



ENGENHARIA: A MÁQUINA QUE CONSTRÓI O FUTURO

VOLUME XVI



EDITORA CONHECIMENTO LIVRE

Frederico Celestino Barbosa

Engenharia: a máquina que constrói o futuro

16^a ed.

Piracanjuba-GO
Editora Conhecimento Livre
Piracanjuba-GO

16ª ed.

Dados Internacionais de Catalogação na Publicação (CIP)

Barbosa, Frederico Celestino
B238E Engenharia: a máquina que constrói o futuro

/ Frederico Celestino Barbosa. – Piracanjuba-GO

Editora Conhecimento Livre, 2023

112 f.: il

DOI: 10.37423/2023.edcl790

ISBN: 978-65-5367-399-1

Modo de acesso: World Wide Web

Incluir Bibliografia

1. desenvolvimento 2. projeto 3. construção I. Barbosa, Frederico Celestino II. Título

CDU: 620

<https://doi.org/10.37423/2023.edcl790>

O conteúdo dos artigos e sua correção ortográfica são de responsabilidade exclusiva dos seus respectivos autores.

EDITORIA CONHECIMENTO LIVRE

Corpo Editorial

MSc Edson Ribeiro de Britto de Almeida Junior

MSc Humberto Costa

MSc Thays Merçon

MSc Adalberto Zorzo

MSc Taiane Aparecida Ribeiro Nepomoceno

PHD Willian Douglas Guilherme

MSc Andrea Carla Agnes e Silva Pinto

MSc Walmir Fernandes Pereira

MSc Edisio Alves de Aguiar Junior

MSc Rodrigo Sanchotene Silva

MSc Wesley Pacheco Calixto

MSc Adriano Pereira da Silva

MSc Frederico Celestino Barbosa

MSc Guilherme Fernando Ribeiro

MSc. Plínio Ferreira Pires

MSc Fabricio Vieira Cavalcante

PHD Marcus Fernando da Silva Praxedes

MSc Simone Buchignani Maigret

Dr. Adilson Tadeu Basquerote

Dra. Thays Zigante Furlan

MSc Camila Concato

PHD Miguel Adriano Inácio

MSc Anelisa Mota Gregoleti

PHD Jesus Rodrigues Lemos

MSc Gabriela Cristina Borborema Bozzo

MSc Karine Moreira Gomes Sales

Dr. Saulo Cerqueira de Aguiar Soares

MSc Pedro Panhoca da Silva

MSc Helton Rangel Coutinho Junior

MSc Carlos Augusto Zilli

MSc Euvaldo de Sousa Costa Junior

Dra. Suely Lopes de Azevedo

MSc Francisco Odecio Sales

MSc Ezequiel Martins Ferreira

MSc Eliane Avelina de Azevedo Sampaio

CAPÍTULO 1	6
ESTIMATIVA DE EVAPOTRANSPIRAÇÃO NO SEMIÁRIDO BRASILEIRO	
Jean Firmino Cardoso	
Ingrid Karoline Vasconcelos da Silva	
Eduarda Luciana Larissa de Lima	
Ana Carolina dos Santos Reis	
Isabelle Sales Alves de Sousa	
João Victor Mendes da Silva	
Priscila Tavares Dias	
Ágnes Sales Alves de Sousa	
Amanda Carolina Alves de Macedo	
Erickson Johny Galindo da Silva	
DOI 10.37423/230908141	
 CAPÍTULO 2	 15
CONSTRUÇÃO DE MODELO PREDITIVO PARA A UMIDADE DO SOLO	
Jean Firmino Cardoso	
Ingrid Karoline Vasconcelos da Silva	
Eduarda Luciana Larissa de Lima	
Ana Carolina dos Santos Reis	
Isabelle Sales Alves de Sousa	
João Victor Mendes da Silva	
Priscila Tavares Dias	
Ágnes Sales Alves de Sousa	
Amanda Carolina Alves de Macedo	
Erickson Johny Galindo da Silva	
DOI 10.37423/230908142	
 CAPÍTULO 3	 26
MODELAGEM DE PRECIPITAÇÃO MENSAL EM CARUARU, PERNAMBUCO, POR MEIO DE REDES NEURAIS ARTIFICIAIS	
Jean Firmino Cardoso	
Ingrid Karoline Vasconcelos da Silva	
Eduarda Luciana Larissa de Lima	
Ana Carolina dos Santos Reis	
Isabelle Sales Alves de Sousa	
João Victor Mendes da Silva	
Priscila Tavares Dias	
Ágnes Sales Alves de Sousa	
Amanda Carolina Alves de Macedo	
Erickson Johny Galindo da Silva	
DOI 10.37423/230908143	

CAPÍTULO 4 38

ANÁLISE FORENSE DE IMAGENS DIGITAIS PARA DETECÇÃO DE FOTOMONTAGEM E DEEPFAKE

Josef Augusto Oberdan Souza Silva

Luís Sérgio Rodrigues Vale

DOI 10.37423/230908154

CAPÍTULO 5 63

GERENCIAMENTO DO PROCESSO DE UTILIZAÇÃO DA FERRAMENTA “LIÇÕES APRENDIDAS” NUMA EMPRESA DO RAMO DE AUTOMAÇÃO AGRÍCOLA

ERIC MARCEL LEMES PISTORI

SILVANA RODRIGUES QUINTILHANO

DOI 10.37423/230908163

CAPÍTULO 6 77

A IMPORTÂNCIA DA CULTURA DE PROTEÇÃO DE DADOS COMO MECANISMO DE FIDELIZAÇÃO DO CLIENTE COM A MARCA DA EMPRESA E SEUS BENEFÍCIOS

Fernanda Cabral Pedreira

Lisleandra Machado

Domingos Sávio da Cunha Garcia

Leonardo Amorim de Araújo

Carlos Arthur Alevato Leal

Arthur Nascimento Assunção

Lucas Mendes Ferreira

Samuel Alves de Freitas

Gustavo José Santiago Rosseti

Silvana Rodrigues Pires Moreira

DOI 10.37423/230908214

CAPÍTULO 7 94

APLICABILIDADE DA LGPD (LEI GERAL DE PROTEÇÃO DE DADOS PESSOAIS) NA SEGURANÇA DA INFORMAÇÃO EM EMPRESAS DE DESENVOLVIMENTO DE SOFTWARE

Roberto Teijeira Claro

Lisleandra Machado

Domingos Sávio da Cunha Garcia

Leonardo Amorim de Araújo

Carlos Arthur Alevato Leal

Arthur Nascimento Assunção

Lucas Mendes Ferreira

Samuel Alves de Freitas

Gustavo José Santiago Rosseti

Silvana Rodrigues Pires Moreira

DOI 10.37423/230908215

Capítulo 1



10.37423/230908141

ESTIMATIVA DE EVAPOTRANSPIRAÇÃO NO SEMIÁRIDO BRASILEIRO

Jean Firmino Cardoso

Universidade Federal de Pernambuco

Ingrid Karoline Vasconcelos da Silva

Universidade Federal de Pernambuco

Eduarda Luciana Larissa de Lima

Universidade Federal de Pernambuco

Ana Carolina dos Santos Reis

Universidade Federal de Pernambuco

Isabelle Sales Alves de Sousa

Universidade Federal de Pernambuco

João Victor Mendes da Silva

Universidade Federal de Pernambuco

Priscila Tavares Dias

Centro Universitário UniFavip - Wyden

Ágnes Sales Alves de Sousa

Universidade Federal de Pernambuco

Amanda Carolina Alves de Macedo

Centro Universitário UniFavip - Wyden

Erickson Johny Galindo da Silva

Universidade Federal de Pernambuco

INTRODUÇÃO

O semiárido brasileiro é caracterizado por um conjunto de períodos com balanços hídricos negativos, resultantes da precipitação média anual inferior a 800 mm, insolação média de 2.800 horas por ano e umidade relativa anual média em torno de 50% (MOURA et al., 2007). Diante deste cenário, se faz necessário a constante preocupação com meios de implementação para um futuro mais imediatista e menos utópico, pois em decorrência variações hidrológicas extremas como estiagem (classificação COBRADE como 1.4.1.1.0) ou chuvas intensas (classificação COBRADE como 1.3.2.1.4) prejudicam a continuidade das famílias que dependem inteiramente do uso constante de água nesta região.

Sendo assim, foram utilizadas métricas para estimativas que caracterizam a quantidade de hídrica para desastres ambientais, de modo que não se tornem ações comuns, e que possam prever esses eventos extremos, por parte dos órgãos gestores e meteorológicos. As precipitações intensas ou precipitações prolongadas causam danos nos municípios, como: erosões no solo, inundações (classificação COBRADE como 1.2.1.0.0), danos à agricultura e sérias dificuldades para o armazenamento de água.

A educação ambiental e a disseminação de técnicas de captação e retenção de água podem contribuir significativamente para enfrentar os desafios impostos pela escassez hídrica na região. A integração de conhecimentos tradicionais com abordagens modernas, como o uso do aprendizado de máquina (machine learning) para previsões meteorológicas, pode resultar em estratégias mais eficazes e adaptativas para lidar com as oscilações climáticas e seus impactos.

Dessa forma, este artigo tem o propósito de estimar a evapotranspiração no semiárido brasileiro utilizando o aprendizado de máquina, aplicando como base mudanças no comportamento pluvial de períodos antecedentes do ambiente de estudo analisado. Validando o método, também pode ser possível estimar a pluviometria na região, sendo necessário apenas a realização de mais testes e a obtenção de mais dados.

METODOLOGIA

Os dados foram coletados na fazenda Buenos Aires, área de pastagem de Serra Talhada-PE, obtidos a partir da torre de monitoramento hidrometeorológicos do Observatório Nacional da Água e do Carbono no Bioma Caatinga (ONDACBC). A série histórica de dados utilizada neste estudo compreende o período entre o ano de 2014 e o início de 2015, tendo 12 mil registros de informações, registrados em intervalos de trinta minutos. Para o cálculo da evapotranspiração de referência foi utilizado o

método de Penman-Monteith-FAO (ET_oPMF) (CONCEIÇÃO, 2006). Para a construção e desenvolvimento do aprendizado de máquina, o modelo de regressão oferecido pela Scikit-Learn, biblioteca gratuita do python (Pedregosa et al., 2011), chamado floresta aleatória (FA).

A princípio, foi necessário calcular as correlações entre todos os registros presentes no conjunto de dados a partir da função da biblioteca “pandas” (MCKINNEY, 2012), na busca de entender quais parâmetros de entrada(s) e saída possuem melhor desempenho, esses valores de correlação podem variar de -1 até 1, sendo 1 e -1 as combinações perfeitas, após o processamento foi possível constatar que a melhor combinação com base nas correlações foram para a evapotranspiração potencial (0.900), a precipitação pluviométrica (0.470) como dados de entrada e a evapotranspiração potencial 60 minutos (1.000) após os registros de entrada como parâmetro de saída. Devido ao longo tempo e consumo de processamento, o modelo preliminar de aprendizado de máquina focou em tentar prever apenas os valores da evapotranspiração potencial no septuagésimo minuto com relação aos registros de entrada.

No modelo da FA foi aplicado o método da validação cruzada (VC) (Cunha, 2019). A VC consiste em encontrar o valor da profundidade ideal, ou seja, um nível em que os resultados dos dados de treino e teste apresentem os melhores resultados em ambos com eficácia, assim evitando que o algoritmo fique viciado para os dados de treino, ou seja, apresenta resultados ótimos para os dados de treino e não satisfatório para os dados de teste.

A avaliação de desempenho do modelo foi realizada considerando as seguintes métricas estatísticas: a Raiz do Erro Quadrático Médio (equação 1), sendo melhor valor 0 no espaço entre $[0, +\infty)$. A eficiência de Nash e Sutcliffe (NASH e SUTCLIFFE, 1970) (equação 2), melhor valor sendo 1 no espaço entre $(-\infty, 1]$. O coeficiente de Kling–Gupta (equação 3), melhor valor é 1 no espaço entre $(-\infty, 1]$. O coeficiente de determinação (equação 4), melhor valor é 1 no espaço entre $[0, 1]$. O coeficiente de correlação de Pearson (equação 5), melhor valor é 1 no espaço entre $[-1, 1]$. O índice de concordância de Willmott (equação 6), melhor valor é 1 no espaço entre $[0, 1]$. Por fim, o índice de desempenho do modelo (equação 7), melhor valor é 1 no espaço entre $(-\infty, 1]$.

$$RMSE(y, \hat{y}) = \sqrt{\frac{\sum_{i=0}^{N-1} (y_i - \hat{y}_i)^2}{N}} \quad (\text{equação 1})$$

$$NSE(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{N-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{N-1} (y_i - \mu_y)^2} \quad (\text{equação 2})$$

$$KGE(y, \hat{y}) = 1 - \sqrt{\left((r(y, \hat{y}) - 1)^2 + \left(\frac{\mu_{\hat{y}}}{\mu_y} - 1 \right)^2 + \left(\frac{\sigma_{\hat{y}}}{\sigma_y} - 1 \right)^2 \right)} \quad (\text{equação 3})$$

$$R^2(y, \hat{y}) = \frac{\left[\sum_{i=0}^{N-1} (y_i - \mu_y) \times (\hat{y}_i - \mu_{\hat{y}}) \right]^2}{\left[\sum_{i=0}^{N-1} (y_i - \mu_y)^2 \times \sum_{i=0}^{N-1} (\hat{y}_i - \mu_{\hat{y}})^2 \right]} \quad (\text{equação 4})$$

$$r(y, \hat{y}) = \frac{\sum_{i=0}^{N-1} (y_i - \mu_y) \times (\hat{y}_i - \mu_{\hat{y}})}{\sqrt{\sum_{i=0}^{N-1} (y_i - \mu_y)^2 \times \sum_{i=0}^{N-1} (\hat{y}_i - \mu_{\hat{y}})^2}} \quad (\text{equação 5})$$

$$d(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{N-1} (\hat{y}_i - y_i)^2}{\sum_{i=0}^{N-1} (|\hat{y}_i - \mu_y| + |y_i - \mu_y|)^2} \quad (\text{equação 6})$$

$$c(y, \hat{y}) = d \cdot r \quad (\text{equação 7})$$

Onde y são os valores medidos, $\hat{y}$ os valores estimados, N o número de elementos, μ_y a média dos valores medidos, $\mu_{\hat{y}}$ a média dos valores estimados, σ_y o desvio padrão dos valores medidos, $\sigma_{\hat{y}}$ o desvio padrão dos valores estimados.

Portanto, para a aplicação do algoritmo de floresta aleatória (FA), foi necessário apenas realizar a importação do algoritmo da biblioteca scikit-learn, realizar a divisão entre os dados de treino e teste, onde 70% dos dados foram utilizados para o treinamento do modelo e 30% dos dados para teste, por fim foi comparado a precisão entre os dados de teste e seus respectivos valores estimados a partir das métricas selecionadas para validação do modelo.

RESULTADOS E DISCUSSÕES

A Figura 1 proporciona uma análise da precisão da estimativa da evapotranspiração potencial ao utilizar o algoritmo de Floresta Aleatória, em comparação com os valores medidos correspondentes obtidos a partir da torre do ONDACBC. A evapotranspiração potencial é uma medida essencial para compreender os processos de perda de água no ambiente. O algoritmo de Floresta Aleatória tem sido amplamente utilizado em estudos hidrológicos devido à sua capacidade de lidar com relações complexas entre variáveis e apresentar resultados confiáveis.

Ao observar a Figura 1, podemos notar a proximidade dos pontos estimados utilizando o algoritmo de Floresta Aleatória em relação aos valores medidos pela torre do ONDACBC. Essa similaridade sugere

que o modelo foi capaz de capturar adequadamente os padrões e tendências da evapotranspiração potencial.

É importante destacar que a escolha da Floresta Aleatória como algoritmo para a estimativa da evapotranspiração potencial está de acordo com estudos anteriores. De acordo com RODRÍGUEZ-GALIANO et al. (2011), a Floresta Aleatória é uma técnica eficaz para estimar a evapotranspiração potencial, pois considera múltiplas variáveis e é capaz de lidar com a não linearidade dos processos hidrológicos.

Além da análise visual na Figura 1, é possível realizar uma avaliação quantitativa utilizando métricas de desempenho para o modelo de Floresta Aleatória. Métricas como Erro Médio Absoluto (MAE), Erro Quadrático Médio (RMSE) e Coeficiente de Determinação (R^2) podem ser calculadas para avaliar a precisão da estimativa.

É fundamental ressaltar que a interpretação dos resultados deve considerar as condições específicas do estudo, como características do local, qualidade dos dados utilizados e a variabilidade temporal e espacial da evapotranspiração potencial.

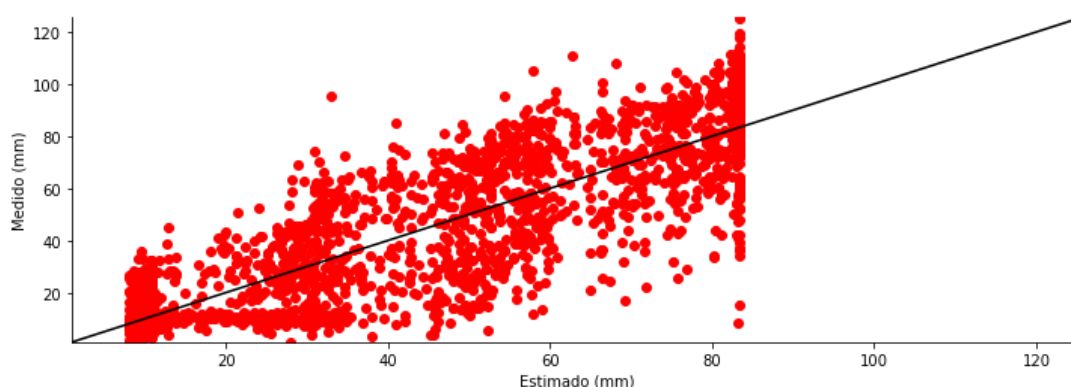


Figura 1: Dispersão gráfica para o modelo FA. Fonte: autoria própria.

Conforme observado na Figura 1, quanto mais pontos vermelhos estiverem alinhados com a diagonal preta no gráfico, maior é a fidelidade do modelo de Floresta Aleatória em relação aos dados reais medidos. No entanto, é importante notar que muitos pontos estão dispersos em relação à linha diagonal, indicando uma menor precisão. Essa dispersão dos pontos sugere que o modelo pode apresentar imprecisões em certas situações. É fundamental ressaltar que a análise da Figura 1 de forma isolada pode levar a interpretações equivocadas. Para uma avaliação mais completa, a Figura 2 apresenta um intervalo temporal dos dados medidos e seus respectivos valores estimados. Essa

visualização permite uma compreensão mais detalhada do desempenho do modelo ao longo do tempo.

Ao observar a Figura 2, é possível verificar que, embora a dispersão dos pontos na Figura 1 sugira uma menor precisão, essa discrepância não é tão intensa quando avaliada em um contexto temporal mais amplo. Essa análise temporal permite identificar que o modelo apresenta dificuldade em interpretar picos de variação da evapotranspiração potencial, levando à subestimação ou superestimação.

Portanto, é fundamental considerar tanto a análise visual quanto as métricas selecionadas para avaliar o desempenho do modelo de Floresta Aleatória na estimativa da evapotranspiração potencial. É recomendado o cálculo de métricas como Erro Médio Absoluto (MAE), Erro Quadrático Médio (RMSE) e Coeficiente de Determinação (R^2), entre outras, para obter uma avaliação mais completa e robusta da precisão do modelo. É importante ressaltar que a interpretação dos resultados deve levar em consideração as características específicas do estudo, incluindo as condições locais, a qualidade dos dados utilizados e a variabilidade temporal da evapotranspiração potencial.

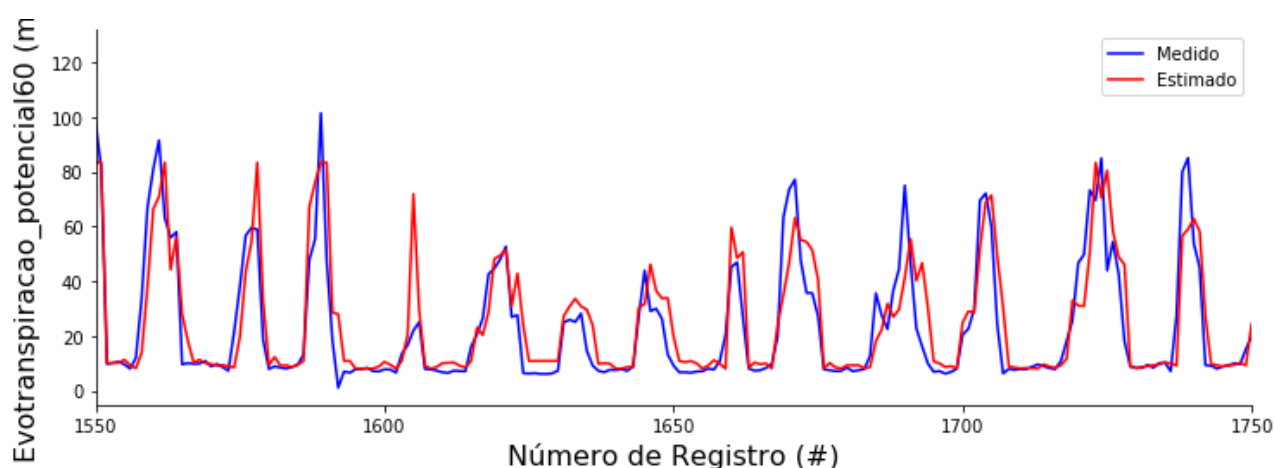


Figura 2: Comparativo entre os dados estimados. Fonte: autoria própria.

Os resultados obtidos para o modelo de Floresta Aleatória desenvolvido são os seguintes: o Índice de Eficiência de Nash-Sutcliffe (NSE) apresentou um valor de 0,816. O Índice de Concordância de Willmott (d) obteve um resultado de 0,947, enquanto o Coeficiente de Correlação de Pearson (r) foi igual a 0,903. O Índice de Desempenho do Modelo (c) resultante foi de 0,855.

De acordo com a tabela de desempenho apresentada por Camargo e Sentelhas (1997), o valor de c indica que o modelo possui uma precisão classificada como "Muito Bom". Essa classificação demonstra que o modelo de Floresta Aleatória apresenta uma boa capacidade de estimativa da evapotranspiração potencial. Ao verificar o coeficiente de Kling-Gupta, observa-se que o cálculo

resultou em 0,859. Essa métrica fornece uma medida adicional do desempenho do modelo, e valores mais próximos de 1 indicam uma melhor aderência às observações.

O coeficiente de determinação (R^2) calculado foi de 0,816, indicando uma forte correlação entre as variáveis de entrada e saída utilizadas no modelo. Esse valor demonstra que o modelo é capaz de explicar aproximadamente 81,6% da variabilidade dos dados medidos. No entanto, é importante mencionar que o Erro Quadrático Médio (RMSE) encontrado foi igual a 12,265. O RMSE é uma métrica que avalia a diferença entre os valores estimados e os valores medidos. O valor mais elevado do RMSE pode ser atribuído às pequenas imprecisões nos valores estimados, bem como aos altos resíduos gerados pela dificuldade do modelo em lidar com os picos de evapotranspiração potencial.

Considerando as métricas avaliadas, podemos concluir que o modelo de Floresta Aleatória apresentou um desempenho geral considerado "Muito Bom" com base nas métricas de avaliação utilizadas. No entanto, é importante considerar as limitações do modelo em relação aos picos de evapotranspiração potencial e as pequenas imprecisões geradas nos valores estimados.

CONCLUSÕES

De forma geral, a estimativa de evapotranspiração potencial pode ser orientada a partir da aplicação de códigos de aprendizagem de máquina e dos conhecimentos técnicos da área de recursos hídricos, no entanto, requer uma demanda maior de dados para aplicabilidade e melhor execução, principalmente com relação aos picos de valores que podem ser corrigidos com base em mais valores de picos no conjunto de dados para treinamento.

Também podemos notar que é necessário sempre realizar uma análise ampla dos resultados, com pelo menos dois gráficos dos dados, além das métricas de validação, para entender a real situação pois dependendo de como o gráfico é desenvolvido pode gerar uma informação superestimada, portanto é necessário ter sempre atenção com essa fase da análise.

REFERÊNCIAS

CAMARGO, AP de; SENTELHAS, Paulo Cesar. Avaliação do desempenho de diferentes métodos de estimativa da evapotranspiração potencial no Estado de São Paulo, Brasil. Revista Brasileira de Agrometeorologia, v. 5, n. 1, p. 89-97, 1997.

CONCEIÇÃO, Marco AF. Roteiro de cálculo da evapotranspiração de referência pelo método de Penman-Monteith-FAO. Embrapa Uva e Vinho. Circular Técnica, 2006.

SANTOS, Sergio Rodrigo Quadros dos; CUNHA, Ana Paula Martins do Amaral; RIBEIRO-NETO, Germano Gondim; Avaliação de dados de precipitação para o monitoramento do padrão espaço - temporal da seca do nordeste do Brasil. Revista Brasileira de Climatologia. Ano 15 –Vol. 25– jul/dez 2019.

MCKINNEY, Wes. Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc.", 2012.

MOURA, M. S. B. de; ESPÍNOLA SOBRINHO, J.; SILVA, T. G. F. da; SOUZA, W. M. de; Aspectos meteorológicos do semiárido brasileiro. Embrapa semiárido. 2019.

NASH, J. Eamon; SUTCLIFFE, Jonh V. River flow forecasting through conceptual models part I - A discussion of principles. Journal of hydrology, v. 10, n. 3, p. 282-290, 1970.

PEDREGOSA, Fabian et al. Scikit-learn: Machine learning in Python. the Journal of machine Learning research, v. 12, p. 2825-2830, 2011.

RODRÍGUEZ-GALIANO, Víctor Francisco et al. Incorporating spatial variability measures in land-cover classification using Random Forest. Procedia Environmental Sciences, v. 3, p. 44-49, 2011.

AGRADECIMENTOS

Os autores agradecem ao apoio do Observatório Nacional da Dinâmica da Água e de Carbono no Bioma Caatinga (ONDACBC) pelos dados fornecidos, à Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), pela concessão da bolsa de pós-graduação, por meio do Processo Nº IBPG-1083-3.01/21. Além do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de iniciação tecnológica, por meio do Processo Nº 181077/2022-1, pela bolsa de iniciação científica, por meio do processo Nº 156553/2022-8.

Capítulo 2



10.37423/230908142

CONSTRUÇÃO DE MODELO PREDITIVO PARA A UMIDADE DO SOLO

Jean Firmino Cardoso

Universidade Federal de Pernambuco

Ingrid Karoline Vasconcelos da Silva

Universidade Federal de Pernambuco

Eduarda Luciana Larissa de Lima

Universidade Federal de Pernambuco

Ana Carolina dos Santos Reis

Universidade Federal de Pernambuco

Isabelle Sales Alves de Sousa

Universidade Federal de Pernambuco

João Victor Mendes da Silva

Universidade Federal de Pernambuco

Priscila Tavares Dias

Centro Universitário UniFavip - Wyden

Ágnes Sales Alves de Sousa

Universidade Federal de Pernambuco

Amanda Carolina Alves de Macedo

Centro Universitário UniFavip - Wyden

Erickson Johny Galindo da Silva

Universidade Federal de Pernambuco

INTRODUÇÃO

A região metropolitana do Recife sofreu com decorrências das fortes chuvas (classificação COBRADE como 1.3.2.1.4) no período de maio à julho de 2022, mais precisamente nas regiões de Areias, Pina, Ibura, Casa Amarela e nos municípios vizinhos como Jaboatão dos Guararapes, Ipojuca, São Lourenço da Mata, Cabo de Santo Agostinho, Camaragibe e Sirinhaém. Devido aos movimentos de massa e alagamentos na região, classificados como 1.1.3.3.1 e 1.2.3.0.0 pela COBRADE, este fato tornou-se o maior desastre natural do século 21 do estado de Pernambuco, pois causaram aproximadamente 130 mortes e 6 mil pessoas desabrigadas, registrando um pouco mais de 70 milímetros de chuva apenas na capital pernambucana. A presença de água no solo é crucial para os fenômenos da evapotranspiração e infiltração, como também para erosão do solo e escoamento superficial, são mecanismos regulatórios de troca de água e energia entre o solo e a atmosfera (DALY e PORPORATO, 2005. TYAGI et al., 2011).

Segundo Reichardt (1987), a umidade do solo é uma variável dependente de vários fatores, como o tempo, variação com ascensão capilar, aumento com a chuva ou com a irrigação do solo e diminuição com a drenagem ou evapotranspiração. Nos métodos indiretos, para estimativa da umidade do solo em campo, a sonda de nêutrons e a reflectometria de domínio do tempo (TDR, SOUZA et al., 2016) ou de frequência (FDR) são destaques entre o grupo. De acordo com Coelho et al. (2006), a determinação encontrada pelo método de reflectometria no domínio do tempo apresenta melhores resultados.

Para o modelo de machine learning, um dos destaques é o algoritmo de máquinas de suporte vetorial (MSV ou SVM), construído de acordo com a teoria estatística de aprendizado e derivam da hipótese estrutural de minimizar o risco empírico e o intervalo de confiança do aprendizado de máquina, com o objetivo de obter uma boa capacidade de extrapolação, têm o potencial de criar uma relação desconhecida entre um grupo de variáveis de entrada e saída.

Na UFPE-CAA um dos grupos de pesquisa almeja o desenvolvimento de algoritmos de previsão de dados futuros do solo e das variáveis hidrológicas temporais, visando a tomada de decisão prévia mais eficiente de desastres como forma de minimizar a ocorrência de danos provocados pelos deslizamentos de solo (classificação COBRADE como 1.1.3.2.1), corridas de solo/lama (classificação COBRADE como 1.1.3.3.1) e os eventos de chuvas intensas (classificação COBRADE como 1.3.2.1.4) através do monitoramento dos dados coletados do local.

Não foi possível obter dados suficientes de Recife-PE, portanto, o modelo preliminar de aprendizado de máquina foi desenvolvido a partir de dados temporais obtidos em Serra Talhada-PE, para a estimativa dos valores da umidade do solo em uma determinada profundidade considerando um intervalo temporal inicial, caso o algoritmo construído seja eficiente na problemática, será aprimorado para intervalos temporais e profundidades cada vez mais maiores e implementado para os dados de Recife-PE assim que possível.

METODOLOGIA

ÁREA DE PESQUISA

O estudo foi realizado na região de pastagem no município de Serra Talhada, encontra-se presente na microrregião do Vale do Pajeú, totalmente posicionada na mesorregião do São Francisco, no Sertão pernambucano. Apresenta clima semiárido, com baixos índices pluviométricos anuais com média de 642 mm e variações de temperatura média do ar com média anual de 25,2° Celsius (Souza, 2014) .

Com área territorial de 2.980 km² está localizado no sertão do estado de Pernambuco, mas precisamente nas coordenadas Latitude Sul: 07° 59' 31", Longitude Oeste: 38° 17' 54" e Altitude: 429 m (Jardim et al., 2019).

EQUIPAMENTOS E MACHINE LEARNING

Os dados de estudo foram coletados por torres de monitoramento de dados hidrometeorológicos mantida pelo projeto ONDACBC - Observatório Nacional da Água e do Carbono no Bioma Caatinga, instaladas nas áreas de pastagem (BR-GST) e de Caatinga (BR-CST) no município de Serra Talhada, continuamente registram os dados essenciais, como os: de umidade do solo, precipitação, temperatura, radiação e a velocidade do vento, tem sensores com equidistância em torno de 2,5 km e uma altura de 2 m, (Alcântara et al., 2021). A série histórica de precipitação analisada possui intervalo entre os anos de 2014 e 2015, totalizando 12 mil registros temporais, apresentando intervalos temporais de trinta minutos entre eles.

Porém sobre o monitoramento dos dados de umidade do solo, foram utilizadas para coleta de dados as sondas TDR CS616 que estão instaladas nas profundidades de 10, 20, 30, 40, 50 cm, essas calibradas baseadas nas recomendações do manual da Campbell.

Para esse método a modelagem com sistema de porosidade simples, utiliza-se a equação de Richards é usada para descrever o fluxo de água no solo, resolvida por meio de elementos finitos na

discretização espacial e diferenças finitas na distribuição temporal. A equação de Van Genuchten-Mualen é empregada para determinar as relações entre as variáveis, permitindo a caracterização da umidade do solo e o potencial de retenção de água.

O desenvolvimento do algoritmo de aprendizado de máquina contou com a utilização da biblioteca Scikit-Learn (Pedregosa et al., 2011), uma plataforma gratuita da linguagem de programação Python, voltada para a implementação de diversos algoritmos e métricas essenciais em projetos que envolvam machine learning.

