meu porto seguro, sustentando-me nos períodos de maior exaustão entre as aulas noturnas e a rotina de dois estágios, um em cada período do dia. Sua compreensão e incentivo nos momentos de grandes dificuldades foram fundamentais para que eu pudesse conciliar a vida acadêmica, profissional e pessoal com equilíbrio.

A todos que, direta ou indiretamente, contribuíram para a realização deste sonho.

RESUMO

LEÃO, M.C. Avaliação de Parâmetros de Ensaio de Compactação a partir de Propriedades Índices do Solo Usando Técnicas de Machine Learning. 2026. 46 f. Trabalho de Diplomação (Graduação em Engenharia Civil) – Faculdade de Engenharia Civil, Universidade Federal do Pará, Belém.

A determinação dos parâmetros de compactação do solo, da umidade ótima ($w_{ót}$) e do peso específico seco máximo ($\gamma_{d,máx}$) por meio do ensaio Proctor é crucial para o controle tecnológico em obras de terraplenagem e pavimentação. Contudo, este método laboratorial é oneroso e demorado, gerando desafios em projetos de grande escala. Diante disso, este trabalho propõe um modelo preditivo inovador, baseado na técnica de machine learning Extreme Gradient Boosting (XGBoost), para estimar esses parâmetros a partir de propriedades de índice do solo de fácil obtenção, visando maior agilidade e eficiência.

A escolha do XGBoost se justifica por sua capacidade de modelar relações não lineares complexas e interações multidimensionais entre variáveis, superando as limitações de modelos estatísticos tradicionais. Para o desenvolvimento, utilizou-se um robusto banco de dados consolidado com 2.730 amostras de solos (74% brasileiras e 26% internacionais), abrangendo limites de Atterberg (LL e LP) e composição granulométrica (frações grossas e finas). A metodologia seguiu um rigoroso fluxo de ciência de dados em Python, incluindo preparação de dados com DMatrix, ajuste de hiperparâmetros via grid search e validação cruzada 5-fold, garantindo a robustez e a generalização do modelo, mitigando o risco de overfitting.

Os resultados demonstraram alta performance preditiva do modelo XGBoost. Para o ($\gamma_{d,máx}$), o coeficiente de determinação (R^2) alcançou 0,870, enquanto para a ($w_{ót}$), o (R^2) foi de 0,891. Esses valores indicam que o modelo explica a maior parte da variabilidade dos dados com reduzida margem de erro, conferindo alta confiabilidade. A análise de importância das variáveis revelou que o Limite de Plasticidade (LP) e o Limite de Liquidez (LL) são os preditores mais influentes, corroborando os princípios da mecânica dos solos. Conclui-se que o XGBoost é uma ferramenta promissora e eficaz para o pré-projeto e controle de qualidade na engenharia geotécnica, otimizando processos, reduzindo custos operacionais e alinhando a geotecnia às modernas abordagens da inteligência artificial.

Palavras-chave: Compactação do Solo. Parâmetros de compactação. Machine Learning. XGBoost. Geotecnia. Engenharia Civil. Modelagem Preditiva.

ABSTRACT

LEÃO, M.C. Evaluation of Compaction Test Parameters from Soil Index Properties Using Machine Learning Techniques. 2026. 46 p. Graduation Thesis (Bachelor in Civil Engineering) – Faculty of Civil Engineering, Federal University of Pará, Belém, 2026.

The determination of soil compaction parameters, specifically optimum moisture content ($w_{ót}$) and maximum dry unit weight ($\gamma_{d,máx}$), through the Proctor test is crucial for technological control in earthworks and paving. However, this laboratory method is costly and time-consuming, posing challenges in large-scale projects. Therefore, this study proposes an innovative predictive model, based on the Extreme Gradient Boosting (XGBoost) machine learning technique, to estimate these parameters from easily obtainable soil index properties, aiming for greater agility and efficiency.

The choice of XGBoost is justified by its ability to model complex non-linear relationships and multidimensional interactions between variables, overcoming the limitations of traditional statistical models. For its development, a robust consolidated database of 2,730 soil samples was used (74% Brazilian and 26% international), covering Atterberg limits (LL and PL) and grain size composition (coarse and fine fractions). The methodology followed a rigorous data science workflow in Python, including data preparation with DMatrix, hyperparameter tuning via grid search, and 5-fold cross-validation, ensuring the model's robustness and generalization, while mitigating the risk of overfitting.

The results demonstrated high predictive performance of the XGBoost model. For ($\gamma_{d,máx}$), the coefficient of determination (R^2) reached 0.870, while for ($w_{ót}$), the (R^2) was 0.891. These values indicate that the model explains most of the data variability with a reduced error margin, providing high reliability. Feature importance analysis revealed that the Plastic Limit (PL) and Liquid Limit (LL) are the most influential predictors, corroborating the fundamental principles of soil mechanics. It is concluded that XGBoost is a promising and effective tool for pre-design and quality control in geotechnical engineering, optimizing processes, reducing operational costs, and aligning geotechnics with modern artificial intelligence approaches.

Keywords: Soil Compaction. Compaction Parameters. Machine Learning. XGBoost. Geotechnics. Civil Engineering. Predictive Modeling.

LISTA DE FIGURAS

Figura 1 - Esquema do solo e suas frações em diferentes estágios de compactação, do não compactado (a) ao avançado (d). (Fonte: Horn, 2003, Adaptado).	3
Figura 2 - Curva de compactação que relaciona a variação do peso específico seco γ (kN/m ³) e o teor de umidade (w%). (Fonte: Magalhães, 2012, Adaptado).....	5
Figura 3 - Curvas de compactação de um solo para diferentes energias. (Fonte: Pinto, 2006).....	6
Figura 4 - Carta de Plasticidade de Casagrande	19
Figura 5 - Carta de Plasticidade de Casagrande - 2730 dados da pesquisa	21
Figura 6 - Carta de Plasticidade de Casagrande - % das frações argilosas	21
Figura 7 - Carta de Plasticidade de Polidori	22
Figura 8 - Carta de Plasticidade de Polidori - % das frações argilosas	23
Figura 9 - Carta de Plasticidade de Polidori - % das frações argilosas	23
Figura 10 - Carta de Plasticidade de Moreno Maroto & Alonso Azcárate.....	24
Figura 11 - Carta de Plasticidade de Moreno Maroto & Alonso Azcárate % frações argilosas.....	26
Figura 12 - Carta de Plasticidade de Moreno-Maroto & Alonso Azcárate - % das frações argilosas.	26
Figura 13 - Histogramas das distribuições dos conjuntos de dados referente à a) γ_d , máx(kN/m ³), b) $w_{ót}$ (%), c) LL(%), d) LP(%), e) G(%) e f) F(%).....	29
Figura 14 - Resultado via análise XGBoost – peso específico seco máximo	31
Figura 15 - Resultado via análise XGBoost – teor de umidade_ ótimo	32

LISTA DE TABELAS

Tabela 1 - Resumo das propriedades consideradas na criação de formulações	15
Tabela 2 - Estatísticas descritivas das variáveis do modelo de estudo	27

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
ASTM	American Society for Testing and Materials
DNER	Departamento Nacional de Estradas de Rodagem
NBR	Norma Brasileira Regulamentadora
ME	Método de Ensaio
EUA	Estados Unidos da América
LL	Limite de Liquidez
LP	Limite de Plasticidade
IP	Índice de Plasticidade
PN	Proctor Normal
PM	Proctor Modificado
GMDH	Group Method of Data Handling
RLS	Regressão Linear Simples
RLM	Regressão Linear Múltipla
RNA	Redes Neurais Artificiais
MED	Mediana
MAE_MA	Erro Absoluto Médio (Média Aritmética)
MAE_MED	Erro Absoluto Médio (Mediana)
MSE	Erro Quadrático Médio
RMSE	Raiz do Erro Quadrático Médio
MAPE_MA	Erro Percentual Absoluto Médio
MAPE_MED	Erro Percentual Absoluto Mediano
MH	Silte de Alta Plasticidade
NPC	Solos Com Argilominerais Laminares
RMSPE	Raiz do Erro Percentual Quadrático Médio
CF	Teor de Argila
CH	Argila de Alta Plasticidade
CL	Argila de Baixa Plasticidade
ML	Siltes de Baixa Plasticidade
OH	Solo Orgânico
SUCS	Sistema Unificado de Classificação de Solos

SVM	Support Vector Machines
L1	Lasso
L2	Ridge
XGBOOST	Extreme Gradient Boosting

LISTA DE SIMBOLOS

$w(\%)$	Teor de Umidade
γ_d	Peso específico Seco
$w_{ót}$	Umidade Ótima
$\gamma_{d,máx}$	Peso específico Seco Máximo
ρ_s	Densidade Real Média dos Grãos
F	Porcentagem de Finos
S	Porcentagem de Areia
G	Porcentagem de Pedregulhos
G_s	Densidade Real dos Grãos
R	Coeficiente de Correlação
R^2	Coeficiente de determinação
Q_c	Valor calculado (previsto)
Q_m	Valor medido
K	Coeficiente de Curtose
N	Número de dados
M	Média
<i>DesvPad</i>	Desvio Padrão
μ_k	Média aritmética dos valores de k
σ_k	Desvio-Padrão dos valores de k.
$\mu_{\ln(k)}$	Média aritmética dos valores de $\ln(k)$
$\sigma_{\ln(k)}$	Desvio-Padrão dos valores de $\ln(k)$
X	Variável (valores de k)
$f(x)$	Função densidade de probabilidade
$\mu(Q_m)$	Média aritmética dos valores de Q_m
E	Coeficiente de Eficiência de Nash-Sutcliffe
D	Índice de Concordância de Willmott

1. Sumário

1. INTRODUÇÃO	1
1.1 Objetivo Geral	2
1.2 Objetivos Específicos	2
2. REVISÃO BIBLIOGRÁFICA.....	3
2.1 Compactação	3
2.2 Estimativas de $wót$ e $\gamma d, máx$ baseados na literatura.....	8
2.3 Extreme Gradient Boosting (XGBoost)	16
3) DADOS (LOCAIS DE ESTUDO).....	19
4) ANÁLISE DOS RESULTADOS	31
5) CONCLUSÕES/ SUGESTÕES.....	33
6) REFERÊNCIAS.....	34

1. INTRODUÇÃO

O solo, um dos principais materiais de construção, possui um comportamento mecânico complexo que é decisivo para a estabilidade e o desempenho de estruturas geotécnicas. A compactação é o processo mecânico pelo qual se busca aprimorar suas propriedades, como aumentar a resistência, reduzir a compressibilidade e diminuir a permeabilidade, por meio da redução do índice de vazios com a expulsão de ar (Holtz & Kovacs, 1981). O controle tecnológico deste processo é tradicionalmente guiado pela relação entre o teor de umidade (w) e o peso específico seco (γ_d), estabelecida por Proctor (1933), que permite identificar um ponto ótimo de compactação onde o solo atinge seu peso específico seco máximo ($\gamma_{d,máx}$) para uma umidade ótima ($w_{ót}$).

A determinação desses parâmetros em laboratório, embora precisos, é um procedimento oneroso e demorado, o que representa um obstáculo em fases de estudos preliminares ou em projetos que exigem o controle de grandes volumes de solo (Das, 2011). Para contornar essa limitação, a literatura geotécnica tem se dedicado ao desenvolvimento de modelos preditivos que estimam os parâmetros de compactação a partir de propriedades de índice, como os Limites de Atterberg (Limite de Liquidez - LL e Limite de Plasticidade - LP) e a composição granulométrica, que são obtidos por ensaios mais simples e rápidos.

As primeiras abordagens para essa estimativa se basearam em correlações empíricas e modelos estatísticos. Nesse contexto, o trabalho de Pantoja e Anjos (2018) representa um marco ao analisar e classificar diversas formulações existentes, propondo novos modelos de Regressão Linear Múltipla (RLM) calibrados para solos brasileiros, o que demonstrou a viabilidade de se obter estimativas razoáveis com ferramentas estatísticas clássicas. Contudo, a natureza frequentemente não-linear das relações entre as propriedades do solo impõe um limite à precisão de modelos lineares.

Buscando superar essa limitação, Álvares Júnior (2019) deu um passo adiante ao aplicar Redes Neurais Artificiais (RNA), uma técnica de inteligência artificial que se destaca pela capacidade de modelar sistemas complexos e não-lineares sem a necessidade de uma função matemática pré-definida (Goh, 1995). Seus resultados indicaram uma melhoria na acurácia das previsões em comparação com os modelos de regressão, evidenciando o potencial do aprendizado de máquina na geotecnia.

Seguindo essa trajetória evolutiva, o presente trabalho propõe a utilização de um algoritmo de machine learning, ou seja, uma técnica mais avançada e robusta, o Extreme

Gradient Boosting (XGBoost). O XGBoost é um método do tipo ensemble baseado em árvores de decisão que tem ganhado notoriedade por sua performance superior em diversas competições de ciência de dados (Chen & Guestrin, 2016). Sua robustez, alta precisão e mecanismos para evitar o superajuste (overfitting) o tornam uma ferramenta promissora para avançar ainda mais na qualidade das previsões dos parâmetros de compactação, oferecendo uma solução potencialmente mais eficiente e confiável para a engenharia geotécnica.

1.1 Objetivo Geral

O objetivo geral deste trabalho é desenvolver e avaliar um modelo preditivo baseado na técnica de machine learning XGBoost para a estimativa dos parâmetros de compactação do solo, umidade ótima ($w_{ót}$) e peso específico seco máximo ($\gamma_{d,máx}$), a partir de suas propriedades de índice, com base no banco de dados de 2730 amostras de solo.

1.2 Objetivos Específicos

Visando atingir o objetivo principal, foram estabelecidos os seguintes objetivos específicos:

Desenvolver um modelo preditivo para os parâmetros de compactação ($w_{ót}$) e ($\gamma_{d,máx}$), utilizando a técnica de machine learning Extreme Gradient Boosting (XGBoost), a partir de um banco de dados consolidado com propriedades de índice do solo (LL, LP, %Grossos e %Finos).

