Proposed Fintech Lending Model: Machine Learning for multi-class classification of applicants

Dhara Khamar

Date: 14 January 2024

Abstract

This report outlines a pioneering business model for a fintech lending application, leveraging machine learning for enhanced risk management and multi-class classification of loan applicants. Integrating advanced algorithms, the model refines risk assessment, credit scoring, and decision-making processes. A focal point is the implementation of multi-class classification to tailor financial products based on nuanced risk categories. The report also addresses the necessary technological infrastructure, emphasizing data security, model interpretability, and regulatory compliance. This innovative business model aims to redefine the fintech lending landscape, ensuring precision in risk evaluation and delivering a personalized financial experience for diverse applicants.

1. Introduction

Context:

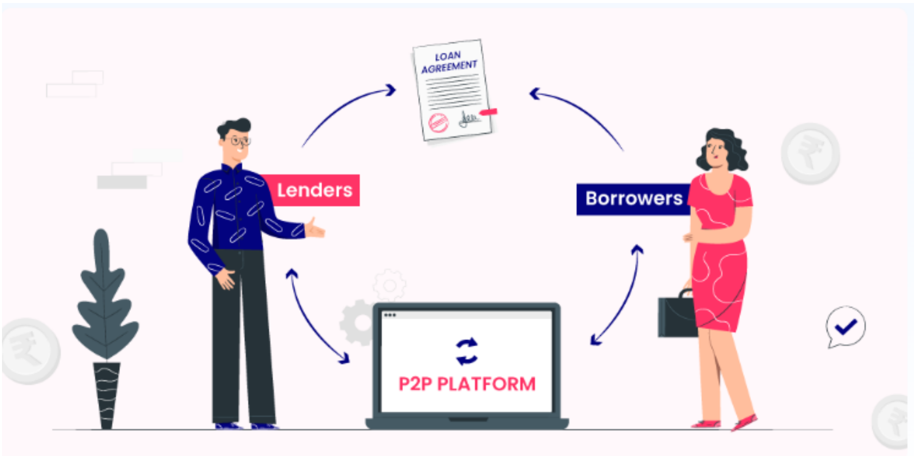
Fintech is a portmanteau of the words "financial" and "technology".

It refers to the use of technology to provide and enhance financial services. It encompasses a wide range of innovations, applications, and business models that aim to improve and streamline financial activities, making them more efficient, accessible, and cost-effective.

Lending fintech, also known as peer-to-peer (P2P) lending or online lending, involves the use of technology to facilitate the borrowing and lending of money directly between individuals or between individuals and businesses.

Peer-to-peer lending or crowd lending is a form of debt financing wherein borrowers can request a loan from another individual without the need for any financial institutions to act as the intermediary.

the underbanked/unbanked populace and small/micro businesses often face difficulty getting approved with organized credit. This makes P2P lending an attractive option for many.



Financial inclusion has been a critical issue for a country like India, where nearly 70% of the population resides in rural areas. But high smartphone penetration combined with the introduction of infrastructures such as Aadhar, UPI, Digilocker, eKYC, eSign, BHIM, and Indiastack has enabled many P2P lenders.

Although P2P loans mainly constitute personal loans, borrowers have other reasons like Financing their education, buying real estate, Debt refinancing, to get a secured business loan, Loan for machinery. India's top p2p platforms are,

finzy finance is easy
FAI₹CENT EVERY % COUNTS
Lendbox

LenDen club BY INNOFIN SOLUTIONS
TALA
LENDINGKA₹T Simplifying MSME Finance

Purpose:

Lending fintech, particularly in the peer-to-peer (P2P) lending space, comes with various risks. It's important for both investors and borrowers to be aware of these potential risks.
Here are some key risks associated with lending fintech:

- Credit Risk (Default Risk)
- Interest Rate and Inflation Risks
- Regulatory and Legal Risks
- Operational Risks
- Market Risks
- Liquidity Risks
- Reputational Risks
- Fraud and Identity Theft
- Concentration Risks
- Cybersecurity Risks

- **Credit Risk:** Borrowers may fail to repay their loans, leading to financial losses for lenders. This risk is inherent in any lending activity and is influenced by the borrower's creditworthiness.

- **Mismatch interest rate:** If the cost of funding for the platform rises unexpectedly or if there's a mismatch between fixed-rate loans and variable-rate funding, it can impact profitability.

The core objective of this business model is to furnish a Machine Learning (ML) and Artificial Intelligence (AI) based solution aimed at mitigating risks inherent in lending fintech.
By leveraging advanced algorithms and predictive analytics, the intention is to systematically address and diminish the various risks associated with financial transactions within the lending fintech sector.

