3.3 Model Limitations and Constraints

The documentation explicitly acknowledges several limitations of the model, which is a strength of the documentation. Key limitations include:

3.3.1 Data Limitations

- **Sample Size:** The 300-loan sample is insufficient for a robust credit risk model, particularly given the complexity of the gradient boosting algorithm.
- **Approved Loans Only:** The model is trained only on approved loans, creating a selection bias that may limit its applicability to the broader applicant population.
- **Binary Outcome:** The simplification to a binary outcome (Default/Fully Paid) does not capture the nuances of loan performance, such as early payoffs or late payments.
- **Temporal Information:** The lack of temporal data prevents proper time-based validation, which is critical for credit risk models.

3.3.2 Model Limitations

- **Binary Classification:** The documentation acknowledges that binary classification simplifies the complex nature of credit risk assessment.
- **Interpretability:** Gradient boosting models have limited interpretability compared to simpler models like logistic regression.
- **Validation Approach:** The random split validation approach, rather than cross-validation or temporal validation, is acknowledged as suboptimal.
- **Feature Interactions:** Limited exploration of feature interactions is noted as a constraint.

Table 3.2: Summary of Model Performance Boundaries

Feature	Optimal Range	Extrapolation Risk
FICO Score	580-850	High risk below 580
DTI	0-150%	High risk above 150%
Annual Income	$20,000-$300,000	Moderate risk outside range
Loan Amount	$1,000-$40,000	High risk above $40,000