Avaliar o desempenho e a robustez do modelo XGBoost por meio da aplicação do Coeficiente de Determinação (R^2), utilizado como principal métrica estatística para mensuração da capacidade explicativa do modelo.

2. REVISÃO BIBLIOGRÁFICA

Este capítulo tem como objetivo apresentar uma fundamentação teórica para a pesquisa desenvolvida. Para isso, abrange inicialmente os conceitos e a teoria da compactação de solos, seguido por uma análise dos principais modelos usados para a previsão dos parâmetros de compactação aplicadas neste trabalho. Por fim, inclui uma introdução concisa à teoria do XGBoost, contextualizando sua aplicação na conjuntura da análise proposta.

2.1 Compactação

O solo é formado por partículas sólidas envoltas por água e ar, sendo os espaços entre essas partículas denominados vazios, mostrado na figura 1. O estado de densificação do solo, quantificado pelo índice de vazios, influencia diretamente seu comportamento mecânico, uma propriedade fundamental para sua utilização em obras de engenharia civil. Para melhorar as características mecânicas do solo, é comum empregar técnicas de estabilização mecânica. A compactação consiste em um processo de densificação do solo, no qual a aplicação de energia mecânica promove a expulsão do ar presente nos vazios, resultando em um aumento da sua resistência e estabilidade. A compactação de um solo qualquer é a redução, mais ou menos rápida, do índice de vazios, por processos mecânicos, onde o aumento do peso de sólidos por unidade de volume é causado pela expulsão do ar dos poros do solo (Holtz & Kovacs, 1981 e Fernandes 1994).

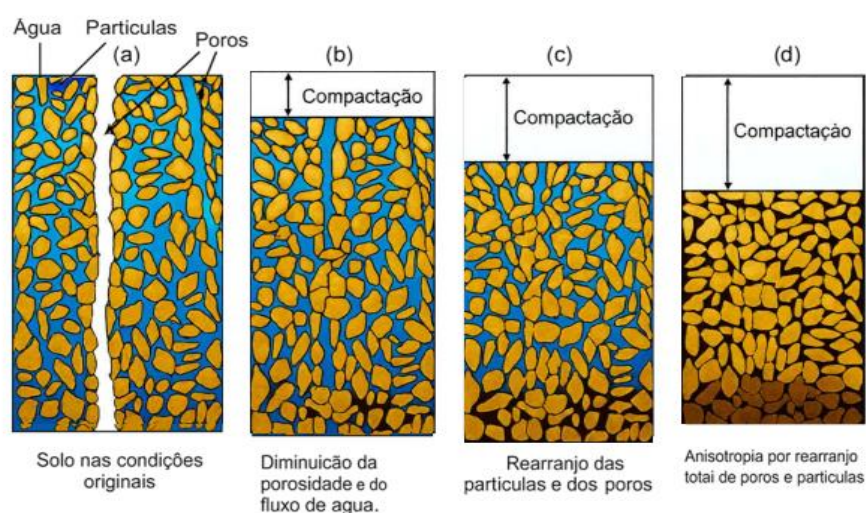


Figura 1 - Esquema do solo e suas frações em diferentes estágios de compactação, do não compactado (a) ao avançado (d). (Fonte: Horn, 2003, Adaptado).

Antes de 1930, a compactação do solo em projetos de engenharia civil era conduzida de maneira empírica, fundamentada na experiência prática dos profissionais, sem uma base teórica consolidada. O crescente investimento em infraestrutura rodoviária nos Estados Unidos impulsionou o desenvolvimento de estudos científicos voltados para a compreensão do comportamento do solo durante a compactação. Nesse contexto, Ralph Proctor destacou-se ao publicar, na década de 30, uma série de trabalhos que estabeleceram a relação entre a umidade do solo e o peso específico seco para diferentes níveis de energia aplicados na compactação (Ribeiro, 2008). Essa relação, conhecida como “Curva de Compactação”, tornou-se fundamental para o controle técnico do processo, permitindo alcançar a densificação máxima do solo. A aplicação desse procedimento resulta em melhorias significativas nas propriedades mecânicas do solo, como o aumento da resistência, além da redução da deformabilidade e da permeabilidade, aspectos essenciais para a estabilidade das estruturas de engenharia civil.

O estado de compactação de um solo depende de diversos fatores, entre eles o peso específico seco, o teor de umidade da amostra, a energia mecânica aplicada durante a compactação e o tipo de solo utilizado. A interação entre esses parâmetros é representada pela curva de compactação, na qual cada combinação distinta gera uma curva e parâmetros específicos. Essa curva é obtida ao se plotar o peso específico seco (γ_d) em função do teor de umidade ($w_{\%}$) de cada amostra, sendo que cada ponto representa um estado particular de compactação. Em geral, a curva de compactação dos solos possui um formato parabólico. Contudo, solos lateríticos apresentam uma característica distinta, exibindo um ramo seco com inclinação mais acentuada (Pinto, 2006).

Para a construção da curva, são necessários, no mínimo, quatro pontos que abranjam tanto o ramo seco quanto o ramo úmido da curva (Holtz & Kovacs, 1981). Cada ponto é obtido por meio da compactação do solo em diferentes teores de umidade, registrando-se o respectivo peso específico. Observa-se que o ramo úmido da curva, que vai do ponto ótimo de compactação até o ponto de maior umidade, apresenta comportamento praticamente paralelo à curva de saturação do solo analisado (Buriticá, 2013).

A curva de compactação (Figura 2) evidencia uma relação não linear entre o teor de umidade e a densificação do solo. Inicialmente, o aumento do conteúdo de água promove redução das forças de atrito interpartículas devido à formação de películas aquosas que atuam como lubrificante, facilitando o rearranjo dos grãos e a expulsão de ar

dos vazios – fenômeno que culmina no pico de peso específico seco máximo (Holtz & Kovacs, 1981). Além desse ponto ótimo, a saturação progressiva dos vazios dificulta a eliminação do ar residual, resultando em diminuição da densidade seca mesmo com incremento de energia de compactação.

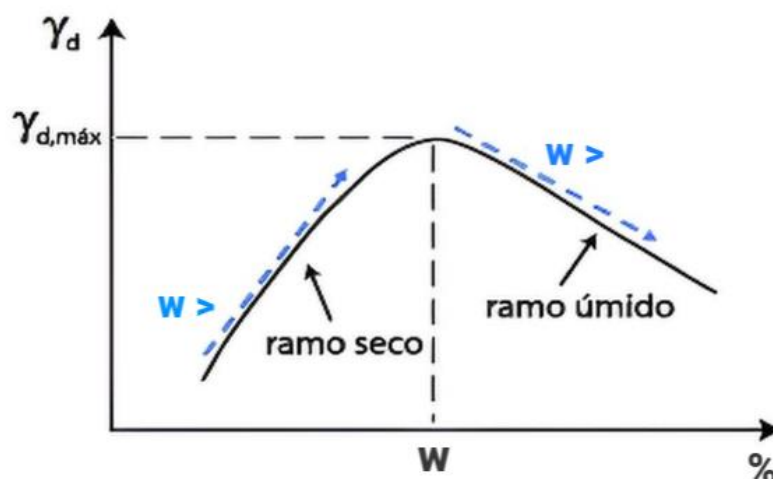


Figura 2 - Curva de compactação que relaciona a variação do peso específico seco γ (kN/m^3) e o teor de umidade ($w\%$). (Fonte: Magalhães, 2012, Adaptado).

Esse comportamento bifásico tem sido explicado por diferentes abordagens teóricas:

1. Teoria da Lubrificação (Proctor, 1933): Destaca o papel da água na redução do atrito Inter partículas, permitindo maior densificação até a umidade crítica.
2. Modelo de Estágios (Hogentogler, 1937): Divide o processo em quatro fases – hidratação (formação de filmes aquosos viscosos), lubrificação (rearranjo otimizado de partículas), inchamento (expansão volumétrica) e saturação (preenchimento total dos vazios por água).
3. Interações Físico-Químicas (Lambe, 1958): Explica a curva através da dinâmica das duplas camadas elétricas em argilas, onde baixas umidades favorecem estruturas floculadas, enquanto altos teores hídricos induzem dispersão das partículas.

Ribeiro (2008) sintetiza que a forma da curva deriva de um equilíbrio complexo entre capilaridade, tensões efetivas e viscosidade da água, sendo influenciada ainda pelas características mineralógicas do solo. Estudos recentes destacam que a energia de

compactação desloca a curva para regiões de maior densidade seca e menor umidade ótima, um parâmetro crucial para projetos geotécnicos.

Cada tipo de solo apresenta uma curva de compactação única, que varia conforme o tipo de solo e a energia mecânica aplicada no processo, geralmente classificada como energia normal, intermediária ou modificada, conforme ilustrado na Figura 2. Para cada nível de energia, existe um teor de umidade específico no qual o solo atinge sua máxima densificação, resultando em diferentes arranjos das partículas. À medida que a energia de compactação aumenta, o teor de umidade ótimo tende a diminuir, permitindo que o solo alcance maior densidade seca com menor quantidade de água. Estudos de Blotz et al. (1993) e Gurtug & Sridharan (2004) demonstram que o peso específico seco do solo compactado está linearmente relacionado ao logaritmo da energia aplicada, evidenciando a influência direta da energia de compactação na eficiência do processo.

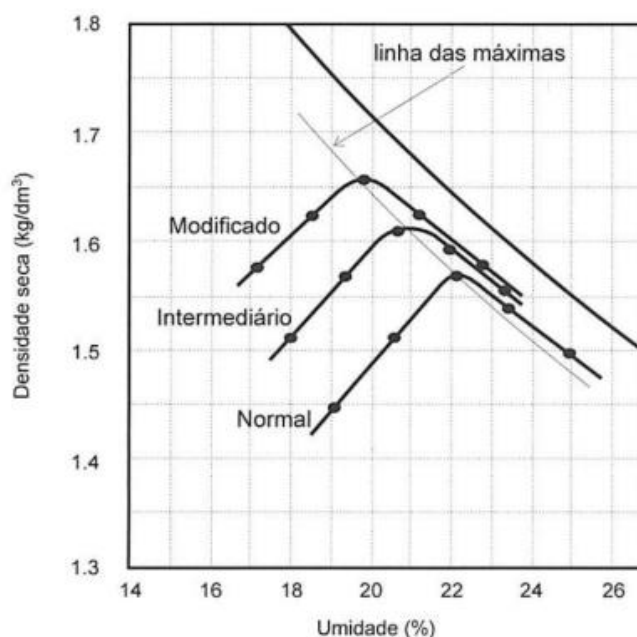


Figura 3 - Curvas de compactação de um solo para diferentes energias. (Fonte: Pinto, 2006)

As curvas de compactação são organizadas em ordem crescente conforme a energia aplicada, sendo que as curvas posicionadas mais acima correspondem geralmente às maiores energias de compactação. Em campo, a energia efetivamente aplicada depende de fatores como o número de passadas do equipamento, o peso da máquina e o volume da área compactada. Já em laboratório, a energia depende do tipo de ensaio realizado. No caso do ensaio por impacto, conhecido como método de Proctor, a energia é calculada

considerando o peso do soquete, a altura da queda, o número de golpes, o número de camadas compactadas e o volume do molde utilizado (Holtz & Kovacs, 1981).

O ponto máximo das curvas de compactação corresponde ao estado de máxima densificação do solo, onde ocorre a condição ideal de equilíbrio entre as fases físicas do solo. Esse ponto é caracterizado por duas coordenadas fundamentais: o peso específico seco máximo e a umidade ótima. Esses parâmetros, determinados em laboratório, são utilizados como referência para o controle do processo de compactação em campo, representando o estado em que o solo apresenta maior resistência e estabilidade para uma dada energia de compactação. O aumento do peso específico seco do solo está diretamente associado à sua resistência mecânica, motivo pelo qual a estabilização mecânica busca atingir o peso específico máximo para uma dada energia de compactação.

A metodologia de Proctor fundamenta-se na importância de conhecer o teor de umidade do solo durante a aplicação da energia de compactação, pois a eficiência na densificação e reorganização estrutural do solo depende diretamente dessa variável. A variação do teor de umidade durante o processo pode impactar significativamente o desempenho das obras de terra. Uma compactação adequada visa aprimorar as propriedades do solo para diversas aplicações, como aterros compactados, construção de barragens de terra, pavimentação de estradas e implantação de loteamentos. Além disso, é essencial para garantir a estabilidade do solo em fundações diretas, terraplanagem e compactação de muros de arrimo, reaterros em valas escavadas a céu aberto e retaludamento em encostas naturais (Massad, 2003).

O ensaio de compactação proposto por Proctor marcou o início do controle sistemático da compactação em obras de engenharia, estabelecendo uma base teórica fundamental para o desenvolvimento e aprimoramento das técnicas laboratoriais aplicadas ao estudo dos solos na construção civil. Tornou-se um método padrão e adotado mundialmente para avaliar a relação entre densidade e teor de umidade dos solos. Na década de 1950, surgiu o “Modified Proctor Compaction Test”, que se diferencia do método original pela aplicação de uma energia significativamente maior durante a compactação, adaptando-se à evolução dos equipamentos de compactação da época (Briaud, 2013). No Brasil, o ensaio Proctor é regulamentado pelas normas ABNT NBR 7182 (2016) e DNER ME 162/94, garantindo sua padronização e confiabilidade. Além disso, variações do método original, que utilizam diferentes volumes de amostra, níveis de energia e técnicas de compactação, têm sido empregadas para atender a requisitos específicos em diversas aplicações geotécnicas.

2.2 Estimativas de $w_{ót}$ e $\gamma_{d,máx}$ baseados na literatura

A evolução das técnicas de análise de dados e o advento da inteligência artificial, especialmente o machine learning, têm ampliado as possibilidades de modelagem e previsão em geotecnia, motivando pesquisadores a desenvolver modelos preditivos para estimar parâmetros do ensaio Proctor a partir de propriedades do índice do solo, reduzindo a dependência de ensaios laboratoriais extensivos (mas não menos importante).