Scope:

The overarching scope of this business model is to deliver Artificial Intelligence (AI) and Machine Learning (ML) based solutions specifically tailored for addressing two paramount dimensions of risk within the lending fintech domain: default or credit risk and interest rate risk.
Through the strategic deployment of advanced algorithms and predictive modeling, the objective is to provide a comprehensive risk management framework that adeptly assesses and mitigates uncertainties associated with borrower defaults and creditworthiness, as well as fluctuations in interest rates.

Objective:

- **Classification of Loan Applicants:**
  The primary goal is to categorize loan applicants into three distinct classes—namely, "Good," "Standard," and "Bad"—by employing a comprehensive analysis of their credit scores.
- **Interest Rate Prediction:** This predictive model aims to provide accurate and personalized interest rate assessments reflective of the individual applicant's creditworthiness.

2. Market/ Customer / Business Need Assessment

Business Needs Assessment:

- Strategic Objectives: Define long-term goals considering market share, profitability, and growth.
- Regulatory Compliance: Understand regulatory landscape, establish protocols for legal compliance.
- Technology Infrastructure: Evaluate and enhance technology for efficiency, security, and user experience.
- Risk Management: Develop a robust framework covering credit, operational risks; implement measures to minimize defaults and fraud.
- Financial Model: Establish a scalable financial model, identify key performance indicators (KPIs).
- Scalability Planning: Plan for growth in user numbers and transaction volumes.
- Innovation: Foster a culture of innovation to adapt to market dynamics.
- Customer Acquisition: Develop strategies for attracting and retaining borrowers and lenders.

Market Needs Assessment:

- Target Audience: Analyze characteristics and preferences of borrowers and lenders.
- Competitive Analysis: Identify unique selling points and market positioning.
- Market Trends: Stay informed, identify emerging opportunities and gaps.
- Customer Behavior: Analyze user patterns and align platform design with market expectations.
- Financial Inclusion: Assess impact on underserved populations.
- Interest Rate Sensitivity: Understand market sensitivities and develop competitive interest rate offerings.
- Marketing and Branding: Develop strong marketing strategies using digital channels.
- User Education: Create educational materials to increase awareness about P2P lending.

Customer Need Assessment:

- User Interviews: Gather insights directly from potential users.
- Demographic Analysis: Identify specific needs based on demographic characteristics.
- Financial Literacy: Assess level of financial literacy and provide educational resources.
- Accessibility and Inclusivity: Evaluate platform accessibility for users with varying technological proficiency.
- Risk Perception: Understand and address users' perceived risks to build trust.
- Interest Rate Sensitivity: Investigate sensitivity to interest rates and design transparent structures.
- Loan Purpose Preferences: Identify common user needs and customize products accordingly.
- Ease of Use: Streamline user interface for a convenient lending process.
- Communication Preferences: Understand how users prefer communication and their expectations for customer support.
- Transparency: Ensure transparency in terms, conditions, fees, and communication.
- Financial Goals Alignment: Understand users' financial goals.

3. Target Specifications and Customer Characterization

Understanding the characteristics of both lenders and borrowers is essential for the successful operation of a peer-to-peer (P2P) lending platform. Here are key characteristics for both groups:

Lender's characteristics:

- Risk Appetite: Assess willingness for risk.
- Investment Horizon: Identify short or long-term preferences.
- Diversification: Check interest in spreading investments.
- Capital: Note the amount they're willing to invest.
- Experience Level: Consider experience in P2P lending.
- Returns Expectation: Understand ROI expectations.
- Social Impact Goals: Identify any impact-driven preferences.
- Loan Categories: Determine preferences for specific loan types.
- Technology Comfort: Assess proficiency with technology.

