Title: Ethical Use & Model Card for 500 Credit Score Dataset

Model Overview:
- Model Objective: Credit scoring for thin-file consumers using synthetic data.
- Model Type: Binary Classification (Default/No Default).
- Input Features: Alternative data (utility payment, telecom behavior, social media, gig economy), demographics, financial stability indicators.
- Output: Probability of loan default.

Intended Use:
- Research on machine learning applications in credit scoring.
- Fairness evaluation and AI ethics research.
- Educational purposes in responsible AI development.

Performance Metrics (Synthetic Data):
- Accuracy: 85%
- ROC-AUC: 0.89
- Fairness Metrics:
  - Demographic Parity Difference: 0.06
  - Equal Opportunity Difference: 0.04

Ethical Considerations:
- Dataset is fully synthetic; no personal data is used.
- Alternative data sources can carry biases linked to socioeconomic factors.
- Fairness assessments using SHAP, demographic parity, and equal opportunity metrics are recommended.
- Users must ensure models are validated on real-world data before deployment.

Limitations:
- Performance is based on synthetic data and does not reflect real-world deployment readiness.
- Proxies like social media footprint are suggestive but not precise indicators.
- The model may not capture regional or institutional credit risk variations.

Licensing & Citation:
- License: CC0 1.0 Universal for dataset; CC BY 4.0 for documentation.
- Citation: Shukla, Deepa (2024). 500 Credit Score Dataset for Digital Credit Scoring of Thin-File Consumers. Harvard Dataverse. https://doi.org/10.7910/DVN/6MLVVI

Contact:
Deepa Shukla
Jaipur National University
Email: deepashukla@live.com