



METODOLOGIA DE CÁLCULO DO IVPD

Índice experimental de Vulnerabilidade para Proteção Digital

Programação, fórmulas, pipeline de dados, validação e interpretação responsável

Iniciativa	Protege.Dados — Observatório de Proteção Digital Infantojuvenil
Autoria	Ana Maria Carvalheiro
Organização	Pretos Na Era Digital Ltda.
Local	Brasília/DF
Data	13 de julho de 2026

Documento técnico para transparência, auditoria e reuso responsável

1. Visão geral

O **IVPD** é um indicador sintético experimental criado para organizar sinais territoriais associados à proteção digital de crianças e adolescentes. A versão submetida combina três componentes: taxa de denúncias do Disque 100, déficit de escolas com internet e déficit de escolas com banda larga.

Importante: o IVPD não mede pessoas, não estima prevalência real, não prevê comportamentos, não identifica crianças ou adolescentes e não deve ser usado como ranking absoluto, punição ou mecanismo automático de distribuição de recursos.

1.1 Componentes e pesos da versão atual

Componente	Peso	Interpretação
Taxa de denúncias do Disque 100	50%	Sinal administrativo de demanda/registro; não equivale à prevalência real.
Déficit de escolas com internet	25%	Lacuna de conectividade básica na infraestrutura escolar.
Déficit de escolas com banda larga	25%	Lacuna de conectividade adequada na infraestrutura escolar.

1.2 Escala final

0 a 100, em que valores mais altos representam maior vulnerabilidade relativa dentro do mesmo conjunto de UFs, período, fontes e versão metodológica.

2. Etapas do cálculo

1

Ingestão — Leitura das bases oficiais e preservação dos arquivos de origem.

2

Padronização — Limpeza de nomes de colunas, códigos territoriais, tipos numéricos, datas e ausências.



3	Agregação — Consolidação dos registros por Unidade da Federação.
4	Construção dos indicadores — Cálculo de taxas, proporções e déficits.
5	Normalização robusta — Conversão dos componentes para uma escala comum entre 0 e 1.
6	Ponderação — Aplicação dos pesos 50%, 25% e 25%.
7	Conversão — Transformação do resultado final para a escala de 0 a 100.
8	Validação e publicação — Testes automáticos, revisão humana e exportação em CSV, JSON e API estática.

2.1 Taxa de denúncias por 100 mil habitantes

$$\text{Taxa de denúncias} = (\text{denúncias da UF} \div \text{população de referência da UF}) \times 100.000$$

O denominador populacional reduz a distorção causada pelo tamanho desigual dos estados. Uma UF com mais habitantes tende a ter maior número absoluto de registros, razão pela qual a comparação utiliza taxa.

2.2 Déficit de internet escolar

$$\text{Déficit de internet} = 100 - \text{percentual de escolas com internet}$$

2.3 Déficit de banda larga escolar

$$\text{Déficit de banda larga} = 100 - \text{percentual de escolas com banda larga}$$

3. Normalização min–max robusta

Os componentes possuem unidades distintas. Para combiná-los, o projeto utiliza uma normalização robusta, limitada pelos percentis 5 e 95. Esse procedimento reduz a influência de valores extremos.

$$N_i = (x_i - P5) \div (P95 - P5)$$

Depois do cálculo, o valor é limitado ao intervalo de 0 a 1: resultados abaixo do percentil 5 recebem 0; resultados acima do percentil 95 recebem 1; os demais são posicionados proporcionalmente.

3.1 Exemplo de normalização

Taxa observada da UF	65
Percentil 5	20
Percentil 95	100

Resultado normalizado	$(65-20)/(100-20)=0,5625$
------------------------------	---

3.2 Fórmula final

$$IVPD = 100 \times [(0,50 \times N \text{ denúncias}) + (0,25 \times N \text{ déficit internet}) + (0,25 \times N \text{ déficit banda larga})]$$

3.3 Exemplo completo

Taxa de denúncias normalizada	0,70
Déficit de internet normalizado	0,40
Déficit de banda larga normalizado	0,60
Média ponderada	$0,50 \times 0,70 + 0,25 \times 0,40 + 0,25 \times 0,60 = 0,60$

IVPD final	60,0
-------------------	-------------

4. Programação utilizada

Linguagem principal: Python.

Tecnologias complementares: pandas, NumPy, HTML, CSS, JavaScript, GitHub Actions, CSV, JSON e API estática.

4.1 Organização do código

Módulo	Função
readers.py	Lê arquivos das fontes oficiais.
processors.py	Limpa, transforma, padroniza e agrega os dados.
indicators.py	Calcula taxas, déficits, normalização e IVPD.
pipeline.py	Orquestra todas as etapas do processamento.
schema.py	Define os campos, tipos e estruturas esperadas.
quality.py	Executa verificações de consistência e qualidade.
utils.py	Reúne funções auxiliares reutilizáveis.
cli.py	Permite executar o pipeline pela linha de comando.