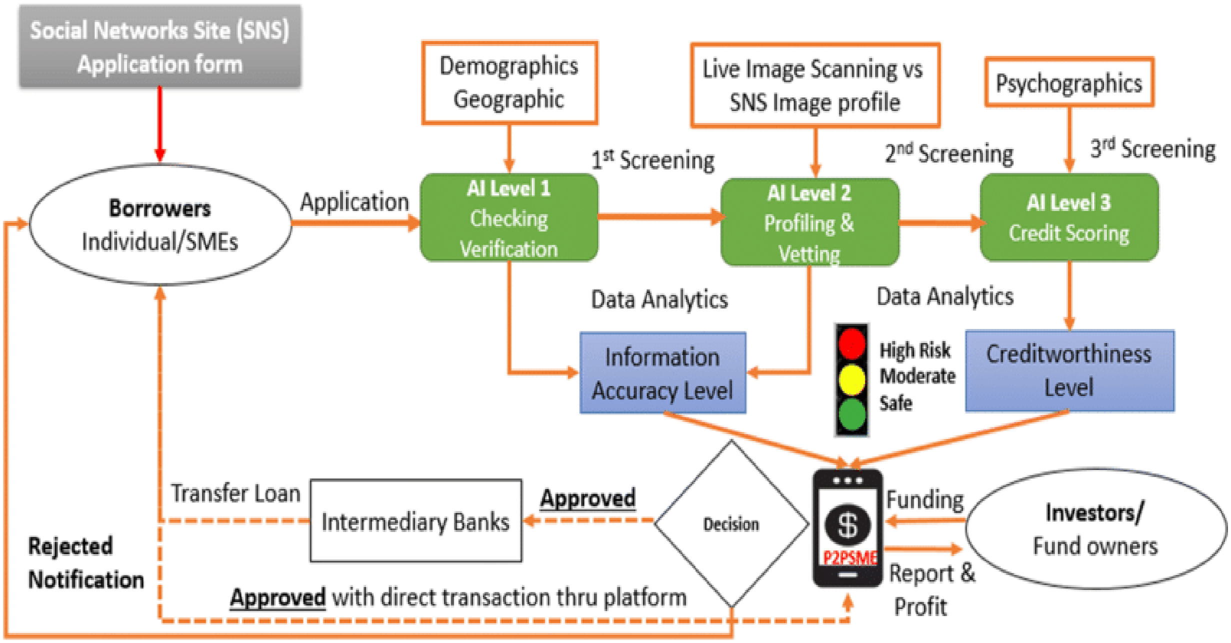
Borrowers' characteristics:

- Creditworthiness: Evaluate credit history and scores.
- Loan Purpose: Understand the purpose of the loan.
- Income Levels: Analyze income for repayment ability.
- Employment Status: Identify employment type for financial stability.
- Loan Amounts: Note preferred loan sizes.
- Loan Tenure: Understand tenure preferences.
- Financial Goals: Gain insights for personalized solutions.
- Technology Adoption: Assess digital comfort.
- Communication Preferences: Understand preferred communication.
- Interest Rate Sensitivity: Investigate sensitivity to rates and fees.
- Repayment Behavior: Analyze past repayment history.

4. External Search

The main challenge for P2P lending is on managing risks. FinTech with artificial intelligence (AI) can be used as a strategic tool in mitigating risks for FinTech companies in assessing creditworthiness of a potential customer. However, AI-enabled assessment has created several ethical issues and dilemmas for the stakeholders in the industry.

An AI-enabled risk assessment will automate processes in understanding potential applicants for P2P lending.



Key application areas of AI/ML in fintech lending

- **Credit Scoring**: Utilize AI for comprehensive creditworthiness assessment with non-traditional data.
- **Fraud Detection**: Implement real-time machine learning models to identify and prevent fraudulent activities.
- **Predictive Analytics for Default Prevention**: Develop early-warning models for potential defaults in lending.
- **Risk Management**: Use predictive models to assess and manage market and operational risks.
- **Personalized Loan Offers**: Employ machine learning to tailor interest rates and terms based on customer behaviour and financial history.
- **Chatbots and Customer Support**: Enhance customer service with AI-powered chatbots for instant support.
- **Behavioural Analytics**: Analyze user behaviour to predict and prevent potential defaults.
- **Dynamic Pricing Models**: Develop machine learning models for real-time adjustment of interest rates based on market conditions and borrower risk.
- **Portfolio Management**: Continuously analyse and adjust investment strategies based on market trends and risks.
- **Customer Segmentation**: Use machine learning to segment customers based on financial behaviour and needs.