Existem diversos algoritmos de aprendizado de máquina que podem ser considerados para resolver diferentes tipos de problemas. É importante testar e comparar vários algoritmos para encontrar a melhor solução para cada caso. Algumas opções que poderiam ser adotadas, são:

1. Árvores de Decisão: Por fácil interpretação, requer pouco pré-processamento, pode lidar com dados numéricos e categóricos;
2. 2.Random Forest: Reduz overfitting, lida bem com grandes conjuntos de dados e alta dimensionalidade
3. Support Vector Machines (SVM): Eficiente em espaços de alta dimensionalidade, funciona bem com dados limitados
4. Redes Neurais Artificiais: Aprende padrões complexos em grandes conjuntos de dados, adequado para problemas de alta dimensionalidade.
5. K-Nearest Neighbors (KNN): Fácil de entender e implementar, funciona bem em problemas com fronteiras de decisão não lineares.
6. Naive Bayes: Eficiente e rápido, fácil de implementar, funciona bem com dados categóricos.

Cada algoritmo tem suas próprias características e adequação para diferentes problemas. Para este estudo em específico, optou-se por empregar o modelo Support Vector Machines (Máquinas de Suporte Vetorial - SVM) como a abordagem selecionada, considerado como um algoritmo poderoso e amplamente utilizado em problemas de classificação e regressão.

O SVM pode lidar com dados complexos e encontrar hiperplanos de separação ótimos, ainda é eficaz mesmo quando o tamanho do conjunto de dados é relativamente pequeno, ele pode ser utilizado para problemas de classificação binária e multiclasse. Além disso, é possível usar diferentes funções de kernel para lidar com diferentes tipos de dados, e oferece também parâmetros de regularização para

controlar o risco de overfitting. Por fim, o SVM pode gerar um hiperplano de separação com função discriminante para ser facilmente visualizada e interpretada.

CÁLCULOS E APLICAÇÕES

Para o cálculo das correlações entre as variáveis disponíveis, foi usada a função nativa da biblioteca “Pandas” gratuita da programação do Python (MCKINNEY, 2012) desenvolvida com base no método de Pearson (BENESTY, 2009).

Nessa análise, observou-se que dentre as diferentes variáveis disponíveis nos registros, as correlações mais significativas foram encontradas entre as variáveis de umidade do solo em suas diferentes profundidades, sendo os valores mais altos de 0,82 para 10 cm x 50 cm, 0,89 para 20 cm x 50 cm, 0,97 para 30 cm x 50 cm, 0,99 para 40 cm x 50 cm e 1,00 para 50 cm x 50 cm de profundidade.

Com base nesses resultados, foi decidido utilizar os dados de chuva registrados (em mm) e a umidade do solo na profundidade de 50 cm (em mm) como entrada para o modelo de análise. Por sua vez, os dados de saída do modelo são as previsões da umidade do solo na profundidade de 50 cm (em mm) considerando o período de 60 minutos após o registro dos dados de entrada. Essas informações são importantes para entender as relações entre as variáveis climáticas e a umidade do solo, auxiliando na previsão e monitoramento das condições hídricas do solo em diferentes profundidades.

Para a avaliação de desempenho do modelo, foram empregadas as seguintes métricas estatísticas relevantes para garantir uma análise abrangente. Entre elas:

Raiz do Erro Quadrático Médio (RMSE - Root Mean Squared Error) (Equação 1): Essa métrica mede a raiz quadrada da média dos erros quadrados entre os valores previstos e os valores reais. Quanto menor o valor resultante, melhor, e o valor ideal é 0.

Eficiência de Nash e Sutcliffe (NASH e SUTCLIFFE, 1970) (Equação 2): Essa métrica avalia o quão bem o modelo se ajusta aos dados observados. Valores próximos a 1 indicam um bom ajuste, enquanto valores negativos significam que o modelo é pior do que uma previsão média. O melhor valor está no intervalo $(-\infty, 1]$.

Coeficiente de Kling-Gupta (Equação 3): Essa métrica compara a variabilidade do modelo com a variabilidade dos dados observados. Valores próximos a 1 indicam um ajuste muito bom, enquanto valores menores que 1 indicam pior desempenho. O melhor valor está no intervalo $(-\infty, 1]$.

Coeficiente de Determinação (R^2) (Equação 4): Essa métrica mede a proporção da variância total explicada pelo modelo em relação à variância dos dados observados. O valor ideal é 1, indicando um ajuste perfeito, enquanto valores mais próximos de 0 indicam menor poder explicativo do modelo.

Coeficiente de Correlação de Pearson (Equação 5): Essa métrica avalia a relação linear entre as previsões do modelo e os valores reais. O valor ideal é 1, indicando uma correlação positiva perfeita, enquanto valores próximos de -1 indicam uma correlação negativa.

Índice de Concordância de Willmott (Equação 6): Essa métrica compara o erro do modelo com o erro de uma previsão média. Valores mais próximos de 1 indicam um bom ajuste, enquanto valores mais próximos de 0 indicam um ajuste pior.

Índice de Desempenho do Modelo (Equação 7): Essa métrica combina várias estatísticas para avaliar o desempenho geral do modelo. Valores próximos de 1 indicam um bom desempenho, enquanto valores menores que 1 indicam desempenho inferior ao esperado.

O uso dessas métricas é essencial para avaliar a qualidade do modelo e sua capacidade de fazer previsões precisas. A seleção adequada das métricas garante uma avaliação completa e confiável do desempenho do modelo de aprendizado de máquina. Todas as equação estão apresentadas a seguir:

$$RMSE(y, \hat{y}) = \sqrt{\frac{\sum_{i=0}^{N-1} (y_i - \hat{y}_i)^2}{N}} \quad (\text{equação 1})$$

$$NSE(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{N-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{N-1} (y_i - \mu_y)^2} \quad (\text{equação 2})$$

$$KGE(y, \hat{y}) = 1 - \sqrt{\left((r(y, \hat{y}) - 1)^2 + \left(\frac{\mu_{\hat{y}}}{\mu_y} - 1 \right)^2 + \left(\frac{\sigma_{\hat{y}}}{\sigma_y} - 1 \right)^2 \right)} \quad (\text{equação 3})$$

$$R^2(y, \hat{y}) = \frac{\left[\frac{\sum_{i=0}^{N-1} ((y_i - \mu_y) \times (\hat{y}_i - \mu_{\hat{y}}))}{\sum_{i=0}^{N-1} (y_i - \mu_y)^2 \times (\hat{y}_i - \mu_{\hat{y}})^2} \right]^2}{\left(\frac{\sum_{i=0}^{N-1} (y_i - \mu_y)^2}{\sum_{i=0}^{N-1} (\hat{y}_i - \mu_{\hat{y}})^2} \right)} \quad (\text{equação 4})$$

$$r(y, \hat{y}) = \frac{\sum_{i=0}^{N-1} ((y_i - \mu_y) \times (\hat{y}_i - \mu_{\hat{y}}))}{\sum_{i=0}^{N-1} (y_i - \mu_y)^2 \times (\hat{y}_i - \mu_{\hat{y}})^2} \quad (\text{equação 5})$$

$$d(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{N-1} (\hat{y}_i - y_i)^2}{\sum_{i=0}^{N-1} (|\hat{y}_i - \mu_y| + |y_i - \mu_y|)^2} \quad (\text{equação 6})$$

$$c(y, \hat{y}) = d \cdot r \quad (\text{equação 7})$$

Onde y é os valores medidos, $\hat{y}$ os valores estimados, N o número de elementos, μ_y a média dos valores medidos, $\mu_{\hat{y}}$ a média dos valores estimados, σ_y o desvio padrão dos valores medidos, $\sigma_{\hat{y}}$ o desvio padrão dos valores estimados.

Na aplicação do algoritmo de máquinas de suporte vetorial (SVM), bastou apenas realizar a importação do algoritmo do *scikit-learn*, dividir os dados em dois conjuntos, 70% dos dados para treinamento do modelo e 30% dos dados para teste, além da aplicação das métricas selecionadas para validação do modelo.

RESULTADOS

Com base na Figura 1, podemos observar a representação dos dados estimados por meio da aplicação do algoritmo de Support Vector Machine (SVM), em comparação com os valores medidos correspondentes, obtidos pela sonda TDR, para a medição da umidade do solo a uma profundidade de 50 centímetros.

A utilização do algoritmo SVM tem se mostrado uma abordagem promissora na área de análise de dados e aprendizado de máquina. Segundo WANG (2005), o SVM é um método de classificação que busca encontrar um hiperplano ótimo para separar dados de diferentes classes, com base na maximização da margem entre elas. Essa técnica tem sido amplamente aplicada em diversos domínios, incluindo a estimativa de parâmetros do solo, como a umidade.

Ao analisar a Figura 1, podemos observar que os dados estimados pelo SVM apresentam uma tendência a seguir os valores medidos pela sonda TDR. Essa concordância sugere que o algoritmo SVM foi capaz de capturar de forma eficiente as informações relacionadas à umidade do solo, permitindo uma estimativa adequada.

É importante ressaltar que a escolha do SVM como algoritmo de estimação da umidade do solo na profundidade de 50 centímetros está em conformidade com estudos anteriores. De acordo com OLDEN et al. (2008), o SVM tem sido amplamente utilizado para a previsão de parâmetros do solo devido à sua capacidade de lidar com problemas complexos e não-lineares.

No entanto, é válido mencionar que, apesar dos resultados promissores apresentados pela Figura 1, a precisão das estimativas pode variar dependendo das condições específicas do solo e do contexto em que o algoritmo é aplicado. Portanto, é necessário considerar cuidadosamente os fatores que podem influenciar a acurácia das estimativas.

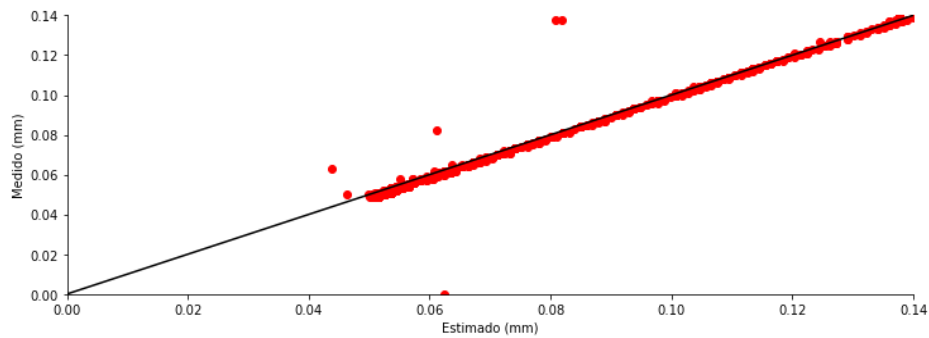


Figura 1: Dispersão gráfica para o modelo SVM. Fonte: autoria própria.

Além da análise visual na Figura 1, é possível realizar uma avaliação mais quantitativa das métricas de desempenho do algoritmo SVM para a estimativa da umidade do solo. A Figura 2 apresenta um recorte comparativo entre os dados medidos e seus respectivos valores estimados, fornecendo uma visualização mais detalhada do ajuste do modelo.

Observa-se que, nos valores mais elevados de umidade, o modelo apresenta uma aderência altíssima, indicando um bom desempenho, especialmente em condições próximas da saturação do solo, momento em que ocorrem com maior frequência os movimentos de massa.

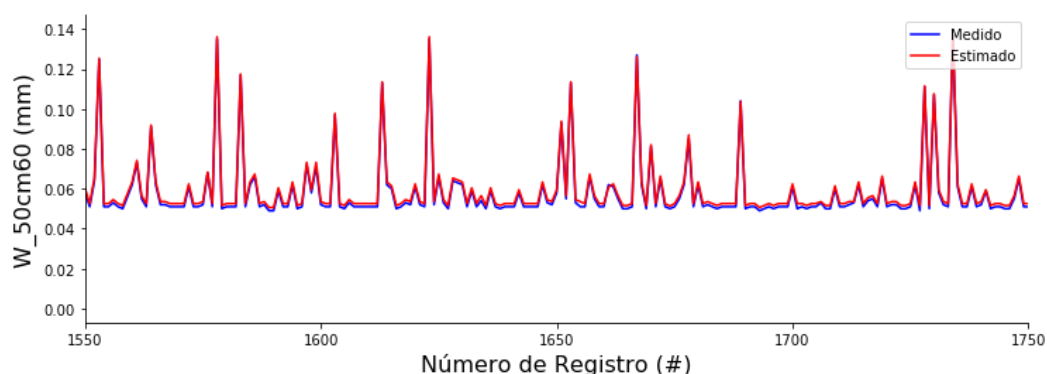


Figura 2: Comparativo entre os dados medidos e estimados. Fonte: autoria própria.

A avaliação das métricas para o método SVM revela os seguintes resultados: o Índice de Eficiência de Nash-Sutcliffe (NSE) apresentou um resultado de 0,979. O Índice de Concordância de Willmott (d) obteve um valor de 0,995, enquanto o Coeficiente de Correlação de Pearson (r) foi igual a 0,993. O Índice de Desempenho do Modelo (c) resultante foi de 0,988. De acordo com a tabela de desempenho proposta por Camargo e Sentelhas (1997), esse valor indica uma precisão classificada como "ótima".

Ao analisar o Coeficiente de Kling-Gupta, observa-se que o cálculo resultou em 0,946. Essa métrica fornece uma medida adicional de desempenho do modelo, com valores mais próximos de 1 indicando uma melhor aderência às observações.

O coeficiente de determinação (R^2) calculado foi de 0,987, o que sugere uma forte correlação entre as variáveis de entrada e saída utilizadas no modelo. Esse valor indica que a umidade estimada pelo algoritmo SVM é capaz de explicar aproximadamente 98,7% da variabilidade dos dados medidos.

Por fim, o Erro Quadrático Médio (RMSE) encontrado foi igual a 0,002, o que indica uma boa precisão na estimativa da umidade do solo. O RMSE é uma medida que avalia a diferença entre os valores estimados e os valores medidos, sendo que valores menores indicam uma maior proximidade entre as estimativas e as observações reais.

Essas métricas fornecem uma avaliação abrangente do desempenho do algoritmo SVM na estimativa da umidade do solo. No entanto, é importante ressaltar que a interpretação desses resultados deve levar em consideração as características específicas do estudo e as condições locais do solo.

CONCLUSÕES

A previsão da umidade do solo é possível a partir da utilização do algoritmo de aprendizado de máquina e de conhecimentos estatísticos aplicados a recursos hídricos, porém necessita de ajustes para aumentar seu intervalo temporal de previsão futura em paralelo com a manutenção do excelente desempenho preliminarmente obtido. Portanto é uma ferramenta que pode ser implementada como apoio para a tomada de decisões técnicas e antecipadas durante o monitoramento de movimentos de massa (1.1.3.3.1 pela COBRADE) ou de qualquer outra análise que envolve a umidade do solo como parâmetro de análise, podendo ser decisivo para a evacuação de área monitoradas que estejam em estado crítico, sendo assim, é de fundamental importância o estudo desse método para outras localidades, principalmente nas localizações onde existe a probabilidade maior para desastres causados por deslizamentos.

Vale destacar também o desempenho obtido pelo coeficiente de Kling-Gupta para o modelo testado, ou seja, o modelo conseguiu prever de forma satisfatória a dinâmica temporal da região da caatinga com relação às variáveis usadas, portanto, é necessário mais dedicação com o intuito de desenvolver o produto de apoio para os locais críticos que devem ser monitorados constantemente.

REFERÊNCIAS

- BENESTY, Jacob et al. Pearson correlation coefficient. In: Noise reduction in speech processing. Springer, Berlin, Heidelberg, 2009. p. 1-4.
- CAMARGO, AP de; SENTELHAS, Paulo Cesar. Avaliação do desempenho de diferentes métodos de estimativa da evapotranspiração potencial no Estado de São Paulo, Brasil. Revista Brasileira de Agrometeorologia, v. 5, n. 1, p. 89-97, 1997.
- COELHO, E. F.; VELLAME, L. M.; COELHO FILHO, M. A.; LEDO, C. A. S. Desempenho de modelos de calibração de guias de ondas acopladas a TDR e a multiplexadores em três tipos de solos. Revista Brasileira de Ciência do Solo. n.30, p.23-30, 2006.
- DALY E.; PORPORATO, A. A review of soil moisture dynamics: from rainfall infiltration to ecosystem response. Environmental Engineering Science, v.22, n.1, p.9–24, 2005.
- Jardim, A. M. da R. F., Queiroz, M., Araújo Júnior, G., Silva, M., & Silva, T. (2019). Estudos climáticos do número de dias de precipitação pluvial para o município de serra talhada-pe. revista engenharia na agricultura- REVENG, 27, 330–337. <https://doi.org/10.13083/reveng.v27i4.875>
- MCKINNEY, Wes. Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc.", 2012.
- PEDREGOSA, Fabian et al. Scikit-learn: Machine learning in Python. the Journal of machine Learning research, v. 12, p. 2825-2830, 2011.
- REICHARDT, K. A água em sistemas agrícolas. São Paulo: Manole, 1987. 188 p.
- Souza, R. M. S. (2014). Dinâmica de energia, água e carbono em área de pastagem no semiárido pernambucano. <http://www.tede2.ufrpe.br:8080/tede2/handle/tede2/6940>
- SOUZA, C. F.; SILVA, C. R.; ANDRADE JUNIOR, A. S.; COELHO, E. F. Monitoramento do teor de água no solo em tempo real com as técnicas de TDR e FDR. Irriga, Botucatu, Edição Especial, Irrigação, p. 26-42, 2016.
- TYAGI, J. V.; RAKESH KUMAR; SRIVASTAVA, S. L.; SINGH, R. D. Effect of micro-environmental factors on natural regeneration of Sal (Shorea robusta). Journal of Forestry Research, v.22, n.4, p.543-550, 2011.
- WANG, Lipo (Ed.). Support vector machines: theory and applications. Springer Science & Business Media, 2005.
- OLDEN, Julian D.; LAWLER, Joshua J.; POFF, N. LeRoy. Machine learning methods without tears: a primer for ecologists. The Quarterly review of biology, v. 83, n. 2, p. 171-193, 2008.

AGRADECIMENTOS

Os autores agradecem ao apoio do Observatório Nacional da Dinâmica da Água e de Carbono no Bioma Caatinga (ONDACBC) pelos dados fornecidos, à Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), pela concessão da bolsa de pós-graduação, por meio do Processo Nº IBPG-1083-3.01/21. Além do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de iniciação tecnológica, por meio do Processo Nº 181077/2022-1, Nº 156553/2022-8.

Capítulo 3



10.37423/230908143

MODELAGEM DE PRECIPITAÇÃO MENSAL EM CARUARU, PERNAMBUCO, POR MEIO DE REDES NEURAIS ARTIFICIAIS

Jean Firmino Cardoso

Universidade Federal de Pernambuco

Ingrid Karoline Vasconcelos da Silva

Universidade Federal de Pernambuco

Eduarda Luciana Larissa de Lima

Universidade Federal de Pernambuco

Ana Carolina dos Santos Reis

Universidade Federal de Pernambuco

Isabelle Sales Alves de Sousa

Universidade Federal de Pernambuco

João Victor Mendes da Silva

Universidade Federal de Pernambuco

Priscila Tavares Dias

Centro Universitário UniFavip - Wyden

Ágnes Sales Alves de Sousa

Universidade Federal de Pernambuco

Amanda Carolina Alves de Macedo

Centro Universitário UniFavip - Wyden

Erickson Johny Galindo da Silva

Universidade Federal de Pernambuco

Resumo: Estudos sobre chuva e clima adquiriram considerável relevância em diversos setores econômicos e sociais, devido à influência decisiva da previsão do tempo sobre o potencial hídrico de uma determinada região. Nesse contexto, o presente trabalho tem como objetivo a construção de um modelo de Deep Learning para a previsão da precipitação mensal, fazendo uso exclusivamente de uma série histórica de dados. Para alcançar este propósito, foram empregados dados de precipitações mensais referentes ao período de 1960 a 2018, coletados no município de Caruaru-PE. Esses dados foram aplicados em um modelo de Redes Neurais Artificiais (RNAs) para o processamento e análise das informações. Diversas configurações de RNAs foram consideradas e avaliadas em conjunto com os dados, com o intuito de identificar a abordagem mais eficaz. Os resultados revelaram a viabilidade do uso de RNAs para abordar a problemática da previsão dos volumes de precipitação mensal, com base em um histórico de dados. No entanto, é importante mencionar que, apesar dos resultados razoáveis obtidos, a quantidade de dados disponíveis pode não ter sido suficiente para alcançar o nível de eficiência desejado no processamento.

Palavras-chave: Machine Learning, Hidrologia estatística, Recursos Hídricos.

INTRODUÇÃO

As previsões de volumes de precipitação pluviométrica podem ser obtidas por meio de duas abordagens principais no campo da meteorologia. A primeira delas envolve a modelagem determinística dos fenômenos hidrometeorológicos. No entanto, essa abordagem é frequentemente considerada pouco viável devido à complexidade dos processos envolvidos na formação da precipitação (dos Santos Sousa et al., 2017).

A segunda abordagem, por outro lado, baseia-se em análises empíricas, utilizando dados de séries históricas. Essa abordagem trabalha com o reconhecimento de padrões, onde variáveis históricas são usadas para identificar padrões periódicos anuais que podem ser relevantes para prever eventos futuros de precipitação.

Dentro das técnicas empíricas, as técnicas de Machine Learning (ML) têm ganhado destaque na previsão de séries temporais. As MLs representam uma abordagem de algoritmos de inteligência artificial que difere da programação tradicional. Enquanto na programação tradicional o programador fornece a lógica e a entrada para obter a saída desejada, na ML o programador fornece a entrada e a saída, permitindo que o algoritmo descubra a relação entre eles (Valente, 2019).

Dentre as abordagens de Machine Learning, as Redes Neurais Artificiais (RNAs) se destacam. Elas são modelos matemáticos complexos que operam de maneira semelhante ao cérebro humano (Carvalho, 2014). As RNAs são consideradas algoritmos poderosos, capazes de resolver problemas lineares e não lineares. De acordo com Zaher et al. (2016), as RNAs consistem em uma estrutura com nós, onde cada processamento de informação está associado a um novo neurônio, respeitando o funcionamento com base em uma função de ativação. Geralmente, um modelo de RNA consiste em três níveis: uma camada de entrada, com um ou mais nós de entrada que variam conforme a quantidade de parâmetros do modelo; uma camada para a função de ativação, que pode ser linear ou não linear; e uma camada de saída, que envia informações para o próximo nó ou para a saída do modelo.

Do mesmo modo, o Deep Learning, uma vertente do Machine Learning, consiste em um conjunto de algoritmos capazes de modelar abstrações de alto nível de dados, usando uma arquitetura de grafo profundo composta por múltiplas camadas de processamento. Nesse contexto, este estudo teve como principal objetivo a construção de um modelo de Deep Learning para prever a precipitação mensal, com base exclusivamente em séries históricas de dados de precipitação. Para atingir esse objetivo, o município de Caruaru-PE foi selecionado como o caso de estudo.

MATERIAIS E MÉTODOS

ÁREA DE ESTUDO

O município de Caruaru (ver Figura 1), situado no agreste pernambucano a 8° 16' 58" de latitude Sul e 35° 58' 33" de longitude Oeste, possui um clima semiárido com amplas variações de temperatura, variando de 20°C no inverno a 38°C no verão. A Figura 1 apresenta a localização dos postos pluviométricos no município de Caruaru.

Para este estudo, foram utilizadas séries históricas de dados de precipitação coletadas pela Agência Pernambucana de Águas e Climas (APAC, 2019). Optou-se por utilizar os dados provenientes do posto pluviométrico do IPA (Instituto Agrônômico de Pesquisa), devido à sua longa e contínua série de dados, que se destaca como a mais abrangente disponível para este trabalho.

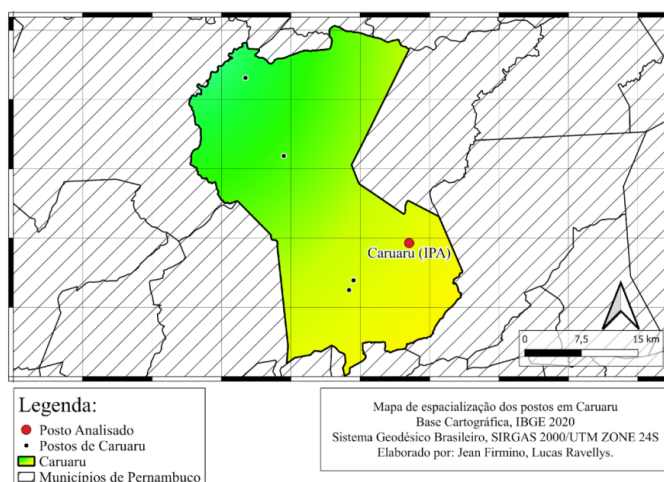


Figura 1 - Localização dos postos de dados pluviométricos no município de Caruaru-PE.

TRATAMENTO DOS DADOS

Foram utilizados dados mensais de precipitação pluviométrica no período de 1960 a 2018, totalizando 708 observações. A Figura 4 apresenta os dados de precipitação pluviométrica mensal acumulada durante esse período. A linha vermelha representa a precipitação média, que foi de aproximadamente 50 mm/mês.

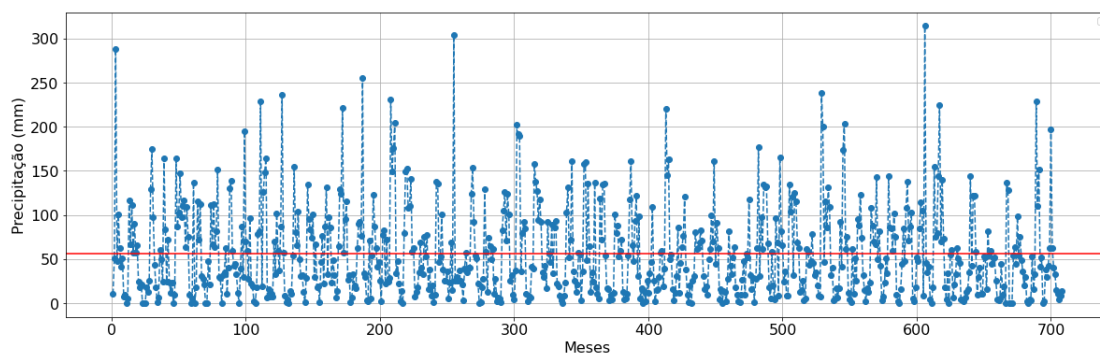


Figura 2. Dados de Precipitação acumulada mensal no período de 1960 até 2018.

Para ajustar o modelo, foi realizada a normalização dos dados utilizando a seguinte Equação (1):

$$P'_i = \frac{\ln(P_i + 1) - \underline{P\ln}}{S\ln} \quad (1)$$

Sendo: P'_i o dado normalizado no tempo i ; P_i a precipitação em mm no tempo i ; $\underline{P\ln}$ a média do $\ln(P+1)$ de treino; e $S\ln$ o desvio padrão do $\ln(P)$ de treino. A Figura 3 apresenta um histograma da precipitação pluviométrica mensal acumulada com os dados brutos e normalizados.

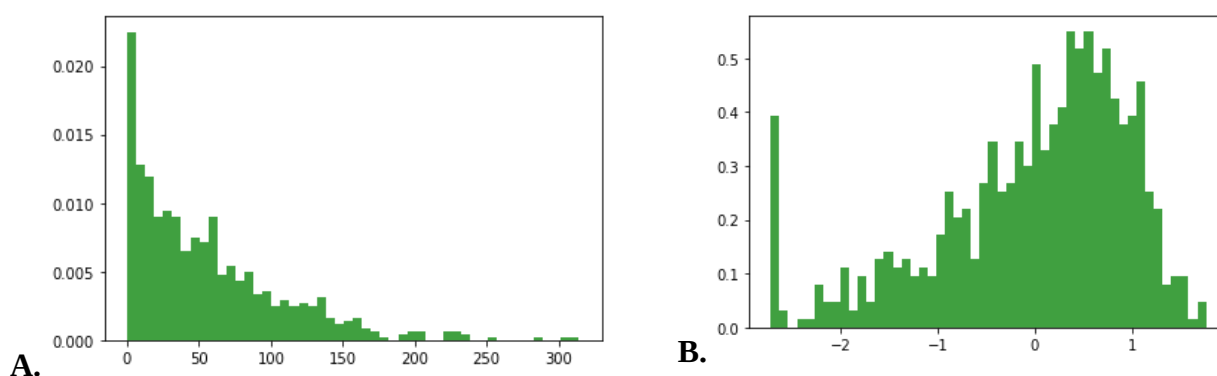


Figura 3. Histograma da precipitação pluviométrica acumulada mensal com os dados brutos (A) e normalizados (B)

CONSTRUÇÃO DO MODELO DE DEEP LEARNING

Para a arquitetura do modelo de Machine Learning, optou-se por utilizar Redes Neurais Artificiais (RNA) com a biblioteca Keras do Python, juntamente com o back-end do TensorFlow 2.0. Como entrada, foi escolhida uma janela de 12 meses de dados, o que significa que, para prever um mês futuro, é necessário fornecer um ano de dados anteriores a esse mês.

Para otimização, foram utilizados tamanhos de batch de 20 e um critério de parada após 1000 iterações. A função de perda escolhida como função objetivo foi a função de Huber, proposta por Huber em 1964. Essa função é comumente usada quando se lida com valores discrepantes nos dados,

sendo empregada para descrever a perda incorrida por um procedimento de estimativa. A função de perda de Huber é definida como:

$$L(a) = \begin{cases} \frac{1}{2} a^2 & \text{para } |a| \leq \delta \\ \delta \left(|a| - \frac{1}{2} \delta \right) & \text{para } |a| > \delta \end{cases} \quad (2)$$

Onde 'a' refere-se ao resíduo, ou seja, à diferença entre os valores observados e os previstos sendo igual a $a = y - f(x)$.

A otimização da função de Huber foi realizada por meio do método do Gradiente Descendente Estocástico, com uma taxa de aprendizado de 10^{-5} e um momento de 0,3. Este método de otimização é amplamente empregado no aprendizado de máquina e tem como objetivo minimizar uma função de perda dada, representada como $L(a)$ (Pinto, 2018). O processo consiste em encontrar um valor $\theta^* \in \mathbb{R}^k$ (onde $k \geq 1$) que minimize a função, começando a busca a partir de um valor inicial θ_0 .

$$\theta_{n+1} = \theta_n - \varepsilon_n L'(\theta_n) \quad (3)$$

Onde 'n' representa o número do passo da iteração, e ' ε_n ' é uma sequência decrescente do tamanho do passo. Assume-se que ' ε_n ' é sempre não negativo e finito (Gemulla et al., 2011).

Foi empregada uma arquitetura de Deep Learning Recursiva para ajustar o modelo. Para o treinamento e validação do modelo, utilizaram-se 70% dos dados (496 meses) e 30% dos dados (212 meses), respectivamente.

As métricas usadas para avaliar a qualidade do ajuste incluem o Coeficiente de Eficiência do Modelo de Nash (NSE) conforme a Equação (4) e o Erro Absoluto Médio (MAE) conforme a Equação (5).

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_i - P_i| \quad (5)$$

No qual: O_i é o valor da precipitação observada em i , P_i é o valor da precipitação predita em i , n é o número de registros, e $\bar{O}$ é a precipitação média da série analisada.

RESULTADOS E DISCUSSÃO

Na Figura 4, são apresentados os Box-Plots das precipitações pluviométricas mensais no período de 1960 a 2018. Observa-se que as maiores precipitações concentram-se entre os meses de março a julho, enquanto as menores precipitações ocorrem de setembro a dezembro. É interessante notar que a precipitação mediana mais alta ocorre no mês de julho. Resultados semelhantes foram obtidos na Bacia do Tapacurá por Silva et al. (2010). No entanto, em outras regiões do Agreste, como observado por Rodrigues et al. (2017), a maior precipitação mensal pode ocorrer no mês de junho.

Também é possível notar que os limites superiores das séries históricas de chuvas apresentam uma maior amplitude entre os meses de fevereiro a abril, diminuindo gradualmente até atingir o menor limite superior em novembro. Essa observação está de acordo com resultados semelhantes relatados por Santos e Montenegro (2012). Por outro lado, os limites inferiores permanecem mais uniformes ao longo dos meses.

Uma das variáveis hidrológicas que podem influenciar na ocorrência de valores atípicos (outliers) ao longo da série histórica é a anomalia na temperatura dos oceanos Atlântico e Pacífico, causada pelo fenômeno global conhecido como La Niña. Conforme mencionado por Rodrigues et al. (2007), entre os anos de 1964 e 2016, 56% dos registros de chuvas intensas foram associados a esse evento. Essas informações destacam a variação temporal heterogênea na distribuição da precipitação no município de Caruaru, corroborando com as constatações de Santos et al. (2010) para a região do Agreste Pernambucano.