4.2 Exemplo de código

```
def calcular_ivpd(df):
    df["taxa_denuncias_100mil"] = (
        df["denuncias"] / df["populacao_referencia"] * 100_000
    )

    df["deficit_internet"] = 100 - df["perc_escolas_internet"]
    df["deficit_banda_larga"] = 100 - df["perc_escolas_banda_larga"]

    df["n_taxa_denuncias"] = normalizacao_robusta(
        df["taxa_denuncias_100mil"]
    )
    df["n_deficit_internet"] = normalizacao_robusta(
        df["deficit_internet"]
    )
    df["n_deficit_banda_larga"] = normalizacao_robusta(
        df["deficit_banda_larga"]
    )
```

```
df["ivpd"] = 100 * (  
    0.50 * df["n_taxa_denuncias"]  
    + 0.25 * df["n_deficit_internet"]  
    + 0.25 * df["n_deficit_banda_larga"]  
)  
  
return df
```

5. Validação, qualidade e reprodutibilidade

O processamento é determinístico: os mesmos arquivos de entrada, a mesma versão do código e os mesmos parâmetros produzem os mesmos resultados. Não há modelo de inteligência artificial, treinamento estatístico ou decisão automatizada.

- ✓ exatamente 27 Unidades da Federação;
- ✓ uma única linha por UF;
- ✓ pesos totalizando 100%;
- ✓ IVPD sempre entre 0 e 100;
- ✓ ausência de valores infinitos e NaN na publicação;
- ✓ JSON estritamente válido;
- ✓ rastreabilidade das fontes e das transformações;
- ✓ testes automatizados antes da publicação;
- ✓ registro de alterações metodológicas em changelog.

5.1 Testes programáticos exemplificativos

```
assert base["uf"].nunique() == 27  
assert not base.duplicated("uf").any()  
assert base["ivpd"].between(0, 100).all()  
assert base["uf"].notna().all()  
assert abs(0.50 + 0.25 + 0.25 - 1.0) < 1e-9
```



5.2 Saídas produzidas

- CSV aberto
- JSON
- API estática
- painel web
- documentação metodológica
- dicionário de dados
- código-fonte e testes

6. Interpretação responsável

O resultado representa vulnerabilidade relativa no mesmo escopo, período e versão. Ele deve ser lido junto com os componentes, a qualidade das fontes, o contexto territorial e evidências qualitativas.

Uso recomendado	Uso vedado
apoio à análise exploratória;	ranking absoluto de estados;
comparação contextual entre UFs;	diagnóstico individual;
identificação de temas para aprofundamento;	inferência causal isolada;
planejamento de pesquisas e políticas;	punição ou sanção;
transparência e controle social.	distribuição automática de recursos;
	previsão de comportamento.

6.1 Limitações fundamentais

- Registros do Disque 100 refletem relatos recebidos, não a prevalência real da violência.
- Taxas podem ser influenciadas por acesso desigual aos canais, campanhas, cultura de denúncia e subnotificação.
- Internet e banda larga declaradas pelas escolas não asseguram qualidade, estabilidade ou velocidade efetiva.
- Mudanças de fonte, período, normalização ou pesos exigem nova versão metodológica.



6.2 Evolução metodológica

A versão atual utiliza três componentes. Versões futuras poderão ampliar o modelo para dimensões adicionais ou para uma matriz de dois eixos (contexto de vulnerabilidade e capacidade territorial de proteção), desde que haja validação, análise de sensibilidade e documentação pública.

7. Texto institucional pronto para cadastro

A metodologia do Protege.Dados é implementada em Python, com uso de bibliotecas de tratamento e análise de dados, especialmente pandas e NumPy. As bases oficiais são lidas, padronizadas, agregadas por Unidade da Federação e integradas por chaves territoriais. Calcula-se a taxa de denúncias por 100 mil habitantes, o déficit de escolas com internet e o déficit de escolas com banda larga. Os componentes são normalizados em escala de 0 a 1 pelo método min-max robusto, com limites nos percentis 5 e 95. O IVPD experimental corresponde à média ponderada desses componentes: 50% para a taxa de denúncias, 25% para o déficit de internet escolar e 25% para o déficit de banda larga. O resultado é convertido para escala de 0 a 100. O processamento é determinístico, documentado, auditável e reproduzível, sem aprendizado de máquina, identificação de pessoas ou decisões automatizadas.

8. Identificação técnica

Projeto

Protege.Dados — Observatório de Proteção Digital Infantojuvenil

Autora e responsável técnica

Ana Maria Carvalheiro

Organização

Pretos Na Era Digital Ltda.

Linguagem principal

Python

Data desta versão

13/07/2026

Finalidade

Transparência, auditoria, reuso de dados abertos e apoio à análise territorial