Jumikis (1946) pioneiramente correlacionou a umidade ótima com os limites de Atterberg e propôs equações para prever o peso específico seco de solos finos. Nas décadas seguintes, as abordagens variaram significativamente: enquanto Đoković et al. (2013) restringiram-se aos limites de Atterberg como variáveis antecipadoras, Gunaydın (2008) e Ardakani & Kordnaeij (2017) exploraram técnicas avançadas como redes neurais artificiais (RNAs) para analisar dados multivariados. Paralelamente, Sridharan & Nagaraj (2005) e Gurtug & Sridharan (2002, 2004) investigaram o impacto da energia de compactação na resposta mecânica do solo.

A diversidade metodológica desses estudos — desde modelos empíricos até algoritmos de machine learning — evidencia o potencial de técnicas computacionais para otimizar a caracterização de solos, tema central desta pesquisa. Segue uma reunião sintetizada das principais formulações e métodos propostos por diversos autores.

Al-Khafaji (1993)

Investigou a relação entre os limites de Atterberg e os parâmetros de compactação sob energia padrão, especificamente o peso específico seco máximo e a umidade ótima, utilizando amostras de solos provenientes do Iraque e dos Estados Unidos. Os dados referentes aos solos dos EUA foram obtidos de outro estudo e apresentaram variações no limite de liquidez (LL) entre 20% e 90%, e no limite de plasticidade (LP) entre 10% e 40%. Já os dados do Iraque foram coletados em quatro diferentes localidades, com LL variando de 18% a 66% e LP entre 10% e 29%. Por meio de técnicas de ajuste de curvas aplicadas a esses conjuntos de dados, o autor desenvolveu equações empíricas que relacionam os limites de Atterberg aos parâmetros de compactação. A partir dessas equações 1 a 4, foram elaborados diagramas que facilitam a aplicação prática na estimativa da umidade ótima ($w_{ót}$) e do peso específico seco máximo ($\gamma_{d,máx}$) contribuindo para a previsão eficiente desses parâmetros essenciais em projetos

geotécnicos. Essa abordagem é especialmente relevante para o desenvolvimento de modelos preditivos baseados em machine learning, alinhando-se ao objetivo do presente trabalho de otimizar a avaliação dos parâmetros de compactação a partir das propriedades do índice do solo.

$$W_{ót_Iraque(\%)} = 0,24.LL + 0,63.LP - 3,13 \quad (1)$$

$$\rho_{d,máx_Iraque(tf/m^3)} = 2,44 - 0,02.LP - 0,008.LL \quad (2)$$

$$W_{ót_EUA(\%)} = 0,14.LL + 0,54.LP \quad (3)$$

$$\rho_{d,máx_EUA(tf/m^3)} = 2,27 - 0,019.LP - 0,003.LL \quad (4)$$

Sridharan & Nagaraj (2005)

De maneira semelhante aos artigos anteriores, o presente estudo visa estabelecer uma correlação entre os limites de consistência — limite de liquidez (LL) e limite de plasticidade (LP) — e os parâmetros de compactação para solos de granulometria fina, utilizando tanto dados extraídos da literatura quanto de seu banco de dados pessoal. A análise revelou que o limite de plasticidade (LP) apresenta uma relação mais precisa com os parâmetros característicos da compactação, como a umidade ótima ($w_{ót}$) e o peso específico seco máximo ($\gamma_{d,máx}$) do que o limite de liquidez (LL). Os dados analisados abrangem uma ampla faixa de variabilidade, com LL variando entre 20% e 110% e LP entre 12% e 55%. Os coeficientes de correlação (R) obtidos foram expressivos, atingindo 0,99 para a equação que relaciona LP à umidade ótima e 0,92 para a equação referente ao peso específico seco máximo, indicando forte poder preditivo dos modelos desenvolvidos, respectivamente.

$$W_{ót(\%)} = 0,92.LP \quad (5)$$

$$\gamma_{d,máx(kN/m^3)} = 0,23.(93,3 - LP) \quad (6)$$

Esses resultados reforçam a importância do LP como variável-chave para a previsão dos parâmetros de compactação, o que é particularmente relevante para o desenvolvimento de modelos de machine learning que busquem estimar esses parâmetros a partir de propriedades índice do solo, objetivo central deste trabalho.

Sivrikaya et alli (2008)

Assim como demonstrado por Blotz et al. (1998) e Gurtug & Sridharan (2002, 2004), o entendimento dos parâmetros de compactação para energias além da energia padrão é fundamental para a correta aplicação e controle da compactação em obras de engenharia. Esses estudos ressaltam que o peso específico seco máximo ($\gamma_{d,máx}$) e a umidade ótima ($w_{ót}$) variam significativamente conforme a energia mecânica aplicada, o que impacta diretamente na qualidade e desempenho das camadas compactadas.

Os autores empregaram regressões multilíneas para modelar a relação entre os parâmetros de compactação, utilizando como variáveis independentes o logaritmo neperiano da energia de compactação e o limite de plasticidade (LP). A preferência foi dada a modelos mais simples, que facilitam a aplicação prática sem perda significativa de precisão. Para a energia padrão de compactação, foram analisados cerca de 150 dados, com LP variando entre 20% e 25%. As equações 7 e 8 são referentes a energia padrão.

$$W_{ót(\%)} = 0,94 \cdot LP \quad (7)$$

$$\gamma_{d,máx}(kN/m^3) = 23,71 \cdot e^{-0,0184 \cdot W_{ót}} \quad (8)$$

Gunaydin (2009)

Gunaydin (2009) explorou diferentes abordagens para modelar e prever os parâmetros de compactação do solo, utilizando um conjunto de dados composto por 126 amostras, cujos limites de liquidez (LL) variaram entre 23% e 57%, e os limites de plasticidade (LP) entre 14% e 30%. Inicialmente, o autor aplicou uma análise de regressão simples (RS), utilizando os limites de consistência como variáveis independentes para estimar os parâmetros de compactação. Posteriormente, adotou a regressão múltipla (RM), incorporando variáveis adicionais relacionadas à granulometria e ao peso específico do solo, além dos limites de consistência. No entanto, devido à complexidade e menor aplicabilidade prática das equações resultantes da regressão múltipla, estas foram consideradas menos úteis para fins práticos.

Na etapa mais avançada do estudo, Gunaydin empregou redes neurais artificiais (RNAs), desenvolvendo uma série de modelos com diferentes configurações para avaliar a influência relativa das variáveis de entrada na previsão dos parâmetros de compactação.

Os resultados indicaram que o limite de liquidez (LL) é a variável com maior impacto sobre a umidade ótima do solo, destacando sua importância na modelagem preditiva. Essa abordagem demonstra o potencial das técnicas de machine learning para capturar relações complexas e não lineares entre propriedades índice do solo e parâmetros de compactação, reforçando a relevância do uso dessas ferramentas no contexto do presente trabalho.

$$W_{ót_RS(\%)} = 0,3802.LL + 2,4513 \quad (r^2 = 0,82) \quad (9)$$

$$W_{ót_RS(\%)} = 0,8442.LP + 0,1076 \quad (r^2 = 0,69) \quad (10)$$

$$W_{ót_RM(\%)} = 0,323.LL + 0,157.LP \quad (r^2 = 0,82) \quad (11)$$

$$\gamma_{d,máx_RS} (kN/m^3) = -0,1008.LL + 21,16 \quad (r^2 = 0,73) \quad (12)$$

$$\gamma_{d,máx_RS} (kN/m^3) = -0,2283.LP + 21,88 \quad (r^2 = 0,64) \quad (13)$$

Matteo et alli (2009)

No estudo em questão, foram desenvolvidas correlações empíricas entre as propriedades de compactação de solos finos e argilosos, utilizando um conjunto de dados composto por 71 amostras, sendo 30 provenientes de trechos rodoviários na Itália e 41 extraídas da literatura especializada. Além dos limites de Atterberg, os autores incorporaram a densidade relativa do solo como variável explicativa, ampliando a capacidade preditiva do modelo. Os valores observados para o limite de liquidez (LL) variaram entre 20% e 82%, o índice de plasticidade (IP) entre 1% e 46%, e o teor de umidade ótima ($w_{ót}$) entre 6,8% e 26,3%.

Para a formulação do modelo empírico, os autores aplicaram técnicas de regressão linear, considerando tanto o método Proctor padrão quanto o método Proctor modificado, este último caracterizado pela aplicação de maior energia durante a compactação. A utilização do método de regressão permitiu estabelecer relações quantitativas robustas entre as propriedades índice do solo e os parâmetros de compactação, destacando a influência significativa da densidade relativa e dos limites de Atterberg no comportamento do solo sob diferentes níveis de energia aplicada. Esse estudo reforça a importância de integrar múltiplas propriedades do solo na previsão dos parâmetros de compactação, fornecendo uma base sólida para o desenvolvimento de modelos preditivos avançados, incluindo aqueles baseados em machine learning, que são o foco principal deste trabalho.

$$W_{\acute{o}t(\%)} = 0,86.LL + 3,04.(LL/G_s) \quad (14)$$

$$\gamma_{d,m\acute{a}x(kN/m^3)} = \left((40,316.w_{\acute{o}t(p)}^{-0,295}).IP^{0,032} \right) - 2,4 \quad (15)$$

$$\gamma_{d,m\acute{a}x_r^2(kN/m^3)} = 35,635.w_{\acute{o}t}^{-0,2564} \quad (16)$$

Doković et alli (2013)

Neste estudo, a estimativa dos parâmetros de compactação foi realizada utilizando exclusivamente os limites de Atterberg como variáveis independentes. Para a construção dos modelos empíricos, foram aplicadas análises de correlação e regressão, buscando estabelecer relações quantitativas entre as propriedades índice do solo e os parâmetros de compactação. O conjunto de dados analisado compreendeu 72 amostras provenientes de barragens com núcleos de argila, cujos limites de liquidez (LL) variaram entre 27% e 71%, e os limites de plasticidade (LP) entre 17% e 31%. Esses intervalos refletem a variabilidade típica de solos argilosos utilizados em obras de engenharia civil, garantindo a representatividade dos modelos para aplicações práticas. A abordagem adotada reforça a importância dos limites de Atterberg como indicadores fundamentais para a previsão dos parâmetros de compactação, contribuindo para a otimização dos processos de projeto e controle em campo.

$$w_{\acute{o}t(\%)} = 0,239.LL + 7,757 \quad (17)$$

$$\gamma_{d,m\acute{a}x(kN/m^3)} = 2,088 - 0,008.LL \quad (18)$$

$$w_{\acute{o}t(\%)} = 0,742.LP + 2,236 \quad (19)$$

$$\gamma_{d,m\acute{a}x(kN/m^3)} = 2,229 - 0,023.LP \quad (20)$$

Ardakani & Kordnaeij (2017)

Um dos estudos mais recentes sobre o tema emprega redes neurais do tipo Group Method of Data Handling (GMDH) combinadas com algoritmos genéticos para desenvolver um modelo preditivo dos parâmetros de compactação, utilizando como variáveis de entrada a composição granulométrica e os limites de Atterberg do solo. A pesquisa foi conduzida com um banco de dados, contendo 212 amostras, o que confere maior confiabilidade e abrangência ao modelo. Além da construção do modelo, o artigo

realiza uma análise de sensibilidade que evidencia a predominância dos limites de Atterberg como os parâmetros mais influentes na equação preditiva. Essa abordagem inovadora demonstra o potencial das técnicas de inteligência artificial para capturar relações complexas entre propriedades físicas do solo e seu comportamento durante a compactação, alinhando-se diretamente ao objetivo do presente trabalho de utilizar machine learning para otimizar a previsão dos parâmetros de compactação a partir de propriedades índice.

$$w_{ót(\%)} = -4,9 + 1,333Y_1 + 0,17Y_2 + 0,0043Y_1^2 + 0,032Y_1^2 - 0,049Y_1.Y_2 \quad (21)$$

$$Y_1 = 9,3 - 0,16LL + 0,104F + 0,0062LL^2 + 0,000009F^2 - 0,0011LL.F$$

$$Y_2 = 15,4 - 0,02LP - 0,143S - 0,0115LP^2 + 0,0007S^2 + 0,001LP.S$$

$$\gamma_{d,máx(kN/m^3)} = 10,68 - 1,376Y_1 + 1,065Y_2 - 0,1027Y_1^2 - 0,1802Y_2^2 + 0,3232Y_1 \quad (23)$$

$$Y_1 = 22,2 - 0,02LL - 0,066F - 0,0012LL^2 + 0,0002F^2 + 0,00043LL.F$$

$$Y_2 = 19,8 - 0,15LP - 0,005S - 0,006LP^2 + 0,0002F^2 + 0,00043LL.F$$

Pantoja & Anjos (2018)

O modelo desenvolvido neste estudo utiliza a regressão linear múltipla (RLM) como ferramenta principal, analisando um extenso banco de dados composto por 1172 amostras de solos, das quais 87% são classificadas como solos de baixa plasticidade ($LL \leq 50\%$). As variáveis independentes incluem propriedades índice do solo, com intervalos representativos das condições geotécnicas brasileiras:

- Limite de Liquidez (LL): $23\% \leq LL \leq 73,5\%$
- Limite de Plasticidade (LP): $10\% \leq LP \leq 51,9\%$
- Fração de Finos (F): $13\% \leq F \leq 100\%$
- Peso Específico Seco Máximo ($\gamma_{d,máx}$): $11,1 \text{ kN/m}^3 \leq (\gamma_{d,máx}) \leq 21,36 \text{ kN/m}^3$
- Umidade Ótima ($w_{ót}$): $7,2\% \leq (w_{ót}) \leq 44,4\%$

As equações propostas por Pantoja & Anjos para estimar os parâmetros de compactação são:

$$w_{ót-RLM(\%)} = 0,023 + 0,231.LP + 0,160.LL + 0,100.F \quad (23)$$

$$\gamma_{d,m\acute{a}x_RLM}(kN/m^3) = 21,902 - 0,035.LL - 0,030.F - 0,079.LP \quad (24)$$

$$w_{\acute{o}t_RD(\%)} = 4,299 + 0,1548.(LP + F) \quad (25)$$

$$\gamma_{d,m\acute{a}x_RD}(kN/m^3) = 21,1467 - 0,0691.LL - 0,0637.LP \quad (26)$$

As formulações apresentadas anteriormente não abrangem a totalidade das pesquisas mencionadas na literatura especializada, tendo sido selecionadas apenas aquelas cujas variáveis estão integralmente disponíveis no banco de dados utilizado neste estudo, de modo a garantir consistência estatística e comparabilidade entre os modelos analisados. A Tabela 1 sintetiza os trabalhos referenciados, indicando os respectivos parâmetros de entrada considerados em suas análises, bem como evidenciando as abordagens predominantes adotadas por diferentes autores. Uma tendência clara observada é a predominância do uso dos limites de Atterberg — limite de liquidez (LL) e limite de plasticidade (LP) — nas correlações para estimativa dos parâmetros de compactação, uma vez que todas as formulações listadas incorporam pelo menos um desses limites. Tal recorrência justifica-se pelo fato desses índices refletirem diretamente o comportamento reológico e a sensibilidade à variação de umidade dos solos finos, sendo amplamente reconhecidos como indicadores representativos da plasticidade e da estrutura interna do material.