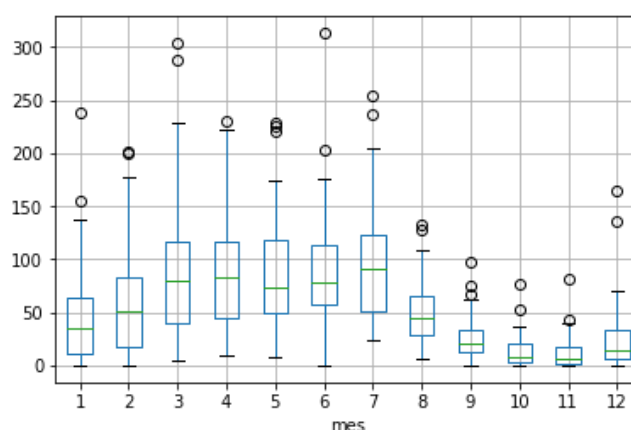


Figura 4. Boxplot dos meses no intervalo entre 1960 até 2018

Na Figura 5A, apresenta-se a precipitação pluviométrica mensal e sua previsão pelo modelo durante o período de validação. Observa-se que o modelo conseguiu capturar as tendências dos dados

medidos, apresentando uma boa previsão quando não há chuva, embora tenha subestimado os picos de chuva. Gorodetskaya et al. (2018) também enfrentaram desafios semelhantes ao utilizar Redes Neurais Artificiais e Random Forest para prever a vazão média diária de uma estação fluviométrica no rio Paraíba do Sul, com um horizonte de previsão de até 7 dias. Assim como no nosso estudo, eles observaram que durante períodos de vazões médias e baixas, a previsão foi satisfatória. Além disso, os autores notaram que os modelos de Redes Neurais apresentaram um desempenho superior na previsão de picos de vazão em comparação com os modelos Random Forest.

No que diz respeito às métricas de avaliação do ajuste adotadas (Figura 5B), é possível identificar várias discrepâncias entre os dados inicialmente processados. Há uma tendência positiva entre os dados medidos e estimados no gráfico de comparação, mas também uma considerável dispersão dos valores. O coeficiente NSE foi de 0,33, e o MAE foi de 0,62. No entanto, ainda não foi possível alcançar uma previsão com eficiência satisfatória para o objetivo, uma vez que almejamos um NSE mais próximo de 1. É importante notar que o MAE calculado pode ter sido influenciado pelas precipitações diárias que se desviam consideravelmente da média anual. Por exemplo, precipitações acima de 50 mm (conforme visto na Figura 2) podem contribuir para afastar o MAE do ideal, que seria próximo de zero.

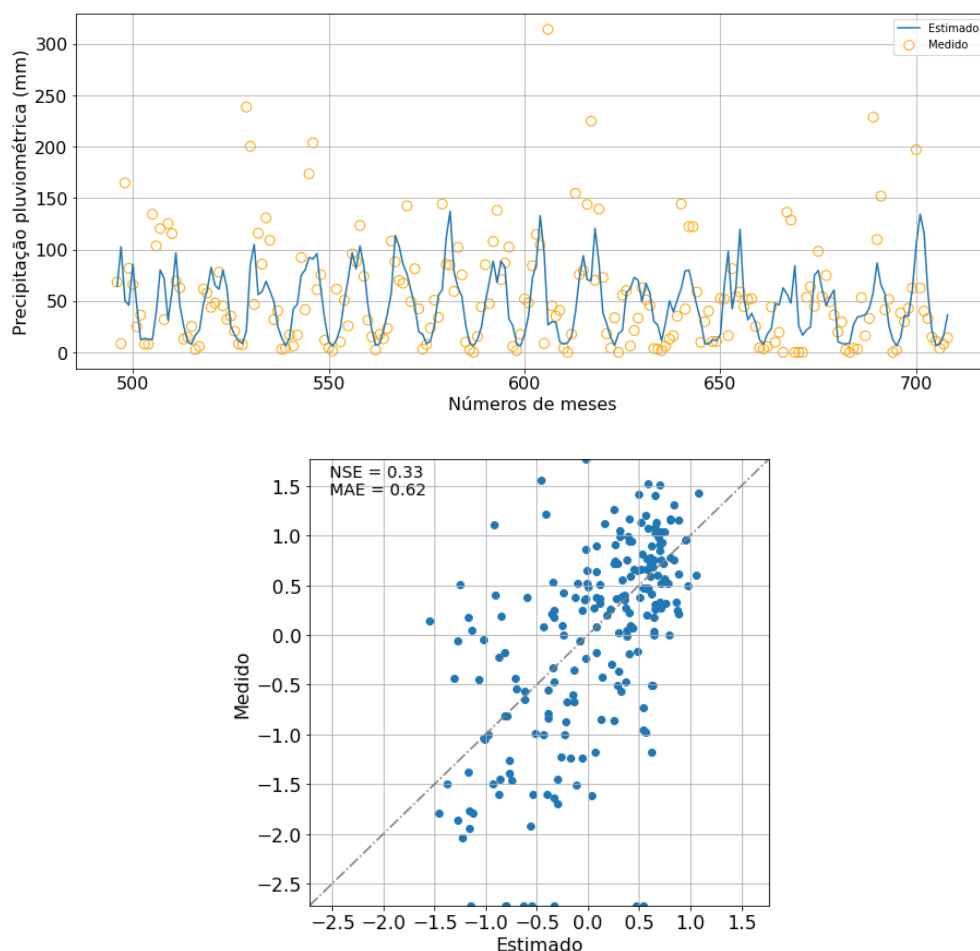


Figura 5. (A) Predição da Precipitação Mensal pelo modelo no período de Validação, (B) Scatter plot dos dados normalizados medidos e estimados pelo modelo no período de validação.

CONCLUSÃO

Com o progresso deste estudo, foi possível aplicar com sucesso as Redes Neurais Artificiais para a previsão das precipitações mensais no município de Caruaru, usando o banco de dados da série histórica pluviométrica disponível. Embora tenhamos obtido resultados promissores, ficou evidente que a quantidade de dados disponíveis não foi suficiente para alcançar o nível de precisão desejado, conforme indicado pelas métricas de avaliação, nomeadamente os coeficientes MAE e NSE.

No entanto, conseguimos identificar padrões notáveis mesmo com essa limitação de dados. Fomos capazes de prever com sucesso os períodos de ausência de chuvas e identificar padrões cíclicos, especialmente durante os meses com baixa precipitação, como novembro. Além disso, pudemos identificar os meses de março a julho como os que apresentam maior probabilidade de eventos de alta precipitação pluviométrica ao longo do ano. Essa descoberta coincide com o período das festas tradicionais anuais, que historicamente celebram a colheita após o final do período chuvoso.

Considerando esses resultados, acredita-se que uma análise multivariada da precipitação, levando em consideração fatores como a temperatura do oceano Atlântico e o efeito El Niño/La Niña, pode contribuir significativamente para melhorar a precisão do nosso modelo de previsão.

AGRADECIMENTOS

Este trabalho foi realizado com o apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo Nº 117603/20201.

REFERÊNCIAS

- AGÊNCIA Pernambucana de Água e Clima APAC. [S.l.], 31 dez. 2018. Disponível em: <http://www.apac.pe.gov.br/>. Acesso em: 10 ago. 2019.
- CARVALHO, H. M. (2014). Aprendizado de máquina voltado para mineração de dados: árvores de decisão.
- DE OLIVEIRA RODRIGUES, L., DE SOUZA, W. M., DE OLIVEIRA COSTA, V. S., & PEREIRA, M. L. T. Influência dos eventos de El Niño e La Niña no regime de precipitação do Agreste de Pernambuco (Influence of El Niño and La Niña events on rainfall of Agreste Pernambuco). *Revista Brasileira de Geografia Física*, 10(6), 1995-2009.
- DOS SANTOS SOUSA, R., SILVA, L., GUEDES, E. B., & DE OLIVEIRA, M. B. L. (2017). Redes neurais artificiais aplicadas à previsão antecipada de precipitações na região central de Manaus. IV Escola Regional de Informática Norte I, 1-9.
- DOS SANTOS, T. E., & MONTENEGRO, A. A. (2012). Erosividade e padrões hidrológicos de precipitação no Agreste Central pernambucano. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 16(8), 871-880.
- GEMULLA, R.; NIJKAMP, E.; HAAS, P.J.; SISMANIS, Y. Large-scale matrix factorization with distributed stochastic gradient descent. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 69–77. ACM, 2011.
- GORODETSKAYA, Yulia et al. Daily streamflow forecasting for Paraíba do Sul river using machine learning methods with hydrologic inputs. In: *Anais do XV Encontro Nacional de Inteligência Artificial e Computacional*. SBC, 2018. p. 162-173.
- KÖPPEN, W.; GEIGER, R. "Klimate der Erde. Gotha: Verlag Justus Perthes". Wallmap 150cmx200cm. 1928.
- MORIASI, D. N., ARNOLD, J. G., VAN LIEW, M. W., BINGNER, R. L., HARMEL, R. D., and VEITH, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *T. ASABE*, 50, 885–900, 2007.
- PEREIRA, D. R., ULIANA, E. M., MARTINEZ, M. A., & SILVA, D. D. (2016). Desempenho de um modelo hidrológico concentrado e de um semidistribuído na predição de vazões diárias. *Irriga*, 21(2), 409-409.
- PINTO, J. C. (2018). Geração de representação distribuída de palavras fatorando a matriz de informação mútua pontual utilizando o gradiente descendente estocástico no modelo de MapReduce.
- RODRIGUES, L. O.; SOUZA, W. M.; COSTA, V. S. O.; PEREIRA, M. L. T. Influência dos eventos de El Niño e La Niña no regime de precipitação do Agreste de Pernambuco. *Revista Brasileira de Geografia Física*, v. 10, n. 6, p. 1995-2009, 2017.
- SANTOS, T. E. M.; SILVA, D. D.; MONTENEGRO, A. A. A. Temporal variability of soil water content under different surface conditions in the semiarid region of the Pernambuco State. *Revista Brasileira de Ciência do Solo*, v.34, p.1733-1741, 2010.

- SCHNITMAN, L. (1998). Controladores Preditivos Baseados em Redes Neurais Artificiais (Doctoral dissertation, Dissertação de Mestrado do curso de Pós Graduação em Eng. Elétrica da–UFBa).
- SHORTRIDGE, J. E., GUIKEMA, S. D., & ZAITCHIK, B. F. (2016). Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrology & Earth System Sciences*, 20(7).
- SILVA, R. M.; PEREIRA E SILVA, L.; MONTENEGRO, S. M. G. L.; SANTOS, C. A. G. Análise da variabilidade espaço-temporal e identificação do padrão da Precipitação na Bacia Do Rio Tapacurá, Pernambuco. *Sociedade & Natureza*, v.22, n.2, p. 357-372, ago. 2010.
- VALENTE, P. O. L. (2019). PROGRAMACAO FUNCIONAL APLICADA AO GERENCIAMENTO DE SISTEMAS DE APRENDIZADO DE MAQUINA (Doctoral dissertation, Universidade Federal do Rio de Janeiro).
- YASEEN, Z. M., ALLAWI, M. F., YOUSIF, A. A., JAAFAR, O., HAMZAH, F. M., & EL-SHAFIE, A. (2018). Non-tuned machine learning approach for hydrological time series forecasting. *Neural Computing and Applications*, 30(5), 1479-1491.

Capítulo 4



10.37423/230908154

ANÁLISE FORENSE DE IMAGENS DIGITAIS PARA DETECÇÃO DE FOTOMONTAGEM E DEEPFAKE

Josef Augusto Oberdan Souza Silva

Instituto Federal Goiano - Campus Ceres

Luís Sérgio Rodrigues Vale

Instituto Federal Goiano - Campus Ceres



Resumo: O presente artigo estuda o uso de técnicas e ferramentas forense digitais para a análise de imagens, com o objetivo de detectar a manipulação de imagens digitais por fotomontagem ou *Deepfake*. A manipulação de imagens é recorrente nos dias atuais e, com o avanço da tecnologia de ferramentas de edição de imagens e de algoritmos de aprendizado de máquina profundo cada vez mais sofisticados, há também o aumento no volume de imagens e vídeos falsos circulando na internet. Este artigo tem como objetivo elucidar a discussão sobre autenticidade de imagens digitais, com o uso de técnicas e abordagens de perícia forense e processamento de imagens para a detecção de fotomontagem e *Deepfake*. Utilizando do repositório público de imagens disponível na internet, o presente estudo tem como foco a técnica de observação direta, análise de coerência perceptiva e análise de histograma nos canais de cor R, G e B. Com os resultados obtidos por meio deste estudo é possível determinar a ilegitimidade, encontrados sinais de adulteração nos detalhes de uma imagem digital.

Palavras-chave: Edição de imagem digital. Detecção de manipulação de imagem. *DeepFake*. Ferramenta forense digital. Fotomontagem. Imagem digital. Processamento de imagem. Perícia.

1. INTRODUÇÃO

A manipulação de imagens, recorrente nas últimas décadas, estava restrita apenas aos especialistas em tecnologia voltados à essa área de estudo, no qual era necessário recortar e colar imagens para recriar um novo cenário. Com o advento de softwares específicos para edição de imagens cada vez mais modernos, houve também o aumento no volume de imagens adulteradas que circulam na internet. Hoje é possível editar uma imagem utilizando um computador pessoal ou por meio de aparelhos celulares. Segundo o estudo de Nightingale et al. (2017), cerca de 40% das imagens manipuladas não são percebidas pela maioria das pessoas.

Com o avanço dos recursos tecnológicos gratuitos para edição de imagens como, por exemplo, aplicativos de edição de imagem, a quantidade de imagens adulteradas digitalmente circulando na internet também aumenta (ARRUDA, 2016). Grande parte das imagens digitais falsas que circulam pelas redes sociais podem gerar notícias falsas (fake news, em inglês) e acarretar problemas financeiros e judiciais (SHARMA, 2016). Embora os estudos nesta área estejam evoluindo, com o surgimento de aplicações como DeepFake, DeepNude e FaceSwap, imagens falsas circulando na internet também aumentaram nos últimos anos (BAKDASH et al., 2018).

Uma forma comumente usada para falsificar digitalmente imagens envolve a criação de uma composição de imagem extraíndo uma parte de uma imagem e inserindo-a em um local específico em outra, também chamado de fotomontagem (splicing, em inglês) (MOHAMADIAN e POUYAN, 2013).

Para identificar imagens adulteradas, o uso de ferramentas forenses é necessário para que sejam identificados vestígios digitais. A computação forense é descrita por Franco et al. (2016:2) como a ramificação da Criminalística, com o objetivo de identificar os vestígios cibernéticos causados por práticas ilegais utilizando-se de ferramentas do meio digital. A definição de forense digital foi descrita a seguir:

O uso de métodos cientificamente derivados e comprovados para a preservação, coleta, validação, identificação, análise, interpretação, documentação e apresentação de evidências digitais derivadas de fontes digitais para fins de facilitar ou promover a reconstrução de eventos considerados criminosos, ou ajudando a antecipar ações não autorizadas que se mostram disruptivas para as operações planejadas (ALTHEIDE e CARVEY, 2011).

Segundo Arruda (2016:208), compreender as técnicas aplicadas nos exames periciais em imagens digitais exige que seja necessário entender as etapas de formação das imagens, como elas são armazenadas nos dispositivos e a aplicação matemática nos estudos voltados para a análise de

imagem. Para o autor, a análise forense de imagens digitais envolve outras áreas como, por exemplo, matemática, computação, física e estatística.

De acordo com Altheide e Carvey (2011), os benefícios de se utilizar ferramentas computacionais de software livre para a análise digital forense corroboram para além das áreas da educação. A flexibilidade, portabilidade, baixo custo e a disponibilidade de verificar correções e alterações entre as versões em softwares de código aberto, também são benefícios encontrados em ferramentas forenses gratuitas. Os autores ainda relatam que a ferramenta forense Sleuth Kit, por exemplo, permite revisar as correções de bugs no software diretamente do código-fonte.

O estudo realizado por Reis (2018) utilizou da análise forense de imagens digitais no estudo de caso da ficha criminal de Dilma Rousseff, mostrando uma abordagem para determinar a ilegitimidade de uma imagem digital publicada por um veículo de imprensa. O estudo realizado pelo autor servirá como fonte conceitual para este artigo.

Os estudos de Mahmood et al. (2016) mostraram que é possível detectar a falsificação de imagens por meio da análise forense em imagens digitais com o uso de ferramentas computacionais gratuitas. Sendo assim, o objetivo deste artigo é analisar e avaliar imagens digitalmente adulteradas por aprendizado de máquina profundo com o auxílio de uma ferramenta forense de software livre para análise de imagens.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 PERÍCIA EM IMAGENS DIGITAIS

De acordo com Parveen et al. (2020), a análise forense em imagens digitais está relacionada a detecção de qualquer forma de falsificação em fotografias. Segundo Gokhale et al. (2020), a análise forense de imagens digitais pode ser dividida em três categorias principais, tais como: a autenticação da fonte ou dispositivo da imagem, a detecção de falsificação da imagem e a diferenciação entre as imagens geradas por computador e as imagens reais.

Freitas (2006:1) descreve a Perícia Forense Computacional como a junção entre o conhecimento da Tecnologia da Informação e técnicas forense para análises periciais. Visto isso, a tecnologia é um mecanismo importante para a perícia forense moderna para a detecção de imagens fraudulentas. Ademais, a escolha e uso das melhores ferramentas para se realizar uma perícia forense adequada se tornam cada vez mais cruciais na qualidade das análises.

2.2 OBSERVAÇÃO DIRETA

Segundo Arruda (2016:222), a técnica de observação direta diz respeito à identificação de detalhes em uma imagem que, a priori, não estão em seu contexto e que parte de uma visão subjetiva, sendo necessária ser utilizada com cautela. Posteriormente, deve-se utilizar das técnicas adequadas para identificar se há alguma adulteração na imagem investigada.

Partindo deste princípio, se faz necessário perceber os indícios que corroboram com a afirmação de que a imagem pode ter sido modificada e iniciar a aplicação das técnicas adequadas que afirmam a descaracterização na imagem. Elementos fora do contexto, distorções na imagem e cores que não se adequam ao restante do cenário podem implicar que a imagem passou por algum processo de falsificação digital.

2.3 TÉCNICA COMPUTACIONAL DE MANIPULAÇÃO FACIAL (DEEPPFAKE)

Deepfake é o termo para imagens ou vídeos manipulados digitalmente com o uso de técnicas de aprendizado profundo de máquina de Redes Adversárias Generativas (GAN's ou StyleGAN, sigla em inglês), podem contém atributos de outra imagem como, por exemplo, cor de pele e contorno do rosto humano de outro indivíduo, capaz de criar uma nova imagem ou alterar a imagem original e confundir quem as observa (ROSSLER, 2019:3).

Na primeira etapa na Figura 1 a seguir, o modelo de aprendizado de máquina profundo cria um codificador comum entre duas imagens de pessoas diferentes. Já nas segunda e terceira etapas, é construído dois decodificadores individualmente para reconstruir, respectivamente, a primeira e a segunda imagem (HUI, 2020).

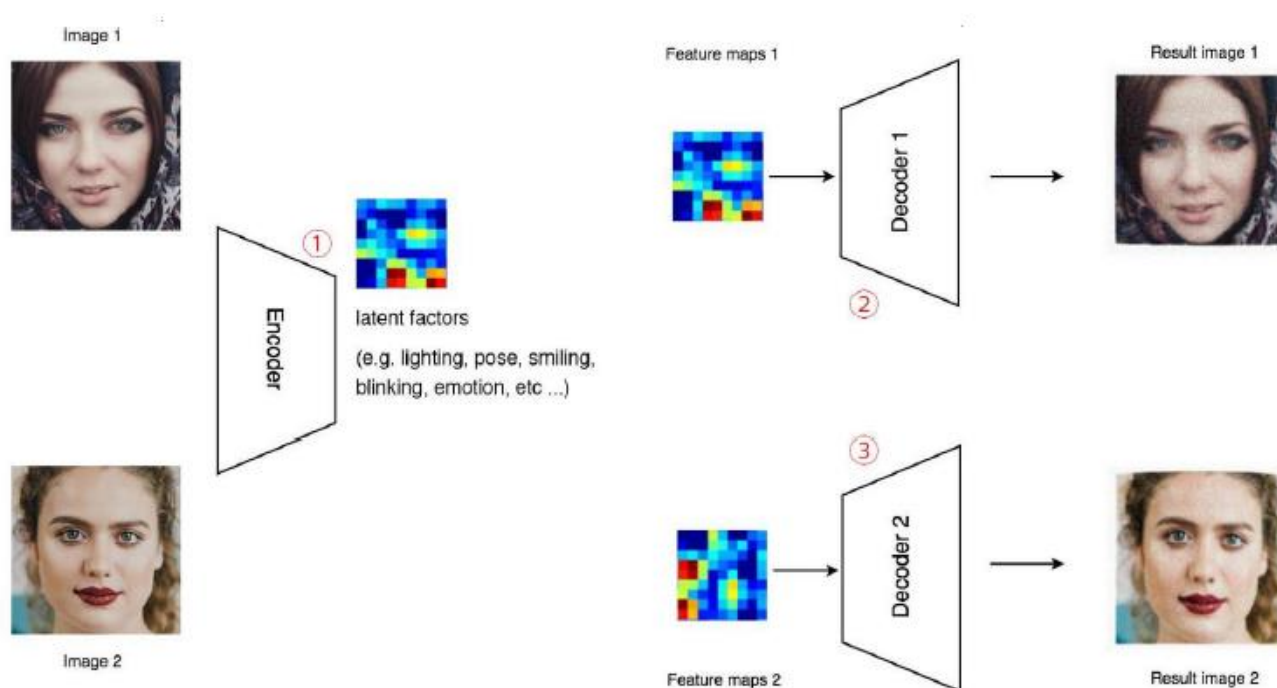


Figura 1 – Diagrama de funcionamento do modelo de Deepfake com codificador e decodificador.

Fonte: Hui (2020)

Segundo Hui (2020), os fatores latentes do rosto de uma das imagens dos indivíduos são capturados pelo codificador e renderizado em outra imagem com o decodificador, copiando a pose, iluminação e expressão facial presentes na imagem. Como mostrado na Figura 2 a seguir, para solucionar o problema que a técnica de fotomontagem apresenta, o decodificador pode utilizar máscaras ou filtros para amenizar os efeitos de recorte visíveis e suavizar outros detalhes, como o tom de pele e a expressão facial, para entregar o melhor resultado.

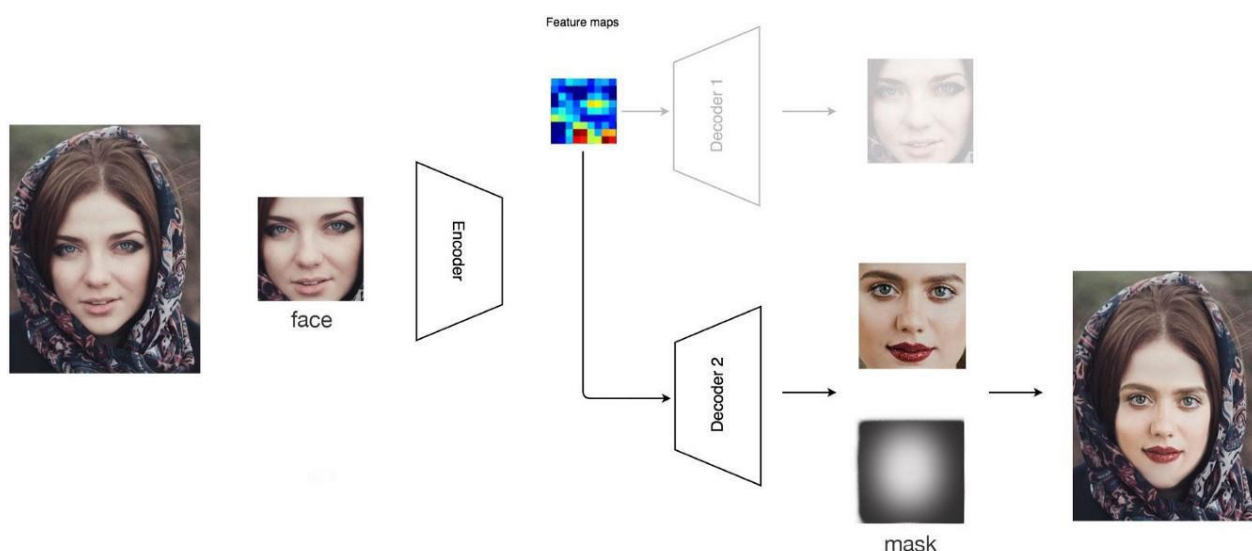


Figura 2 – Diagrama de funcionamento do modelo de Deepfake com decodificador utilizando máscaras nas imagens.

Fonte: Hui (2020)

Ao amenizar os efeitos visíveis da imagem final gerada pelo decodificador do algoritmo de aprendizado de máquina, é possível que haja ainda alguns detalhes que podem não ter sido cobertos pelos filtros e máscaras utilizados. Esses detalhes podem também ser visíveis e, conseqüentemente, revelar que a imagem não é legítima.

2.4 ANÁLISE DE COERÊNCIA PERCEPTIVA

A análise de coerência perceptiva está diretamente relacionada com a percepção sensorial em um exame de imagem digital que envolve perceber a continuidade lógica, inconsistências físicas, efeitos de sombra e luz e as dimensões. Este tipo de análise vai além da observação direta, pois utiliza-se de técnicas para a análise e se baseia na ciência da Fotogrametria.

Segundo Arruda (2016:236-237), a definição de fotogrametria aborda um conjunto de técnicas que possibilitam capturar informações sobre as dimensões de pessoas ou objetos em uma imagem bidimensional, executando os princípios de perspectiva cônica. Este termo se faz presente no ramo pericial como forma de estimar a altura de um indivíduo, por exemplo.

2.5 ANÁLISE DE HISTOGRAMA NOS CANAIS DE COR RGB

De acordo com Reis (2018:4), o processo de formação das imagens digitais se baseia na premissa da interpretação de um cenário analógico para um cenário digital. A construção deste cenário é realizada

por meio de sensores sensíveis a luz, disposto em uma matriz que é composta, por sua vez, por um elemento de imagem, denominado pixel. Uma imagem digital então é constituída por um conjunto de pixels.

A análise de histograma nos canais de cor RGB consiste em analisar a disposição e intensidade individualmente em cada canal de cor da imagem a ser investigada. O conceito de histograma nos canais de cor RGB é definido a seguir:

Uma vez que o histograma é uma distribuição de determinados valores dentro de uma classificação previamente estabelecida, podemos inferir que um histograma dos canais de cor RGB é uma disposição, onde cada uma das cores será representada dada sua intensidade/distribuição de acordo com uma classificação prévia. No caso de imagens digitais é comum a utilização da escala de 0 a 255 (8 bits) para a representação da luminosidade de um determinado pixel, onde zero seria uma intensidade de brilho nulo e 255 a intensidade máxima (REIS, 2018:4).

Pode-se notar que o histograma dos canais de cor RGB permite analisar, em um exemplo prático, se o tom de pele do rosto de um indivíduo em uma imagem digital condiz ou não com o busto da pessoa presente na imagem. Outros detalhes podem ser percebidos como, por exemplo, a perda na tonalidade ou na qualidade da imagem, e/ou desfoque na região dos olhos.

3. DESENVOLVIMENTO

3.1 FERRAMENTA DE ANÁLISE

Para a análise das imagens geradas pela técnica da fotomontagem e pelo método Deepfake, é importante perceber que se deve utilizar uma ferramenta que possa analisar alguns detalhes que levem indícios de que a imagem não é legítima. Visto isso, foi utilizado neste estudo o software de código aberto ImageJ®.

O software de código aberto ImageJ® foi idealizado e implementado por Wayne Rasband em 1987. O software é amplamente utilizado em análises científicas diversas, principalmente no campo da nanotecnologia. A principal vantagem de se utilizar o software ImageJ® se concentra em sua escalabilidade, podendo adicionar novos plug-ins ampliando sua capacidade técnica de análise de imagem digital.

3.2 OBSERVAÇÃO DIRETA E ANÁLISE DE COERÊNCIA PERCEPTIVA EM IMAGEM COM FOTOMONTAGEM

Para a realização do experimento proposto neste estudo foi utilizado uma base de imagens, coletadas no repositório público de imagens Unsplash®, presente no seguinte endereço de acesso: <https://unsplash.com/>.

Neste repositório, as imagens disponíveis são gratuitas, de alta qualidade de resolução, com rosto de indivíduos diferentes entre si, e com autoria de diversos fotógrafos (UNSPLASH, 2022). A aplicação web PhotoFunia®, no seguinte endereço de acesso:

https://photofunia.com/pt/effects/face_swap/.

Foi utilizada para realizar a troca dos rostos nas imagens. Foram utilizadas 10 imagens no total, com diferentes rostos, escolhidos ao acaso. As imagens foram separadas em cinco pares e, em seguida a técnica de fotomontagem (splicing, do inglês) foi aplicada pelo algoritmo. As figuras a seguir mostram como as imagens foram geradas por fotomontagem.

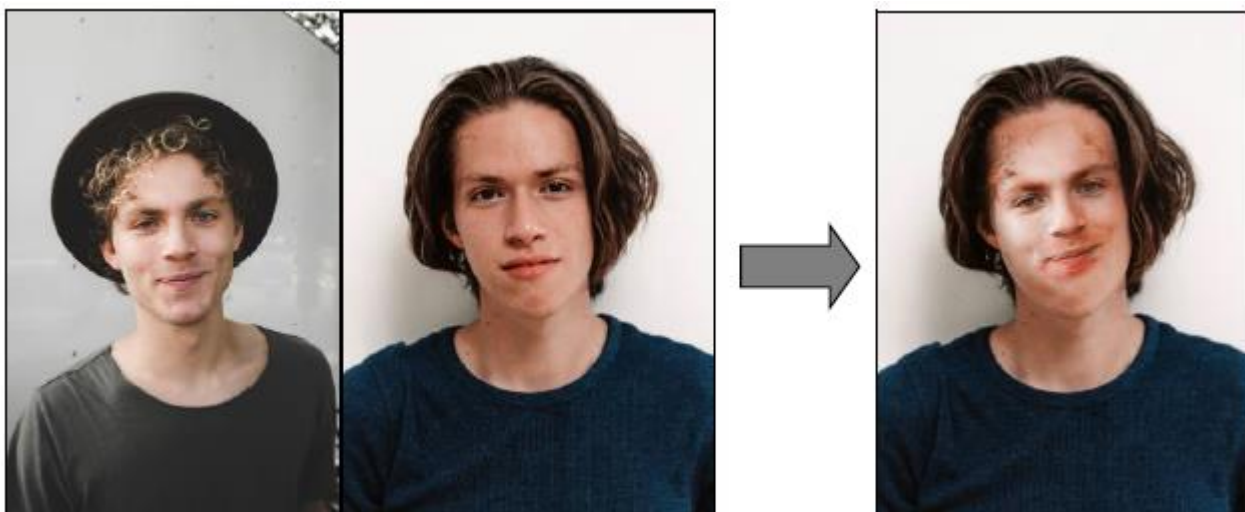


Figura 3 – Algoritmo de fotomontagem facial.

Fonte: Adaptado de Unsplash® (2022)

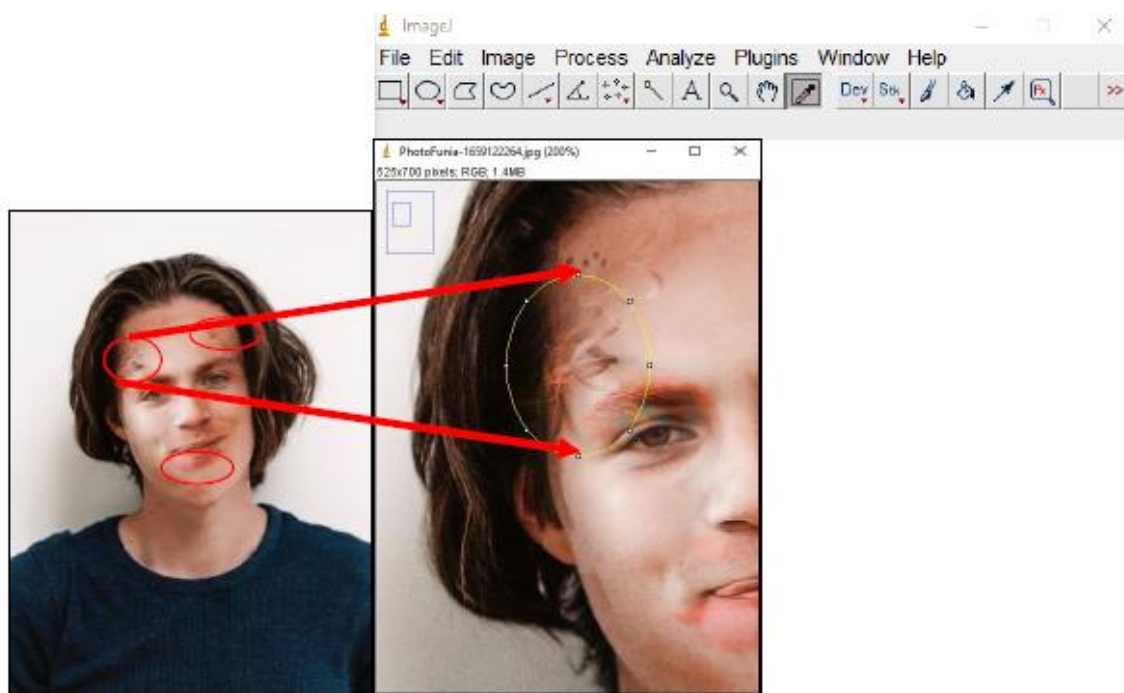


Figura 4 – Análise de fotomontagem facial usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

Na técnica de observação direta foi feita a identificação de detalhes nas imagens que, a priori, não fazem parte do contexto da cena na imagem feita por fotomontagem. Na Figura 4 gerada pelo algoritmo é possível notar alguns efeitos perceptíveis que distinguem o rosto dos demais detalhes da cabeça e do pescoço do rapaz na imagem como, por exemplo, a diferença no tom de pele (indicado com os círculos em vermelho). É possível ainda perceber que o rosto não se enquadrava da melhor forma e que há distorções, gerando um efeito de queixo duplo sobrepondo as imagens.