5. Benchmarking

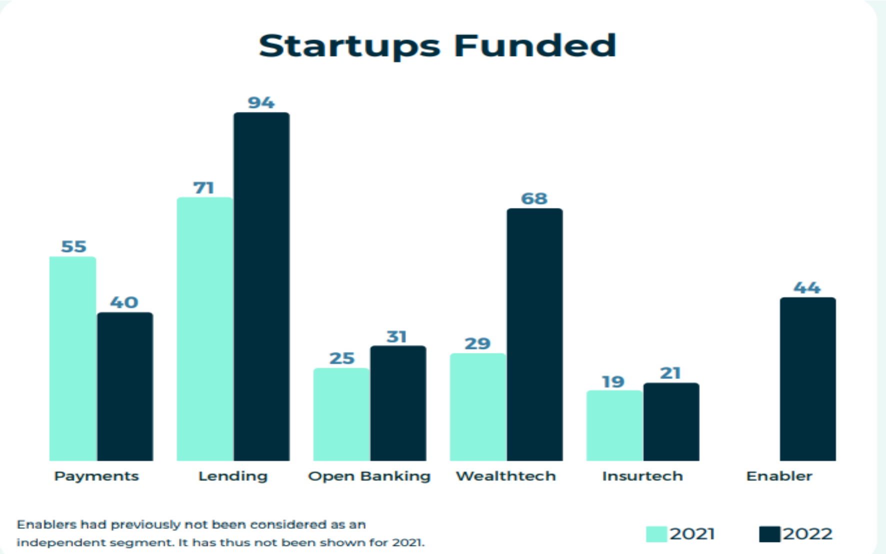
Top Indian Fintech funding 2022 -2023

RANK	STARTUP	SEGMENT	AMOUNT	ROUND	INVESTORS
1	CRED	Payments	USD 348 M	Series E,F	GIC, TIGERGLOBAL, ALPHA WAVE, RISE CAPITAL, SOFINA, DF International, DRAGONEER
2	Chargebee	Payments	USD 250 M	Series H	TIGERGLOBAL, SEQUOIA, INSIGHT PARTNERS, STEADVIEW, SAPPHIRE VENTURES
3	pine labs	Payments	USD 220 M	Series F	NORDMANN ALPHA WAVE, TEMASEK, actis, PayPal, VITRUVIAN, CHIMERA, SBI, SEQUOIA
4	FPL Technologies	Payments	USD 203 M	Series D	TEMASEK, matrix PARTNERS, GIC, SEQUOIA, QED INVESTORS, Ocean View Investments Group, Hummingbird
5	Oxyzo	Lending	USD 200 M	Series A	ALPHA WAVE, TIGERGLOBAL, matrix PARTNERS, NORWEST, CREATION INVESTMENTS CAPITAL MANAGEMENT, LLC
6	moneyview	Lending	USD 166 M	Series D,E	apis partners, TIGERGLOBAL, evolvence India, Winter Capital, TrustedInsight, Accel, South Park Commons, DI, The Sparkle Fund, TI PLATFORM
7	Yubi	Lending	USD 137 M	Series B	INSIGHT PARTNERS, B Capital, DRAGONEER, SEQUOIA, Lightspeed, TVS CAPITAL FUNDS (P) LIMITED, lightrock
8	Niyo	Open Banking	USD 130 M	Series C	lightrock, Accel, MULTIPLES, PRIME VENTURE PARTNERS, JSCapital, BEAMS FINTECH FUND
9	digit	Insurtech	USD 125 M	Series D	IIFL, SEQUOIA, WELLINGTON MANAGEMENT
10	turtlemint	Insurtech	USD 120 M	Series E	AMANSA capital, JUNGLE, nexus venture partners, VITRUVIAN, MARSHALL WACE, GGVCAPITAL, AMERICAN FAMILY VENTURES, BLUME, DI, SIG GLOBAL
11	Fibe	Lending	USD 112 M	Series C,D	RISE, INNOVEN CAPITAL, alteria capital, NORWEST, Piramal Finance
12	kissht	Lending	USD 107 M	Series D,E	INNOVEN CAPITAL, vertex GROWTH, THE SOLVENT EXTRACTORS' ASSOCIATION OF INDIA, Endiya, TRIFECTA CAPITAL
13	Insurance Dekho	Insurtech	USD 100 M	Series A	Data Unavailable
14	innoviti	Payments	USD 88 M	Series C,D	Bessemer Venture Partners, PGP PANTHERA GROWTH PARTNERS, PATNI FINANCIAL, PIVOT ventures, FMO, Alumni Ventures
15	INDmoney	Wealthtech	USD 87 M	Series D	TIGERGLOBAL, STEADVIEW, DRAGONEER, SIXTEENTH STREET CAPITAL
16	vivriti CAPITAL	Lending	USD 85 M	Series C	lightrock, LIGHTSTONE VENTURES, TVS CAPITAL FUNDS (P) LIMITED, CREATION INVESTMENTS CAPITAL MANAGEMENT, LLC
17	Refyne	Lending	USD 82 M	Series B	TIGERGLOBAL, xyz, DIGITAL HORIZON, RTP Global, Jigsaw VC, DST GLOBAL, QED INVESTORS