Além dos limites de Atterberg, outras variáveis frequentemente incluídas nos estudos são a energia mecânica aplicada durante a compactação e os dados granulométricos, especialmente no que se refere à distribuição das partículas finas, os quais apresentam influência significativa na determinação da umidade ótima ($w_{\acute{o}t}$) e do peso específico seco máximo ($\gamma_{d,m\acute{a}x}$). A incorporação desses parâmetros permite capturar efeitos associados à reordenação das partículas e à resposta do solo sob diferentes níveis de esforço de compactação.

A densidade real média dos grãos (ρ_s) também é considerada em parte das formulações, embora seu valor isolado apresente pouca variação entre solos de mesma origem mineralógica, resultando em impacto direto limitado nas correlações. Contudo, sua relevância está associada à relação com a curva de saturação do solo e com o índice de vazios, influenciando indiretamente os parâmetros de compactação e contribuindo para uma representação mais fiel do comportamento físico do material. Dessa forma, observa-se que a combinação dessas variáveis busca equilibrar simplicidade operacional e

representatividade geotécnica, reforçando a importância de modelos que utilizem parâmetros de fácil obtenção em laboratório sem comprometer a precisão das estimativas.

Tabela 1 - Resumo das propriedades consideradas na criação de formulações

Autores	LL	LP	IP	F	S	G	G_s	$w_{ót}$	E	PN
Al-Khafaji (1993)	X	X								X
Sridharan & Nagaraj (2005)		X							X	X
Sivrikaya et ali (2008)		X						X	X	X
Gunaydin (2009)	X	X		X	X	X	X			X
Matteo et ali (2009)	X		X				X	X		X
Đoković et ali (2013)	X	X								X
Ardakani & Kordnaeij (2017)	X	X		X	X	X				X
Pantoja & Anjos_RLM (2018)	X	X		X						X
Pantoja & Anjos_RD (2018)	X	X		X						X

Onde: LL – Limite de Liquidez; LP – Limite de Plasticidade; IP – Índice de Plasticidade; F – Porcentagem de Finos; S – Porcentagem de Areia; G – Porcentagem de Pedregulhos; G_s – Densidade Real dos Grãos; $w_{ót}$ – Umidade Ótima; E – Energia de Compactação; PN – Proctor Normal.

Para a estimativa do peso específico seco máximo ($\gamma_{d,máx}$), é comum a inclusão da variável peso específico (γ), que pode ser o valor verdadeiro medido ou um valor previsto conforme o método adotado em cada estudo. Essa abordagem permite a formulação de correlações mais robustas para determinar o par ordenado que representa o ponto de máxima compactação do solo, ou seja, o peso específico seco máximo e a umidade ótima correspondentes a uma determinada energia de compactação. Essa análise das variáveis de entrada reforça a importância de selecionar parâmetros representativos e disponíveis para a construção de modelos preditivos, especialmente quando se pretende aplicar técnicas de machine learning. A escolha criteriosa das variáveis impacta diretamente na qualidade e na generalização dos modelos, alinhando-se ao objetivo deste trabalho de desenvolver ferramentas eficientes para a previsão dos parâmetros de compactação a partir das propriedades índice do solo.

2.3 Extreme Gradient Boosting (XGBoost)

O avanço da Transformação Digital na Engenharia Civil tem impulsionado a transição de métodos puramente empíricos para abordagens baseadas em dados (Data-Driven Engineering). Neste cenário, o XGBoost (eXtreme Gradient Boosting) consolidou-se como uma das ferramentas mais avançadas e eficientes no campo do aprendizado de máquina (Machine Learning). Sua aplicação é particularmente valiosa em problemas geotécnicos que demandam robustez, precisão e elevada capacidade de generalização (sobretudo em solos) materiais inerentemente heterogêneos e complexos.

Diferente de modelos lineares simples, a estrutura do XGBoost baseia-se no princípio de Boosting por gradiente. Esse método constrói sequencialmente árvores de decisão, onde cada nova unidade é treinada especificamente para minimizar os erros residuais das árvores anteriores. Esse processo iterativo resulta em modelos altamente acurados e resilientes a ruídos, permitindo que o algoritmo "aprenda" padrões sutis em bases de dados que métodos tradicionais de regressão frequentemente ignoram.

Na prática da geotecnia, a previsão de parâmetros de compactação — como a Umidade Ótima e o Peso Específico Seco Máximo é vital para o controle tecnológico de aterros, barragens e obras viárias. Tradicionalmente, esses valores são obtidos via ensaios de Proctor, que demandam tempo e volume considerável de amostras. A proposta deste trabalho utiliza o XGBoost para prever esses parâmetros a partir de propriedades índice de obtenção mais rápida e econômica, como os Limites de Atterberg (LL, LP) e a Granulometria (frações de areia, silte e argila).

O algoritmo destaca-se por sua capacidade intrínseca de capturar relações não lineares e interações multidimensionais entre essas variáveis. Por exemplo, a interação entre o teor de finos e a plasticidade da argila exerce uma influência complexa na compactação que dificilmente é descrita por uma equação linear simples. O XGBoost não apenas modela essas complexidades, mas também realiza uma seleção automática de atributos, identificando quais variáveis possuem maior poder preditivo e descartando redundâncias (Oliveira et al., 2021; Silva, 2022).

A superioridade do XGBoost sobre outros métodos, como Random Forest ou Support Vector Machines (SVM), reside em sua arquitetura otimizada. Ele implementa uma função objetivo que combina uma função de perda (loss function), para medir a proximidade entre a previsão e o valor real, com um termo de Regularização (L1 - Lasso e L2 - Ridge).

A regularização é um diferencial crítico: ela penaliza a complexidade excessiva do modelo, mitigando o risco de overfitting (sobreajuste), onde o modelo "decora" os dados de treino, mas falha ao ser aplicado em dados novos. Além disso, o XGBoost utiliza o método de Newton para otimizar a função de perda, permitindo uma convergência mais rápida e precisa para o mínimo global do erro.

Outro diferencial técnico é a eficiência digital. O algoritmo utiliza uma estrutura de dados denominada DMatrix, que otimiza o uso da memória e permite o processamento paralelo, utilizando todos os núcleos do processador. Isso possibilita o processamento de grandes inventários de dados geotécnicos em frações de segundo, tornando-o viável para aplicações em tempo real no canteiro de obras (Souza, 2023).

Bases de dados de sondagens e ensaios laboratoriais frequentemente apresentam lacunas ou valores ausentes (missing values). Enquanto muitos algoritmos exigem a exclusão dessas amostras ou imputações estatísticas complexas, o XGBoost possui um algoritmo de aprendizado de direção esparsa. Ele aprende automaticamente a melhor direção para lidar com dados faltantes com base na redução da perda, tornando o modelo robusto mesmo diante de campanhas de investigação incompletas.

Estudos comparativos demonstram que o XGBoost frequentemente apresenta o melhor desempenho em métricas de erro fundamentais, como o Erro Médio Absoluto (MAE), a Raiz do Erro Quadrático Médio (RMSE) e o Coeficiente de Determinação (r^2).

A implementação do XGBoost em ambiente Python segue um fluxo rigoroso de ciência de dados. A etapa de ajuste de hiperparâmetros (Hyperparameter Tuning) é onde o modelo é refinado. Parâmetros como taxa de aprendizado: controla a magnitude das correções a cada nova árvore; profundidade máxima: limita o crescimento da árvore para evitar complexidade desnecessária; Subamostragem: utiliza frações aleatórias dos dados para treinar cada árvore, aumentando a robustez. Esses ajustes são validados por meio de Validação Cruzada (k-fold Cross-Validation), garantindo que o desempenho do modelo seja consistente em diferentes subconjuntos da base de dados, equilibrando o viés e a variância.

Um dos maiores benefícios do XGBoost para a engenharia é a transparência. Através da análise de Feature Importance, o modelo quantifica a contribuição de cada variável para a previsão final. Além da eficiência preditiva, essa funcionalidade permite uma análise profunda da lógica de decisão, fornecendo o que chamamos de consistência fenomenológica. No contexto deste trabalho, essa transparência é fundamental para validar que o modelo não está apenas realizando correlações estatísticas aleatórias, mas

sim capturando a física real do solo. Ao atribuir pesos significativos aos Limites de Atterberg, o algoritmo reconhece matematicamente que a plasticidade — reflexo da morfologia laminar dos argilominerais e da dinâmica da dupla camada difusa — é o principal controlador da macro-resposta de compactação, corroborando teorias clássicas da mecânica dos solos (Holtz & Kovacs, 1981; Das, 2011).

A robustez do XGBoost frente à heterogeneidade dos solos tropicais também se deve ao seu algoritmo de Weighted Quantile Sketch. Diferente de outros métodos de boosting que tratam todos os dados de forma uniforme, o XGBoost utiliza um sistema de pesos que permite encontrar pontos de divisão (splits) otimizados mesmo em distribuições de dados altamente assimétricas ou platocúrticas, como as observadas nas frações granulométricas deste estudo.

Essa capacidade de lidar com a variabilidade inerente aos materiais naturais garante que o modelo mantenha uma elevada precisão tanto em solos predominantemente finos quanto em composições bimodais complexas. Esse refinamento matemático assegura que o processo de poda (pruning) das árvores ocorra de forma a preservar apenas as interações que possuem relevância física real, minimizando a propagação de erros em bases de dados diversificadas.

Por fim, a integração do XGBoost no fluxo de trabalho da engenharia civil representa um salto qualitativo na transição para a Engenharia 4.0. A capacidade de gerar previsões instantâneas a partir de ensaios de caracterização rápida transforma o modelo em uma ferramenta de controle tecnológico preventivo.

Em grandes obras de terraplenagem, onde a variação do teor de umidade em jazidas pode comprometer o cronograma e a qualidade das camadas, a utilização desta abordagem permite a antecipação de alvos hídricos e densidades máximas com horas ou dias de antecedência em relação aos laudos laboratoriais convencionais. Essa agilidade não apenas mitiga riscos operacionais, mas também fundamenta o desenvolvimento de sistemas de monitoramento contínuo e autônomo, essenciais para a modernização e sustentabilidade da infraestrutura terrestre.

O uso do Python como plataforma de integração e do XGBoost como motor de cálculo está em total sintonia com as tendências globais de digitalização da infraestrutura, oferecendo uma base sólida para a tomada de decisões estratégicas em projetos de engenharia civil (Oliveira et al., 2021; Silva, 2022; Souza, 2023).

3) DADOS (LOCAIS DE ESTUDO)

Esta seção apresenta a análise estatística dos resultados de amostras de solos, totalizando 2.730 resultados laboratoriais utilizados na caracterização dos solos, que permitiu compreender a distribuição dos parâmetros geotécnicos estudados. Em relação aos dados, 74% foram retirados de amostras brasileiras e 26% de amostras internacionais. As variáveis avaliadas incluem: peso específico seco máximo ($\gamma_{d,máx}$), teor de umidade ótima ($w_{ót}$), limite de liquidez (LL), limite de plasticidade (LP) e composição granulométrica (frações grossas e finas).

Os dados foram analisados com o uso de Cartas de Plasticidade com o trabalho de Casagrande (1948) iniciando. Em seguida têm-se os trabalhos de Polidori (2007) e Moreno-Maroto & Alonso-Azcárate (2017).

A carta de plasticidade de Casagrande (1948) é um marco na Engenharia Geotécnica com quase 80 anos de uso. O gráfico de é o pilar do Sistema Unificado de Classificação de Solos (SUCS). Ele utiliza a relação entre o Limite de Liquidez (LL) e o Índice de Plasticidade (PI) para categorizar solos finos. O gráfico segue conforme Figura 4 e dá ênfase a linha “A” em que cabe melhor observação mais adiante no trabalho.