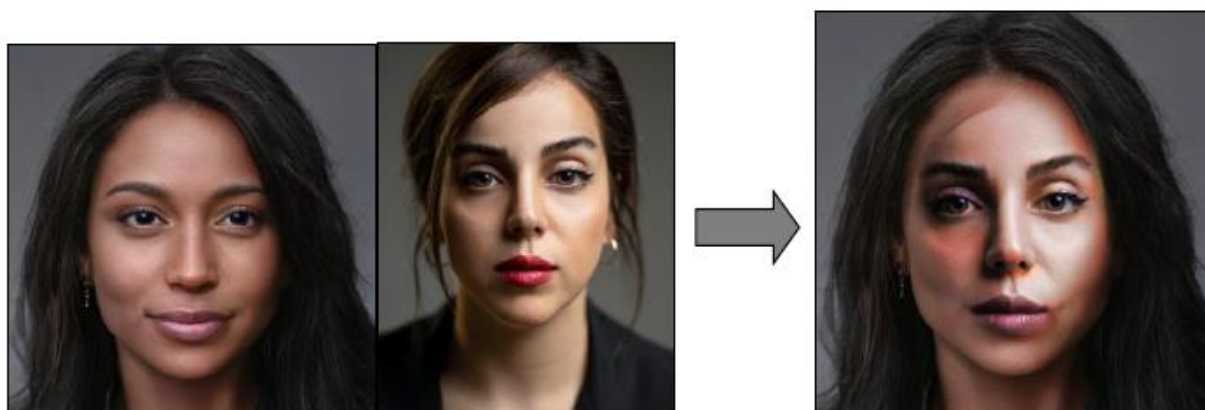


Figura 5 – Algoritmo de fotomontagem facial.

Fonte: Adaptado de Unsplash® (2022)

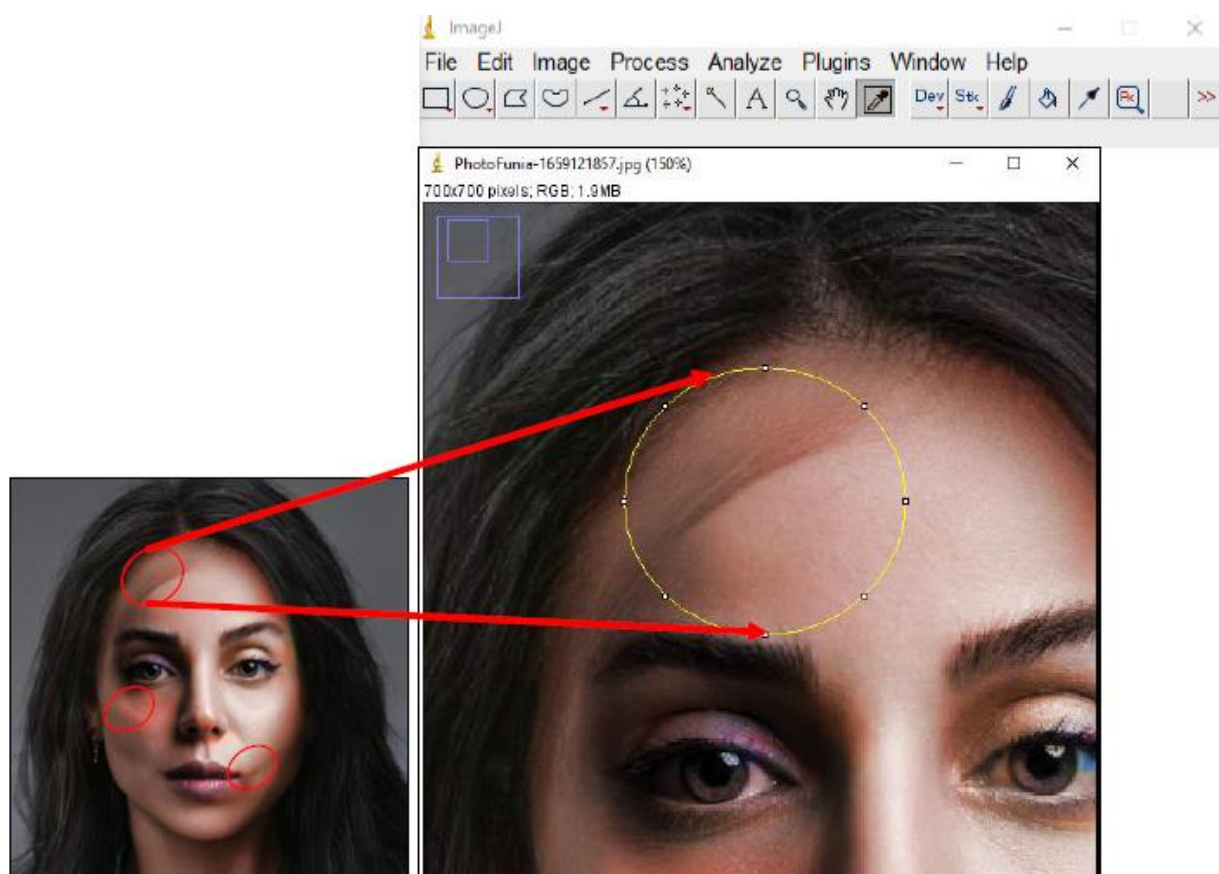


Figura 6 – Análise de fotomontagem facial usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

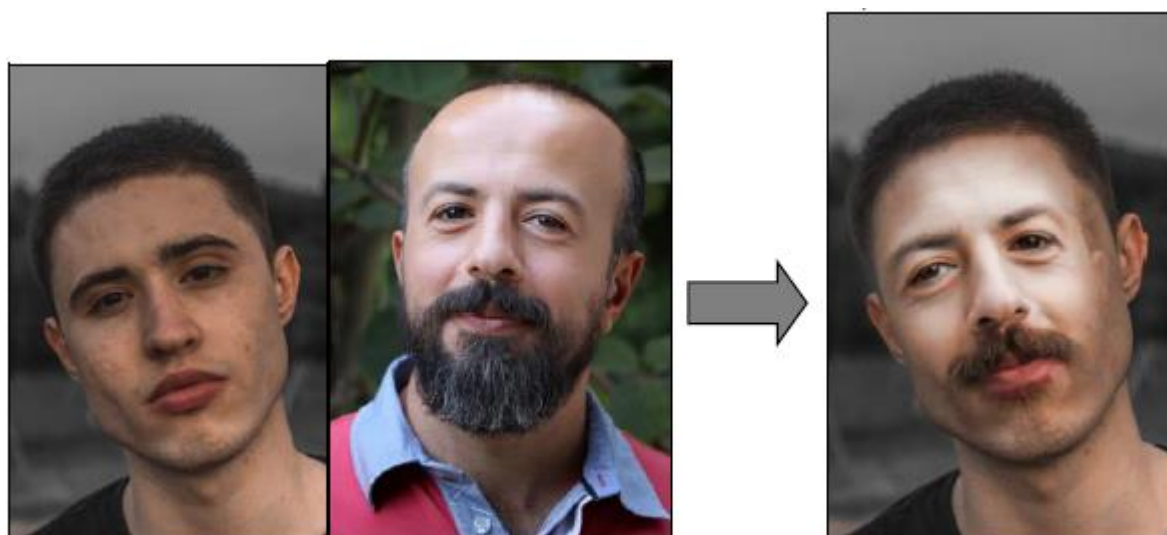


Figura 7 – Algoritmo de fotomontagem facial.

Fonte: Adaptado de Unsplash® (2022)

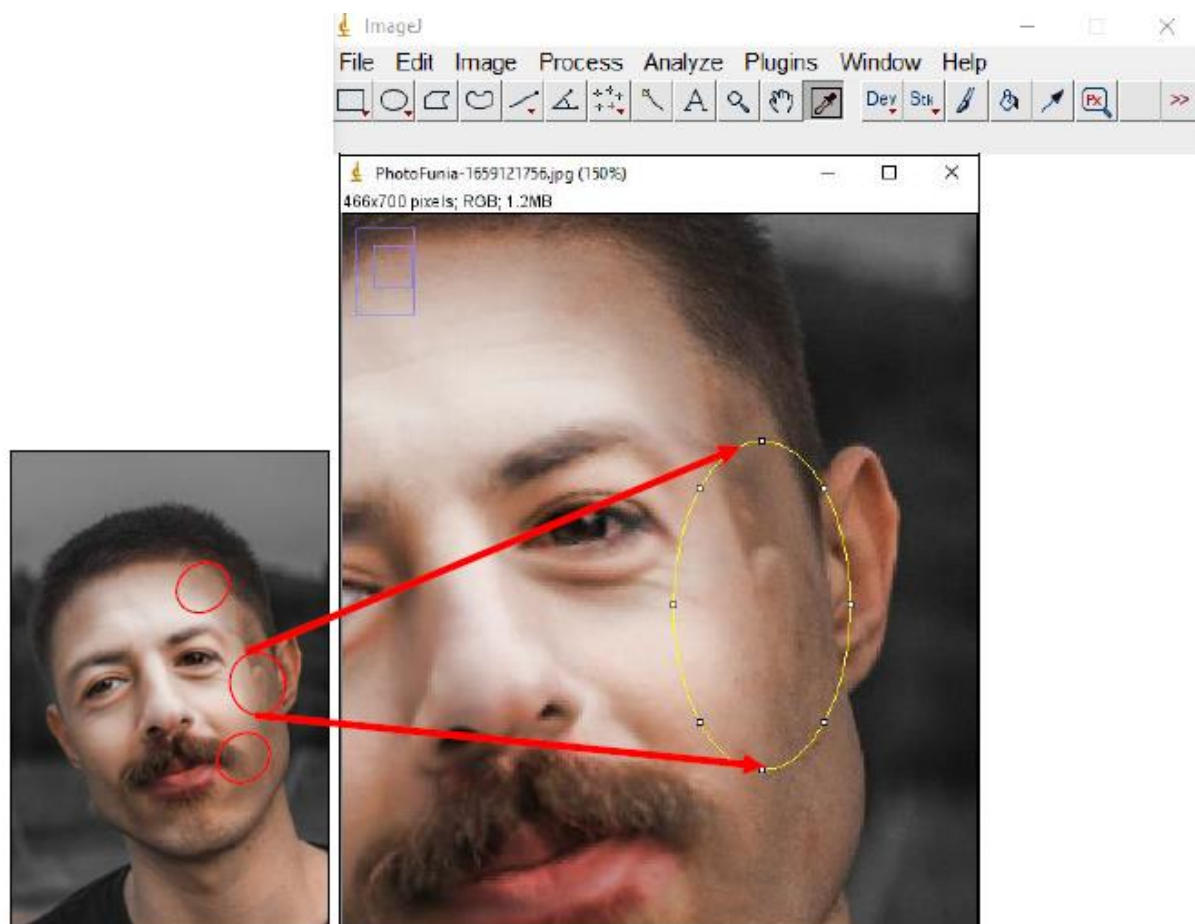


Figura 8 – Análise de fotomontagem facial usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

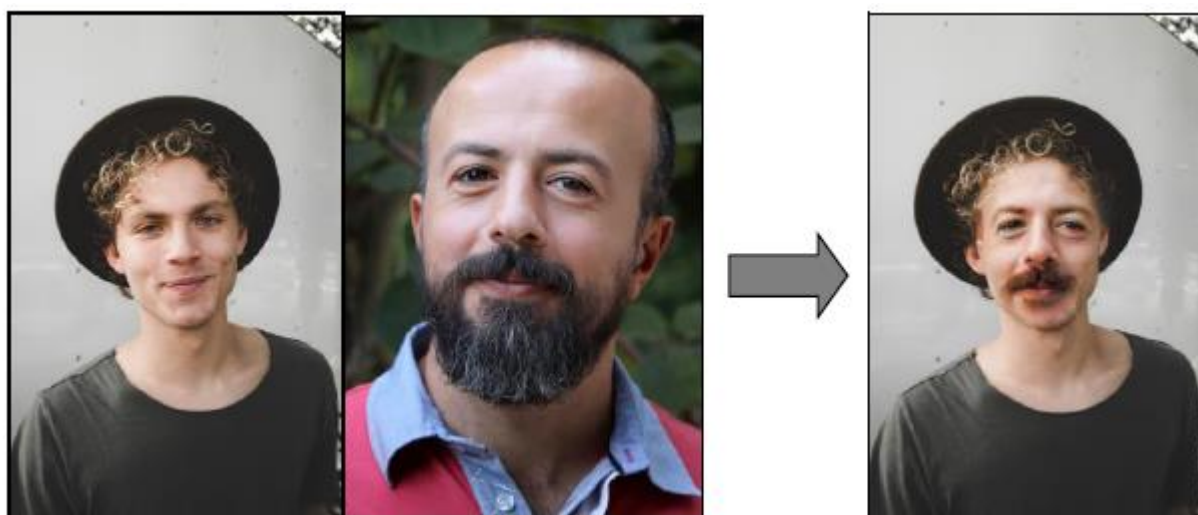


Figura 9 – Algoritmo de fotomontagem facial.

Fonte: Adaptado de Unsplash® (2022)

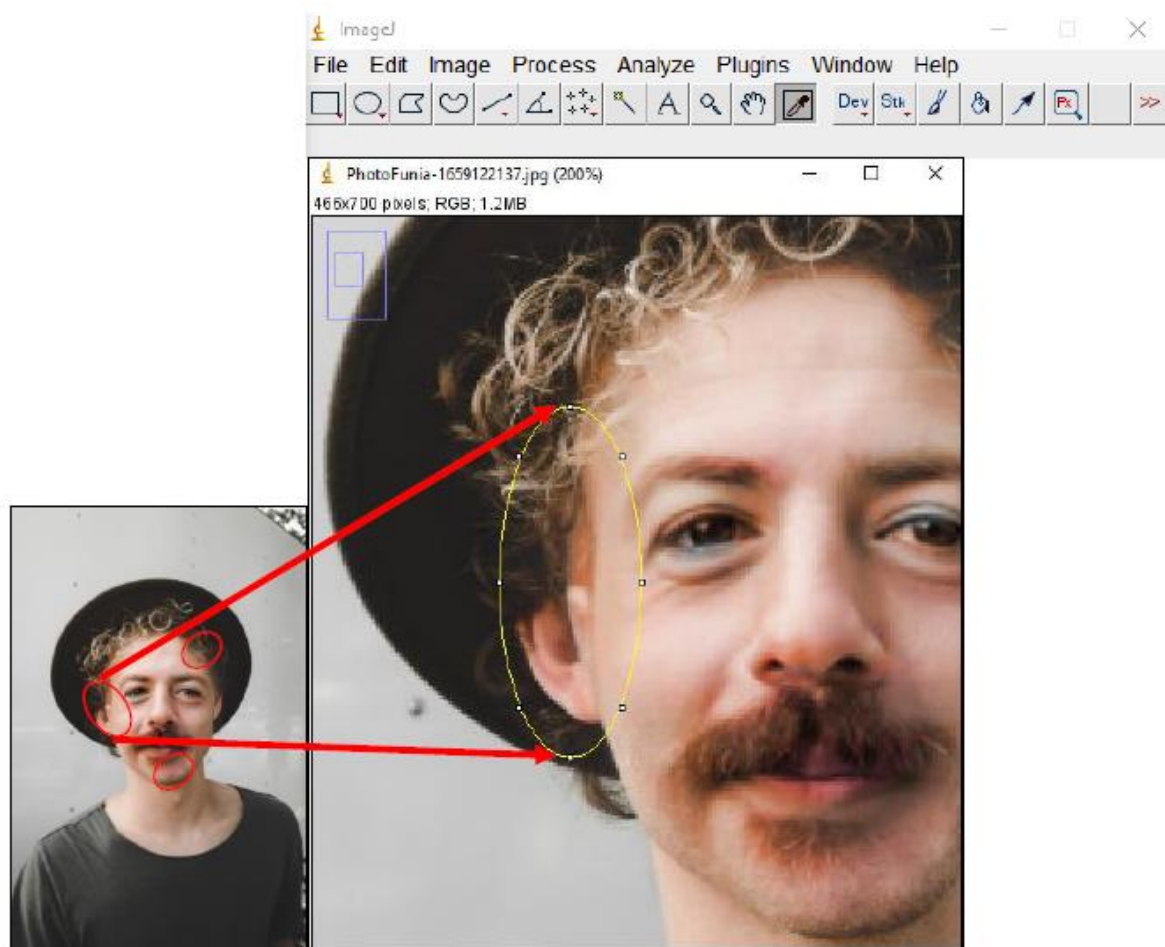


Figura 10 – Análise de fotomontagem facial usando o software ImageJ®.

Fonte: Adaptado de Unsplash® (2022)

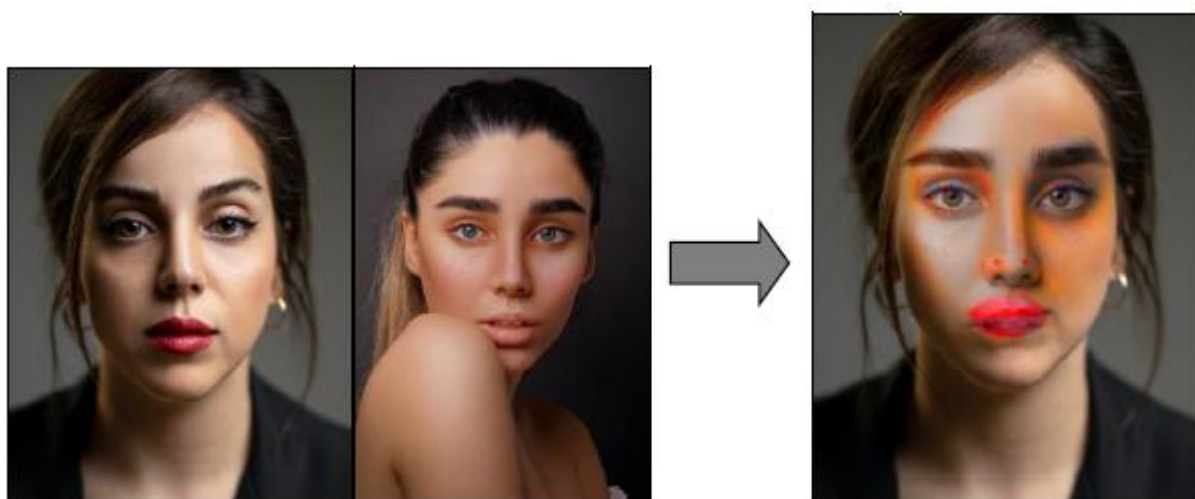


Figura 11 – Algoritmo de fotomontagem facial.

Fonte: Dados produzidos pelo o autor (2011)

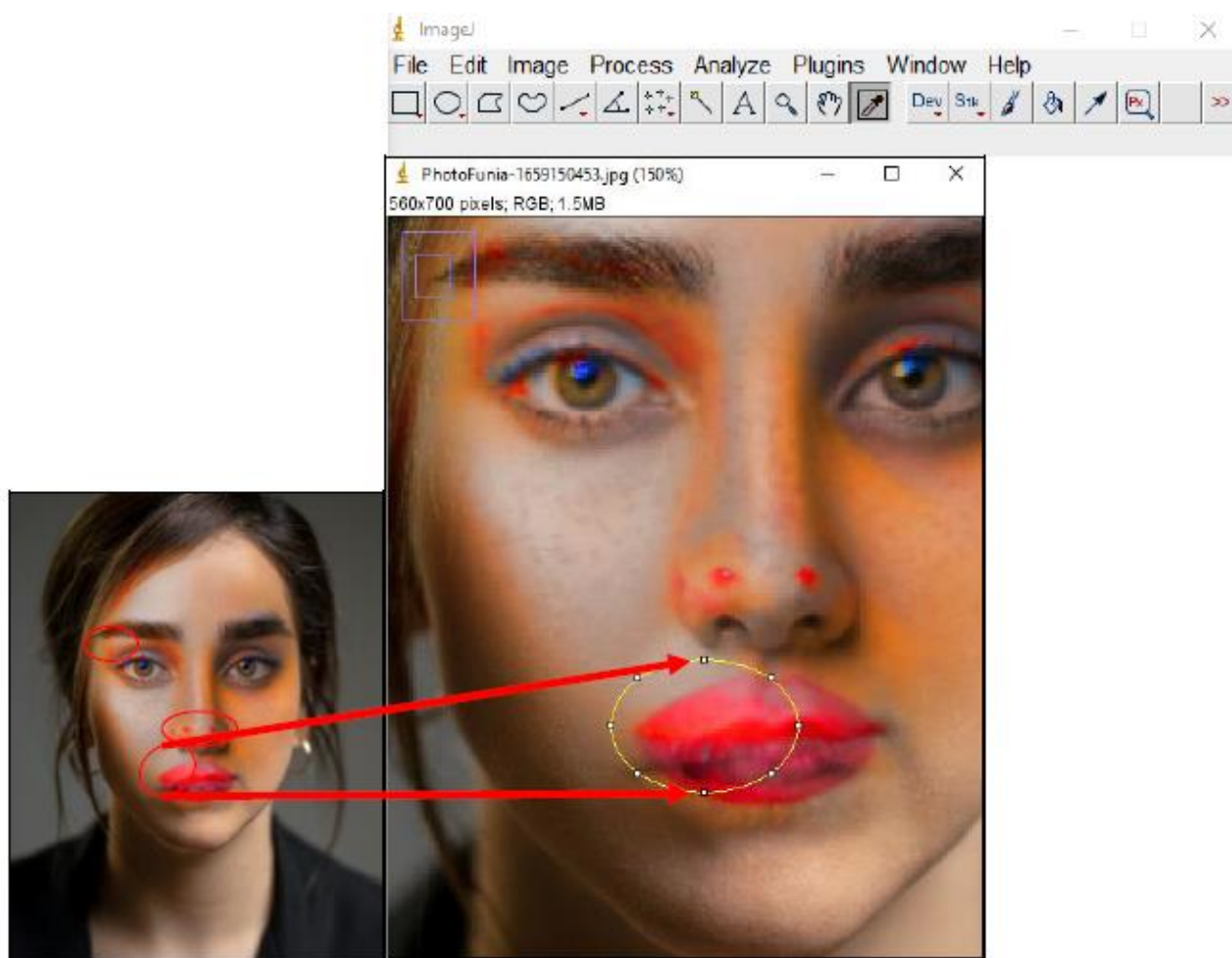


Figura 12 – Análise de fotomontagem facial usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

Para a análise de coerência perceptiva foi necessário perceber a continuidade lógica, inconsistências físicas, efeitos de sombra e luz e as dimensões. Nas Figuras 6 e 12 é possível notar que há dois tons de pele na testa, além do efeito embaçado na imagem final gerada pelo algoritmo de fotomontagem.

Há detalhes perceptíveis nos tons da bochecha e próximo à boca, na maioria das imagens geradas, que não condiz com restante dos rostos dos indivíduos mostrados. As imagens finais geradas também apresentam distorções que, possivelmente, foram causadas pela diferença entre as dimensões do rosto entre as duas imagens geradoras, revelando a ilegitimidade das imagens geradas.

3.3 OBSERVAÇÃO DIRETA E ANÁLISE DE COERÊNCIA PERCEPTIVA EM IMAGEM COM DEEPPFAKE

A aplicação web Random Face Generator® (this person does not exist), presente no seguinte endereço de acesso <https://this-person-does-not-exist.com/en>, foi utilizada para gerar aleatoriamente imagens

digitais de rostos de pessoas que não existem. Foram geradas três imagens aleatoriamente pelo algoritmo de aprendizado de máquina profundo (*Deepfake*).

De acordo com a página do desenvolvedor, o sistema é capaz de gerar rostos de pessoas aleatórias e é fundamentado no algoritmo StyleGAN, da fabricante NVIDIA. O sistema é um modelo pré-treinado, composto por sistema de desdobramento e interpolação, que permite encontrar fatores ocultos de variação em uma imagem e passar por cada camada convolucional, construindo uma nova imagem.

A arquitetura do algoritmo StyleGAN segue alguns atributos para gerar imagens de alta qualidade como, por exemplo, gênero, faixa etária, posição do rosto, penteado, personalidade de cada indivíduo, sardas, dentre outros. As figuras a seguir mostram mais informações sobre as imagens geradas por *Deepfake*.

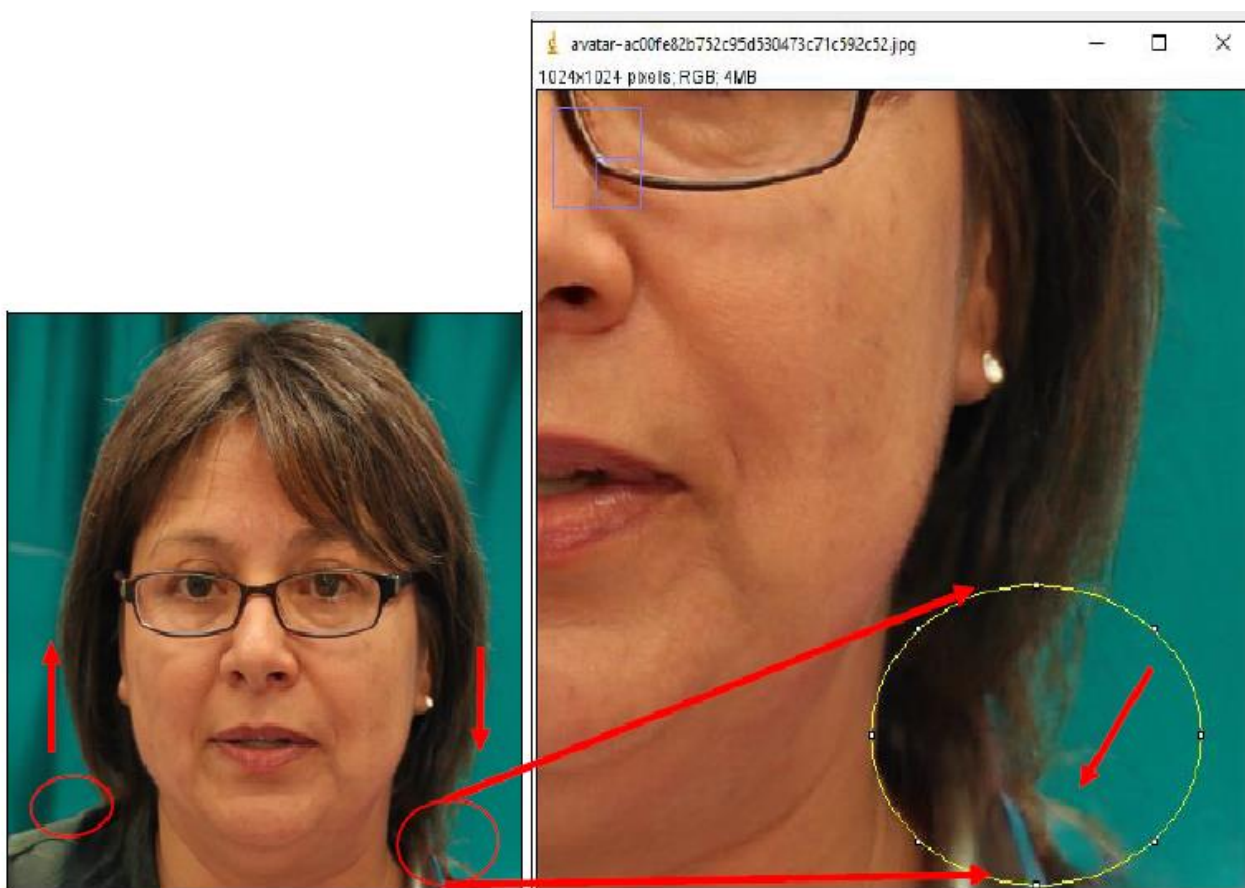


Figura 13 – Análise de Deepfake usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

Para a técnica de observação direta foi feita também a identificação de detalhes nas imagens que, a priori, não fazem parte do contexto da cena nas imagens geradas por Deepfake. Na Figura 13 é possível perceber que há falhas de “sangramento de cor” ao aplicar o método de normalização de instância

usado pelo algoritmo StyleGAN, visto no ombro direito da imagem. A posição dos ombros do indivíduo na imagem também está desalinhada em relação ao pescoço e à cabeça, revelando que a imagem é falsa.

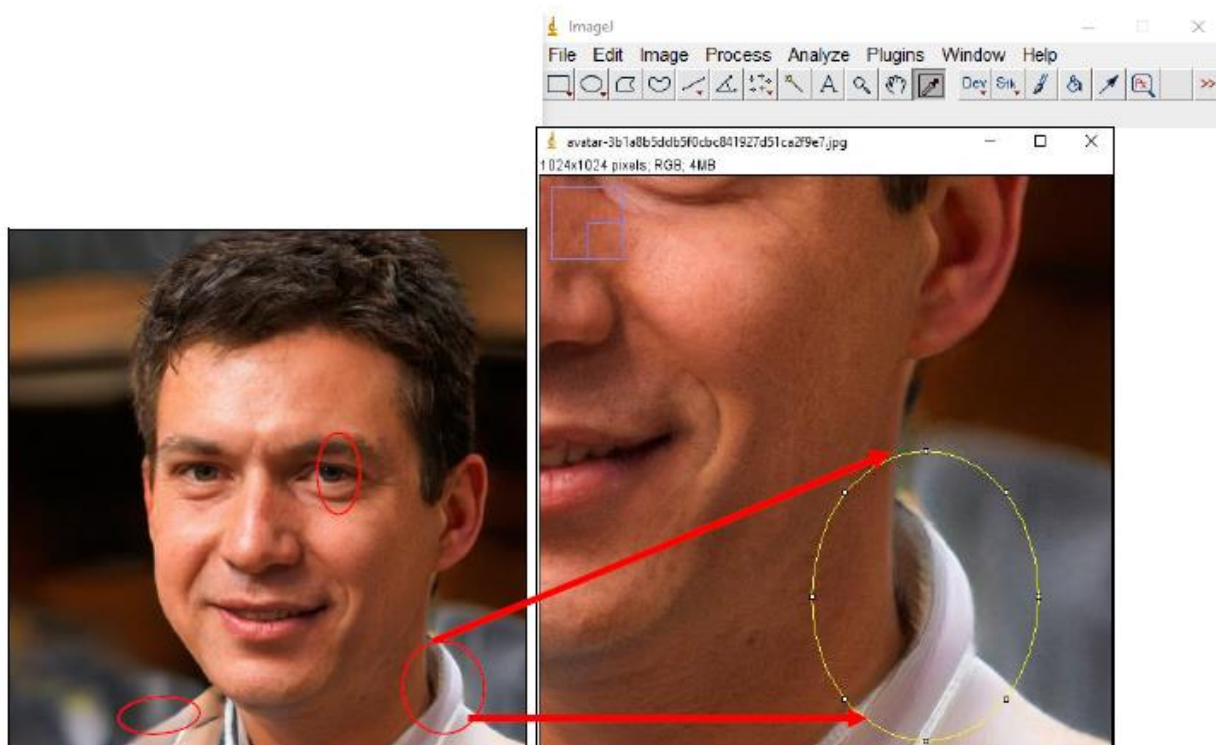


Figura 14 – Análise de Deepfake usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

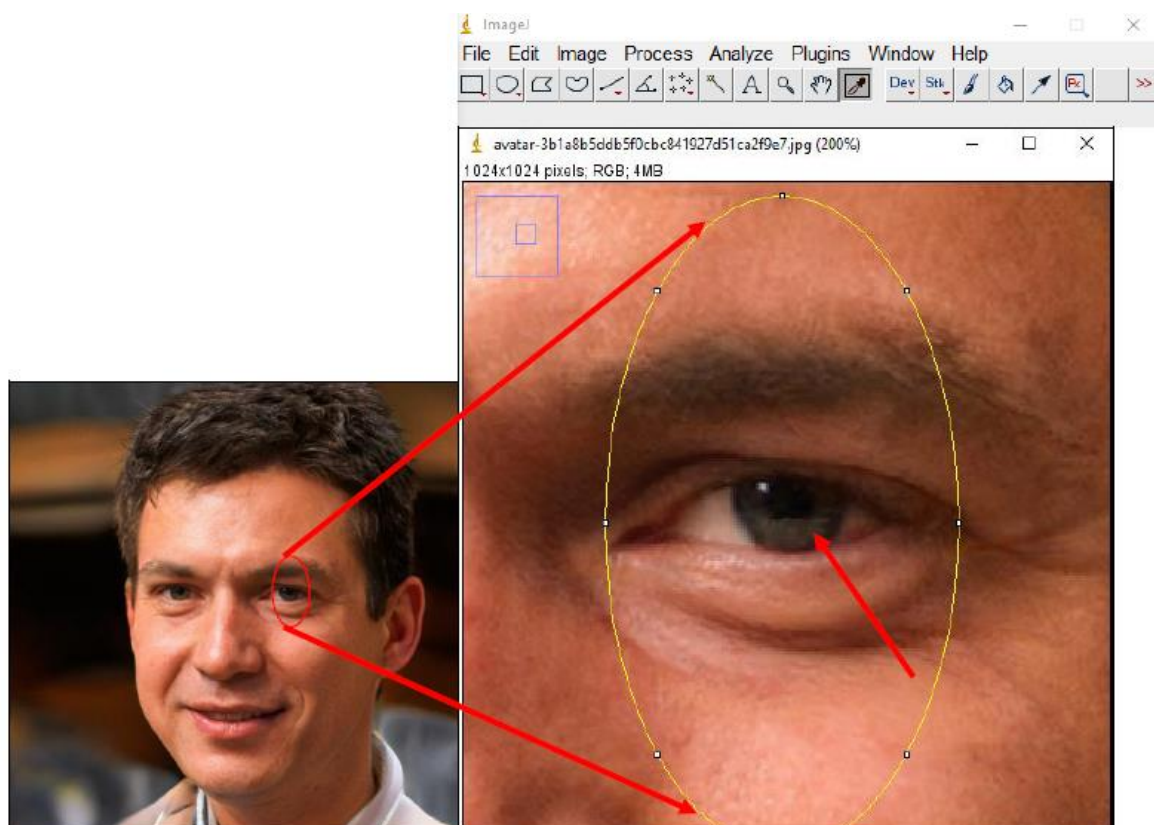


Figura 15 – Análise de Deepfake usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

Na análise de coerência perceptiva foi necessário também perceber a continuidade lógica, inconsistências físicas, efeitos de sombra e luz e as dimensões. Na Figura 14 é perceptível que o tom de cor da roupa do indivíduo na imagem está ofuscado e com uma aparência borrada (efeito blur, do inglês). Este efeito é comum em imagens criadas digitalmente por *Deepfake*, por causa da máscara aplicada pelo algoritmo de aprendizado de máquina profundo.

