**Real-Time Credit Risk Assessment Using Alternative Data**

The Objective of this project is to develop a system that assesses credit risk in real time by analysing traditional financial data and alternative data sources like social media, utility payments, and spending habits.

**Objectives**

**1. Develop a data pipeline:**

Objective: Collect and integrate the traditional and alternative data sources in real-time.

**Data Sources:**

Traditional: credit scores, income statements, loan repayment history

**Alternative:**

Social media sentiment: Scan posts and activity to determine financial stability.

**Geolocation data:** Mode of moving to evaluate lifestyle and employment consistency.

**Utility Payment Records:** Record of timely payments either from utilities, rents, or subscriptions.

**Break down:**

* Design and implement a robust data collection pipeline.
* Centralize and clean data using a data lake or warehouse.
* Generate relevant features from both traditional and alternative datasets.
* Ensure compliance with data privacy regulations like GDPR or CCPA.

**2. Train a machine learning model**

Objective: Predict loan default risk based on the mix of the two data types.

Approach: Preprocess data that has missing values and normalize features**.**

**Choose the models from:**

Ensemble models (Random Forest, Gradient Boosting).

Neural networks for more complex non-linear relations.

**Asses model performance testing against important metrics:**

Accuracy to reflect general model correctness.

AUC-ROC for classification quality.

Precision-recall for an imbalanced dataset.

Hyperparameter tuning must be executed to improve performance.

**Deliverables:** A model capable of providing real-time credit risk assessments.

**3. Ensure explainability**

**A goal:** Credit risk score that is interpretable and understandable by lenders.

**Ways:** Employ tools such as SHAP (Shapley Additive Explanations) or LIME (Local Interpretable Model-Agnostic Explanations) to justify decisions of the models.Spotlight on attributes affecting credit worthiness.

**Generate a visualization and reporting structure enough to present results to lenders.**

**Deliverables:** An interface/dashboard showing the credit risk scores and some insight behind it.

**Output Expectations:** A robust and timely credit risk assessment system that bridges traditional and alternative data.Increased credit access for individuals with little to no credit histories.Transparent and elucidated credit risk scores that can instill trust in lenders.

References:

1. **Alternative Data for Credit Risk Management: An Analysis of the Literature** [**Jan Roeder, *University of Goettingen, Faculty of Business and Economics, Goettingen, Germany***](https://aisel.aisnet.org/do/search/?q=bp_author_id%3A%22da406baa-4bb0-4c87-9b6a-528113b8698e%22%20OR%20%28author%3A%22Jan%20Roeder%22%20AND%20-bp_author_id%3A%5B%2A%20TO%20%2A%5D%29&start=0&context=509156)
2. **Enhancing Credit Scoring with Alternative Data**
3. **AI & Alternative Data: Redefining Credit Scoring**