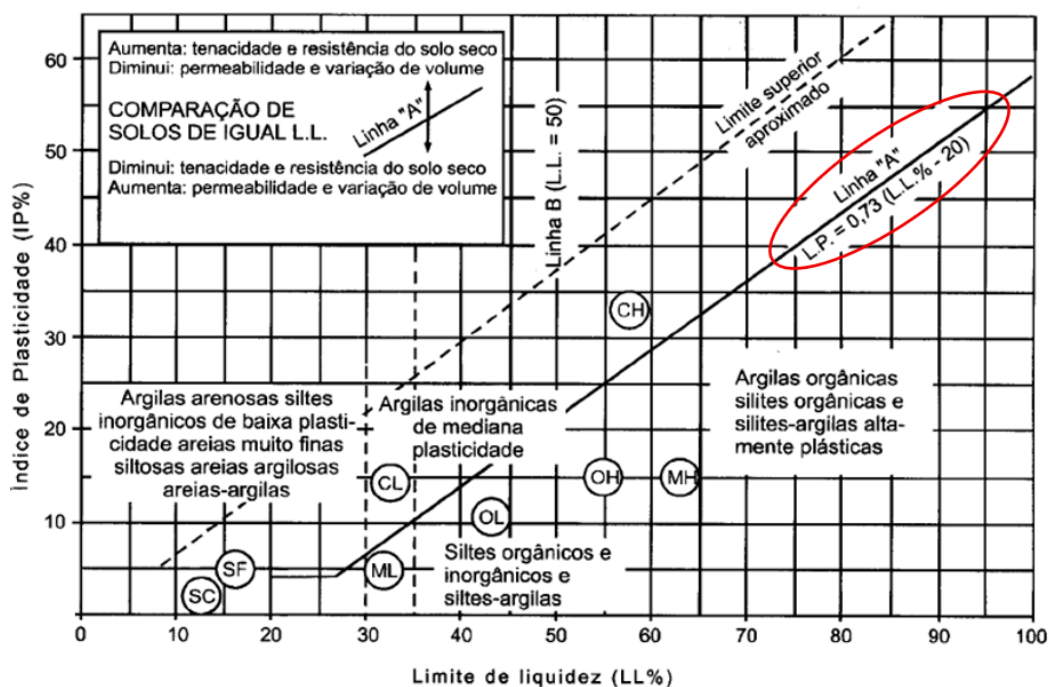


Figura 4 - Carta de Plasticidade de Casagrande

O gráfico de Casagrande (1948) enfrenta uma crítica técnica severa quanto à sua fundamentação matemática. O ponto central é a natureza estatística de seus eixos: o eixo

X representa o Limite de Liquidez (LL), enquanto o eixo Y representa o Índice de Plasticidade (IP). Como o IP é definido pela equação $IP = LL - LP$, o termo LL aparece em ambos os eixos. Estatisticamente, isso cria uma correlação espúria (artificial). Ao plotar IP x LL o gráfico força uma tendência linear que é, em grande parte, um artifício matemático e não necessariamente uma reflexão da realidade física ou mineralógica do solo. Essa dependência intrínseca pode induzir ao erro de concluir que um solo está se tornando "mais plástico" apenas porque o seu LL aumentou. Na verdade, essa percepção pode ser apenas uma consequência da forma como os dados são organizados no gráfico, mascarando o comportamento real do material. Gutiérrez (2006) argumenta que o sistema de coordenadas original apresenta uma "realidade falsa". Para uma classificação fisicamente mais representativa, ele propõe que Casagrande deveria ter utilizado o Limite de Plasticidade (LP) como variável independente (eixo Y). Se o gráfico fosse estruturado como LL x LP, a Linha A (que separa siltes de argilas) teria sido traçada sobre variáveis estatisticamente independentes. É altamente provável que essa nova abordagem resultasse em uma inclinação diferente para a Linha A, oferecendo uma fronteira mais precisa e fundamentada na física dos solos para distinguir siltes e argilas.

Apesar dessa falha matemática, a carta de Casagrande continua sendo o padrão da indústria devido à sua conveniência prática e ao enorme volume de correlações de engenharia (como compressibilidade e resistência) que foram construídas em anos de uso.

Nas Figuras 5 e 6 tem-se os resultados dos 2730 dados aplicados a Carta de Plasticidade de Casagrande com a representação na própria carta de Casagrande e, em seguida, na forma de porcentagem das frações de solo. As porcentagens encontradas foram: 47% para CL- ML (argilas e siltes de baixa plasticidade), 36,5% para CL (argilas de baixa plasticidade), 8,3 % para CH (argilas de alta plasticidade), 5,8% de MH - OH (Silte de alta plasticidade/ solo orgânico) e 2,3% de ML (silte de baixa plasticidade).

As frações CL e CL-ML dominam a quantidade desse banco de dados com 83,5% no total. Solos CL exibem boa compactação em ensaios Proctor, com Wot entre 12-18% devido à retenção de água pelas partículas laminares, alcançando $\gamma_{d,max}$ de 17-19 kN/m³; são sensíveis a excesso de umidade, podendo formar agregados plásticos que reduzem $\gamma_{d,max}$. Já os CL-ML, com características transitórias, apresentam Wot menor (10-15%) e maior $\gamma_{d,max}$ (18-20 kN/m³) por influência siltosa, compactam mais uniformemente, mas demandam controle fino para evitar segregação de finos e perda de trabalhabilidade.

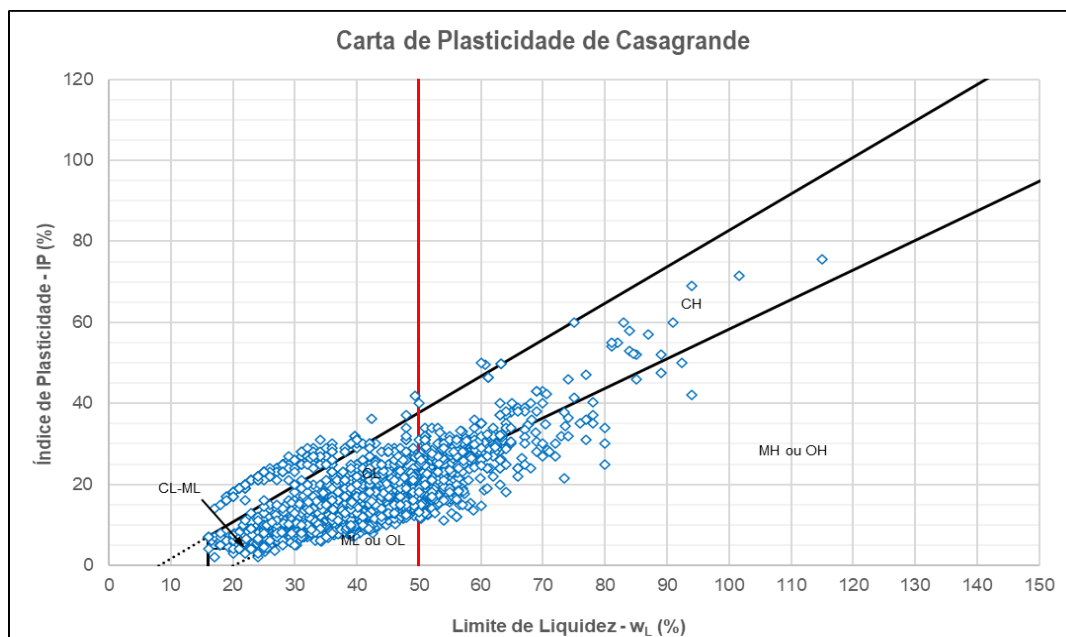


Figura 5 - Carta de Plasticidade de Casagrande - 2730 dados da pesquisa

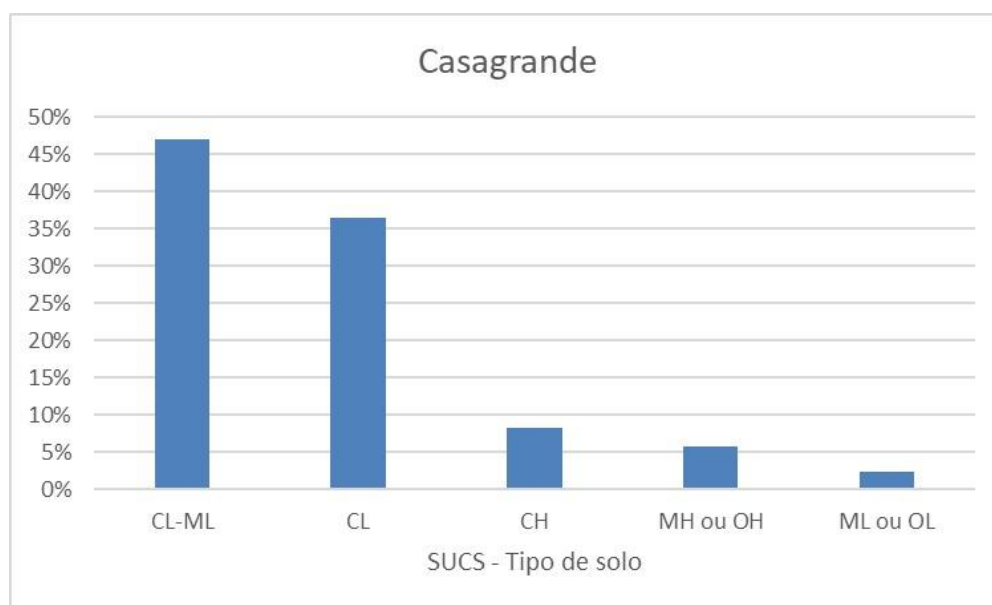


Figura 6 - Carta de Plasticidade de Casagrande - % das frações argilosas

O método de Polidori (1997) baseou-se na análise laboratorial de seis solos inorgânicos naturais e suas respectivas misturas com areia sílica fina, variando controladamente o teor de argila. Polidori (1997) observou que, para solos com minerais argilosos laminares (platy clay minerals), existe uma relação linear direta entre o aumento do teor de argila e os valores de LL e LP (exceto em teores muito baixos de argila). O autor derivou equações que vinculam a plasticidade à composição granulométrica,

permitindo recalcular os limites de Atterberg e definir zonas de classificação baseadas na física da mistura, reduzindo a subjetividade puramente comportamental do método de Casagrande. Lembra-se que argilominerais laminares (platy clay minerals), como a caulinita, a illita e a montmorilonita, são partículas microscópicas em forma de placas ou lamelas que constituem a fração mais ativa dos solos finos. Sua geometria e físico-química são os verdadeiros "motores" das propriedades que medimos nos limites de Atterberg. A morfologia laminar é fundamental porque confere ao solo uma elevada área superficial específica e a presença de cargas elétricas negativas em suas faces. No contexto do comportamento geotécnico, essa estrutura permite a formação da camada dupla difusa, onde a água é fortemente atraída e retida entre as lamelas. Essa característica é o que possibilita a plasticidade: a forma achatada permite que as partículas deslizem umas sobre as outras sem perder a coesão, enquanto a água retida atua como lubrificante e agente de ligação. Assim, a predominância desses minerais dita a alta compressibilidade, a baixa permeabilidade e a anisotropia da resistência ao cisalhamento, diferenciando mecanicamente uma argila de um silte (que possui partículas mais esferoidais e menos ativas).

A proposta de Polidori (2007) é um gráfico de IP em função de LL mas incorporando o teor de argila (CF, fração $< 2 \mu\text{m}$) como parâmetro interdependente, permitindo classificar solos inorgânicos com minerais argilosos laminares e CF não muito baixa. A metodologia permite classificação alternativa a Casagrande, priorizando o controle físico da fração argilosa na plasticidade e assim, útil para solos onde granulometria fina é incerta. O gráfico é mostrado na Figura 7.

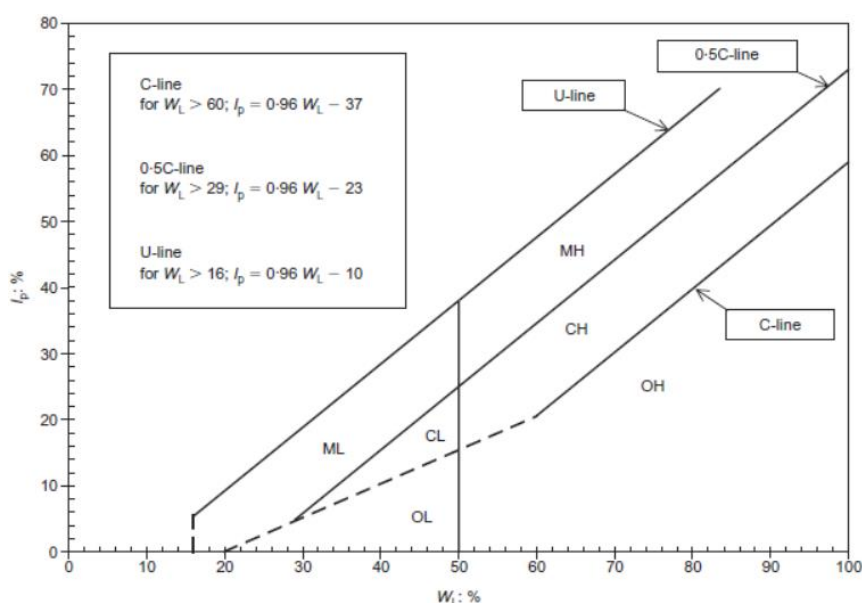


Figura 7 - Carta de Plasticidade de Polidori

Nas Figuras 8 e 9 tem-se os resultados dos 2730 dados aplicados a sugestão de Polidori (2007) com a representação na própria carta de Polidori e, em seguida, na forma de porcentagem das frações de solo. As porcentagens encontradas foram: 55,7% para CL (argila de baixa plasticidade), 30,1% para ML (silte de baixa plasticidade), 10 % para CH (argilas de alta plasticidade), 2,1% de MH (silte de alta plasticidade) e 2,1% de OH-NPC (solo orgânico). NPC são solos com argilominerais laminares (platy clay minerals), como a caulinita, a illita e a montmorilonita. As frações CL e ML totalizam 85,8% no total.

