**Home Credit Default Risk Prediction**

**Project Proposal Report**

**Submitted by**

Albin Jose (PES2PGE23MB032)

Bindushree R (PES2PGE23MB078)

Pooja L (PES2PGE23MB089)

Shrikant P Yalakki (PES2PGE23MB068)

Tarun Raj (PES2PGE23MB074)

Course

Executive Master of Business Administration (MBA)

**Institution**

PES University

**Date of Submission**

[Insert Date]

**Introduction**

This project aims to explore the Home Credit Default Risk dataset to identify patterns and factors influencing loan defaults. Financial institutions are increasingly relying on data-driven decision-making to minimize credit risk. This proposal lays the foundation for developing a predictive model that assists in classifying high-risk applicants.

**Problem Statement**

Home Credit aims to make credit available to underbanked populations. However, assessing the risk associated with loan applicants is challenging due to limited and diverse data. The problem is to predict whether a loan applicant will default, based on demographic, employment, and financial attributes.

**Project Objective**

* To analyse key factors influencing loan repayment behaviour.
* To build a classification model predicting the likelihood of default.
* To evaluate the model using accuracy, recall, precision, and F1-score.
* To provide actionable insights to help reduce loan losses.

**Expected Outcomes**

The project is expected to deliver:

* A cleaned and well-understood dataset.
* A predictive model that identifies high-risk applicants.
* Insightful visualizations highlighting trends in loan defaults.
* Business recommendations based on data patterns.

**Dataset Description**

The dataset used in this project is the Home Credit Default Risk dataset, sourced from Kaggle. It contains historical loan application data with both demographic and financial details of applicants. The primary CSV file includes the following variables:

|  |  |  |
| --- | --- | --- |
| Sl. No | Column Name | Description |
| 1 | SK\_ID\_CURR | Unique ID for each loan application. |
| 2 | TARGET | Binary indicator (1: default, 0: repaid). |
| 3 | CODE\_GENDER | Gender of the applicant. |
| 4 | FLAG\_OWN\_CAR | Whether the applicant owns a car. |
| 5 | AMT\_INCOME\_TOTAL | Total annual income |
| 6 | AMT\_CREDIT | Loan credit amount. |
| 7 | NAME\_CONTRACT\_TYPE | Type of loan. |
| 8 | NAME\_INCOME\_TYPE | Source of income. |
| 9 | DAYS\_EMPLOYED | Number of days employed. |
| 10 | OCCUPATION\_TYPE | Applicant's profession. |

The dataset is structured and spans multiple files, but the core features reside in the application data. These attributes help in understanding client behaviour and building a reliable model.