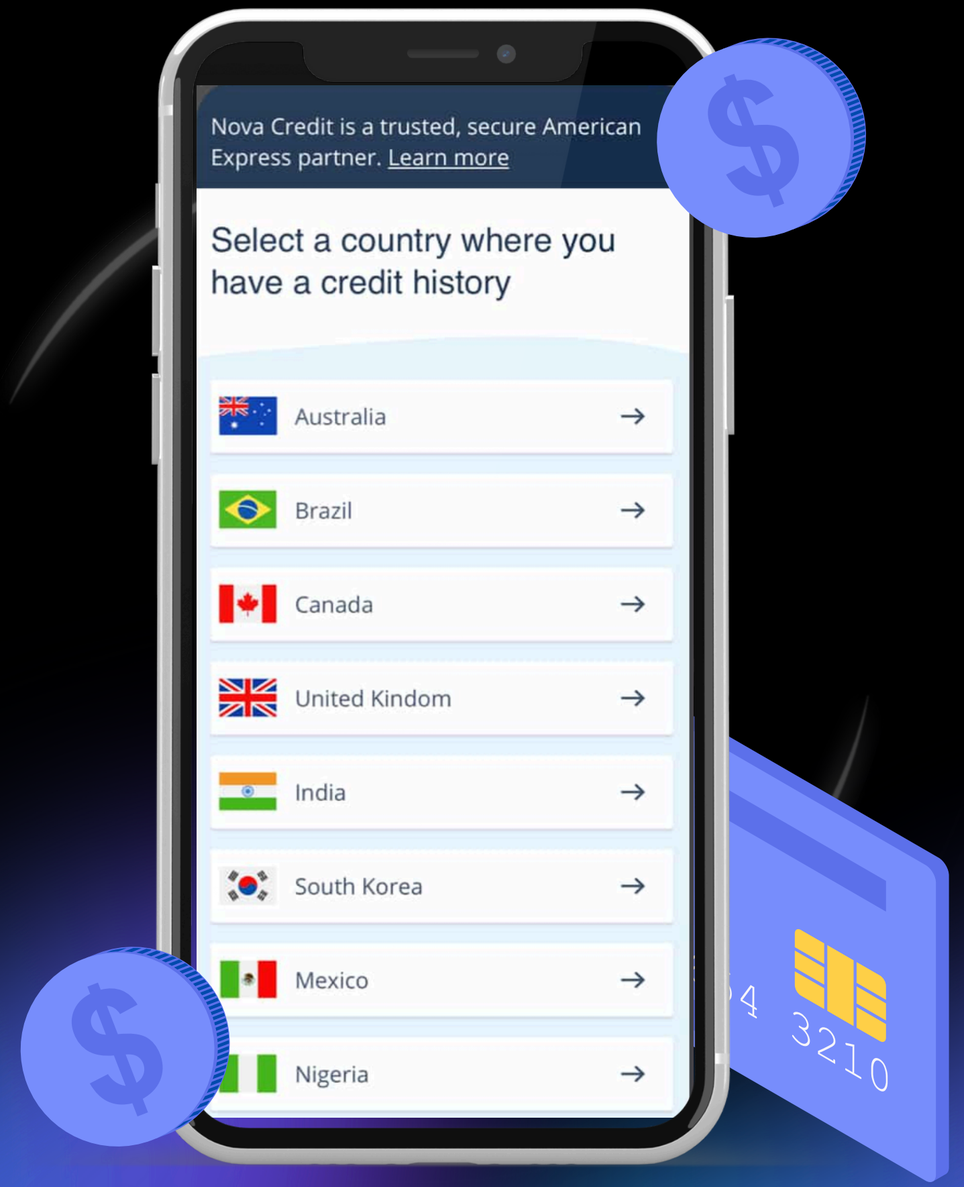
CreditDekho –

‘Score ki duniya, ab aapki muthi mein’

Inspiration: Nova Credit

Get Started →



The Power to Underwrite Anyone

What is Nova Credit?

Nova Credit is a U.S.-based fintech startup that enables immigrants and newcomers to carry their credit history across borders. Its Credit Passport® converts international credit data into a format local lenders can understand and use.

How It Works?

- Connects to 20+ global credit bureaus via API
- Translates foreign credit history into a U.S.-equivalent score
- Lenders use this to evaluate new residents just like local applicants

Business Model

- Credit Report Access Fees
- Commission from product onboarding
- API-based tools like Cash Atlas™ & Income Navigator

CreditDekho

PROBLEM:

Over 100 million Pakistanis lack access to formal credit due to no credit history or informal employment.

SOLUTION:

CreditDekho – a smart credit scoring platform using mobile behavior, digital activity, and demographic data.

OBJECTIVE:

Provide alternative credit scoring to financial institutions and fintechs to enhance financial inclusion.

Target Audience: Freelancers, gig workers, students, and informal sector workers.

