**Home Credit - Credit Risk Model Stability Project Proposal**

**Introduction**

The project proposal aims to address the challenge of predicting loan default risk for clients with limited credit history. By leveraging machine learning techniques, we seek to develop a robust and stable model that consumer finance providers can use to make more informed lending decisions. This proposal follows the structure outlined in Jeremy Jordan's Machine Learning Projects Guide.

**Dataset+competition** link: <https://www.kaggle.com/competitions/home-credit-credit-risk-model-stability/data>

**Motivation**

The motivation behind this project is to improve financial inclusion by enabling consumer finance providers to assess default risk accurately. Many individuals with little or no credit history face challenges in accessing loans due to traditional risk assessment methods. By developing a predictive model, we aim to bridge this gap and make loans more accessible to those who may benefit from them.

**Methodology**

Our approach involves the following steps:

**Data Collection:** We will utilize the Home Credit dataset from the Kaggle competition "Home Credit Credit Risk Model Stability."

**Data Preprocessing:** This step includes data cleaning, feature engineering, handling missing values, and scaling features as needed.

**Model Selection:** We will explore a range of machine learning algorithms such as Random Forest, Gradient Boosting, Logistic Regression, LightGBM and XGBoost.

**Model Training and Evaluation:** Models will be trained on historical data and evaluated using metrics such as accuracy, precision, recall, and the gini stability metric.

**Hyperparameter Tuning:** We will optimize model performance through techniques like grid search or Bayesian optimization.

**Model Interpretation:** Emphasis will be placed on creating interpretable models to understand the factors contributing to loan default risk.

**Model Deployment:** The final model will be deployed in a scalable and production-ready environment for real-time risk assessment.

**Intended Experiments**

Our experiments will focus on:

* Comparing the performance of different machine learning algorithms.
* Investigating feature importance and its impact on model predictions.
* Assessing model stability over time using the gini stability metric.
* Evaluating the trade-offs between model stability and performance.

Evaluation

The primary evaluation metric will be the gini stability metric, which considers the predictive stability of the model over time. We will also consider traditional evaluation metrics such as accuracy, precision, recall, and area under the ROC curve (AUC) to ensure a comprehensive assessment of model performance.

**Dataset and Prior Research**

The Home Credit dataset provides a rich source of information for training and testing our models. Prior research in credit risk assessment, feature engineering, model stability, and interpretability will guide our approach and help us leverage existing knowledge in the field.

**Project Impact**

By participating in this competition and developing an effective loan default risk prediction model, we aim to:

Enable consumer finance providers to make more informed lending decisions.

Improve financial inclusion by offering loans to individuals with limited credit history.

Contribute to the ongoing research and development of machine learning applications in finance.

**Conclusion**

In conclusion, this project proposal outlines our plan to tackle the challenge of predicting loan default risk using machine learning techniques. We are excited about the potential impact of our work and look forward to contributing to the field of financial technology and inclusion.