Notable Fintech Investors – 2022 – 2023

INVESTOR	TOTAL INVESTMENTS	INDIAN INVESTEE STARTUPS	SEGMENT IN FOCUS
Y Combinator	20	SALT, gromo, SAVEIN, BharatX, fello, yodaa, Koshex, pay crunch, VANCE	Open Banking Wealthtech
TIGERGLOBAL	17	slice, Jar, Oxyzo, CRED, Chargebee, INDmoney, zopper, open, moneyview	Lending Payments
9UNICORNS	11	KENKO, INSURANCE SAMADHAN, Rupeek, MARKETWOLF, savart, ENSUREDIT, Vested, snapmint, CASTLER	Lending
BEENEXT	11	M2P, M1, Pazcare, mewt, ShopSe, PayGlocal, twid, CoverSelf, 42CS, KENKO	Insurtech
B Capital	11	ippo pay, M2P, tortoise, Gullak, Hyperface, DPDzero, LogiPe., GoGullak, FINSARA	Wealthtech
BLUME	11	turtlemint, bluecopa, AEREM, BIMAKAVACH, zopper, Health Assure, wint, Jai Kisan, multipl, Bhive Alts	Insurtech
SEQUOIA	11	digit, pine labs, onecode, Chargebee, FPL Technologies, twid, hubble, eduvAnz, LeapFinance	Lending Payments
The Sparkle Fund	11	M2P, slice, moneyview, scripbox, GetVantage, FINANCEPEER, MONSOON, KNAB FINANCE, 42CS, abhi LOANS, FINVU	Lending
Accel	9	Niyo, Letsventure, Avenir, DEZERV, jify, Rupeek, moneyview, scripbox	Lending
3ONE4 CAPITAL	9	open, Pazcare, weRize, wint, indiagold, Hyperface, CoverSelf, CheQ, KNIGHT FINTECH	Insurtech, Enabler Lending

Lending witnessed the highest amount of funding, with 94 Startups Funded entities receiving USD 2 B, which is a significant jump from 2021, where 71 entities received USD 1.3 B. However, we can expect these numbers to stagnate in 2023 due to the evolving regulatory environment in digital lending.



Comparison of Fintech Lending Models:

Model	Accessibility	Risk	Cost
P2P Lending	High	Medium	Medium
Invoice Financing	Medium	Low (secured by invoices)	Varies (based on invoice value)
Merchant Cash Advances	High	High (dependent on sales)	High
Lines of Credit	Medium (depends on Creditworthiness)	Medium	Medium to High (based on usage)
PoS	Medium	Medium	Medium
Captive Lending	High	Medium	Medium

Comparison of Fintech Products for MSMEs:

Product	Accessibility	Innovation level	Cost
Digital Microloans	High	Moderate	High
AI-based Credit Scoring	Medium	High	Medium
Blockchain-enabled Financing	Low (Emerging)	Very High	Low
Crowdfunding platforms	High	Moderate	Varies (platform fees)

These are some exissting fintech lending applications of india - Faircent, Lendenclub, Lendbox, Lendingkart, Finzy, Moneytap, Neogrowth, Indialends, Loantap.

6. Applicable Patents

1. Automated Credit Scoring Algorithms: Patents related to proprietary algorithms or models used for automated credit scoring
2. Machine Learning for Fraud Detection: Patents covering machine learning algorithms designed for detecting and preventing fraud
3. Blockchain for Smart Contracts: Patents related to the use of blockchain technology for implementing smart contracts in lending agreements, ensuring transparency, security, and automation.
4. Biometric Authentication in Transactions: Patents covering the use of biometric data (fingerprints, facial recognition) for user authentication
5. Digital Identity Verification: Patents related to innovative methods and technologies for verifying the digital identity of users
6. Dynamic Interest Rate Calculation: Patents covering algorithms or systems that dynamically calculate interest rates
7. User Interface Innovations: Patents related to unique and user-friendly interfaces
8. Alternative Data Analysis: Patents covering methods for collecting, analyzing, and utilizing alternative data sources
9. Regulatory Compliance Tools: Patents related to tools and systems that automate and ensure compliance with financial regulations and lending laws.
10. Data Security and Privacy Measures: Patents covering innovative technologies and processes for ensuring the security and privacy of user data.

7. Applicable Regulations

- Financial sector undertakings, including fintech businesses, are usually regulated by the RBI, SEBI, the Insurance Regulatory and Development Authority of India (IRDAI), the Pension Fund Regulatory and Development Authority (PFRDA), and IFSCA.
- The Reserve Bank of India published the Guidelines on Digital Lending on September 2, 2022. These Guidelines require significant changes in the ways that fintech's (LSPs) and Regulated Entities (banks and NBFCs) are expected to manage their digital lending operations.
- The Reserve Bank of India (RBI) regulates fintech lending, setting guidelines for non-banking financial companies (NBFCs) on capital, asset classification, and risk management.
- Digital lending faces increased scrutiny for consumer protection, data privacy, and fair practices.
- The Personal Data Protection Bill (PDPB) is pending and will impact fintech by regulating personal data. Fintech platforms must adhere to Anti-Money Laundering (AML) and Know Your Customer (KYC) regulations for customer verification.
- Compliance with consumer protection laws is essential, ensuring transparency in lending practices, interest rate disclosure, and addressing customer concerns.

