The probability of default model leverages data from a Kaggle dataset which aims to improve the accuracy of predicting financial distress within a two-year period. Here's a brief overview of how this model works and the performance achieved:

Data and Variables
The dataset consists of historical data on 250,000 borrowers, with various attributes that influence creditworthiness, including:
- **Revolving Utilization of Unsecured Lines**: Ratio of unsecured revolving credit to total credit available.
- **Age**: Borrower's age.
- **Number of Times 30-59 Days Past Due**: Instances of being 30-59 days late on a payment in the last two years, but not worse.
- **Debt Ratio**: Ratio of monthly debt payments to monthly gross income.
- **Monthly Income**: Borrower's monthly income.
- **Number of Open Credit Lines and Loans**: Total open loans and credit lines.
- **Number of Times 90 Days Late**: Instances of being 90 or more days late on a payment.
- **Number of Real Estate Loans or Lines**: Total mortgage and real estate loans, including home equity lines.
- **Number of Times 60-89 Days Past Due**: Instances of being 60-89 days late on a payment in the last two years, but not worse.
- **Number of Dependents**: Number of family dependents, excluding the borrower.

Model Selection
A **Random Forest Classifier** was chosen for its ability to handle both classification and probability estimation. This method is beneficial for its robustness and ease of handling various data types and correlations without extensive parameter tuning.

Strategy and Performance
The approach primarily uses quantiles to categorize credit risk efficiently and effectively. The model achieved a commendable accuracy rate of 80.05% on Kaggle's test set. This rate is notable given the use of a straightforward quantile-based approach without parameter optimization.

Conclusion
This model demonstrates a solid foundation for predicting the likelihood of a borrower facing severe financial distress within two years, using a relatively simple yet effective machine learning strategy. It highlights the balance between complexity and performance, making it a practical solution for real-world applications in credit scoring.