Outro detalhe que se destaca na Figura 15 é o formato da pupila que não está consistente com um formato normal conhecido e o brilho exagerado encontrado em algumas imagens. Isto ocorre porque dificilmente o algoritmo consegue recriar com perfeição esse detalhe, sendo um defeito comumente encontrado em imagens geradas por *Deepfake*.

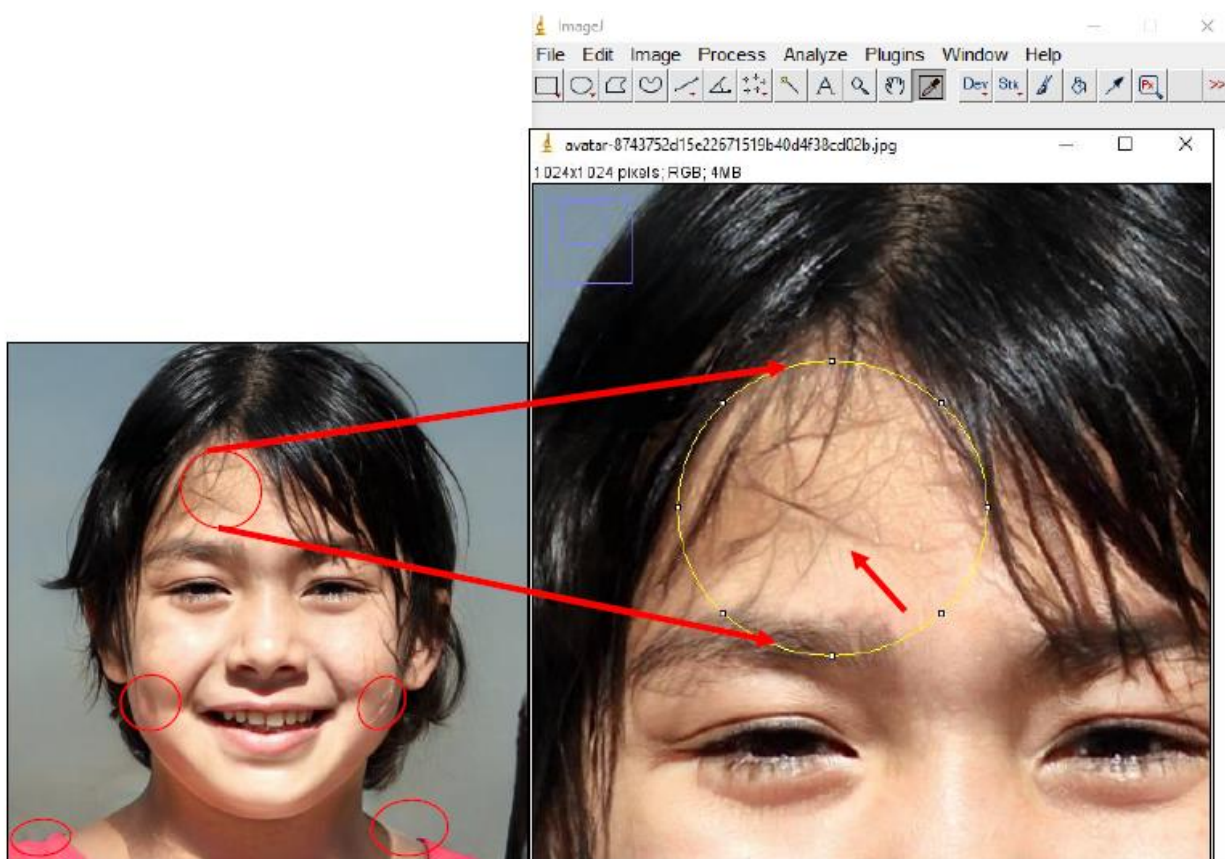


Figura 16 – Análise de Deepfake usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

Na Figura 16 é possível notar que há inconsistências na região das bochechas, bem como distorções no formato dos ombros do indivíduo na imagem. Há ainda uma incoerência na região capilar, próximo à testa, podendo ser percebido um problema de continuidade nesta região.

3.4 ANÁLISE DE HISTOGRAMA NOS CANAIS DE COR RGB EM IMAGENS COM FOTOMONTAGEM

Nesta análise de histograma nos canais de cor RGB, a hipótese inicial levantada é de que as imagens à direita foram adulteradas com fotomontagem. Para sustentar essa hipótese, amostras de pixels na imagem investigada foram coletadas observando os valores de intensidade nos canais de cor vermelho, verde e azul.

Para esta análise foi utilizado a ferramenta de forense de software livre ImageJ® e foram utilizadas duas imagens geradas por fotomontagem. O objetivo desta análise foi comparar os histogramas de cor RGB da imagem original com a imagem gerada e comprovar, por meio de uma ferramenta forense, que a imagem à direita foi adulterada. Nas Figuras 17 e 21 a seguir mostram a região de interesse (ROI, do inglês) no qual a análise foi realizada, marcadas por um retângulo amarelo.

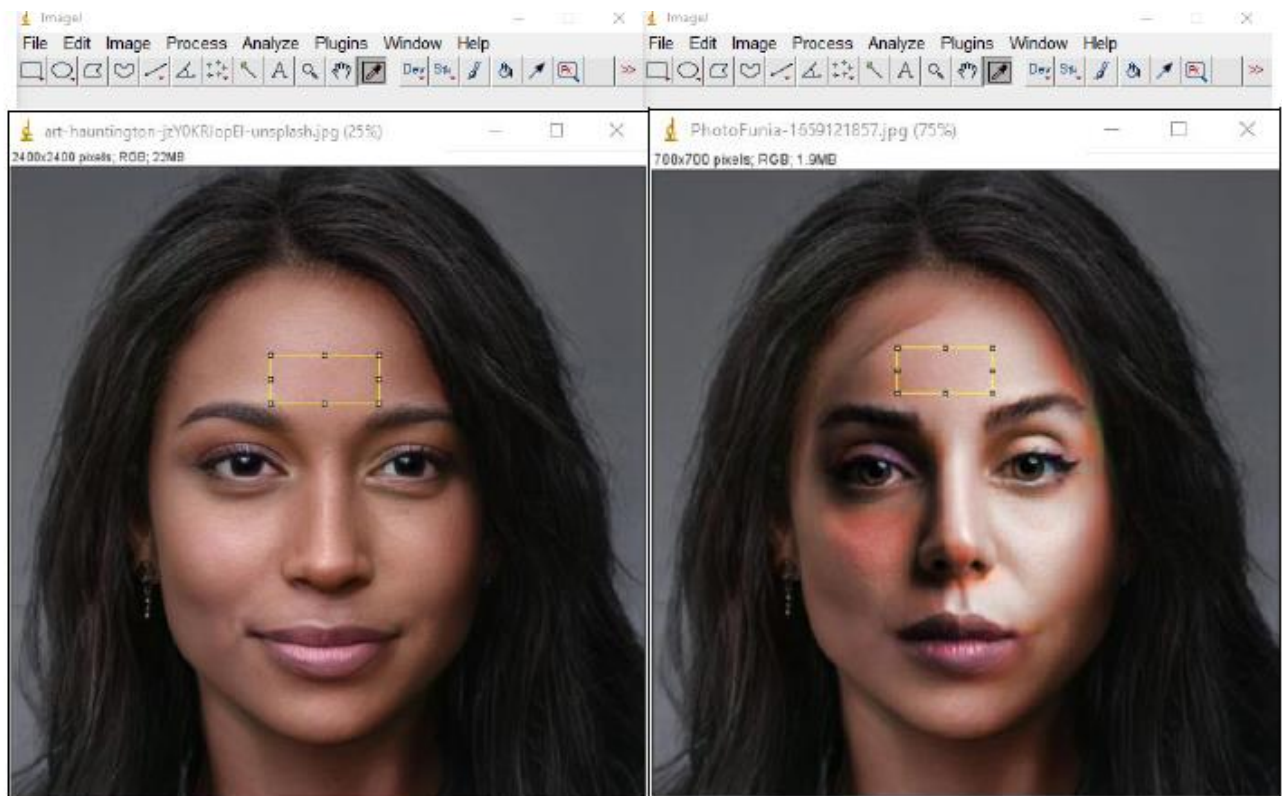


Figura 17 – Análise de fotomontagem usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

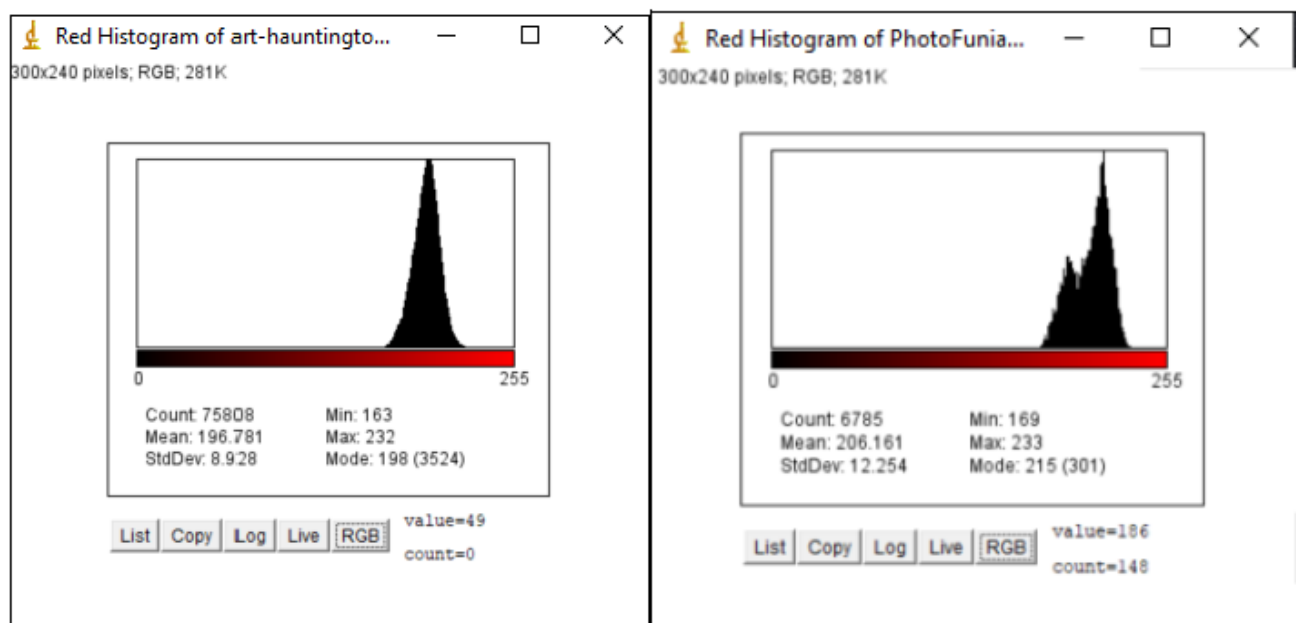


Figura 18 – Análise comparativa de histogramas de canais de cor vermelho usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

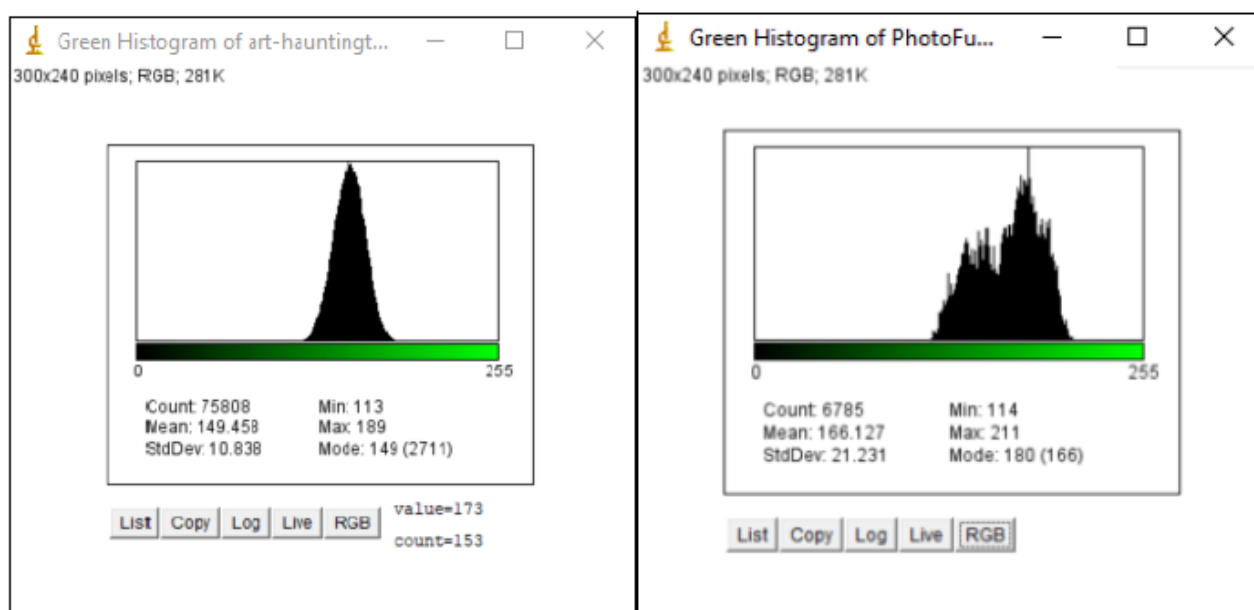


Figura 19 – Análise comparativa de histogramas de canais de cor verde usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

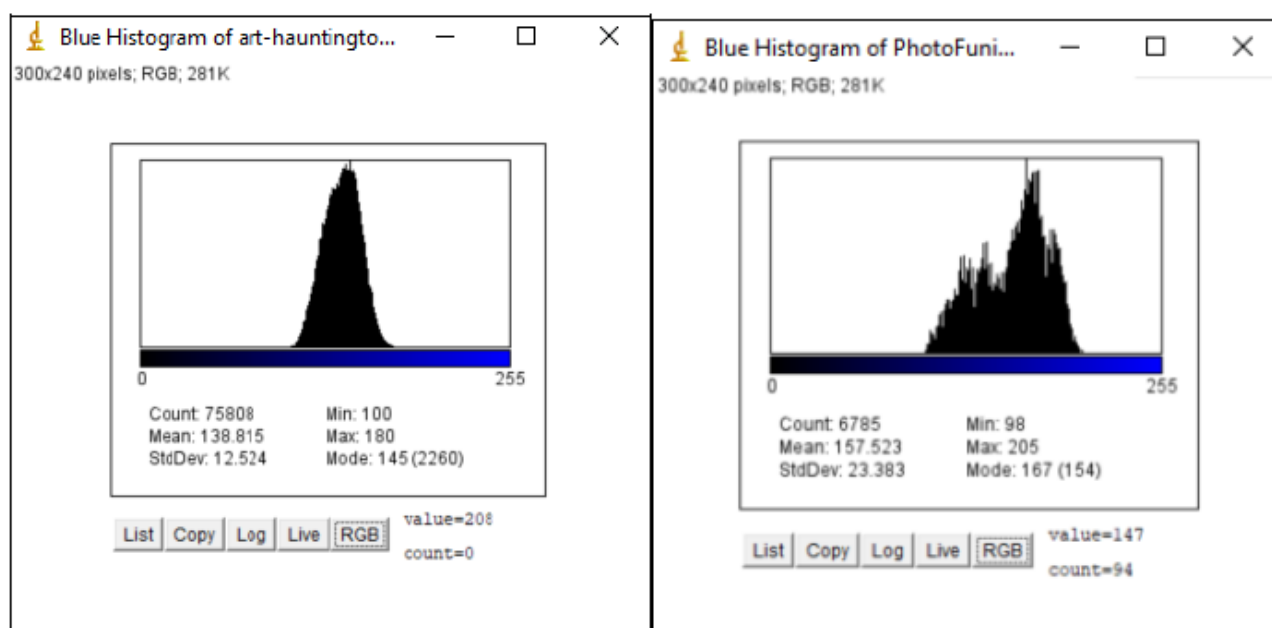


Figura 20 – Análise comparativa de histogramas de canais de cor azul usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

RESULTADOS:

- Valores divergentes de intensidade nos canais RGB na mesma região de interesse ao comparar as imagens originais com as imagens geradas;

- b) Histogramas divergentes nos três canais de cor ao comparar as imagens originais com as imagens geradas.

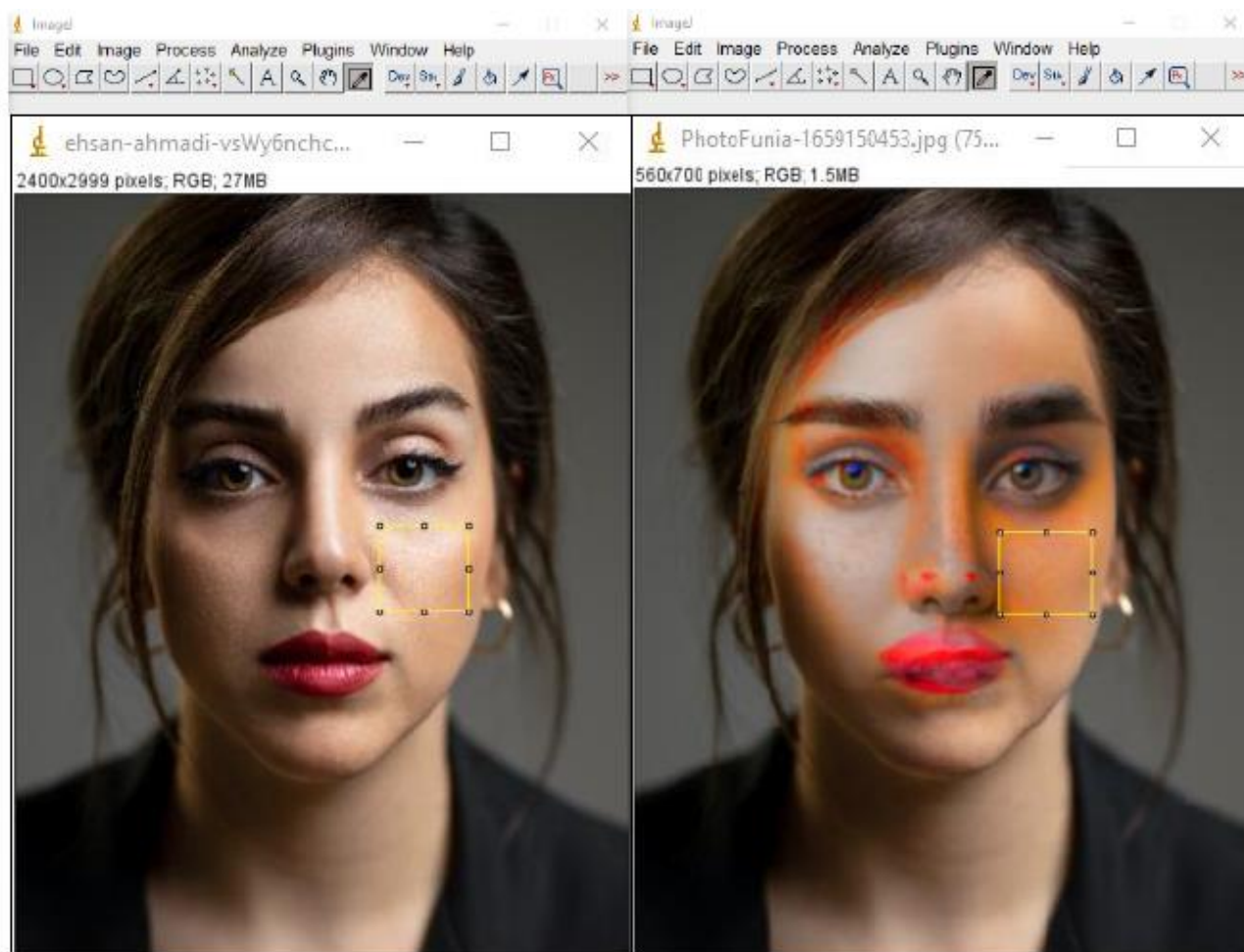


Figura 21 – Análise de fotomontagem usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

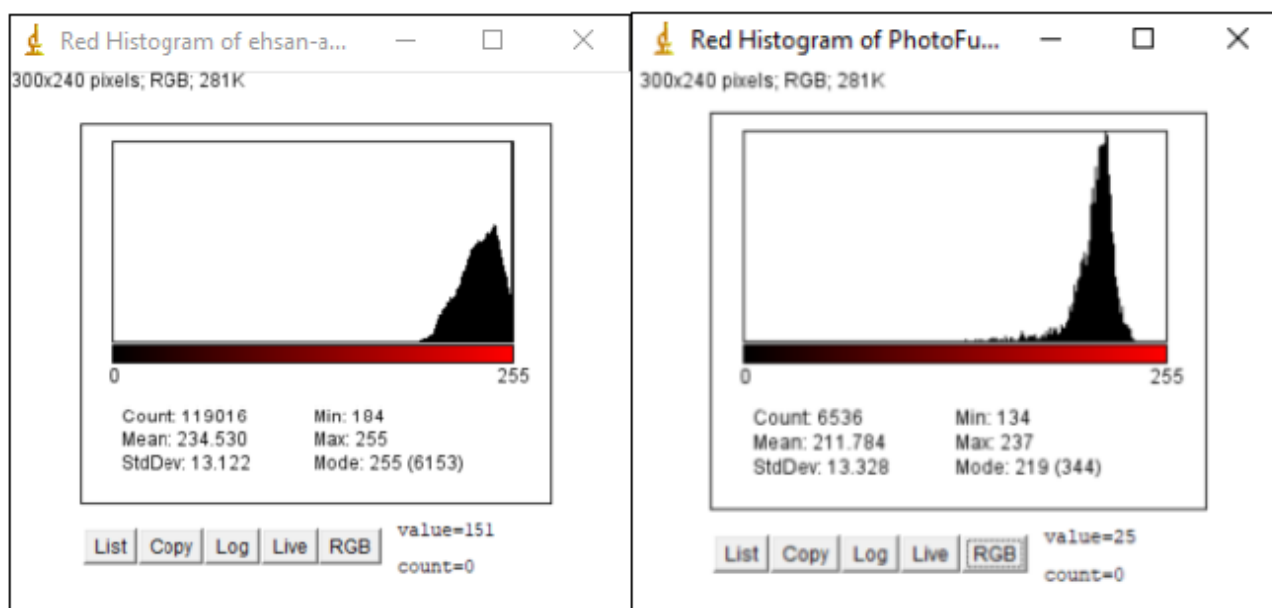


Figura 22 – Análise comparativa de histogramas de canais de cor vermelho usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

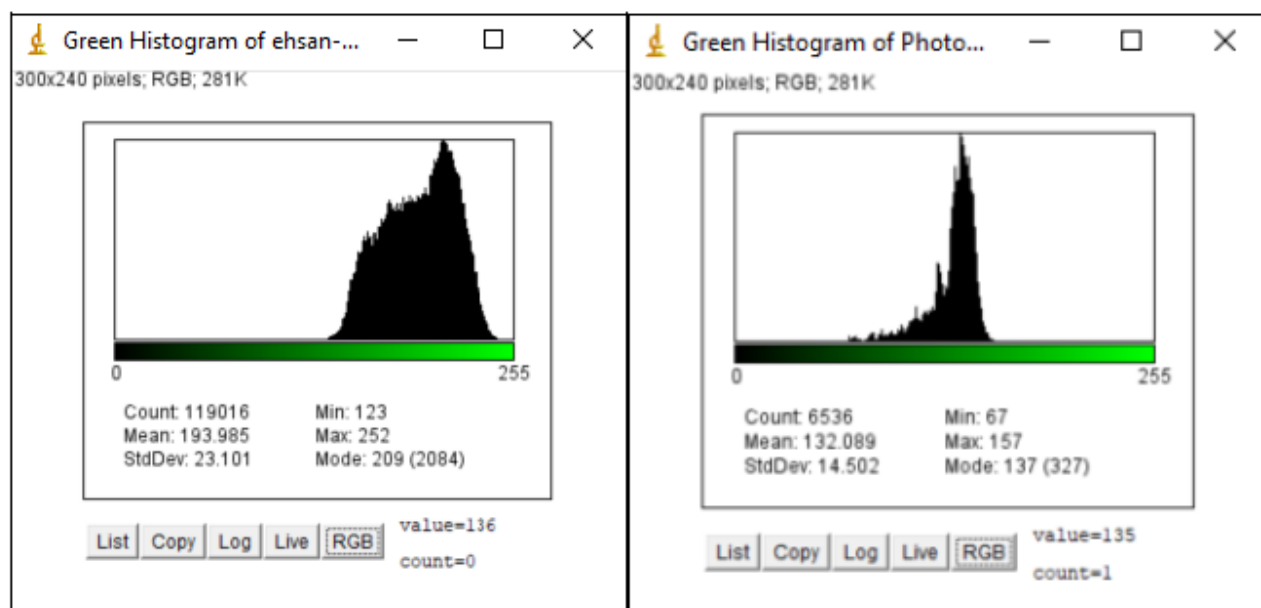


Figura 23 – Análise comparativa de histogramas de canais de cor verde usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

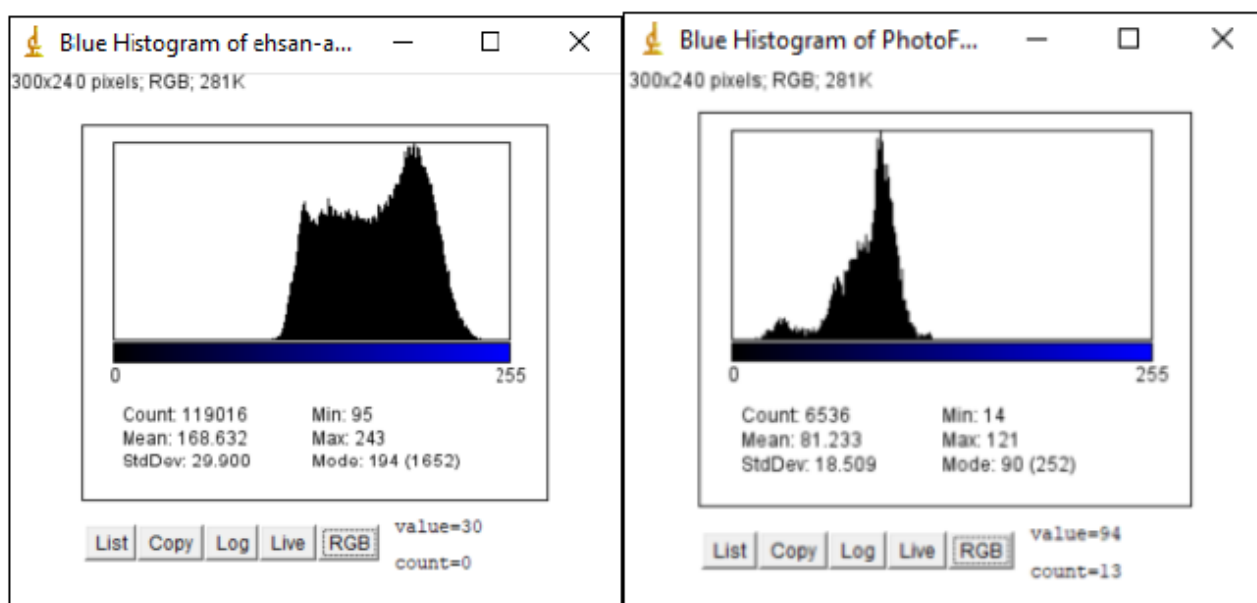


Figura 24 – Análise comparativa de histogramas de canais de cor azul usando o software ImageJ®.

Fonte: Adaptado de ImageJ® (2022)

RESULTADOS:

- Valores divergentes de intensidade nos canais RGB na mesma região de interesse ao comparar as imagens originais com as imagens geradas;
- Histogramas divergentes nos três canais de cor ao comparar as imagens originais com as imagens geradas.

4. CONCLUSÃO

Com base nos resultados extraídos nas análises é possível determinar que as imagens geradas por algoritmos de fotomontagem e aprendizado de máquina profundo, neste estudo, são ilegítimas e que uma ferramenta forense de software livre é capaz de identificar indícios de adulteração nestes tipos de imagens.

Este estudo mostrou resultados promissores com o uso de uma ferramenta forense de software livre para detecção de imagens criadas por *Deepfake*. A técnica de observação direta conjuntamente com a análise de coerência perceptiva e a análise de histograma de canais de cor RGB mostraram-se promissoras em seus resultados, quando comparado a imagem original com a imagem gerada pelo *Deepfake*, sendo possível chegar nas seguintes conclusões:

- a) As imagens geradas por fotomontagem, em sua grande maioria, apresentaram distorções nas formas e nos contornos faciais, sobreposição de imagens causando distorção, e perda de foco na região dos olhos e boca;
- c) As imagens geradas por *Deepfake* apresentaram distorções no formato dos ombros, ou incoerência perceptiva na região capilar, ou incoerência perceptiva na região dos olhos;
- d) Ao comparar as imagens originais com as imagens geradas por fotomontagem, as duas imagens analisadas apresentaram valores divergentes de intensidade nos canais RGB na mesma região de interesse e histogramas divergentes nos três canais de cor ao comparar as imagens originais com as imagens geradas.

No ponto de vista científico, observa-se poucos trabalhos voltados ao tema e um número limitado de ferramentas capazes de identificar adulteração de imagens criadas por algoritmos de aprendizado de máquina profundo. Ademais, na área de perícia forense, os trabalhos com prática em análise de imagens adulteradas digitalmente por *Deepfake* ainda são escassos, dificultando ser realizado uma análise minuciosa comparando os resultados.

Em suma, estabelece-se que as técnicas e abordagens utilizadas neste estudo foram coerentes para a análise forense das evidências digitais. Espera-se que este estudo possa contribuir com as técnicas de análise forense de imagem digitais para detecção de fotomontagem e *Deepfake*.