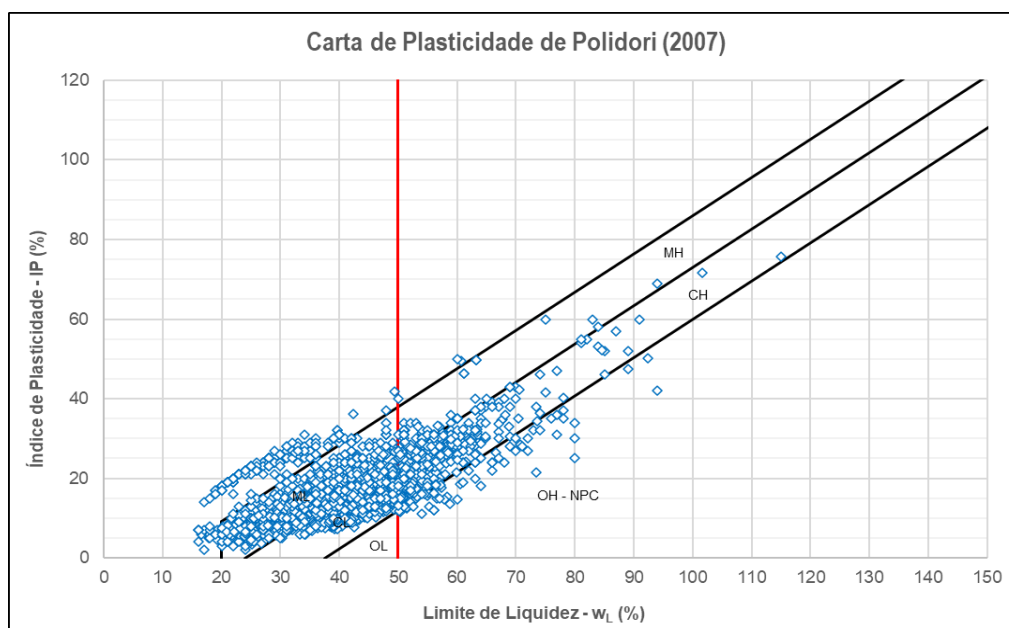


Figura 8 - Carta de Plasticidade de Polidori - % das frações argilosas

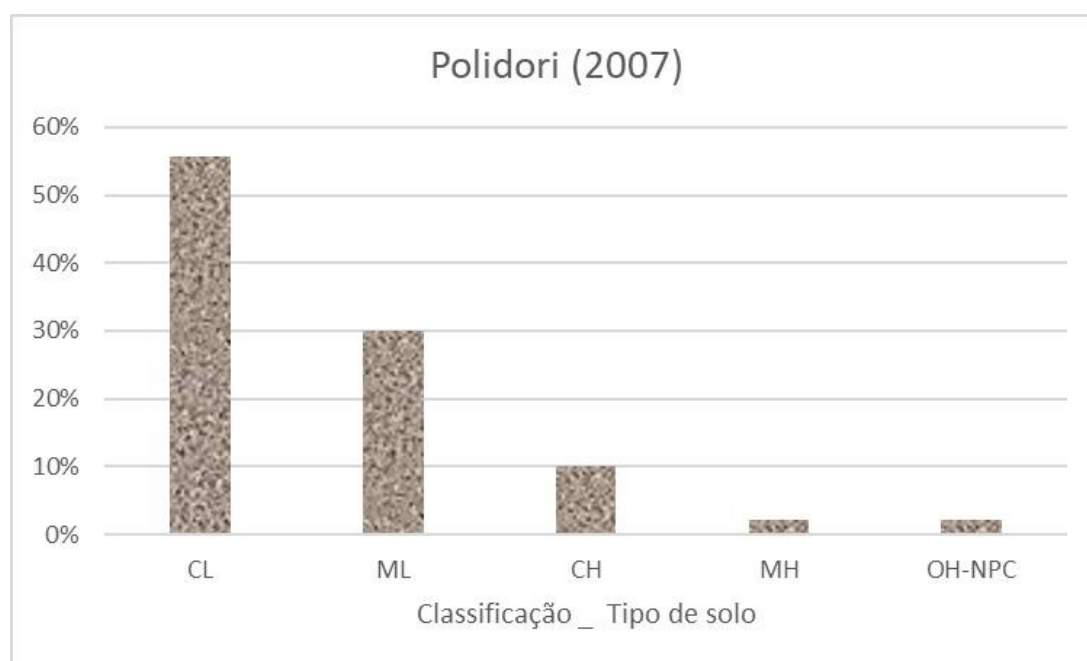


Figura 9 - Carta de Plasticidade de Polidori - % das frações argilosas

O trabalho do Moreno-Maroto & Alonso-Azcárate (2017) propõe uma forma alternativa de obter parâmetros de consistência (limite de plasticidade – LP) usando um método de flexão (bending) e usa esses resultados para sugerir um “novo chart” (gráfico) de classificação/interpretação da plasticidade dos solos. Lembra-se que o limite de liquidez LL, não é afetado pelas análises. O “novo chart” continua com a lógica de cartas tipo Casagrande (IP x LL) mas reinterpretando as faixas/zonas de plasticidade com base em um critério de ensaio menos dependente do operador. De acordo com o trabalho de Moreno-Maroto (2017) uma nova proposta de gráfico de classificação dos solos é apresentada na Figura 10. Decidiu-se manter os eixos com LL e IP (o IP é um parâmetro tradicional, bem conhecido e utilizado mundialmente) mas com a adição de duas novas linhas para separar argilas e siltes e são representadas: a linha C e a linha M. A linha C tem relação entre IP e LL conforme a equação $IP = 0,4855 \times LL$, ao passo que na linha M é $IP = 0,3311 \times LL$. Ainda, decidiu-se preservar a nomenclatura tradicional para solos: C para argilas (clays), M para siltes (silts - mjala), L para baixa compressibilidade e H para alta compressibilidade. Em solos orgânicos, os termos M e C devem ser substituídos por O, em consonância com os requisitos da ASTM D2487-11. Potencialmente, a relação espúria entre IP e LL foi retirada com a nova definição das retas do gráfico de classificação de Moreno-Maroto & Alonso-Azcárate (2017)

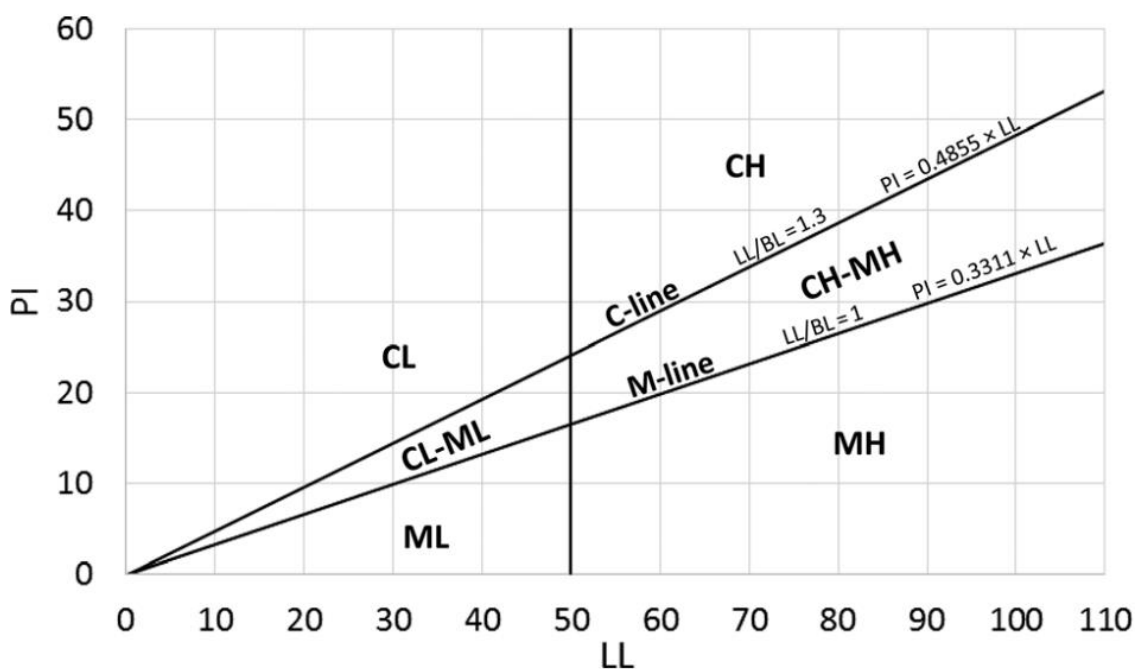


Figura 10 - Carta de Plasticidade de Moreno Maroto & Alonso Azcárate

Um aspecto interessante dessa nova abordagem é que o grupo de solos com características intermediárias entre siltes e argilas não fica restrito apenas a uma pequena área com valores muito baixos de LL, como descrito no gráfico tradicional de Casagrande. Assim, na nova proposta, esse tipo de solo se localiza ao longo da área delimitada entre a linha C e a linha M. Nesse caso, o grupo CL-ML se estende até valores de LL de 50; e, quando $LL > 50$, define-se um novo grupo, CH-MH, que corresponde a argilas siltosas ou siltes argilosos de alta plasticidade/alta compressibilidade.

Na opinião dos autores da metodologia, essa nova visão é mais realista do que a tradicional, porque, na prática, solos com características sobrepostas entre siltes e argilas podem ser encontrados não apenas quando $LL < 25$, mas ao longo de todos os valores possíveis de LL (por exemplo, há muitos solos silto-argilosos cujos LL são maiores que 50 e que são descritos simplesmente como CH ou MH, quando, na verdade, apresentam propriedades intermediárias de ambos os grupos; assim, uma designação CH-MH parece ser mais adequada nesses casos). Por fim, a percepção dos autores é de que a nova proposta de classificação é mais realista e, de fato, a nomenclatura está mais bem ajustada às características do solo usados na avaliação criteriosa da nova abordagem de classificação de solos argilosos e assim, a proposta de classificação apresentada no trabalho pode ser altamente proveitosa para se obter uma melhor compreensão das propriedades do solo.

Nas Figuras 11 e 12 tem-se os resultados dos 2730 dados aplicados a sugestão de Moreno-Maroto & Alonso-Azcárate (2017) com a representação na própria carta dos autores e, em seguida, na forma de porcentagem das frações de solo. As porcentagens encontradas foram: 39,6% para ML (silte de baixa plasticidade), 33,3% para CL-ML (argila e silte de baixa plasticidade), 12,9 % para CL (argilas de baixa plasticidade), 6,4% de CH-MH (argila e silte de alta plasticidade), 5,4% de CH (argila de alta plasticidade) e 2,4% de MH (silte de alta plasticidade). As frações ML e CL-ML totalizam quase 73% no total de dados analisados. A carta de plasticidade de Moreno-Maroto & Alonso-Azcárate (2017) ajusta adequadamente uma lacuna de solos que estão na parte superior da carta de plasticidade tanto de Casagrande quanto de Polidori que ficam “órfãs” de classificação. Moreno-Maroto & Alonso-Azcárate admitem que essa zona os solos que ali estão são classificados como CL (argila de baixa plasticidade). As frações CL e ML totalizam aproximadamente 73% no total de dados. Diante das três propostas vê-se que valores entre 73% a 86% classificam como CL, ML ou ambos.

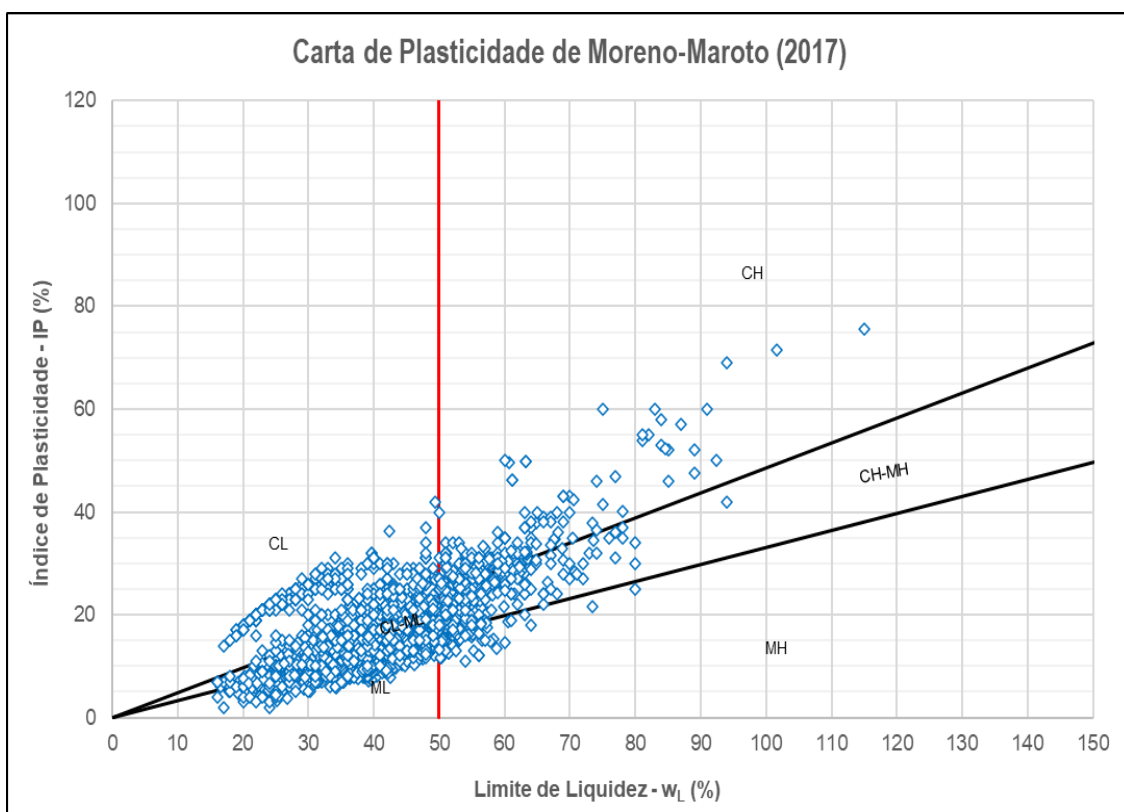


Figura 11 - Carta de Plasticidade de Moreno Maroto & Alonso Azcárate % frações argilosas



Figura 12 - Carta de Plasticidade de Moreno-Maroto & Alonso Azcárate - % das frações argilosas.