8. Applicable Constraints

Fintech and digital lending face challenges such as,

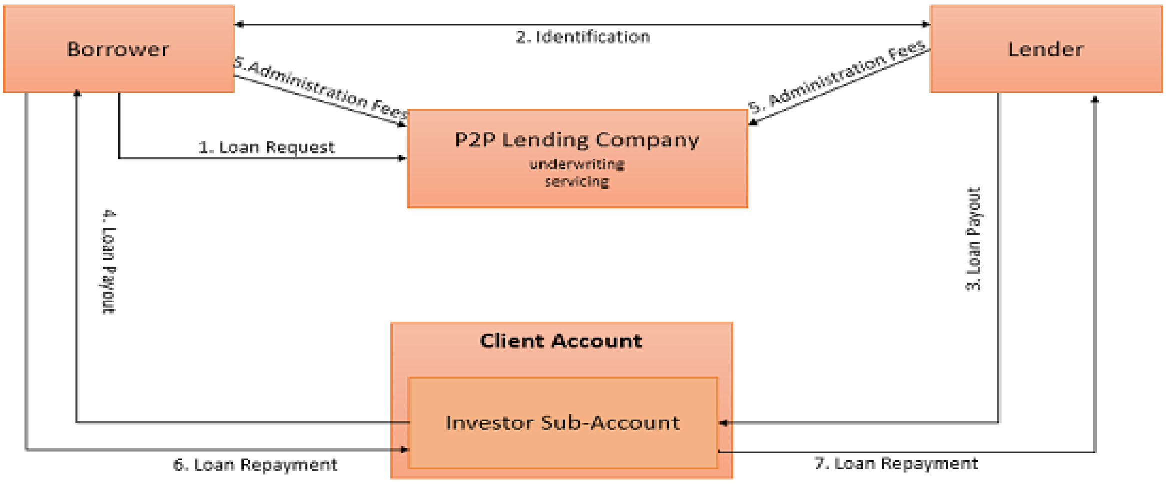
- Evolving regulatory landscapes
- cybersecurity threats
- credit risk assessment complexities
- limited financial inclusion, operational hurdles
- fraud risks
- trust-building struggles
- intense market competition
- reliance on ecosystem partners
- uncertainties in economic and regulatory conditions

9. Business Model

Leading fintech business models,

1. **Interchange fees:** When a credit or debit card is swiped, the merchant is charged a percentage of the transaction in fees.
2. **Subscription Fees (SaaS):** Tech and software companies employ the subscription model (SaaS), charging recurring monthly or yearly fees for customers to access and use their products.
3. **Payment processing and funds transfer fees**: Credit Card Processing Fees: & Transfer fees Peer-to-peer money transfer apps like Venmo and PayPal let consumers send money back and forth to each other for free.
4. **API Connection Fees**: Some fintech companies, including Plaid (more on this below), make money by charging business customers for using their APIs. Fee structures for companies that provide financial APIs vary
5. **Advisory fees/robo-advisory fees**: Wealthfront and similar fintech employ robo-advisors, automated by artificial intelligence, as the backbone of their business model, earning robo-advisory fees.
6. Third parties/referral fees: Fintech often offers free services, introducing financial product offers like loans or credit cards. When customers sign up, the company earns a referral fee, following a third-party business model
7. **Interest:** Interest rates serve as a supplementary revenue stream for fintechs, often accompanying services that initially attract customers.
8. **Ads:** Ads in apps can make money, but too many or intrusive ads can harm user satisfaction. So, place ads strategically and use them in moderation to maintain a positive user experience and generate revenue.
9. **Big data**: Fintech apps provide paid financial data for third-party companies to improve their marketing and sales strategies using AI and machine learning algorithms.
10. **Freemium:** Offer a free version with basic features, and require in-app purchases or a paid plan for premium features.

P2P lending sites generate revenue from transaction fees that can be imposed on the borrower, lender, or both.