5. REFERÊNCIAS

- ALTHEIDE, C.; CARVEY, H. DAVIDSON, R. Digital Forensics with Open Source Tools. Elsevier, 2015.
- ARRUDA, Gustavo Henrique Machado de et al. Exames em Imagens Digitais. In: VELHO, Jesus Antonio (Org.). Tratado de Computação Forense. Campinas: Millenium Editora, 2016. cap. 6, p. 208-243.
- BBCNews.DeepfakepresidentesusedinRussiaUkrainewar.Disponívelem:<https://www.bbc.com/news/technology-60780142>. Acesso em 22 jul de 2022.
- FRANCO, Deivison Pinheiro et al. Introdução à Computação Forense. In: VELHO, Jesus Antonio (Org.). Tratado de Computação Forense. Campinas SP: Millenium Editora, 2016. cap. 1, p. 1-15.
- FREITAS, Andrey Rodrigues. Perícia forense aplicada à informática. Rio de Janeiro: Brasport, 2006. 240 p.
- GOKHALE, A. et al. A Bibliometric Analysis of Digital Image Forensics. Science and Technology Libraries, v. 00, n. 00, p. 1–18, 2020.
- HUI, J. Detect AI-generated Images & Deepfakes (Part 1). Disponível em: <https://jonathan-hui.medium.com/detect-ai-generated-images-deepfakes-part-1-b518ed5075f4>. Acesso em: 21 jun 2022.
- NIGHTINGALE, S. J.; WADE, K. A.; WATSON, D. G. Can people identify original and manipulated photos of real-world scenes? Cognitive Research: Principles and Implications, v. 2, n. 1, 2017.
- PARVEEN, A.; KHAN, Z. H.; AHMAD, S. N. Classification and evaluation of digital forensic tools. Telkomnika (Telecommunication Computing Electronics and Control), v. 18, n. 6, p. 3096–3106, 2020.
- PHOTOFunia.Disponívelem:https://photofunia.com/pt/categories/all_effects/face_swap?e=no_faces. Acesso em 20 jul de 2022.
- RANDOM Face Generator (This Person Does Not Exist). Disponível em: <https://this-person-does-not-exist.com/algorithm/en>. Acesso em 19 jul de 2022.
- SHARMA, M. Foto transformada de segunda guerra mundial com novo símbolo tricolor de patriotismo. The Times of India, 2016.
- REIS, S. W. Análise forense de imagens digitais no estudo de caso da ficha criminal de Dilma Rousseff. Revista Especialize On-line IPOG-Goiânia-Ano, v. 9, p. 16, 2018.
- UNSPLASH. Beautiful Free Images. Disponível em: <https://unsplash.com>. Acesso em: 18 jun 2022.

Capítulo 5



10.37423/230908163

GERENCIAMENTO DO PROCESSO DE UTILIZAÇÃO DA FERRAMENTA “LIÇÕES APRENDIDAS” NUMA EMPRESA DO RAMO DE AUTOMAÇÃO AGRÍCOLA

ERIC MARCEL LEMES PISTORI

*Universidade Tecnológica Federal do Paraná
UTFPR - Campus Londrina*

SILVANA RODRIGUES QUINTILHANO

*Universidade Tecnológica Federal do Paraná
UTFPR - Campus Londrina*



Resumo: Ao longo do ciclo de vida de um projeto é possível aprender lições e descobrir oportunidades de aperfeiçoamento. O uso das lições aprendidas é um componente fundamental de uma cultura organizacional desempenhada em potencializar níveis de melhoria contínua. Portanto, o objetivo desse artigo se aplica à metodologia Lições Aprendidas numa empresa de automação agrícola, analisando a utilização do processo das lições para a diminuição de gastos e esforços desnecessários.

Palavras-chave: Gerenciamento de Projeto. Metodologia Lição Aprendida. Melhoria Contínua. Agricultura Digital.

1 INTRODUÇÃO

O avanço da tecnologia no campo nas últimas décadas desencadeou o crescimento exponencial de empresas tecnológicas, as quais geram soluções e ferramentas digitais que beneficiam o setor. Existe uma certa complexidade em aplicar tecnologias avançadas no agronegócio brasileiro, não apenas pela falta de mão de obra qualificada, mas por variáveis que desafiam o processo, correlatado às dificuldades de comunicação em tempo real devido áreas com ausência de sinal para transmissão de dados, sensoriamento, modelagem de lógica computacional, estrutura de equipamentos, treinamentos, gerenciamento dos processos de implantação pelos stakeholders e manutenibilidade do software ou hardware. A gama de intempéries desafia organizações deste segmento, impactando diretamente na implementação dos projetos.

A empresa, objeto desse artigo, atua na melhoria de serviços e produtos, cria ações de melhoria contínua em processos, métodos e estratégias, permitindo assim, o aprimoramento dos projetos. O gerenciamento das lições aprendidas potencializa níveis de excelência e registra atividades nas fases de iniciação, planejamento ou execução dos projetos. Esta prática cria ações e benefícios para a organização em destaque, registrando o que deu certo e, principalmente, o que deu errado. Contudo, implementar a metodologia lições aprendidas requer cautela e embasamento teórico, desde a estruturação do banco de dados até a escolha correta da ferramenta a se utilizar, tendo em vista que a inserção desta metodologia deve atender as necessidades da empresa e o respectivo portfólio de projetos.

Nesse sentido, o objetivo desse artigo será apresentar a metodologia lições aprendidas, visando a aplicabilidade da ferramenta numa empresa do ramo de automação agrícola.

2 GERENCIAMENTO DE PROJETOS

Para ter sucesso na gestão de projetos, somente o uso de boas ferramentas e o conhecimento das melhores práticas não são suficientes para garantir uma elevada qualidade no planejamento e execução de projetos, faz-se necessário seguir uma metodologia adequada à sua equipe e cultura organizacional. O gerenciamento de projetos é a aplicação de conhecimentos, habilidades, ferramentas e técnicas às atividades do projeto a fim de cumprir seus requisitos (Guia PMBOK® Sexta Edição).

De acordo com Vargas (2009, p. 3),

Para atender as demandas de maneira eficaz, em um ambiente caracterizado pela velocidade das mudanças, torna-se indispensável um modelo de gerenciamento baseado no foco de prioridades e objetivos. Por essa razão, o gerenciamento de projetos tem crescido de maneira tão acentuada no mundo nos últimos anos.

Organizações estão compreendendo cada vez mais a importância da estruturação delimitada, com funções e ações bem definidas para o sucesso de um projeto. O estudo de *Benchmarking* em gerenciamento de projetos realizado em 2013 pela PMSURVEY, com amostragem do cenário brasileiro, identifica a necessidade de aperfeiçoamento das organizações quanto a utilização de ferramentas para gerenciamento de projetos.

Tabela 1 – Benchmarking em Gerenciamento de Projetos (amostra Brasil, somente).

Posição	Problema	% de empresas
1	Problemas de comunicação	68,1%
2	Escopo não definido adequadamente	59,6%
3	Não cumprimento dos prazos	54,9%
4	Mudanças de escopo constantes	52,5%
5	Recursos humanos insuficientes	47,5%

Fonte: PMSURVEY, 2013.

Conforme Vargas (2009) o gerenciamento de projetos é um conjunto de ferramentas gerenciais que permite que a empresa desenvolva um conjunto de habilidades, incluindo conhecimento e capacidades individuais, destinados ao controle de eventos não repetitivos, únicos e complexos, dentro de um cenário de tempo, custo e qualidade predeterminados.

2.1 ESCRITÓRIO DE PROJETOS

Define-se Escritório de Gerenciamento de Projetos (EGP) ou *Project Management Office* (PMO) como uma estrutura de gerenciamento, que padroniza os processos de governança relacionados com o projeto e facilita o compartilhamento de recursos, metodologias, ferramentas e técnicas (Guia PMBOK® Sexta Edição). Repõe as melhores práticas em planejamento, estimativa, avaliação dos riscos, rastreamento de habilidades, tempo, manutenção e apoio de métodos e padrões (CRAWFORD, 2006). Logo, o PMO é visto como uma unidade organizacional que centraliza e coordena o gerenciamento de projetos sob seu domínio (PMI, 2014, p. 17).

3 METODOLOGIA LIÇÕES APRENDIDAS

Caracteriza-se projeto todo esforço temporário implementado na criação de um produto, serviço ou resultado único. A compreensão do ciclo de vida de um projeto está diretamente vinculada aos cinco grupos de processos de Gerenciamento de Projetos, delimitados como processo de iniciação, processo de planejamento, processo de execução, processo de monitoramento de controle e processo de encerramento. Ao longo do ciclo de vida de um projeto é possível aprender lições e descobrir oportunidades de aperfeiçoamento. O uso das lições aprendidas é um componente fundamental de uma cultura organizacional desempenhada em potencializar níveis de melhoria contínua.

As lições aprendidas são informações documentadas que regem experiências positivas ou negativas adquiridas no projeto, os métodos garantem que informações valiosas sejam disseminadas em novos planejamentos, processos de trabalho e atividades. Segundo Schindler e Eppler (2003), lições aprendidas podem ser definidas como as experiências chave que contêm certa relevância de negócios para projetos futuros.

Baaz, A. et al. (2010) expõe que as lições aprendidas são um meio de explicitar o conhecimento, desenvolvê-lo, aumentar o compartilhamento do mesmo no projeto e entre projetos, aumentar a satisfação com o trabalho, melhorar a relação entre os participantes e contribuir para o aprendizado.

O registro das lições aprendidas são usadas para melhorar o desempenho do projeto e evitar a repetição de erros. O registro ajuda a identificar onde definir regras ou diretrizes para alinhar as seções das equipes (Guia PMBOK® Sexta Edição).

3.1 PROCESSO DE LIÇÕES APRENDIDAS

A maioria das organizações sabem da importância de capturar lições aprendidas, contudo, esta prática atinge diferentes níveis de utilização no mercado. Alguns comportamentos são marcados pelo fato de organizações não dispor de um processo de captura definido, ou então, por possuir uma coleta regular dentre as fases do projeto e ainda assim não proporcionar métricas tangíveis para absorção de conhecimento e melhoria contínua. O objetivo de um processo de lições aprendidas é definir as atividades necessárias para capturar e usar com êxito as lições. De acordo com Schindler e Eppler (2003), o processo de lições aprendidas deve ser feito de forma sistemática, permitindo que a organização compare seus distintos projetos metodicamente, documentando o seu mecanismo de forma eficaz para a resolução de problemas.

Os processos usados para coletar, compartilhar e disseminar as lições aprendidas podem sofrer variações, mas em geral é composto por cinco elementos principais, sendo eles a definição do projeto, coleta de informações, verificação, aplicabilidade, armazenamento e disseminação. Na Figura 1 está ilustrado o processo básico de lições aprendidas.

Figura 1 - Processo básico de lições aprendidas



Fonte – Adaptado de White e Cohan, 2005.

3.2 COMO ARQUITETAR O PROCESSO DE LIÇÕES APRENDIDAS

A primeira etapa do processo de lições aprendidas está compreendida em identificar a equipe envolvida, quais serão as lições coletadas e de que forma deverá ser estabelecida. Peszynski et al (2008) ressalta o comprometimento organizacional como fator extremamente importante para o sucesso deste esforço. Deve-se destacar que ao estabelecer os objetivos específicos, as partes interessadas e o público alvo do produto passam a empregar a eficiência e integração das lições, sua linguagem de representação deve proporcionar meios rigorosos quanto ao vocabulário utilizado, analisando os modelos ou raciocinando diretamente sobre eles (SCHREIBER, 2000). Logo, a equipe deve estipular o template que o produto será exposto e sua duração de apresentação, a coleta e análise de dados, bem como pesquisas e workshops. Esses fatores são imprescindíveis para a estratégia de divulgação do fluxo de atividades e resultados do processo.

A segunda etapa do processo compõe-se da coleta, a captura de informações através de estruturas delineadas, com formulários escritos e reuniões.

A estruturação e obtenção das informações da base histórica do projeto pode ser provido da fase final do projeto, ou ao longo do ciclo de vida do projeto. Schindler e Eppler (2003) apontam vantagens a uma revisão periódica durante o progresso do projeto, os eventos são mais recentes e os aprendizados subsequentes podem ser recordados facilmente. Huemann e Hanbari (2007) apontam que as lições

aprendidas são uma investigação sistemática sobre o mérito da gestão técnica, de processos e critérios de desempenho. Ela ajuda a identificar as causas raiz de sucesso ou fracasso e destaca melhorias e oportunidades.

Verificar e sintetizar, esta etapa do processo verifica a precisão e aplicabilidade das lições coletadas. O grau de conhecimento dos especialistas ou coordenadores envolvidos caracterizam a importância das lições, o efeito ou relevância para um projeto em específico, coletivo e impactante aos demais setores ou à organização como um todo.

Denomina-se Loja a etapa do processo de armazenagem das lições em um repositório. Na grande maioria das organizações utiliza-se um compartilhamento mútuo de documentações e projetos, empregando pastas compartilhadas entre os envolvidos do projeto, esta ação dificulta o acesso eficiente e assertivo da documentação Lição Aprendida, portanto deve-se criar um banco de dados específico para a disseminação destas informações, facilitando o acesso ao banco de dados e proporcionando a identificação por meio de pesquisa com palavra-chave. As palavras-chave são, em última análise, um dos fatores determinantes para o sucesso de utilização das lições aprendidas (PRICHARD,1997 apud ROWE; SIKES, 2006).

O último elemento do processo e mais importante é a disseminação das lições aprendidas, entretanto pode-se ter pouco valor agregado se não ocorrer uma distribuição e utilização adequada por pessoas que se beneficiarão delas. Conforme White e Cohan, (2005) a divulgação pode incluir a revisão de um processo de trabalho, treinamento ou rotina de difusão através dos meios de comunicação definidos.

3.3 COLETA DAS LIÇÕES APRENDIDAS

Nesta seção apresenta-se duas abordagens diferentes para capturar as lições aprendidas, todavia cada organização deve atribuir o formato que melhor se enquadra no perfil de seus projetos. As abordagens são brevemente detalhadas nas seções a seguir.

3.3.1 AVALIAÇÃO APÓS AÇÃO

Criada pelo exército americano em 1993 esta abordagem incorpora lições quando um evento é relevante para as atividades pertinentes ao projeto, ou quando é estabelecido a resolução de um problema. Seu objetivo é coletar erros e acertos de forma rápida, portanto quaisquer ações que se enquadrem nestas condições devem ser imediatamente inseridas na base de dados. Piscopo (2014) complementa que a avaliação após ação tem sido proposta na literatura em diferentes formatos,

desde reuniões de brainstorming de 20 minutos até uma sessão de discussão de duas horas. Toda a equipe do projeto mais as partes interessadas que possuem participação no evento devem participar da reunião, logo deve-se definir um moderador para coordenar a condução da pauta.

Schindler e Eppler (2003) complementam que para este método delimita-se quatro questões principais:

- O que era esperado acontecer?
- O que realmente aconteceu?
- Por que existiram diferenças?
- O que você pode aprender desta experiência?

Após as perguntas serem respondidas pelos participantes, faz-se necessário definir uma medida de conduta para projetos futuros, documentando a ação e o procedimento adotado.

3.3.2 HISTÓRIAS DE APRENDIZAGEM

Proposto por Roth e Kleiner em 1998, este método captura as lições aprendidas após o término do projeto. Segundo Piscopo (2014) uma história consiste dos principais eventos de um projeto ordenados cronologicamente. O comprimento do documento resultante pode estar entre 20 e 100 páginas, seu conteúdo deve apresentar o registro de experiências relevantes, proporcionando um contexto estruturado e rico de históricos e lições vivenciadas no ciclo de vida do projeto.

As histórias de aprendizagem se caracterizam inicialmente por três perguntas:

- Como podemos julgar essa experiência?
- Como o resto da organização pode se beneficiar desta experiência?
- Quais foram as ações que determinaram nossa falha ou sucesso?

Dentre a estruturação dos eventos no template das histórias, observa-se a citação dos participantes da entrevista na coluna da direita, os historiadores do projeto que fornecem mais contexto ou observações estão caracterizados na coluna da esquerda, eles são os detentores do conhecimento tácito e interligam as declarações. Entre as colunas são estabelecidos comentários adicionais que delatam detalhes relevantes do projeto (Piscopo, 2014).

De acordo com Piscopo (2014, p.74) os entrevistados são citados de forma anônima, apenas identificados para o leitor através de seus cargos (por exemplo, “gerente de projeto”), o que torna

mais fácil admitir erros. Uma vez compiladas, as histórias de aprendizagem são validadas em discussões com os envolvidos.

Este método firma um valioso processo de disseminação das lições pelos workshops, agrega mecanismos que permitem à organização o conhecimento de suas principais fraquezas e possibilita discussões assertivas para novas estratégias.

4 APLICAÇÃO DA METODOLOGIA LIÇÕES APRENDIDAS NO SETOR DE COSTUMER SUCESS DE UMA EMPRESA DO SETOR AGRÍCOLA

Líder em agricultura digital a empresa em destaque transforma dados em aliados de eficiência, desenvolve e produz equipamentos, tecnologias e sistemas informáticos para a automatização de ativos móveis e processos de logística. Entre os produtos e tecnologias que a empresa dispõe ao agronegócio mundial estão computadores de bordo, sistemas de telemetria, softwares para a gestão e controle de ativos móveis, aplicativos, sensores especiais, dentre outros. A organização se destaca pela contínua busca de “soluções”, cujo sucesso alcançado no mercado é devido ao foco empresarial voltado a “solucionar os problemas de seus clientes”, da forma mais integral possível, combinando as mais avançadas tecnologias com o profundo conhecimento sobre o processo produtivo e logístico do agronegócio. Os índices de comercialização contemplam mais de 3.500 computadores de bordo por ano para o setor sucroenergético, com um portfólio de 30 mil equipamentos sendo monitorados com suas soluções, atingindo assim, uma expansão territorial de 6 milhões de hectares de área de cobertura.

4.1 COSTUMER SUCESS (SUCESSO DO CLIENTE)

Para que a empresa consiga entregar com qualidade as soluções descritas na seção 4, faz-se necessário manter estruturas e equipes de alta performance, fundamentadas de conhecimentos técnicos, de Gerenciamento de Projetos e de fluxos bem definidos pelo Escritório de Projetos (EP). Os processos de iniciação, planejamento, execução, controle e encerramento se interligam para assegurar a base histórica dos projetos. O setor responsável por executar e implementar as soluções desenvolvidas pela organização compete a equipe de serviços e projetos, bem-apresentada como Customer Success (Sucesso do Cliente).

4.2 INVESTIGAÇÃO DE CENÁRIO

O crescimento exponencial da empresa e a constante demanda de solicitações e liberações de novos desenvolvimentos proporcionaram competências em análises de erros e assertos pelas partes interessadas, despertando olhares acurados na direção dos desvios de prazo, retrabalho e custos de operação. Uma empresa que lidera o mercado e disponibiliza tecnologia e soluções agrícola, se coloca diretamente frente aos erros, entretanto compreende os valores positivos e negativos gerados, principalmente aqueles que influenciam na evolução do produto e serviço disponibilizado pela empresa. A partir de uma análise qualitativa de aprimoramento, ficou compreendido o potencial de implementação dos métodos de gerenciamento das lições aprendidas.

O primeiro levantamento realizado envolveu uma análise na base histórica dos projetos e soluções implantadas no setor sucroenergético, contemplando a faixa de ano entre 2018 e 2019 pela organização. Na Tabela 2 está indicado a quantidade de projetos dotados do documento “Base Histórica do Projeto” e o total de implantações correspondentes ao ano.

Tabela 2 – Quantidade de Base Histórica dos Projetos x Quantidade de implantações (sucroenergético).

Quantidade total de projetos implantados no ano de 2018	Quantidade total de projetos contendo a Base Histórica no ano de 2018	Quantidade total de projetos implantados no ano de 2019	Quantidade total de projetos contendo a Base Histórica no ano de 2019
44	16	49	21

Fonte: Autoria própria

Observa-se que a coluna correspondente a quantidade de projetos implantados em 2018 e 2019 não é representada de forma proporcional quanto a quantidade de documentos elaborados na Base Histórica dos Projetos, dado perceptível pois o banco de dados não está definido e dedicado somente para este processo de análise. Importante destacar que esta metodologia não se aplica com vigor na organização.

O segundo tópico analisado refere-se aos maiores índices de recorrência encontrados nas ocorrências apreciadas nas implementações de 2018 e que novamente foram notadas no ciclo de vida dos projetos de 2019. Na Tabela 3 é apresentado o histórico de ocorrências por palavras-chave, relevantes para os projetos de 2019 e a Tabela 4 faz-se referência aos projetos de 2018.

Tabela 3 – Fatores de risco e impactos informados no ciclo de vida dos projetos no ano de 2019.

FIRMWARE	INSTALAÇÃO	FERRAMENTA DE TRABALHO CLIENTE	FERRAMENTAS DO AMBIENTE WEB	IHM APLICATIVO - INTERFACE HOMEM MÁQUINA	CHAMADOS - TI	RELATÓRIOS WEB	SENSOR/TAG BASCULAMENTO	FALTA DE MATERIAL
6	2	5	7	4	5	7	2	3

Fonte: Autoria própria.

Tabela 4 – Fatores de risco e impactos informados no ciclo de vida dos projetos no ano de 2018

FIRMWARE	INSTALAÇÃO	FERRAMENTA DE TRABALHO CLIENTE	FERRAMENTAS DO AMBIENTE WEB	IHM APLICATIVO - INTERFACE HOMEM MÁQUINA	CHAMADOS - TI	RELATÓRIOS WEB	SENSOR/TAG BASCULAMENTO	FALTA DE MATERIAL
5	4	2	3	4	3	2	1	3

Fonte: Autoria própria.

5 RESULTADOS E DISCUSSÃO

Os processos metodológicos da aplicação das Lições Aprendidas e as literaturas apresentadas construíram uma visão mais detalhada do cenário real da organização em destaque, a partir dos conhecimentos teóricos adquiridos observou-se reais necessidades do enquadramento sublime nas aplicações das Lições Aprendidas no cotidiano da companhia.

Ao longo das coletas históricas do ciclo de vida dos projetos (2018 e 2019) evidenciou-se a necessidade de estabelecer uma metodologia sólida e regulamentada, fatores estes influenciados principalmente pela permanência de eventos contínuos por aproximadamente 2 anos. A necessidade de ações rápidas para a melhoria do produto, serviço e principalmente de custos e despesas, promoveu o embasamento adequado para a implementação da avaliação após ação (literatura apresentada na seção 3.3.1) no fluxo do processo organizacional da empresa. É clara a necessidade de uma resposta rápida para a captura das ações diagnosticadas no percurso das implantações. As coletas deverão proporcionar um banco de dados dedicado, capaz de promover lições e históricos estruturados de forma assertiva, potencializada com ações imediatas para padronização de acertos e melhoria de erros.

Visando diminuir gastos e esforços desnecessários no atual mercado, este artigo irá potencializar a inserção da metodologia lições aprendidas na companhia, de modo a respeitar o perfil da organização, atendendo principalmente o exigente e concorrido mercado de soluções agrícolas.

6. CONCLUSÃO

Diante da sistematização das Lições Aprendidas, conclui-se que, quando bem monitorado os “erros” dentro de um projeto tornam-se parâmetros de condutas para projetos posteriores, contribuindo para a qualidade dos processos, bem como a melhoria dos serviços organizacionais. Evidentemente, a assertividade na resolução dos problemas, torna-se um diferencial de mercado.

REFERÊNCIAS

- BAAZ, A.; AB, E.; Holmberg, L.; Sandberg, A. B.; AB, E. Lessons learned Appreciating Lessons Learned. IEEE Software 27(4): 72-79. 2010.
- HUEMANN, M., ANBARI, F.T., Project auditing: a tool for compliance, governance, empowerment, and improvement. Journal of Academy of Business and Economics (in press), 2007.
- PESZYNSKI, K., Cooper, V. and Molla, A., Developing a Knowledge Management Strategy: Reflections from an Action Research Project. RMIT University, GPO Box 2476V, Melbourne, VIC, 3001, Austrália, 2008.
- PISCOPO, Marcos Roberto. Métodos para captura de lições aprendidas: em direção a melhoria contínua na gestão de projetos. Revista de Gestão e Projetos: GeP, Z, v. 5, n. 1, p. 71-83, jan/abr. 2014.
- PROJECT MANAGEMENT INSTITUTE. Um Guia do Conjunto de Conhecimentos em Gerenciamentos de Projetos: Guia PMBOK. Pennsylvania. 4. ed. Global Standard, 2008.
- ROWE, S. F. & SIKES, S. (2006). Lessons learned: taking it to the next level. Paper presented at PMI® Global Congress 2006—North America, Seattle, WA. Newtown Square, PA: Project Management Institute.
- SCHINDLER, M., & EPPLER, M. J. (2003). Harvesting project knowledge: a review of project learning methods and success factors. International Journal of Project Management, 21(3), 219–228.
- SCHNEIDER, K. (2000). LIDs: a light-weight approach to experience elicitation and reuse. In Product Focused Software Process Improvement (pp. 407- 424). Springer Berlin Heidelberg
- SCHREIBER, G., Akkermans H., Anjewierden A., Hoog R.D., Shadbolt N., Velde W.V.D., Wielinga B., Knowledge Engineering and Management, The MIT Press, Cambridge, 2000.
- VARGAS, Ricardo Viana. Gerenciamento de Projetos: estabelecendo diferenciais competitivos. 7. ed. Rio de Janeiro: Brasport, 2009. 276 p.
- WHITE, M., COHAN, A. A guide to capturing lessons learned. Conservation Gateway - The Nature Conservancy. 2010.

SESSÃO TEMÁTICA

- ☐ Educação na Engenharia de Produção
- ☒ Gestão da Produção
- ☐ Pesquisa Operacional
- ☐ Interdisciplinar

Capítulo 6



10.37423/230908214

A IMPORTÂNCIA DA CULTURA DE PROTEÇÃO DE DADOS COMO MECANISMO DE FIDELIZAÇÃO DO CLIENTE COM A MARCA DA EMPRESA E SEUS BENEFÍCIOS

Fernanda Cabral Pedreira

*Pós Graduada do Curso de DIGITAL
BUSINESS - 2022 do MBA USP/ESALQ*

Lisleandra Machado

*IFSUDESTEMG-Instituto Federal de
Educação Ciência e Tecnologia do Sudeste
de Minas Gerais
UNEMAT - Universidade do Estado do Mato
Grosso*

Domingos Sávio da Cunha Garcia

Leonardo Amorim de Araújo

*IFSUDESTEMG-Instituto Federal de
Educação Ciência e Tecnologia do Sudeste
de Minas Gerais*

Carlos Arthur Alevato Leal

*IFSUDESTEMG-Instituto Federal de
Educação Ciência e Tecnologia do Sudeste
de Minas Gerais*

Arthur Nascimento Assunção

*IFSUDESTEMG-Instituto Federal de
Educação Ciência e Tecnologia do Sudeste
de Minas Gerais*

Lucas Mendes Ferreira

*IFSUDESTEMG-Instituto Federal de
Educação Ciência e Tecnologia do Sudeste
de Minas Gerais*

Samuel Alves de Freitas

*IFSUDESTEMG-Instituto Federal de
Educação Ciência e Tecnologia do Sudeste
de Minas Gerais*

Gustavo José Santiago Rosseti

*IFSUDESTEMG-Instituto Federal de
Educação Ciência e Tecnologia do Sudeste
de Minas Gerais*

Silvana Rodrigues Pires Moreira

*IFSUDESTEMG-Instituto Federal de
Educação Ciência e Tecnologia do Sudeste
de Minas Gerais*

Resumo: A Lei Geral de Proteção de Dados (Lei nº 13.709/18 - LGPD) buscou alinhar o Brasil ao *standard* mundial sobre proteção de dados, trazendo para nossa realidade social uma mudança de perspectiva acerca da forma de pensar e agir quando falamos sobre dados pessoais. Este trabalho busca trazer uma breve contextualização a respeito da utilização de dados na sociedade atual em paralelo com a existência da LGPD, demonstrando empiricamente a relevância da matéria para a manutenção da relação das empresas com potenciais consumidores, e como o desenvolvimento da cultura de proteção de dados dentro da organização pode influenciar diretamente na relação consumidor/empresa como mecanismo de fidelização do cliente com a marca. Desse modo, foram observados o conhecimento de potenciais consumidores sobre a LGPD e a importância dada à proteção de seus dados pessoais. Buscou-se ainda entender o comportamento destes diante de casos práticos, e os benefícios que as empresas podem agregar quando da aplicação da cultura de proteção de dados e da implementação do programa de privacidade na sua organização.

Palavras-chave: LGPD; *compliance* digital; cultura.

INTRODUÇÃO

Não é preciso ser um especialista na área de privacidade para compreender a importância que o tema proteção de dados tem nos dias de hoje. A preocupação com a privacidade e a proteção de dados não nasceu em 2018 com a Lei Geral de Proteção de Dados (Lei nº 13.709/18), a famosa LGPD, mas, foi se desenvolvendo ao longo dos anos acompanhada pelas evoluções sociais e tecnológicas.

Em termos filosóficos, as raízes da privacidade foram fincadas na antiga Grécia, através de Aristóteles, com a distinção entre esfera pública e privada/doméstica, *polis e oikos* (Stanford Encyclopedia of Philosophy, 2022). Até que em 1890, com a publicação, do “*The right to privacy*”, na Harvard Law Review (BRANDEIS, L. D., WARREN, S. D., 1890), estabeleceram um ponto de partida emblemático para o desenvolvimento do termo privacidade.

Durante esse lapso temporal muitas mudanças aconteceram, e, ao longo do tempo, em termos doutrinários, o conceito de privacidade atrelado ao da proteção de dados foi se aperfeiçoando (Stanford Encyclopedia of Philosophy, 2022).

A evolução da informática e da telecomunicação, desde o início da década de 70, atrelado ao advento dos computadores e da formação de blocos econômicos, possibilitou o aumento expressivo do processamento de dados e do conjunto de dados utilizados (hoje conhecido como Big Data) (Gartner Glossary, 2022), em especial dos dados pessoais (aquelas informações capazes de identificar alguém, ou tornar alguém identificável) (BRASIL, 2018), atrelado ao compartilhamento de dados pessoais em larga escala (decorrência do IoT) (CLARK, 2016), inclusive em níveis internacionais (VAINZOF, 2021).

Esse período marcado pelo trinômio do Big Data, Internet das Coisas e Inteligência Artificial, permitiu que a execução de ações automatizadas trouxesse reflexos para diversos setores da economia. Esse cenário proporcionou significativo aumento na produtividade e eficiência de empresas e governos, aumentando a preocupação com a proteção de dados e privacidade dos indivíduos. (VAINZOF, 2021).

Diante disso, leis vinculadas ao tratamento de dados pessoais e à proteção da privacidade foram criadas, em busca do equilíbrio entre a proteção das liberdades individuais e a preservação de um fluxo de dados pessoais capaz de sustentar o livre comércio internacional. Nesse sentido, a evolução regulatória histórica em torno da questão revelou uma clara utilização de um juízo de ponderação, onde parâmetros como a necessidade e proporcionalidade foram utilizados para modelar e contrabalancear o interesse privado e o interesse público. (VAINZOF, 2021).

Toda a lógica por trás desse mecanismo de proteção envolve dar ao titular os meios para sua autodeterminação informativa: o poder de controle de seus próprios dados. (VAINZOF, 2021).

A conexão entre o consumidor e a proteção de dados se torna cada dia mais forte em uma economia da informação, na medida em que as empresas buscam ao máximo a personalização dos seus produtos e serviços de acordo com o seu consumidor. Nas novas relações de consumo, os dados pessoais atrelados às novas tecnologias se transformam em recursos essenciais e valiosos, tanto para a redução de riscos empresariais, como, principalmente para a fidelização do consumidor. (BRASIL, 2010). Mas, para isso, é necessário entender não apenas sobre a importância da implementação da cultura de proteção de dados dentro das empresas, mas como isso representa uma capitalização de recursos em decorrência da fidelização do consumidor à marca da empresa.

Esse tipo de cultura não cresce magicamente dos valores de uma companhia, mas decorre dos *stakeholders* e seus líderes empresariais pensarem a cultura de forma mais holística. É preciso entender que a cultura pode mudar fundamentalmente a maneira como uma empresa toma decisões, de uma maneira que sempre resulte em melhores resultados para o negócio. (LINS, 2021).

MATERIAL E MÉTODOS

Esse trabalho se caracteriza como de natureza aplicada. Tal natureza é responsável pela promoção de conhecimentos através da obtenção de resultados com reflexos práticos para a aplicação destes à solução de problemas ou objetivos específicos. Ainda, essa natureza decorre da relevância de objetivos comerciais para o desenvolvimento de novos processos ou produtos orientados para as necessidades do mercado em razão do desenvolvimento da cultura de proteção de dados. (TURRIONI, J. B.; MELLO, C. H. P., 2012).