Data & Features

TRADITIONAL CREDIT DATA (BASELINE)

Give Me Some Credit – Kaggle
150,000+ real borrower records with 10 classic credit features
Ground truth for modeling delinquency and establishing risk patterns

ALTERNATIVE DATA SOURCES

1. Mobile Device Usage & Behavior Dataset
- Inspired telecom features: mobile bill payments, usage patterns, recharge consistency

2. US Census Demographic Data
- Informed social features: activity level, friend network, engagement metrics

AUGMENTED TRAINING DATASET (10,000 ROWS)

- Combined real + synthetic records
- Ensured diversity, statistical realism, and balance
- Modeled edge-case scenarios using telecom & social media signals

Machine Learning Models

1. **Random Forest** **2. XGBoost** **3. CatBoost** **4. LightGBM** **5. SVM** **6. KNN**

Aspect	Insight
Class Distribution	Imbalanced (71% vs 29%) – focus on recall/precision for class 0
Missing Values	Primarily in `MonthlyIncome` – impute using median
Outliers	Present in income and debt ratio – consider filtering or capping
Feature Redundancy	App behavior and delinquency variables are highly correlated
Feature Separation	Age and Debt Ratio show strongest separation with creditworthiness

Results & Justification

Top Models by LogLoss and Class 0 Recall

Rank	Model	Preprocessing	LogLoss	Recall_0	ROC_AUC
1	RandomForest	Raw + NoOutlier + Mean + NoBalance + NoScale	0.167	0.909	0.972
2	XGBoost	Raw + NoOutlier + Mean + NoBalance + NoScale	0.244	0.898	0.961
3	RandomForest	FE + IQRRemove + Median + NoBalance + NoScale	0.302	0.731	0.915

These models performed the best in terms of generalization, achieving low log loss while maintaining high recall for the defaulter class.

- Raw data with no balancing or scaling consistently outperformed oversampling-based techniques in this case.
- Models using SMOTE or SMOTEENN generally had worse log loss and tend to overfit the majority or minority class.
- Feature-engineered data, while helpful in some configurations, did not universally improve blind performance.

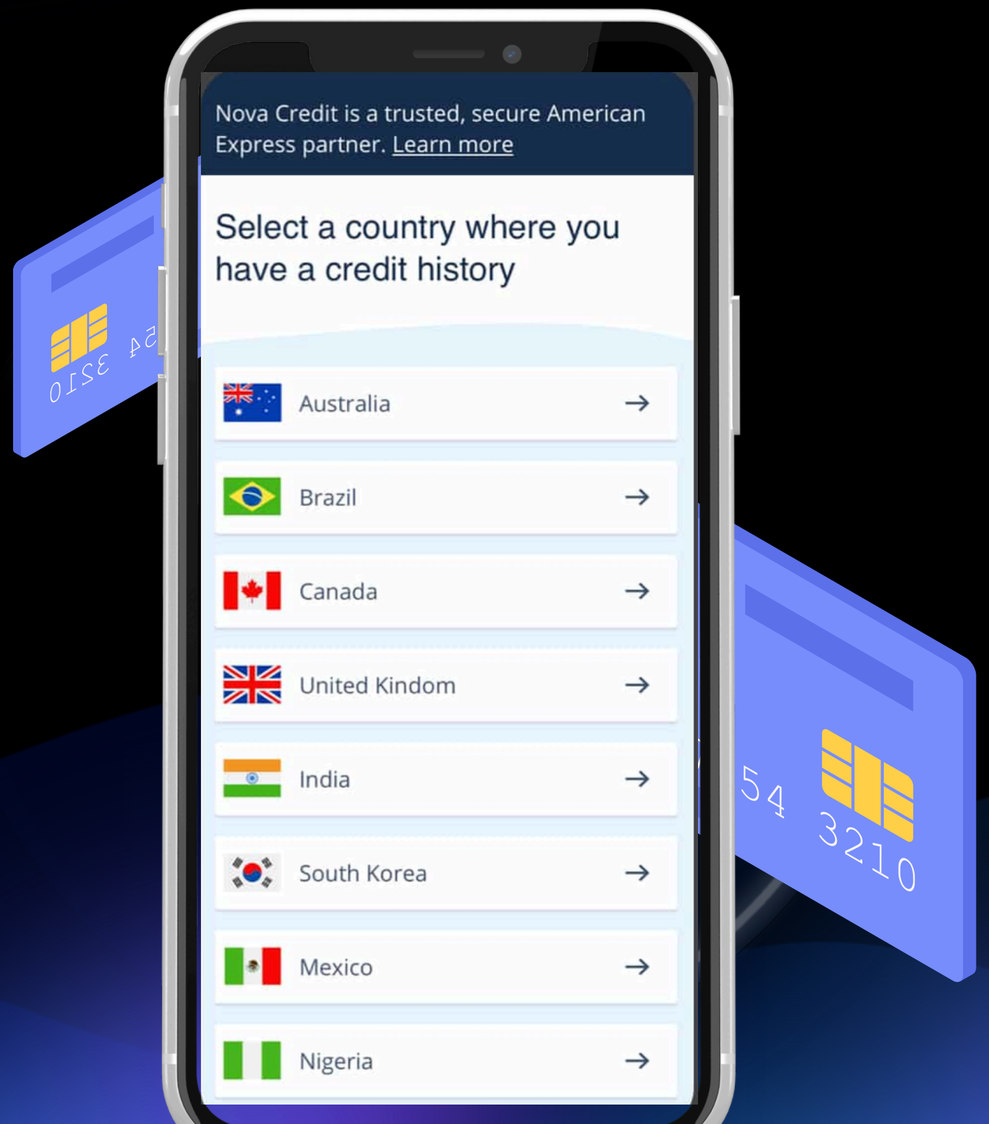
Real-World Impact & Defensibility

Use Cases

- Microloans to riders, freelancers
- Credit integration with platforms like Easypaisa, Bykea
- Student and women inclusion in credit economy

Defensibility:

- Local data partnerships (telcos, utilities, NADRA)
- Custom scoring algorithms tailored to Pakistani users
- Scalable nationwide with urban to rural roadmap



‘Score ki duniya, ab aapki muthi mein’

Thank You!