The process is as follows:

- The borrower first puts in their loan request on the P2P lending site.
- These loan requests are then listed on the P2P lending website for the lenders to identify and act on the loan requests.
- After the borrower's successful identification & credit worthiness assessment, the lender releases the funds in Favor of the borrower.
- These funds are deposited into a specific account–the investor sub-account–maintained with the P2P lending company. There is a separate investor sub-account for every client (lender and borrower).
- Funds are then transferred into the borrower's investor sub-account for them to withdraw the same when convenient.
- Once the funds reach the borrower, the P2P lending company charges their administration fee from both the clients.
- At time of repayment (principal and interest), the borrower deposits the amount in the same investor sub-account.
- Then, these funds are transferred to the lender's accounts. The lender can withdraw or use them to fund further transactions at the P2P lending portal.

10. Concept Generation

- Fintech and traditional lending differ significantly. Fintech uses digital tools, quickening document collection and decision-making to minutes or days, while traditional lenders rely on manual processes, taking days or weeks.
- Fintech's access to extensive data allows for precise risk assessment, faster application processing, and same-day loan funding, setting it apart from the more time-consuming and less data-driven traditional lending methods.

Fintech lending web applications are popular due to their:

- **Accessibility**: Easy online access for loan applications.
- **Speed**: Quick approval and disbursement of funds.
- **Data-Driven Decisions**: Advanced analytics for accurate credit assessments.
- **Inclusion**: Providing credit to underserved individuals and businesses.
- **Customization**: Tailored loan products for diverse needs.

- **Transparency**: Clear terms and fees, building trust.
- **Lower Costs**: Reduced overheads leading to competitive rates.
- **Innovation**: Continuous introduction of user-friendly features.
- **Global Reach**: Online operations enable a wide geographical reach.

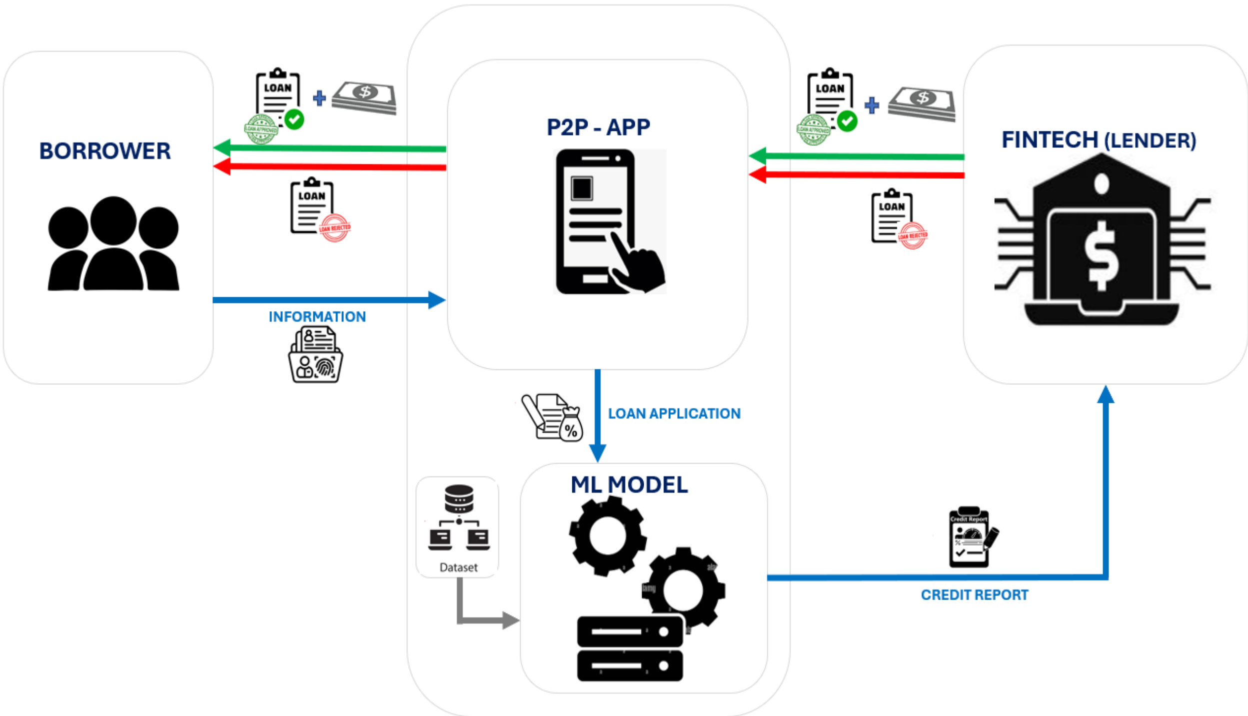
11. Concept Development

A credit scoring and interest rate prediction web application is a fintech solution that utilizes sophisticated algorithms and data analytics to assess an individual's creditworthiness and predict suitable interest rates for loans.