Dentre as variáveis do problema apresentado quer seja: limites de Atterberg (liquidez e plasticidade) porcentagens de finos (silte+argila) e de grossos (areia e pedregulhos) além dos parâmetros teor de umidade ótima e peso específico seco máximo (ensaio Proctor normal) mostra-se na Tabela 2 as estatísticas básicas do problema.

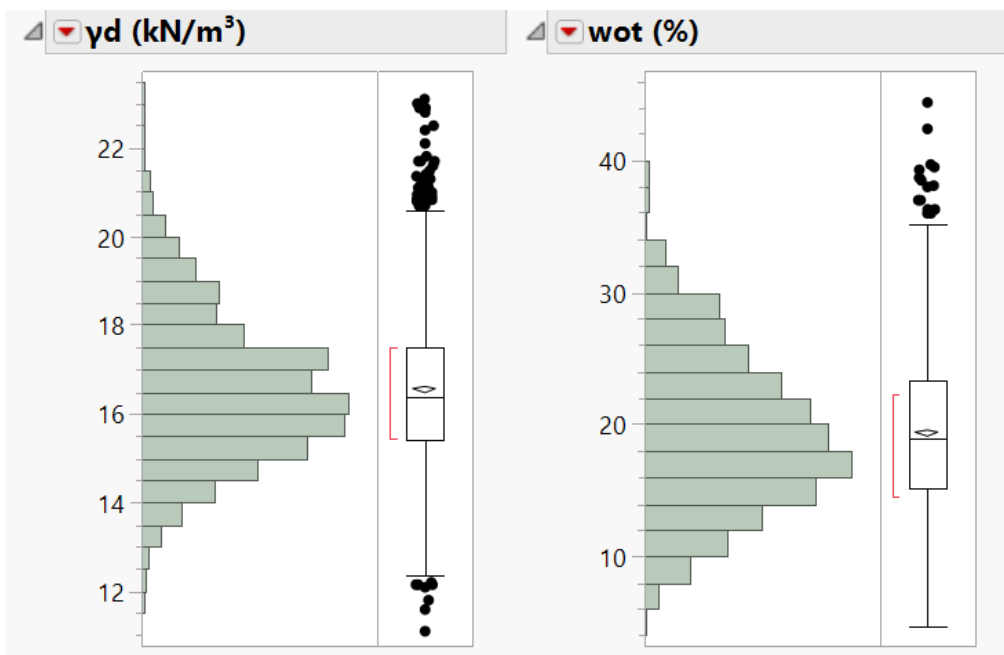
Tabela 2 - Estatísticas descritivas das variáveis do modelo de estudo

Variável	γ_d (kN/m ³)	w_{ot} (%)	LL (%)	LP (%)	Grossos (%)	Finos (%)
Média	16,5561	19,3702	41,4171	25,2198	42,4554	57,545
EP Média	0,0318908	0,113293	0,203584	0,153192	0,41523	0,415241
Mediana	16,38	18,885	40,13	27,115	41,035	58,965
Moda	15,19	20	36	20	55	45
DesvPad	1,66628	5,91948	10,6371	8,00418	21,6955	21,6961
Variância	2,77648	35,0403	113,149	64,0669	470,696	470,721
Curtose	0,37	-0,09	3,17	0,98	-0,28	-0,28
Assimetria	0,44	0,36	0,91	-0,82	0,33	-0,33
Amplitude	12	39,7	99	53	97	97
Mínimo	11,1	4,7	16	2	0	3
Máximo	23,1	44,4	115	55	97	100

Ainda, complementam as informações estatísticas, os histogramas de cada uma das variáveis e dos parâmetros resultantes do ensaio de compactação conforme Figura 13. Essas representações gráficas permitem visualizar de forma objetiva os padrões de concentração da média, assimetria e dispersão dos dados, complementando as estatísticas descritivas previamente conforme Tabela 2. Além disso, fornecem embasamento visual para a avaliação da consistência do banco de dados e para a identificação de possíveis tendências ou valores extremos, reforçando a confiabilidade do conjunto amostral adotado e subsidiando as análises subsequentes de correlação entre os parâmetros geotécnicos. Em cada conjunto de histogramas têm-se o Box-Plot equivalente em que os valores espúrios (*outliers*) são evidenciados em cada resultado. Finos e grossos são os de menor quantidade de *outliers*.

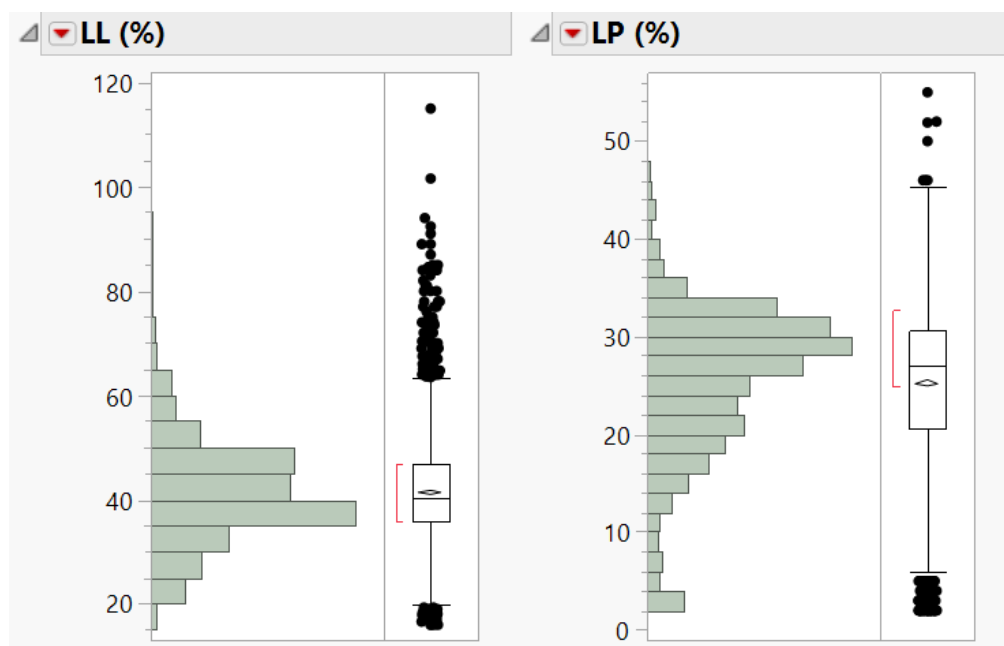
Média e EP Média mostram o valor central e a precisão da estimativa. A mediana e moda mostra as posições centrais robustas a outliers (úteis em solos heterogêneos) enquanto o desvio padrão, variância e amplitude medem dispersão (alta variância indica solos mais não homogêneos). A assimetria e curtose avaliam formato da distribuição (assimetria >0 sugere cauda direita, comum em finos; curtose alta indica picos extremos).

Os valores mín/máx são os limites absolutos para detectar anomalias. Os histogramas visualizam a distribuição de frequência, revelando normalidade (gaussiana ideal para modelagem), multimodalidade (mistura de solos) ou assimetria prática. Essas métricas validam dados (ex.: rejeitar outliers em $LL > "x" \%$), suportam modelagem (XGBoost usa variância para priorizar features), otimizam projetos (amplitude alta exige mais sondagens) e quantificam incertezas (desvio padrão define intervalos de confiança), elevando a confiabilidade em compactação e estabilidade.



a)

b)



c)

d)

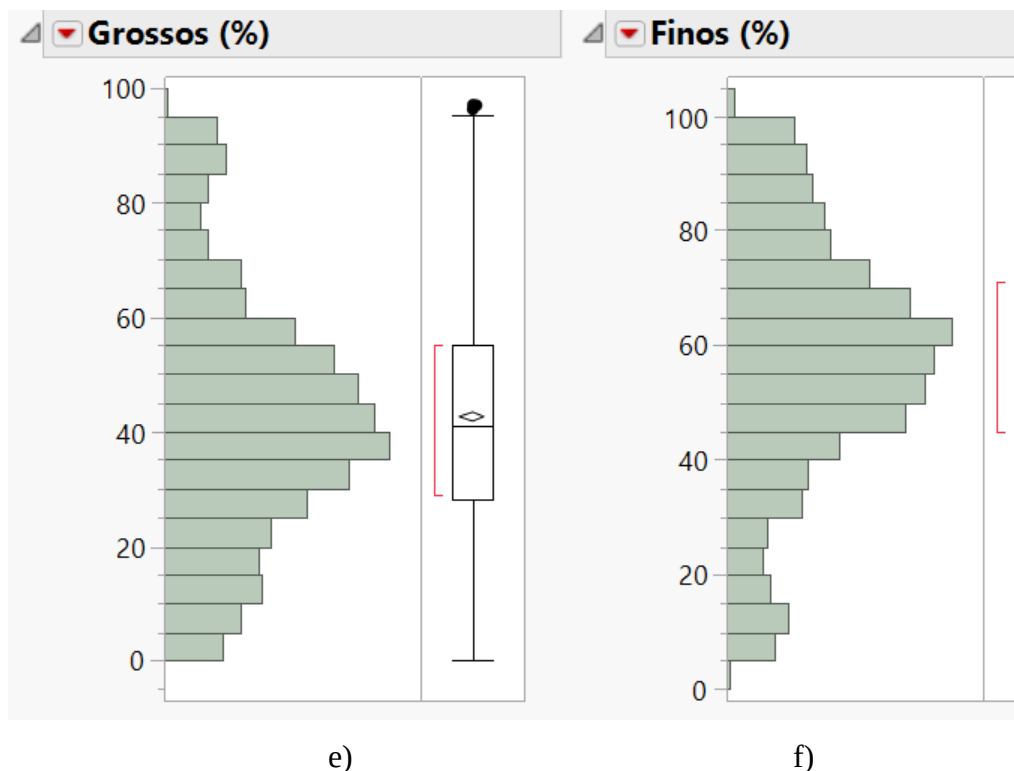


Figura 13 - Histogramas das distribuições dos conjuntos de dados referente à a) $\gamma_{d,m\acute{a}x}$ (kN/m³), b) $w_{\acute{o}t}$ (%), c) LL(%), d) LP(%), e) G(%) e f) F(%).

Em tempo, a assimetria indica para qual lado a "cauda" do seu histograma está se alongando. Ela nos diz se os dados estão concentrados em valores baixos ou altos. Assimetria Positiva (Direita): A cauda se estende para os valores maiores (à direita). Isso é comum no Limite de Liquidez (LL): a maioria dos solos tem LL baixo/médio, mas alguns poucos solos muito plásticos puxam a média para cima ($Média > Mediana > Moda$). Assimetria Negativa (Esquerda): A cauda se estende para os valores menores (à esquerda). No seu caso, o Limite de Plasticidade (LP) apresentou assimetria negativa (-0,82), o que significa que há alguns valores muito baixos "puxando" a média para a esquerda ($Moda > Mediana > Média$). Simétrica: Os dois lados são espelhados (como a Distribuição Normal).

A curtose não mede apenas o "pico", mas sim a concentração de dados no centro em relação às caudas (a presença de outliers); Leptocúrtica ($K > 0$): O gráfico é muito "pontiado". Os dados estão muito concentrados em torno da média. Na Tabela 2, o LL teve curtose de 3,17, indicando que a maioria dos valores está muito próxima, mas existem caudas com valores extremos (outliers); Platocúrtica ($K < 0$): O gráfico é "achatado". Os dados estão mais espalhados e a curva é mais larga. Suas variáveis de Grossos e Finos (-0,28) são platocúrticas, o que faz sentido, pois você tem solos de todos

os tipos (de 0% a 100%), sem uma concentração única tão forte; Mesocúrtica ($K \approx 0$): É a distribuição normal perfeita. Uma assimetria ou curtose muito alta pode indicar que um ensaio foi feito errado ou que você tem uma amostra de solo completamente diferente das outras (um outlier).

Por fim, uma forma comentada da Tabela 2 e dos Histogramas na Figura 13 é dada a seguir:

γ_{dmax} (kN/m³): tendência central moderada (média 16,56; mediana 16,38; moda 15,19), típica de solos compactados finos/mistos; baixa dispersão (Desv.Pad 1,67; variância 2,78), indicando consistência em ensaios Proctor; assimetria positiva leve (0,44) e curtose mesocúrtica (0,37), distribuição ligeiramente assimétrica à direita; amplitude =12 (11,1-23,1), faixa realista para solos argilosos/siltosos;

wot (%): média alta (19,37%; mediana 18,89%; moda 20%), sugestiva de solos finos com retenção elevada de água; alta dispersão (Desv.Pad 5,92; variância 35,04), refletindo variabilidade natural em compactação. Assimetria positiva (0,36) e curtose platocúrtica leve (-0,09), cauda direita; amplitude ampla=39,7 (4,7-44,4%), indicando solos heterogêneos (arenosos a argilosos);

LL (%): média 41,42% (mediana 40,13%; moda 36%), compatível com solos CL/CH de baixa-média plasticidade; alta dispersão (Desv.Pad 10,64; variância 113,15), comum em solos variáveis; forte assimetria positiva (0,91) e curtose leptocúrtica (3,17), picos e cauda longa à direita; amplitude extrema=99 (16-115%), sinal de outliers ou solos orgânicos/altamente plásticos;

LP (%): média 25,22% (mediana 27,12%; moda 20%), consistente com solos plásticos moderados; dispersão moderada-alta (Desv.Pad 8,00; variância 64,07); Assimetria negativa forte (-0,82) e curtose mesocúrtica (0,98), cauda esquerda; com a amplitude igual a 53 (2-55%), variabilidade em plasticidade por teor de argila;

Grossos (%) - Fração Grossa: média 42,46% (mediana 41,04%; moda 55%), solos mistos (arenosos/pedregosos predominantes em alguns); alta dispersão (Desv.Pad 21,70; variância 470,70), heterogeneidade granulométrica; Assimetria positiva (0,33) e curtose platocúrtica (-0,28), distribuição achatada; amplitude= 97 (0-97%), ampla gama de solos;

Finos (%) - Fração Fina: média 57,55% (mediana 58,97%; moda 45%), solos siltosos/argilosos predominantes; Alta dispersão (Desv.Pad 21,70; variância 470,72), complementar aos grossos; Assimetria negativa (-0,33) e curtose platocúrtica (-0,28), espelhada aos grossos; amplitude= 97 (3-100%), confirma composição bimodal (finos vs. grossos).