Considerando o objetivo da pesquisa exploratória e explicativa, se buscou entender a percepção dos potenciais consumidores sobre a cultura de proteção de dados, a relevância para estes acerca do tema, além da demonstração da cultura de proteção de dados como diferencial competitivo para fidelização do cliente com a marca e a apresentação dos benefícios da promoção da cultura de proteção de dados para o negócio. Ainda, considerando a relevância atual do tema, se buscou proporcionar maior familiaridade com o problema visando torná-lo explícito e construir hipóteses de causa e consequência, analisando as vantagens auferidas pelo investimento na matéria em questão. Neste ponto foram realizados levantamentos bibliográficos; entrevistas com pessoas; e análise de exemplos. A respeito do objetivo explicativo, se procurou identificar os fatores que contribuem para a ocorrência

do fenômeno de fidelização do cliente, em razão do investimento na cultura de proteção de dados. (TURRIONI, J. B.; MELLO, C. H. P., 2012).

A pesquisa foi desenvolvida por meio da abordagem combinada, se utilizando ora de forma quantitativa, ora qualitativa, através da técnica de questionário para a coleta de dados de forma objetiva, pela análise percentual, e subjetiva, por meio da disponibilização de resposta livre.

A abordagem quantitativa nesta pesquisa objetivou entender o conhecimento do consumidor a respeito das normas e diretrizes de proteção de dados, a relevância dada à matéria e a tendência do consumidor diante de casos práticos. Do ponto de vista qualitativo, visou conhecer quais possíveis dificuldades práticas eram enfrentadas pelo consumidor quanto ao acesso às informações apresentadas no corpo da pesquisa. Ambos feitos por meio da técnica de questionário e pesquisa documental.

A pesquisa documental analisou as principais referências encontradas sobre o tema levantando, reunindo e sintetizando os resultados de diversos estudos primários. Foram utilizados livros, sítios na internet, revistas e periódicos científicos (nacionais e internacionais), permitindo a elaboração da contextualização, problematização e validação do quadro teórico utilizado na pesquisa.

Foi avaliado o conhecimento e a opinião de possíveis consumidores quanto a cultura de proteção de dados e seus reflexos para fidelização do cliente com a marca da empresa, foi aplicado um questionário.

Na execução desta etapa, para construção do questionário, aplicação e posterior coleta dos resultados foi utilizada a ferramenta do *Google Forms*, plataforma promovida pela Google Inc. (*Mountain View, California*, Estados Unidos), de aplicação gratuita para elaboração de formulários online. Sua utilização justifica-se na flexibilidade de acesso, facilidade no gerenciamento de dados, praticidade na coleta de dados, além de ser amplamente utilizado para prática acadêmica e pedagógica quando é necessário avaliar com públicos diversos. Foram utilizadas respostas de sim ou não, respostas de graduação de relevância de 1 a 4, múltiplas escolhas e respostas escritas.

Buscou-se a resposta de forma indistinta por toda e qualquer pessoa que se interessasse por responder o questionário, formando um universo de 205 pessoas.

Foram realizados 9 (nove) questionamentos com objetivo principal de explorar, após consolidação de todas as respostas ao questionário, a abordagem quantitativa e qualitativa para avaliação que busca

a utilização de medidas e resultados quantificáveis para que as informações tenham tratamento estatístico posterior via operacionalização das análises. (DA COSTA, M.A.; PAZ, A.M.O, 2019).

RESULTADOS E DISCUSSÃO

Para a etapa da coleta de dados, foi enviado um questionário virtual (Google Forms) para pessoas/profissionais/empresas que atuam diretamente e são impactadas pela legislação, as quais responderam entre os dias 15/08/2022 e 19/08/2022. Tais pessoas/profissionais/empresas tiveram os seus perfis analisados a partir de seu histórico e da filtragem de respostas submetidas via formulário. Foram obtidas duzentas e cinco respostas.

Em razão do tema em questão a ser analisado, considerando que qualquer pessoa pode se tornar cliente de uma marca, o universo avaliado buscou alcançar a maior variedade de pessoas dispostas a responder o questionário. Dentro desse universo foram avaliados perfis distintos de consumidores como: estudantes, profissionais em cargos de alta e baixa gestão, desempregados, entre outros.

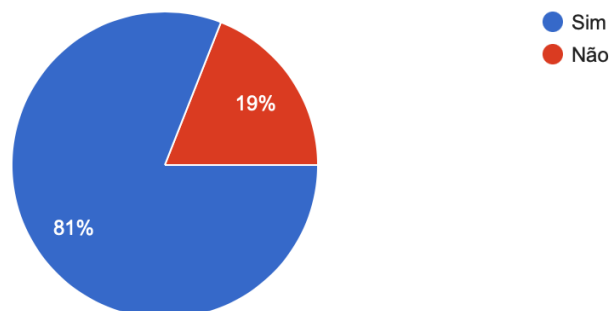
Inicialmente, a pesquisa buscou entender o nível de maturidade de possíveis consumidores à respeito do conhecimento acerca da legislação brasileira que trata sobre a proteção de dados pessoais. Após, foram introduzidos questionamentos para o levantamento da importância dada ao tema e o quanto esse universo se interessa pelos aspectos de proteção de seus próprios dados pessoais. Para isso, foram realizados três questionamentos iniciais: (i) o primeiro buscando entender qual o conhecimento prévio daquele universo à respeito da Lei Geral de Proteção de Dados (LGPD); (ii) o segundo através da análise da importância sobre a proteção dada aos seus dados pessoais e sensíveis tratados pelas organizações que a amostra tinha algum tipo de vínculo (seja como consumidor, fornecedor, colaborador, etc.), divididos entre: insignificante, pouco importante, importante e muito importante; e (iii) o terceiro no que tange ao interesse da amostra sobre o tema da proteção de seus dados.

Considerando a amostragem, para o primeiro ponto levantado foi observado que 81% do rol da amostra, que representam potenciais consumidores, já tinham ouvido falar na LGPD, contra 19% que não tinham conhecimento da legislação.

Figura 1. Respostas à pergunta referente acerca do conhecimento prévio sobre a LGPD

Antes desse questionário, você já tinha ouvido falar na Lei Geral de Proteção de Dados (Lei 13.709/18), conhecida como LGPD?

205 respostas



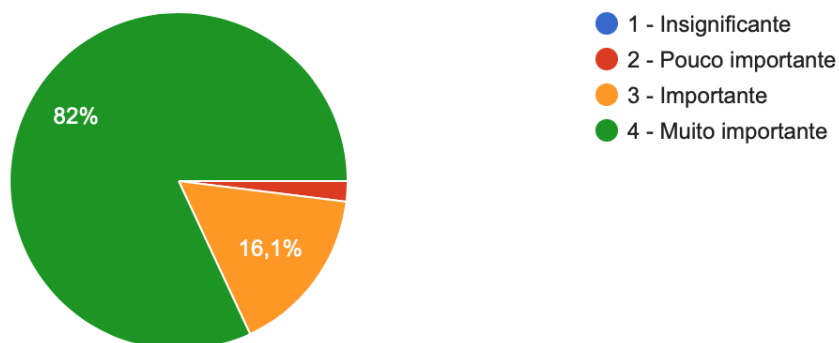
Fonte: Elaborado pelo autor (2022)

Ainda, ao serem questionados da importância que o tema tinha em relação a proteção de seus dados pessoais fornecida pelas empresas que eles tinham algum tipo de vínculo, 82% das pessoas consideraram o tema em questão muito importante, com 16,1% considerando importante, contra 2% que consideraram pouco importante.

Figura 2. Respostas ao questionamento referente à importância dada pela amostra quanto a proteção de dados pessoais

De 1 a 4 quão importante você considera a proteção dada aos seus dados pessoais e dados sensíveis (nome, RG, CPF, número de telefone, dado biométrico, dados sobre sua saúde) pelas organizações que você tem algum tipo de relação (seja como consumidor, fornecedor, colaborador, etc.)? Seja o mais sincero possível.

205 respostas



Fonte: Elaborado pelo autor (2022)

Isso demonstra que, apesar do universo de pessoas que conhecem a legislação ser de 81%, número considerável, mais de 98% da amostra considera pelo menos importante a adoção de medidas, pelas empresas, para que protejam os seus dados pessoais.

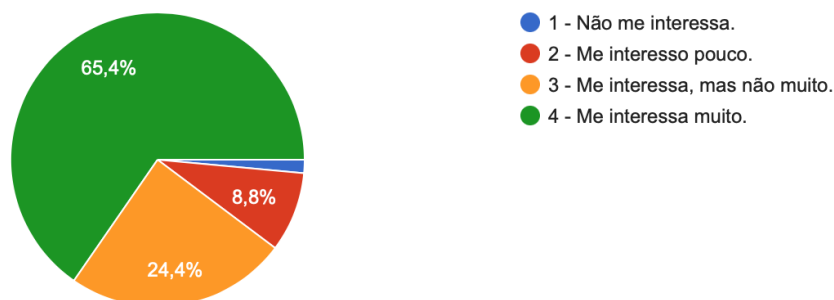
Quanto ao interesse dos mesmos sobre o tema relacionado à proteção de dados, a amostra destaca que 65,4% demonstra muito interesse pelo tema, 24,4% apresenta interesse, contra 8,8% que indicaram pouco interesse, e apenas 1,5% que não se interessavam pela matéria.

Figura 3. Respostas ao questionamento referente ao interesse da amostra sobre a proteção de seus dados

De 1 a 4, quanto o tema sobre a proteção de seus dados te interessa (seja o mais sincero possível)?

 Copiar

205 respostas



Fonte: Elaborado pelo autor (2022)

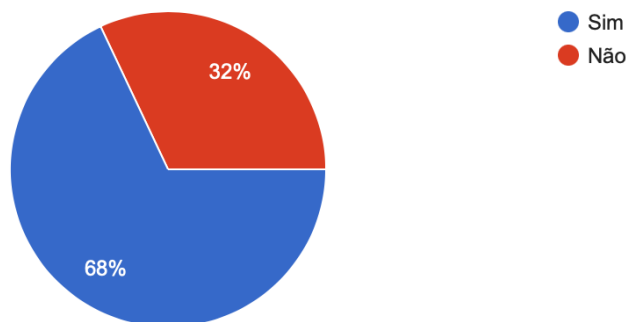
Ou seja, apesar da amostra considerar o tema como importante (98%), existe uma parcela, ainda que pequena, que não se interessa pelo tema (1,5%).

Após análise do conhecimento da amostra sobre o tema, sua relevância e interesse daquele universo, se buscou entender se aqueles possíveis consumidores sabiam ou não das consequências práticas que decorrem de possíveis incidentes de segurança, em especial aqueles com divulgação de dados na internet. Foi observado que, no universo de 203 pessoas, 68% tinham ciência dos possíveis resultados, contra 32% que desconheciam sobre a matéria. Isso demonstra que, apesar da amostra entender a relevância acerca do tema, mais de 1/3 não compreende o que na prática pode vir a acontecer caso seus dados pessoais estejam dispostos na rede.

Figura 4. Respostas ao questionamento a respeito das consequências na hipótese da divulgação de dados pessoais na internet

Você sabe quais consequências podem ocorrer caso seus dados sejam divulgados na internet?

203 respostas



Fonte: Elaborado pelo autor (2022)

Demonstrou-se, portanto, que apesar de cerca de quase 1/3 das pessoas não conhecerem de fato as consequências práticas de um incidente, todas elas consideraram a matéria importante em algum grau quanto da relação estabelecida entre elas e as organizações com as quais tenham vínculo.

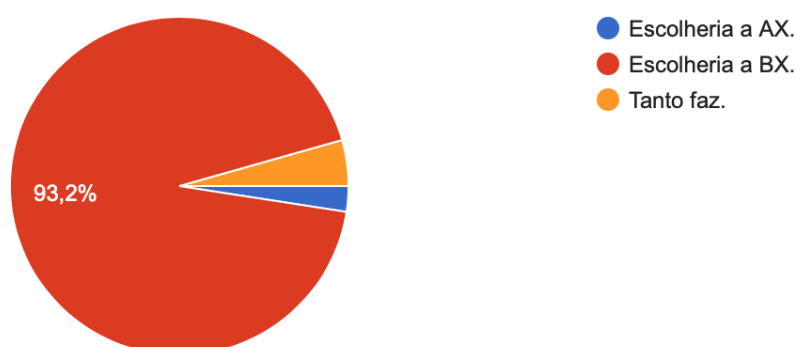
Nas três perguntas subsequentes, se buscou avaliar, sob o ponto de vista prático, como os consumidores se comportariam frente à prestação ou oferta de um produto ou serviço, mas em hipóteses distintas. A primeira hipótese abarcou como se daria a escolha entre empresas com a oferta do mesmo produto/serviço, com a mesma qualidade e com o mesmo valor, onde uma delas (AX) sofre um ataque cibernético ocasionando um vazamento no banco de dados, com o compartilhamento de diversos dados pessoais na internet.

Observou-se que, a grande maioria (93,2%) dos potenciais consumidores optaram por empresas que ainda não tinha sofrido (BX) o ataque, em detrimento daquela que sofreu o ataque (AX), enquanto que para 4,4% não faria diferença entre uma e outra, e 2,4% ainda sim optariam pela empresa que sofreu com o vazamento de dados.

Figura 5. Respostas ao questionamento hipotético sobre a escolha a respeito da situação exemplo 1

Você é o consumidor de um produto/serviço, e algumas empresas ofertam o que você está procurando. Dentre essas empresas estão a AX e a BX, que oferecem o **mesmo produto/serviço, com a mesma qualidade e com o mesmo valor**. Durante uma breve pesquisa de mídia sobre cada uma dessas empresas, você descobre que a empresa AX sofreu um ataque cibernético, ocasionando um vazamento no banco de dados da empresa, onde diversas informações dos seus clientes foram disponibilizados na internet como: nome, CPF, e-mail e número de telefone. Nesse caso, como se daria a sua escolha entre as empresas?

205 respostas



Fonte: Elaborado pelo autor (2022)

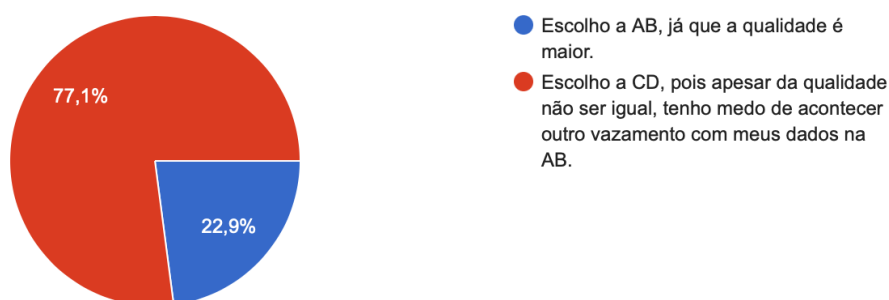
No segundo momento, a hipótese proposta apresentava à amostra uma empresa que oferecia produto/serviço superior à outra. A empresa com oferta do produto de qualidade superior sofria um ataque cibernético ocasionando o vazamento no banco de dados, na internet. Para 77,1% dos respondentes escolheriam uma empresa com qualidade de produto ou serviço inferior, pois teriam medo da ocorrência de um outro vazamento de dados. Isso demonstra que, a incidência de um vazamento de dados por uma empresa com produto em destaque no mercado, traria efeitos negativos à imagem desta, inviabilizando que uma nova relação seja firmada entre a instituição e o consumidor, afetando diretamente a fidelização do consumidor em decorrência de uma mídia negativa frente à cultura de proteção de dados.

Figura 6. Respostas ao questionamento hipotético sobre a escolha a respeito da situação exemplo 2

Agora você é o consumidor de um produto/serviço, e dentre as empresas que oferecem o produto/serviço está a empresa AB e a CD, sendo **a qualidade da AB superior a da CD**. Durante sua pesquisa de mídia sobre as empresas, você descobre que a empresa AB sofreu um ataque cibernético ocasionando o vazamento no banco de dados onde dados de diversos clientes foram divulgados na internet. Nesse caso, como se dará a sua escolha entre as empresas?



205 respostas



Fonte: Elaborado pelo autor (2022)

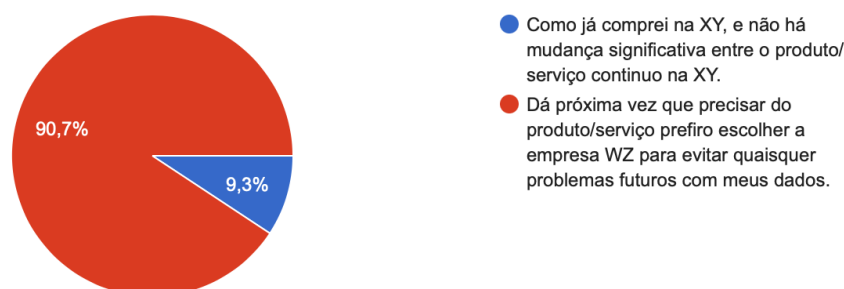
Na terceira e última hipótese, a oferta do produto ou serviço se apresenta com qualidade e preço semelhante. Após a primeira compra o consumidor descobre que a empresa estaria vendendo seus dados para outra empresa sem a sua autorização. Nesse caso, se buscou entender como se daria a sua próxima compra. Mais de 90% das pessoas indicaram que da próxima vez escolheriam outra empresa para evitar quaisquer problemas futuros envolvendo seus dados, frente a 9,3% que permaneceria na mesma empresa, por não haver mudança significativa entre o produto ou serviço em comparação com as demais.

Figura 7. Respostas ao questionamento hipotético sobre a escolha a respeito da situação exemplo 3

Você é o consumidor de um produto/serviço, e algumas empresas ofertam o você está procurando. Dentre essas empresas está a empresa XY e a WZ, que oferecem o mesmo **produto/serviço com qualidade e preços semelhantes**. Na sua primeira compra você opta pela empresa XY, porém, após 2 dias da compra você descobre que a XY está vendendo seus dados para outra empresa sem a sua autorização. Nesse caso, como se dará a sua escolha entre as empresas para a **próxima compra**?



205 respostas



Fonte: Elaborado pelo autor (2022)

Ora, por essa análise pode-se perceber que para todo e qualquer evento de mídia negativa envolvendo vazamento de dados pessoais, os consumidores têm a tendência significativa em simplesmente optar por outras empresas que não possuem esse histórico, mesmo que o produto ou serviço ofertado seja de qualidade inferior ao prestado pela empresa atacada.

Passando para a análise da perspectiva de como o consumidor analisa a promoção da cultura de proteção de dados dentro das empresas, foi observado que para 51,2% das pessoas, mesmo sendo o valor do produto ou do serviço mais caro, o investimento seria um diferencial para a escolha da empresa, com 0,5% acredita ser obrigação da empresa manter sigilo dos dados.

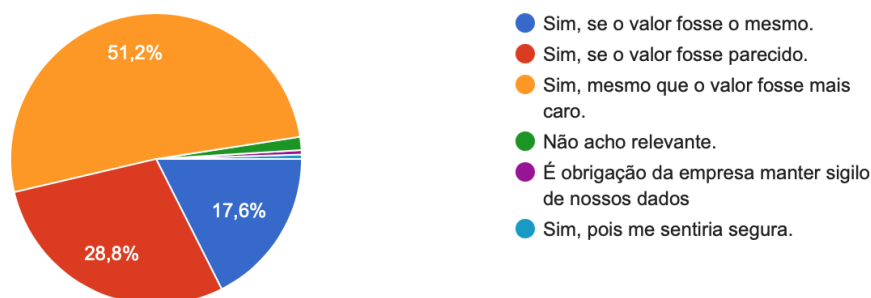
Para 28,8%, a escolha ainda seria um diferencial se o valor fosse parecido, enquanto que para 17,6% seria um diferencial apenas se o valor fosse o mesmo. Ou seja, consumidores tendem a pagar por produtos/serviços mais caros quando o assunto é a proteção dos seus dados. Para os demais, 1,5% não considera como relevante.

Figura 8. Respostas ao questionamento sobre a promoção da cultura de proteção de dados no momento da escolha para relação com as empresas

Você acredita que se as empresas promovessem uma cultura de proteção de dados sobre a forma de gerir dados seria um diferencial para você no momento da escolha da empresa?

 Copiar

205 respostas



Fonte: Elaborado pelo autor (2022)

Por fim, foram analisadas quais seriam as maiores dificuldades encontradas pelos consumidores para entender sobre a forma que as empresas tratam os dados pessoais coletados, onde foi elencado como principal fator as empresas não produzirem conteúdo informativo à respeito sobre como os dados são tratados (64,6%), os conteúdos produzidos não estarem facilmente acessíveis (31,8%), o conteúdo possuir difícil leitura/entendimento (23,1%), e outros como não darem importância para falar sobre o tema, não ser possível a identificação do local que está disposto o conteúdo e não se ter certeza de que os dados são realmente tratados com a proteção informada.

Buscando compreender o impacto da privacidade de dados dentro das empresas e em seus negócios, a Cisco realizou um estudo no final de 2018, publicado em 2019 no Data Privacy Benchmark Study, com mais de 3.200 profissionais de segurança familiarizados com processos de privacidade e o impacto em suas organizações, em 29 entrevistas aprofundadas com líderes de privacidade em todo o mundo. Os resultados da pesquisa demonstraram o impacto da privacidade de dados com fatores como: atrasos nas vendas (causados por problemas ou desalinhamento durante o processo de contratação do fornecedor em razão de práticas ou políticas de privacidade abaixo da média, ou pelo não compartilhamento de suas práticas atuais, com atrasos na assinatura do contrato ou até mesmo levando a um redesenho do produto), bem como a correlação entre a GDPR (Regulation (EU) 2016/679 - General Data Protection Regulation), a probabilidade e custos de uma violação de dados. (CISCO, 2019).

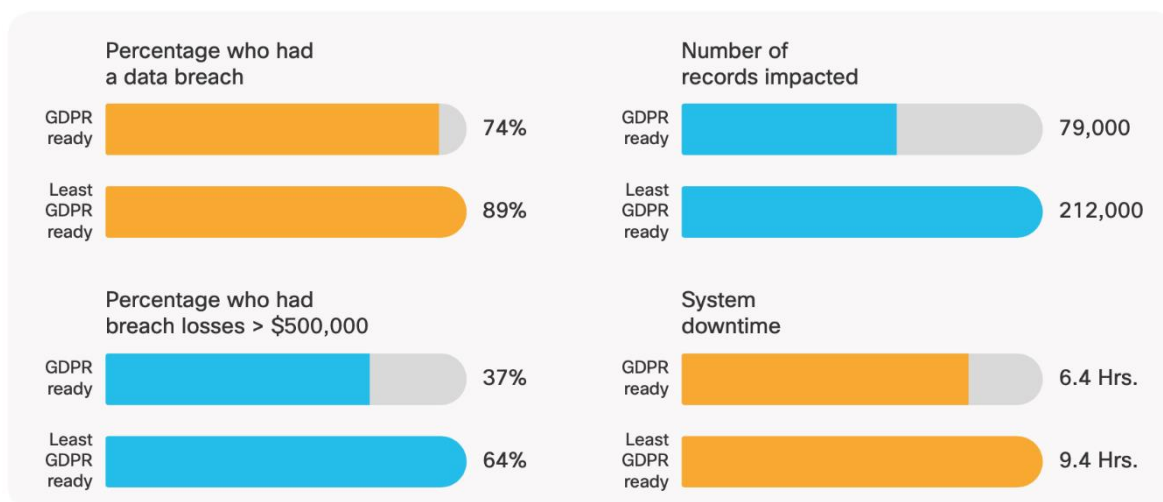
O estudo evidenciou que 97% das empresas globais estão obtendo pelo menos um ou mais benefícios com investimentos relacionados à privacidade como maior agilidade e inovação (42%), eficiência operacional (41%), vantagem competitiva (41%), além de terem sofrido menos com violações de dados pessoais (39%), redução no atraso com vendas (37%) e interesse por investidores (36%). (CISCO, 2019).

A pesquisa demonstra ainda que a depender do setor em que a empresa esteja inserida, a visão do consumidor frente ao negócio e sua relação com a privacidade de dados pode ser vista tanto como um diferencial ou requisito básico, a depender da área que este esteja conectado. (CISCO, 2019).

No setor de varejo, por exemplo, foi indicado que os clientes tendem a ser inconstantes e geralmente não se importam em trocar facilmente de fornecedor, com base em preocupações relacionadas à proteção de dados. Enquanto isso, no setor de saúde, estes estão menos propensos a ver uma abordagem centrada na privacidade como um diferencial, dada a expectativa inerente à área de que fabricantes e fornecedores atuem com a proteção de dados sensíveis. (CISCO, 2019).

O desenvolvimento da cultura de proteção de dados implementado através de um programa de privacidade não apenas ajuda a proteger uma empresa de uma violação, mas quanto mais maduro for o programa, menor será a perda financeira na hipótese de um incidente de segurança da informação. Foi demonstrado que as empresas que investiram para se adequarem à GDPR sofreram menos violações e tiveram menos gastos. Cerca de 74% das empresas adequadas à GDPR sofreram violação de dados no ano passado, contra 89% das empresas menos preparadas, apresentando uma redução de 15% em relação aos incidentes. Ainda, na hipótese de violação, as empresas adequadas à GDPR tiveram menos registros (79.000 versus 212.000), menor tempo de inatividade do sistema (6,4 horas versus 9,4 horas), e apenas 37% delas sofreram uma perda de mais de US\$ 500.000 no último ano de violações de dados comparado com 64% das empresas menos adequadas para GDPR.

Figura 09. GDPR-ready companies have fewer and less costly breaches



Fonte: Cisco 2019 Data Privacy Benchmark Study

A pesquisa aponta ainda que a privacidade é capaz de gerar insights valiosos de dados existentes para seus controladores, assim como é capaz de levar a um uso mais eficiente e eficaz. Um programa de privacidade bem elaborado está apto para reduzir o risco do uso inapropriado dos dados dos clientes, garantindo que os funcionários os usem dentro dos limites apropriados, como por exemplo em campanhas de marketing. Assim, líderes de privacidade cumprem em reiterar a ideia de que os programas de privacidade desenvolvidos criam valor para a organização, especificamente em torno de insights de dados e uso de dados. Na América Latina, o chefe de privacidade de um banco multinacional destacou que a privacidade de dados permitiu que o banco funcionasse de forma mais eficaz e com total confiança. Assim, um programa de privacidade forte em vigor permite que as empresas obtenham maiores insights, inovação e sucesso nos negócios. (CISCO, 2019).

Durante a fase inicial de “mapeamento de dados”, para o estabelecimento de um programa de privacidade, as empresas geralmente descobrem ineficiências em seus fluxos de trabalho atuais ao rastrear o fluxo dos dados dentro da sua organização. Isso possibilita que as empresas tenham a capacidade de aumentar a sua eficiência operacional, o que é de valor tangível para esta. A existência de políticas de privacidade claras e processos estabelecidos para lidar com problemas complexos de privacidade permite que as organizações minimizem possíveis atrasos que uma organização menos madura em privacidade poderia enfrentar. (CISCO, 2019).

Além da mitigação de violações de dados, fidelização do cliente e demais vantagens competitivas já apontadas, os líderes de privacidade em todo o mundo apontam que a existência de programas de

privacidade pode atrair investimentos em dólares. Em razão do número de violações de dados de alto perfil que ocorreram nos últimos anos, os investidores se preocupam mais do que antes e buscam saber exatamente o que as empresas estão fazendo para se proteger contra essa preocupação iminente. Isso vale para todos os setores e regiões geográficas. Portanto, quanto menor a preocupação da empresa com a privacidade de seus dados menor o entusiasmo dos investidores para aporte monetário em uma empresa que provavelmente sofrerá uma violação de dados de proporções significativas. (CISCO, 2019).

Embora seja difícil prever o que o futuro reserva, o mercado provavelmente imaginará e perceberá maneiras novas e inovadoras de processar informações pessoais. Essas novas abordagens serão desenvolvidas com a privacidade em foco, proporcionando mais valor e diferenciação de mercado (CISCO, 2019), correndo concomitantemente à propagação da cultura de proteção de dados na sociedade brasileira, atrelada ao desenvolvimento do conhecimento dos usuários, consumidores e clientes, que demandarão tal atitude das empresas. Caso contrário, a manutenção desse vínculo corre grave risco de ficar significativamente estremecida, assim como ficarem suscetíveis de receberem diversos tipos de sanções previstas nas mais diversas legislações (Constituição Federal (BRASIL, 1988), Código Civil (BRASIL, 2002), Código de Defesa do Consumidor (BRASIL, 1990), Marco Civil da Internet (BRASIL, 2014) e LGPD (BRASIL, 2018)).

CONSIDERAÇÕES FINAIS

Conforme demonstrado, observamos que os benefícios da implementação de práticas e sistemas de privacidade sólidos para o desenvolvimento de uma cultura de proteção de dados vão muito além do atendimento aos requisitos de conformidade dispostos na LGPD.

O amadurecimento da cultura de privacidade e proteção de dados é boa tanto para os negócios internamente, como para suas relações externas. Por isso, líderes de privacidade e suas organizações reconhecem esse fato e vêm se beneficiando diariamente.

Embora as empresas ainda enfrentem desafios significativos quando se trata de mudança na sua cultura organizacional, garantir a alocação adequada de recursos à privacidade e aumentar a conscientização sobre o valor da privacidade é o primeiro passo para ajudar as organizações a tomarem melhores decisões, mantendo consigo o que deve ser de maior valor: seus clientes (e os dados pessoais do seu cliente).

REFERÊNCIAS

BRASIL. Constituição da República Federativa do Brasil. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>. Acesso em: 18 de set de 2022.

BRASIL. Lei nº 10.406, de 10 de janeiro de 2002. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/2002/l10406compilada.htm>. Acesso em: 18 de set de 2022.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. art. 5º, I. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>. Acesso em: 17 de ago de 2022.

BRASIL. Escola Nacional de Defesa do Consumidor. A proteção de dados pessoais nas relações de consumo: para além da informação creditícia. vol. 02 Disponível em: <https://www.gov.br/mj/pt-br/assuntos/seusdireitos/consumidor/defesadoconsumidor/Biblioteca/manuaisupload/vol_2_protecao_de_dados_pessoais.pdf>. Acesso em: 10 de ago de 22.

CISCO. Privacy Gains: Business Benefits of Privacy Investment. Disponível em: <https://www.cisco.com/c/dam/en_us/about/doing_business/trustcenter/docs/privacygainsbusinessbenefitsofprivacyinvestment-white-paper.pdf>. Acesso em: 12 de jun de 2022. p. 4

CLARK, Jean. What is the Internet of Things (IoT)?. Disponível em: <<https://www.ibm.com/blogs/internet-of-things/what-is-the-iot/>>. Acesso em: 19 de ago de 2022.

DA COSTA, M.A.; Paz, A.M.O. A produção de questionários de pesquisa como gênero acadêmico. Mandinga – Revista de Estudos Linguísticos. 2019, p. 106-127.

GARTNER GLOSSARY. Big Data. Disponível em: <<https://www.gartner.com/en/informationtechnology/glossary/big-data>>. Acesso em: 17 de jul de 2022.

LINS, Augusto. The direct link between company culture and customer loyalty. Disponível em: <<https://www.fastcompany.com/90698946/the-direct-link-between-company-culture-and-customer-loyalty>>. Acesso em: 19 de ago de 2022.

MALDONADO, Viviane Nóbrega; OPICE BLUM, Renato. Revista dos Tribunais. LGPD comentada. 3ª ed. rev. at. e ampl. Brasil, 2021, p. 22.

STANFORD ENCYCLOPEDIA OF PHILOSOPHY. Privacy. Disponível em: <<https://plato.stanford.edu/entries/privacy/#pagetopright>>. Acesso em: 03 de jul de 22.

TURRIONI, João Batista; MELLO, Carlos Henrique Pereira. Metodologia De Pesquisa Em Engenharia De Produção, estratégias, métodos e técnicas para condução de pesquisas quantitativas e qualitativas. Brasil, 2012, p. 80.

WARREN Samuel D.; BRANDEIS Louis D. The right to privacy. Disponível em: <<https://www.cs.cornell.edu/~shmat/courses/cs5436/warren-brandeis.pdf>>. Acesso em: 03 de jul de 22.

Capítulo 7



10.37423/230908215

APLICABILIDADE DA LGPD (LEI GERAL DE PROTEÇÃO DE DADOS PESSOAIS) NA SEGURANÇA DA INFORMAÇÃO EM EMPRESAS DE DESENVOLVIMENTO DE SOFTWARE

Roberto Teixeira Claro

Pós Graduada do Curso de DIGITAL BUSINESS - 2021 do MBA USP/ESALQ

Lisleandra Machado

*IFSUDESTEMG-Instituto Federal de Educação Ciência e Tecnologia do Sudeste de Minas Gerais
UNEMAT - Universidade do Estado do Mato Grosso*

Domingos Sávio da Cunha Garcia

Leonardo Amorim de Araújo

IFSUDESTEMG-Instituto Federal de Educação Ciência e Tecnologia do Sudeste de Minas Gerais

Carlos Arthur Alevato Leal

IFSUDESTEMG-Instituto Federal de Educação Ciência e Tecnologia do Sudeste de Minas Gerais

Arthur Nascimento Assunção

IFSUDESTEMG-Instituto Federal de Educação Ciência e Tecnologia do Sudeste de Minas Gerais

Lucas Mendes Ferreira

IFSUDESTEMG-Instituto Federal de Educação Ciência e Tecnologia do Sudeste de Minas Gerais

Samuel Alves de Freitas

IFSUDESTEMG-Instituto Federal de Educação Ciência e Tecnologia do Sudeste de Minas Gerais

Gustavo José Santiago Rosseti

IFSUDESTEMG-Instituto Federal de Educação Ciência e Tecnologia do Sudeste de Minas Gerais

Silvana Rodrigues Pires Moreira

IFSUDESTEMG-Instituto Federal de Educação Ciência e Tecnologia do Sudeste de Minas Gerais

Resumo: Com a entrada em vigor da LGPD (Lei Geral de Proteção de Dados Pessoais) o Brasil segue os países desenvolvidos e em desenvolvimento, tendo uma regulamentação moderna que busca proteger os dados pessoais e sensíveis. A lei 13.709 de 14 de agosto de 2018, entrou em vigor em 18 de setembro de 2020, no auge da pandemia. Este trabalho procurou analisar a adesão das empresas de desenvolvimento de *software* à nova legislação, sendo dividido em duas etapas, onde, na primeira etapa foi realizado um estudo de caso analisando a aderência à legislação por uma empresa de tecnologia da baixada Santista, e a segunda etapa foi realizada uma pesquisa exploratória a fim de comparar os resultados, os quais foram analisados por diversos aspectos, tratando os pontos positivos e negativos que tal lei apresenta para essas empresas, bem como a conformidade na criação de procedimentos, papéis e responsabilidades.

Palavras-chave: (LGPD; GDPR; segurança da informação; proteção de dados;).

INTRODUÇÃO

No dia 18 de setembro de 2020, passou a vigorar a lei 13.709 conhecida como LGPD (Lei Geral de Proteção de Dados Pessoais). Trata-se de um marco no que tange ao armazenamento, processamento e proteção dos dados pessoais, é uma legislação moderna, que segue as diretrizes da GDPR (*General Data Protection Regulation* – Regulamento Geral sobre a Proteção de Dados), lei europeia publicada em 2016 e entrando em vigo em 25 de maio de 2018. A LGPD aborda pontos em que normas e padrões de mercado já discorrem há algum tempo, como por exemplo a família ISO 27000, a qual trata da segurança da informação, gestão de segurança, gestão de riscos e seus controles (LIMA, 2021).

A estrutura da LGPD se apoia em três pilares: Confidencialidade, Integridade e Disponibilidade, onde, os dados podem ser considerados como seguros se, estão acessíveis somente por quem deve acessá-lo, estar íntegro (não vulnerável), e ter sua disponibilidade assegurada. Desta forma, os dados devem estar protegidos de qualquer manipulação (edição ou exclusão) não requerida (GARCIA, 2020).

A LGPD em seu artigo 52 discorre sobre as sanções administrativas, que vão desde uma simples advertência a multas de 2% do faturamento da pessoa jurídica (limitado a um total de 50 milhões de reais), bloqueio dos dados pessoais ou suspensão do exercício de atividades de tratamento dos dados pessoais (PLANALTO, 2018).

No Brasil, *startups* e empresas PMEs de tecnologia, enfrentam grandes desafios ao lidar com a LGPD, visto que nesses segmentos de empresas de tecnologia (*techs*), existe uma forte ligação com dados, sendo na captação, armazenamento ou tratamento. Estas empresas geram valor e receita baseadas no processamento de grandes volumes de dados, escalando e gerando inteligência de mercado, de forma a se tornarem competitivas e valiosas. Porém em sua fase inicial passam pelos entraves regulatórios, altíssimos custos de implantação e adequação à LGPD, e conviver com a incerteza de aceitação e validação do modelo de negócio perante o mercado (LIMA, 2021).

Por outro lado, para que as empresas de tecnologia possam ser competitivas em um mundo globalizado, pandêmico e pós-pandemia, é necessário que sejam adotadas medidas para adesão à um modelo, que demonstre segurança e confiança de que os dados estão sendo tratados e protegidos. Neste cenário a LGPD se faz extremamente necessária, o Brasil está atrasado em alguns anos em relação a seus vizinhos, e ao resto do mundo, quanto a ter uma regulamentação específica para tratar da proteção dos dados. Sem uma legislação específica, as empresas nacionais correm o risco de limitarem-se a entrega de serviços dentro do território nacional ou a países sem uma lei de proteção

de dados, visto que, 71% dos países já contam uma regulamentação em vigor conforme dados da Conferência das Nações Unidas para o Comércio e Desenvolvimento (2020). Os países que possuem uma lei específica geram uma barreira de entrada para empresas que não em conformidade com as normas em vigor (UNCTAD, 2022).

Conforme relatório apresentado pela CISCO (*Cisco Data Privacy Benchmark Study 2020*), o retorno médio das empresas no Brasil que investem em privacidade aumenta em 3,3 vezes, esse valor varia conforme o país e a média geral foi de 2,7 vezes, transformando dessa forma, a segurança dos dados, em uma potencial fonte de benefícios, considerando a eficiência, redução de riscos e fidelização de clientes.

Com base no exposto, este trabalho tem como objetivo analisar a aderência das empresas de desenvolvimento de *softwares* à LGPD, e entender como consultorias de desenvolvimento de *software*, *startups* e *techs*, estão em conformidade a implementação das políticas e procedimentos de proteção dos dados, orientação de seus colaboradores, treinamentos, auditorias e gestão dos acessos aos dados pessoais e sensíveis.

MATERIAL E MÉTODOS

Este trabalho foi dividido em duas partes, sendo a primeira um estudo de caso no qual foi analisado o *compliance* à LGPD, abordando as justificativas e o contexto do cenário em que se encontra a empresa analisada perante a legislação. Na segunda parte, foi realizada uma pesquisa exploratória não probabilística com o intuito de verificar a se o cenário encontrado na empresa analisada se trata de um caso isolado ou representa uma incidência maior dentro da amostragem coletada.

Como linha mestre para todo o estudo realizado, foi utilizado o texto da lei federal 13.709/2018. Sendo realizada também a pesquisa bibliográfica em livros, artigos e estudos sobre às boas práticas, métodos de implementação e os impactos (dificuldades, processos e o esforço em tempo e dinheiro) para a adequação das rotinas das empresas e a atenção ao tratamento de dados pessoais.

ESTUDO DE CASO

O estudo de caso para este trabalho foi elaborado a partir da análise da aplicação e conformidade à LGPD em uma empresa de consultoria em desenvolvimento de *software* de médio porte (com mais de 220 colaboradores diretos) com sede no litoral de São Paulo. Tal empresa possui uma carteira de clientes diversificada, que atuam nos setores de indústria, portos, comércio exterior, logística, setor

financeiro, e agronegócio. Sendo que os clientes estão distribuídos em todo o território nacional, com maior concentração nos estados de São Paulo, Rio de Janeiro e Bahia. Foram realizadas entrevistas com os gestores dos departamentos de pessoas, financeiro, operações e vendas, onde foi possível identificar o grau de entendimento da LGPD, e verificar a conformidade dos processos e adequação de normas e procedimentos.

A entrevista foi realizada individualmente de forma online via *Microsoft Teams*®, entre os dias 04 e 08 de julho de 2022, onde foi aplicado o questionário de pesquisa como base para a coleta dos dados, que foram posteriormente analisados e tabulados, para proporcionar um melhor entendimento e facilitar a análise qualitativa e posterior comparação com as respostas ao questionário aplicado na segunda parte do estudo, realizado de forma mais ampla, abrangendo outras empresas do mesmo segmento.

COLETA DE DADOS

Para a segunda parte do estudo, foi realizada uma pesquisa de amostragem por julgamento, sendo enviado questionário virtual via *Google Forms*®, para 128 (cento e vinte e oito) profissionais que atuam na área de tecnologia da informação, os dados foram coletados entre os dias 04 e 29 de julho, e foram computadas as respostas de 53 (cinquenta e três) participantes, sendo que 31 (trinta e um) deles (58,5%) trabalham em empresas na área de desenvolvimento de software.

O questionário (vide Apêndice) que foi elaborado para a pesquisa, está constituído de 18 (dezoito) questões estruturadas, dentre essas 5 (cinco) questões foram para realizar a classificação dos entrevistados de acordo com a formação, área de atuação e posição dentro da hierarquia na empresa, sendo as respostas apresentadas em forma de alternativas tabuladas. As demais questões visaram identificar o conhecimento do entrevistado referente à LGPD e sua implementação dentro da empresa a qual trabalham, elas foram elaboradas buscando identificar o nível de aderência que a empresa possui em relação aos aspectos pontuados da legislação, possibilitando dessa forma mapear o cenário em que se encontra diante da legislação.

Visando manter um padrão e ter uma forma comparativa, as respostas referente à aderência à LGPD foram estruturadas em uma escala de classificação, indo de 0 (zero) a 5 (cinco), onde zero corresponde a “pouco ou nada importante” e cinco indica “extremamente importante”. O questionário foi constituído também de questões onde as respostas procuravam entender se determinado aspecto da legislação foi implementado, com alternativas diretas, “Sim”, “Não” ou “Não sei responder”.

O público-alvo foi determinado e restrito aos profissionais que atuam em empresas de tecnologia, procurando dessa forma traçar um comparativo uniforme, e poder classificar o resultado, obtendo um menor desvio padrão. O período de coleta de dados, foi estabelecido com a finalidade de ter tempo hábil para que os entrevistados pudessem responder com tranquilidade, sobretudo em um momento em que a LGPD estrou em fase de sanção.

RESULTADOS E DISCUSSÃO

O mundo vive uma fase econômica bastante desafiadora, passando por uma pandemia de COVID-19 iniciada em 2020, e em 2022 teve o início de uma guerra entre Rússia e Ucrânia, com uma desaceleração global e possível recessão econômica. Em meio a esse cenário um tanto quanto caótico, entrou em vigor no dia 18 de setembro de 2020 a Lei Geral de Proteção de Dados Pessoais (LGPD), a qual impõe que sejam aplicados instrumentos para controle e proteção dos dados pessoais, com sanções aplicadas a partir de agosto de 2021. Com toda essa dificuldade econômica as empresas precisam se adequar, revisitando seus processos, atualizando ou contratando sistemas e treinando os colaboradores, todo esse esforço geram custos, que no caso de *startups*, pequenas e médias empresas de desenvolvimento de *software*, tornam-se grandes limitadores, gerando um enorme impacto nos investimentos, desta forma analisei os dados coletados e o estudo de caso sob essa ótica, traçando um paralelo entre os custos para adequação e manutenção à LGPD, e o retorno de investimento na segurança dos dados e qualidade no serviço prestado ao cliente.

Com base no questionário aplicado e da amostra gerada, 58,5% dos entrevistados responderam que atuam em empresas de desenvolvimento de software, conforme apresentado na figura 1. Dentro desse grupo conforme a figura 2, em uma escala de 1 a 5 onde 1 representa nenhum conhecimento e 5 um conhecimento profundo sobre a LGPD, a maioria (52%) responderam um conhecimento mediano, 19% um ótimo conhecimento e 6% um profundo conhecimento.

Está claro que o assunto proteção de dados pessoais é de interesse geral, de forma que os entrevistados em sua maioria (acima de 80%) entendem como extremamente importante a proteção dos dados, conforme demonstrado na figura 3, em que consideram seus dados e de terceiros.



Figura 1: Percentual de entrevistados que trabalham empresas de TI

Fonte: Elaborado pelo autor (2022)

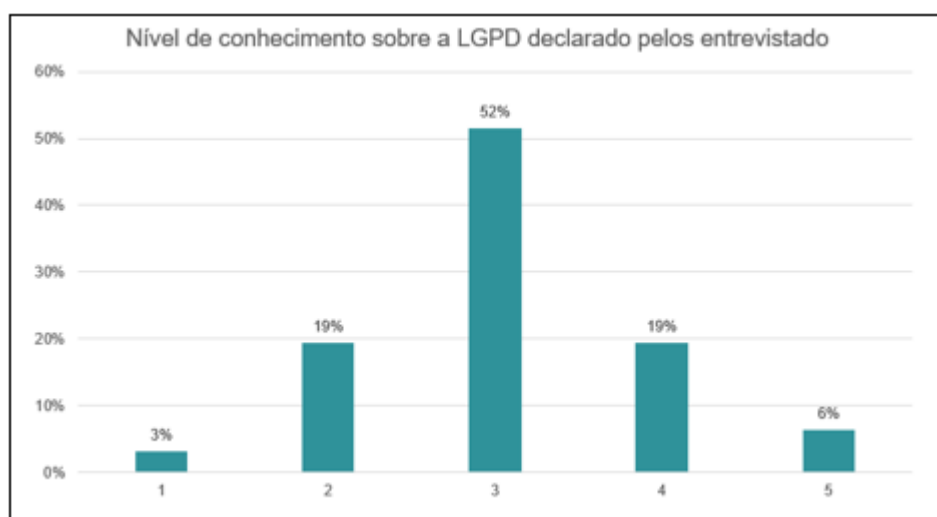


Figura 2: Nível de conhecimento sobre a LGPD declarado pelos entrevistados

Fonte: Elaborado pelo autor (2022)

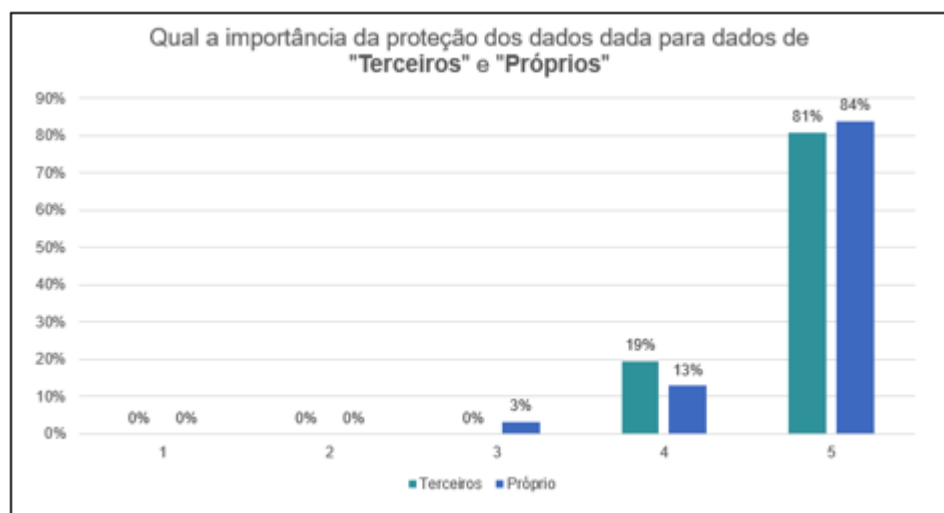


Figura 3: Importância da proteção dos dados

Fonte: Elaborado pelo autor (2022)

Em continuação foi analisada a conformidade à LGPD, onde as questões abordaram principalmente a orientação dos colaboradores, estabelecimento de procedimentos e normas. Aqui o cenário fica equilibrado, conforme podemos visualizar no gráfico da figura 4, onde 45% respondeu que sabe identificar o controlador dos dados e 48% tem ciência de quem é o responsável no papel de operador, por outro lado, somando-se os que não sabem e os que não souberam responder, temos uma maioria que indica que os procedimentos não estão sendo devidamente aplicados, sendo apresentado na figura 5, onde 58% dos entrevistados informou que não tem participado de treinamentos e 55% não tem acesso a procedimentos e normas para o tratamento dos dados.

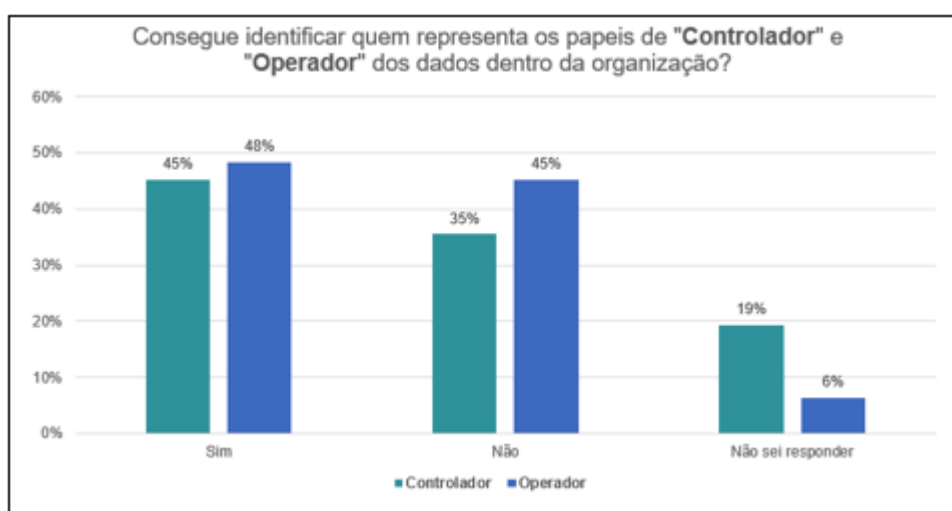


Figura 4: Responsáveis pelos papéis de Controlador e Operador dos dados dentro da organização

Fonte: Elaborado pelo autor (2022)

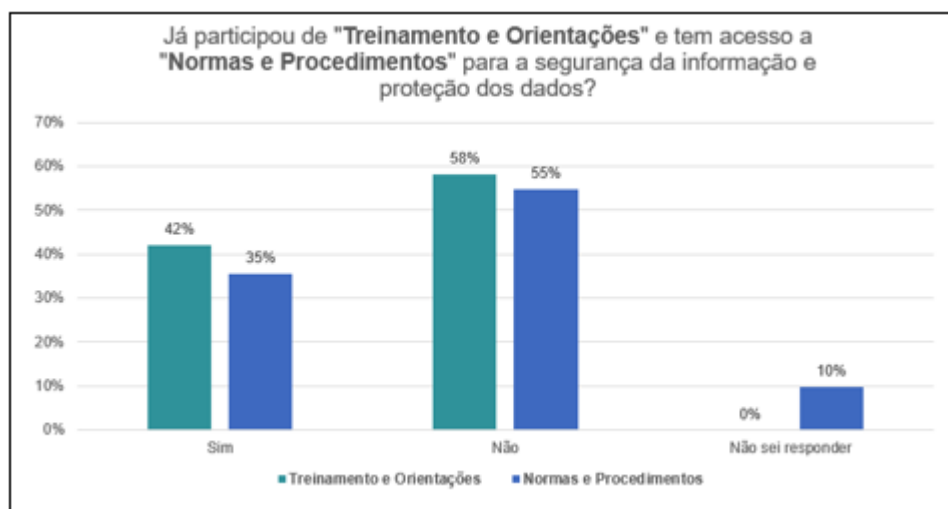


Figura 5: Participou de treinamento e conhece as normas e procedimentos da companhia

Fonte: Elaborado pelo autor (2022)

Finalizando a coleta de dados, foi questionado se o colaborador possui acesso à dados sensíveis, caracterizado conforme a lei, onde 29% responderam que sim, possuem acesso a dados como origem racial, dados médicos, financeiros, dentre outros.

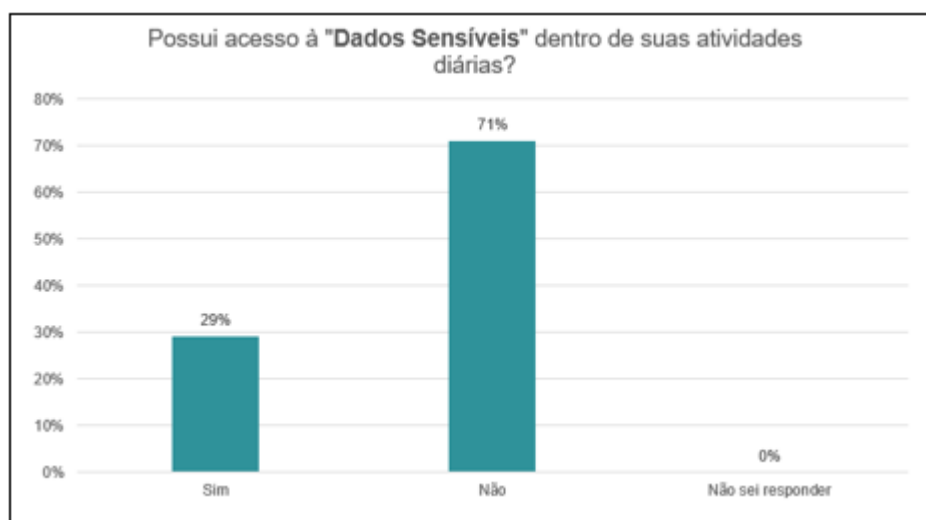


Figura 6: Possui acesso à dados sensíveis

Fonte: Elaborado pelo autor (2022)

No estudo de caso, foram realizadas entrevistas, utilizando como base esse mesmo questionário, e adicionalmente colhendo os *feedbacks* e aprofundando na discussão do tema. Ficou claro nas entrevistas que os gestores não possuem a visão clara de quais são os procedimentos e normas adotados internamente para garantir o tratamento dos dados. Por se tratar de uma empresa na área

de desenvolvimento de *software*, esta consultoria lida com dados próprios, relacionados a atividade de gestão dos colaboradores, bem como dados oriundos de clientes, os quais utilizam suas bases de dados para realizar testes dos *softwares* e aplicativos desenvolvidos.

Com essa visão, temos dois cenários dentro da empresa:

A – Tratamento dos dados de funcionários – gestão dos funcionários, controle de acessos, políticas de segurança, acessibilidade aos dados – neste cenário controlador e operador são responsáveis por acompanhar e fiscalizar o tratamento dos dados tal como dita a LGPD.

B – Tratamento dos dados de clientes – acesso a base de dados compartilhadas ou disponibilizada pelo cliente para servir como massa de dados para testes, simulações e validações dos *softwares* desenvolvidos.

Para o cenário A, segue o processo principal, onde o controlador deve determinar os procedimentos, cuidar que as medidas sejam implementadas e executadas, e certificar-se através de relatórios e auditorias que os procedimentos estão sendo cumpridos.

A LGPD define em seu Art. 5º o papel de encarregado, cujas responsabilidades pode-se destacar como um canal de comunicação entre o controlador, operador e a ANPD (Autoridade Nacional de Proteção de Dados), durante as entrevistas, os gestores não souberam indicar os responsáveis pelos papéis de controlador, operador, e do encarregado dos dados.

Foram identificados sistemas que fazem uso dos dados pessoais, os quais possuem uma política de acesso e backup, garantido nesta fase a segurança da informação, porém carece de procedimentos de validação da integridade dos dados e os processos desenhados para serem auditados conforme prevê a LGPD, bem como acesso controlado aos dados.

No cenário B, surge a figura do subcontratado do operador, que tem iguais responsabilidades que o operador dos dados, devendo seguir as orientações do controlador do cliente (contratante), o qual deve firmar contratos que estabeleçam regras claras e precisas para o procedimento de tratamento e manipulação dos dados, bem como o procedimento e comunicação em caso de incidentes, sendo este um cenário delicado e que requer grande atenção.

Neste cenário, não foram identificados contratos ou procedimentos para o tratamento dos dados pessoais envolvidos na atividade fim (desenvolvimento de software). Durante a prestação de serviço realizada pela empresa a título de desenvolvimento de sistemas, aplicações ou soluções, o controlador da contratante deve estabelecer claramente as regras para o tratamento dos dados, indicando os

procedimentos e auditorias necessários que o fornecedor deve cumprir. Por sua vez, o fornecedor, que está na figura do subcontratado, deve zelar pelo cumprimento dos procedimentos e gerar seus próprios processos para garantir o compliance.

É importante que exista um instrumento que defina papéis e responsabilidades entre as partes, um mapa dos sistemas, acordo de confidencialidade e procedimentos contendo boas práticas para serem adotadas, desta forma fica claro as obrigações de cada parte e sanções em caso do não cumprimento.

CONSIDERAÇÕES FINAIS

Analisando todo o panorama, entrevistas e respostas aos questionários, é possível concluir que a empresa analisada busca atender a legislação vigente, porém existe muito trabalho a ser feito, sobretudo no que se refere a procedimentos e treinamento dos colaboradores, para que tenham orientação e conhecimento compatível durante o tratamento de dados. Neste sentido, devem ser desenvolvidas rotinas de treinamento, instrumentos de acompanhamento e auditoria para verificar a conformidade das medidas adotadas. Este estudo foi restrito à amostra e dados coletados, os quais foram tabulados e analisados, de acordo com as respostas ao questionário aplicado, deve ser destacado que não se trata de um caso isolado, outras empresas no segmento de desenvolvimento de software, tampouco vem cumprindo seu papel adequadamente, deixando margem a falhas de conformidade à LGPD. Portanto, apesar da LGPD ser uma lei moderna e traz grande segurança à sociedade, é necessário que as empresas enxerguem como um benefício que irá trazer segurança jurídica e qualidade no atendimento ao cliente, para que vejam o compliance como um importante marco do qual podem se beneficiar e não somente como um ônus, gerando maiores custos e transtornos.

AGRADECIMENTOS

Agradeço primeiramente à minha esposa, pelo apoio e incentivo, principalmente nos momentos mais difíceis. Agradeço à minha orientadora Lisleandra “Lis” Machado, pela paciência e dicas fundamentais para a elaboração deste trabalho.

REFERÊNCIAS

BRASIL2018LGPDLeiGeraldeProteçãodeDadosPessoais<http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm> - acessado em 06 de abril de 2022.

CartilhadeSegurançaparaInternetCert.Br.Versão3.1de2006<https://cartilha.cert.br/sobre/old/cartilha_seguranca_3.1.pdf> - acessado em 08 de abril de 2022.

CISCO – From Privacy to Prot: Achieving Positive Returns on Privacy InvestmentsCisco Data Privacy BenchmarkStudy2020https://www.cisco.com/c/dam/global/en_uk/products/collateral/security/2020-data-privacy-cybersecurity-series-jan-2020.pdf - acessado em 14 de junho de 2022.

GARCIA, Lara Rocha et al. Lei Geral de Proteção de Dados (LGPD): guia de implantação. São Paulo: Blucher, 2020. Ebook. (1recursoonline). ISBN9786555060164. Disponível em: <https://integrada.minhabiblioteca.com.br/books/9786555060164> - acessado em: 10 abril. 2022.

GDPR – 2016 – General Data Protection Regulation - <https://gdpr.eu/tag/gdpr/> - acessado em: 08 de abril de 2022.

LIMA, Ana Paula Moraes Canto de; CRESPO, Marcelo; PINHEIRO, Patricia Peck (coord.). LGPD aplicada. São Paulo: Atlas, 2021. ISBN 9788597026931.

PINHEIRO, Patricia Peck; Proteção de Dados Pessoais. São Paulo: Saraiva Jur, 2021. ISBN 9786555595116.

UNCTAD-Data Protection and Privacy Legislation Worldwide<https://unctad.org/page/data-protection-and-privacy-legislation-worldwide>

APÊNDICE

A seguir segue reprodução do questionário (*Google Forms*®) aplicado na fase de coleta de dados.



The image shows a Google Form titled "Pesquisa de análise de aderência das empresas à LGPD". The header features a dark blue banner with icons representing data protection (locks, Wi-Fi, fingerprint, USB) and the text "LGPD LEI GERAL DE PROTEÇÃO DE DADOS". The main content area is white with a light blue border. It contains an introductory paragraph about the research's purpose, a statement on data confidentiality, instructions for respondents, and a thank you message from Roberto Teixeira Claro. At the bottom, there is a section for the respondent's email, marked as required with a red asterisk.

Pesquisa de análise de aderência das empresas à LGPD

Esta pesquisa visa gerar uma amostragem qualitativa quanto ao nível de aderência, entendimento e *compliance* em relação à LGPD (Lei Geral de Proteção de Dados) dentro da organização e de seus colaboradores.

Os dados aqui gerados, são de caráter sigiloso e tem por finalidade servir de base para os estudos, não sendo nenhum dado utilizado para fins comerciais.

Ao responder o questionário, tenha em mente os aspectos práticos da regulamentação e aplicação da LGPD, seja sincero ao responder cada questão, lembre-se, você não está sendo avaliado.

Agradeço antecipadamente sua colaboração!
Roberto Teixeira Claro

rtclaro@gmail.com [Alternar conta](#)

*Obrigatório

E-mail *

Seu e-mail

Qual sua posição dentro da organização? *

- ☐ Direção
- ☐ Gerência
- ☐ Liderança
- ☐ Analista ou Assistente
- ☐ Outra

Qual a área de atuação de sua organização? *

- ☐ Saúde
- ☐ Indústria
- ☐ Comércio
- ☐ Serviços de TI
- ☐ Serviços (outros)
- ☐ Outra área de atuação

Você possui formação ou conhecimentos na de Tecnologia da Informação? *

- ☐ Sim
- ☐ Não
- ☐ Não sei responder

Qual seu nível de conhecimento sobre a LGPD (Lei Geral de Proteção de Dados)? *

1 2 3 4 5
Nenhum conhecimento ☐ ☐ ☐ ☐ ☐ Profundo conhecimento

Na sua visão, a LGPD é : *

- ☐ Uma lei que trata somente dos dados de usuários na internet
- ☐ Uma regulamentação que protege as empresas do vazamento de dados
- ☐ Uma lei que obriga as empresas a guardar os dados dos clientes por até 5 anos
- ☐ Uma lei que regula o tratamento, privacidade e proteção dos dados pessoais
- ☐ Não sei responder

Quão importante você considera a proteção e privacidade dos seus dados? *

1 2 3 4 5
Pouco ou nada importante ☐ ☐ ☐ ☐ ☐ Extremamente importante

Quão importante você considera a proteção e privacidade os dados de terceiros? *

1 2 3 4 5
Pouco ou nada importante ☐ ☐ ☐ ☐ ☐ Extremamente importante

Você conhece e entende a diferença entre dados públicos, pessoais e sensíveis? *

- ☐ Sim
- ☐ Não
- ☐ Não sei responder

Sabe dizer quem a LGPD define como titular dos dados? *

- ☐ Sim
- ☐ Não
- ☐ Não sei responder

Dentro de sua organização, você consegue identificar quem possui o papel de "Controlador" perante a LGPD? *

☐ Sim

☐ Não

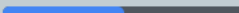
☐ Não sei responder

Você conhece quais são as multas ou sanções aplicadas com base na LGPD? *

☐ Sim

☐ Não

☐ Não sei responder

Próxima  Página 1 de 2 [Limpar formulário](#)



Pesquisa de análise de aderência das empresas à LGPD

rtclaro@gmail.com [Alternar conta](#)

*Obrigatório

Dentro de sua organização, você consegue identificar quem possui o papel de "Operador" perante a LGPD? *

☐ Sim

☐ Não

☐ Não sei responder

Dentro de sua organização, você consegue identificar quem possui o papel de "Encarregado de dados" perante a LGPD? *

☐ Sim

☐ Não

☐ Não sei responder

Você já participou em sua empresa, de treinamentos informativos, de orientação, formação ou reciclagem em "Segurança da Informação"? *

☐ Sim

☐ Não

☐ Não sei responder

Com base na questão anterior, qual a periodicidade que são realizados os treinamentos? *

☐ Não sei dizer

☐ 1 vez ao ano

Você tem acesso a uma cartilha ou manual, seja impresso ou online, com os processos, políticas e normas, estabelecidas dentro da sua organização visando atender e orientar em relação à proteção de dados pessoais? *

☐ Sim

☐ Não

☐ Não sei responder

Dentro de suas atividades diárias, você tem acesso a dados de terceiros, tais como: dados médicos, genéticos ou biométricos, financeiros, origem racial, étnica, convicção religiosa ou filosófica, opiniões políticas ou sindicais? *

☐ Sim

☐ Não

☐ Não sei responder

Já participou (ativamente ou passivamente) de algum processo de auditoria interna em sua companhia? *

☐ Sim

☐ Não

☐ Não sei responder

☐ Enviar uma cópia das respostas para o meu e-mail.

[Voltar](#) [Enviar](#) Página 2 de 2 [Limpar formulário](#)

ENGENHARIA: A MÁQUINA QUE CONSTRÓI O FUTURO

VOLUME XVI

 conhecimentolivre.org/home
 contato@conhecimentolivre.org
 [editoraconhecimentolivre](https://www.instagram.com/editoraconhecimentolivre)



EDITORA CONHECIMENTO LIVRE