4) ANÁLISE DOS RESULTADOS

Os resultados obtidos nas análises feita com a técnica de machine learning XGBoost para a estimativa dos parâmetros de compactação do solo, umidade ótima ($w_{ót}$) e peso específico seco máximo ($\gamma_{d,máx}$) são apresentados em seguida.

Na Figura 14 têm-se o resultado dos valores medidos (*actual*) e calculados (*predicted*) a partir das análises desenvolvidas com 2730 que foram usados nas análises. O modelo alcançou um coeficiente de determinação igual a 0,870, o que denotou uma elevada capacidade preditiva (tipicamente valores acima de 0,8 indicam alta correlação). O valor do coeficiente de determinação significa que 87% da variabilidade dos dados é explicada pelo modelo, deixando apenas 13% não explicado (resíduos). Esse resultado atende uma vaga gama de projetos e, sobretudo, ao que se pretende com a pesquisa aqui deste trabalho em que pese, não é a de retirar a importância dos ensaios padrão de compactação. Os 13% não explicados (via resíduos) são geralmente associados a fatores dos ensaios (fatores humanos ou não). Com a quantidade de dados empregado na análise, o resultado em si é muito atrativo sob vários pontos de vista e uso na Engenharia de pavimentos com energia de compactação Proctor Normal e, desde que se cumpra com o uso dos parâmetros de entrada conforme Tabela 2, sua utilização se mostra vantajosa para um pré-projeto.

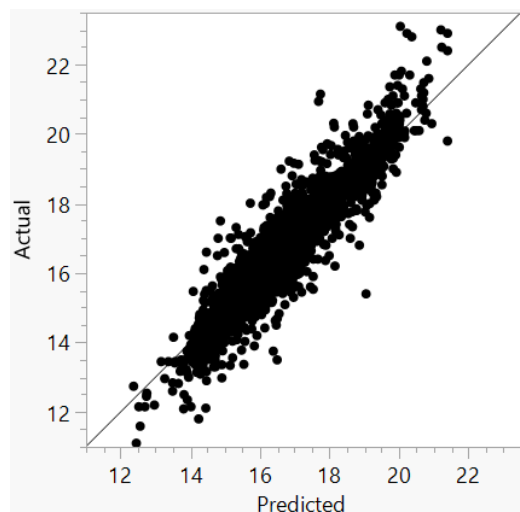


Figura 14 - Resultado via análise XGBoost – peso específico seco máximo

O modelo permite avaliar a importância de cada uma das variáveis do problema com o LP, LL, Finos e Grossos associados a 38%, 25, 25 e 12% de importância,

respectivamente. $\gamma_{d,m\acute{a}x}$ aumenta (em média) à medida que Grossos aumentam e cai com o aumento de LL, LP e Finos. Esse resultado é coerente com a mecânica de solos.

Para a umidade ótima ($w_{ót}$) têm-se Figura 15 o resultado dos valores medidos (*actual*) e calculados (*predicted*) a partir das análises desenvolvidas com 2730 que foram usados nas análises. Um coeficiente de determinação igual a 0,891 foi avaliado e indica também uma adequada medida com relação ao desempenho (maior que o valor obtido para $\gamma_{d,m\acute{a}x}$). O valor do coeficiente de determinação significa que 89,1% da variabilidade dos dados é explicada pelo modelo, deixando apenas 10,9% não explicado pelos resíduos.

Esse resultado também atende a praticamente todos os projetos e, sobretudo, ao que se pretende com a pesquisa aqui deste trabalho em que pese, não é a de retirar a importância dos ensaios padrão de compactação. Os 10,9% não explicados (via resíduos) são invariavelmente associados a fatores com erros associados aos ensaios (fatores humanos ou não). Da mesma forma que para o $\gamma_{d,m\acute{a}x}$, com a quantidade de dados empregado na análise, o resultado em si é extremamente atrativo no uso da Engenharia de pavimentos com energia de compactação Proctor Normal desde que se cumpra com a faixa de uso dos parâmetros de entrada conforme Tabela 2.

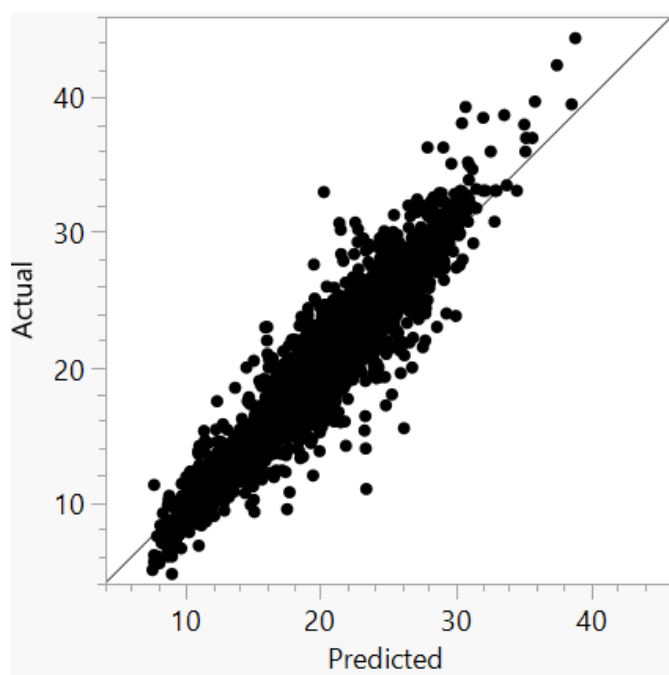


Figura 15 - Resultado via análise XGBoost – teor de umidade_ ótimo

O modelo permite avaliar a importância de cada uma das variáveis do problema com o LL, LP, Finos e Grossos associados a 42%, 26, 20 e 12% de importância, respectivamente. A ($w_{ót}$) aumenta (em média) à medida que LL, LP e Finos aumentam.

A quantidade de Finos no modelo tem discreta representatividade em relação ao seu aumento no modelo e por fim, à medida que Grossos aumentam, diminui a umidade ótima que também é coerente com a mecânica de solos.

5) CONCLUSÕES/ SUGESTÕES

Este estudo foi realizado com grande quantidade de dados e usou técnica de machine learning para a resolução de um problema de determinação do peso específico seco máximo e da umidade ótima a partir das informações das quantidades em porcentagem de LL, LP, Finos e Grossos.

Foram resultados de amostras de solos, totalizando 2.730 resultados laboratoriais utilizados na caracterização dos solos, que pode ser distribuído em 74% foram amostras brasileiras (ensaio reais em diversos estados Brasileiros) e 26% de amostras internacionais (literatura técnica).

Uma ampla gama (range) dos parâmetros de entrada foi registrado em que a Tabela 2 indica os valores médios, o erro padrão da média, a mediana, moda, desvio padrão (DesvPad), Variância, curtose, Assimetria, Amplitude, mínimo e máximo dos parâmetros de entrada como os limites de Atterberg (LL, LP) e as frações granulométricas advindas do ensaios de granulometria completa: Finos=silte + argila e Grossos=areia + Pedregulho além das saídas (respostas das análises) como peso específico seco máximo e teor de umidade ótima. Ainda, a utilização da metodologia deve seguir o específico uso de cada parâmetro no *Range* de cada uma das entradas.

Em relação aos resultados, tanto o peso específico seco máximo quanto o teor de umidade ótima foram avaliados em termos de técnica de machine learning (XGBoost) como promissora com valores do coeficiente de determinação igual a 0,870 para $\gamma_{d,m\acute{a}x}$ e 0,891 para $(w_{\acute{o}t})$. A pesquisa marca um aumento do número de dados do banco de dados desde o trabalho de Pantoja (2018) e Álvares Júnior (2019).

6) REFERÊNCIAS

AL-KHAFAJI, A. N. Estimation of Proctor compaction parameters. *International Journal of Surface Mining, Reclamation and Environment*, v. 7, n. 2, p. 43-51, 1993.

ALMUAYTHIR, S. et al. Predicting soil compaction parameters in expansive soils using machine learning. *Scientific Reports*, v. 15, n. 1, p. 1-15, 2025.

ÁLVARES JÚNIOR, J. J. Aplicação de Redes Neurais Artificiais para a previsão de parâmetros de compactação de solos. 2019. Trabalho de Conclusão de Curso (Graduação em Engenharia Civil) – Faculdade de Engenharia Civil, Universidade Federal do Pará, Belém, 2019.

Anjos, G. J. M. Acervo pessoal de dados. 2026.

ARDAKANI, A.; KORDNAEIJ, A. Prediction of compaction parameters of soils using GMDH-type neural network optimized by genetic algorithm. *International Journal of Geotechnical Engineering*, v. 13, n. 6, p. 557-565, 2019.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 7182: Solo - Ensaio de compactação. Rio de Janeiro: ABNT, 2016.

BLOTZ, L. R.; BENSON, C. H.; BOUTWELL, G. P. Estimating optimum water content and maximum dry unit weight for compacted clays. *Journal of Geotechnical and Geoenvironmental Engineering*, v. 124, n. 9, p. 907-912, 1998.

BRIAUD, J. L. *Geotechnical Engineering: Unsaturated and Saturated Soils*. Hoboken, NJ: John Wiley & Sons, 2013.

BURITICÁ, J. A. D. Estudo da compactação de um solo residual de gnaiss sob diferentes energias. 2013. Dissertação (Mestrado em Engenharia Civil) – Universidade Federal de Viçosa, Viçosa, 2013.

CASAGRANDE, A. Classification and identification of soils. *Transactions of the American Society of Civil Engineers*, v. 113, n. 1, p. 901-930, 1948.

CHEN, T.; GUESTRIN, C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA: ACM, 2016. p. 785-794.

DAS, B. M. *Fundamentals of Geotechnical Engineering*. 4. ed. Stamford, CT: Cengage Learning, 2011.

DEPARTAMENTO NACIONAL DE ESTRADAS DE RODAGEM. DNER-ME 162/94: Solos - ensaio de compactação utilizando amostras não trabalhadas. Rio de Janeiro: DNER, 1994.

DOKOVIĆ, N. et al. Estimation of compaction parameters of fine-grained soils by using Atterberg limits. *Journal of the Serbian Chemical Society*, v. 78, n. 11, p. 1823-1834, 2013.

FERNANDES, M. *Mecânica dos Solos: Introdução à Engenharia Geotécnica*. Porto: FEUP Edições, 1994.

GOH, A. T. C. Empirical design in geotechnics using neural networks. *Géotechnique*, v. 45, n. 4, p. 709-714, 1995.

GUNAYDIN, O. Estimation of compaction parameters of soils using statistical analyses and artificial neural networks. *Environmental Geology*, v. 57, n. 1, p. 203-215, 2009.

GURTUG, Y.; SRIDHARAN, A. Prediction of compaction characteristics. *Proceedings of the Institution of Civil Engineers - Geotechnical Engineering*, v. 157, n. 1, p. 27-35, 2004.

HOLTZ, R. D.; KOVACS, W. D. *An Introduction to Geotechnical Engineering*. Englewood Cliffs, NJ: Prentice-Hall, 1981.

HORN, R. Stress-strain effects in structured unsaturated soils on macro- and microscale. 2003. (Special Lecture).

JUMIKIS, A. R. *Physical Characterization of the Soil*. New Brunswick: Rutgers University Press, 1946.

LAMBE, T. W. The structure of compacted clay. *Journal of the Soil Mechanics and Foundations Division*, v. 84, n. 2, p. 1-34, 1958.

MAGALHÃES, M. S. *Manual de Laboratório de Mecânica dos Solos*. Belém: UFPA, 2012.

MATTEO, L. D. et al. Predictability of compaction curves of fine-grained soils. *Géotechnique*, v. 59, n. 6, p. 537-545, 2009.

MORENO-MAROTO, J. M.; ALONSO-AZCÁRATE, J. What is clay? A new definition of "clay" based on plasticity and its impact on soil classification. *Applied Clay Science*, v. 161, p. 126-136, 2018.

OLIVEIRA, R. S. et al. *Digitalização da Infraestrutura e Machine Learning na Geotecnia*. São Paulo: Editora Técnica, 2021.

PANTOJA, B. F.; ANJOS, G. J. M. *Análise da estimativa dos parâmetros de compactação de solos baseado em propriedades índices*. 2018. Trabalho de Conclusão de Curso (Graduação em Engenharia Civil) – Faculdade de Engenharia Civil, Universidade Federal do Pará, Belém, 2018.

PINTO, C. S. Curso Básico de Mecânica dos Solos. 3. ed. São Paulo: Oficina de Textos, 2006.

POLIDORI, E. Proposal for a new classification chart for inorganic soils. Canadian Geotechnical Journal, v. 44, n. 2, p. 215-226, 2007.

PROCTOR, R. R. Design and control of compacted earth fills. Engineering News-Record, v. 111, n. 9, p. 245-248, 1933.

RIBEIRO, L. F. M. Compactação de Solos. Brasília: UnB, 2008.

SILVA, A. M. Seleção Automática de Atributos em Modelos de Machine Learning aplicados à Geotecnia. Rio de Janeiro: COPPE/UFRJ, 2022.

SIVRIKAYA, O. et al. Prediction of compaction parameters of fine-grained soils using Atterberg limits and soil class. Environmental Geology, v. 54, n. 4, p. 745-759, 2008.

SOUZA, F. L. Otimização de Modelos XGBoost para Aplicações em Tempo Real em Obras de Terra. Belo Horizonte: UFMG, 2023.

SRIDHARAN, A.; NAGARAJ, H. B. Plastic limit as a predictor of compaction characteristics of fine-grained soils. Proceedings of the Institution of Civil Engineers - Geotechnical Engineering, v. 158, n. 2, p. 109-114, 2005.