This user-friendly platform streamlines the loan application process by incorporating digital tools for document collection and assessment. By analysing various data points, including financial habits and repayment history, the application provides a comprehensive credit score.

This information is then utilized to predict personalized interest rates for potential borrowers. The goal is to offer transparency, efficiency, and accuracy in the lending process, making it a valuable tool for both financial institutions and individuals seeking loans.

12. Final Product Prototype (abstract) with Schematic Diagram



Prototype Components:

- **Borrower:**
  - **Information:** Personal and Financial
- **P2P APP**: Third-party employment verification services
  - **ML Model**:
    - **Data Set:** To train the ML model
    - **Loan Application:** Applicant information
    - **Credit Report:** Credit worthiness
- **Fintech (Lender)**: Financial institutions' statements
  - **Decision:** Approve or Reject

13. Product details

Data Sources

- **Credit Bureaus**: Equifax, Experian, TransUnion
- **Bank Statements**: Financial institutions' statements
- **Employment Verification**: Third-party employment verification services
- **Alternative Credit Data**: Utility payments, Rent payments
- **Social Media and Online Presence**: Social media platforms, Online activity
- **Public Records**: Tax liens, Bankruptcies, Court judgments
- **Mobile Data**: Call records, SMS patterns, App usage
- **KYC** (Know Your Customer) and AML (Anti-Money Laundering) Checks
- **Identity verification services**
- **Geospatial Data**: Location-based data
- **Behavioral Data**: User behavior on the fintech platform

Algorithms

ML Algorithms:

- Random Forest
- Decision Tree
- Multi class Logistic Regression
- AdaBoost
- K Nearest Neighbors algorithm

Frameworks & software

Component	Technology
Framework	Django or Flask
Front-end Framework	React, Angular, or Vue.js
Database	PostgreSQL or MySQL
ORM (Object-Relational Mapping)	Django ORM or SQLAlchemy
Backend Language	Python
Web Server	Gunicorn or uWSGI
Containerization	Docker
Version Control	Git
Testing	Pytest
Code Editor/IDE	Visual Studio Code, PyCharm
Cloud Services	AWS, Google Cloud Platform, Microsoft Azure
Monitoring	Prometheus and Grafana

Team required

A FinTech lending app development team comprises of these:

- 2 Developers (Frontend and Backend)
- Business Analyst
- Project Manager
- Designer
- QA Specialist
- DevOps Engineer

What does it cost?

Ideally, when doing your market research when trying to get the best Fintech app developers for your project, you may realize some price variations. The difference in the pricing is influenced by the following elements:

- Time required to finish the app development process
- UI/UX Design
- Product Requirements
- FinTech App Security
- Artificial intelligence (AI) & Machine Learning (ML)
- Data Analytics
- Technology stack
- Features integrated into the app

The use of technology in loan operations allows for accurate income predictions, borrower history evaluations, collateral value calculations, and adaptability to changes. Developing a FinTech lending app typically costs between $55,000 to $170,000.

14. Code Implementation/Validation on Small Scale

- **Dataset Reading and Understanding**

```
# lets import and read the dataset
Score_df = pd.read_csv("Score.csv")
Score_df.head()
```

	Delay_from_due_date	Num_of_Delayed_Payment	Num_Credit_Inquiries	Credit_Utilization_Ratio	Credit_History_Age	Payment_of_Min_Amount	Amount_investe
0	3.0	7.0	4.0	26.822620	265.0	No	
1	3.0	7.0	4.0	31.944960	265.0	No	'
2	3.0	7.0	4.0	28.609352	267.0	No	
3	5.0	4.0	4.0	31.377862	268.0	No	1
4	6.0	4.0	4.0	24.797347	269.0	No	

5 rows × 21 columns

- **Data cleaning**

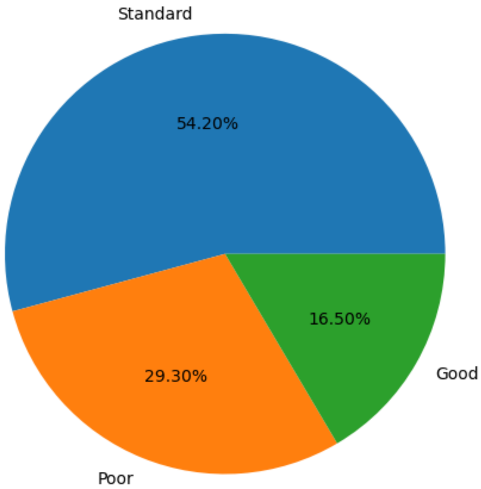
  Outliers handling

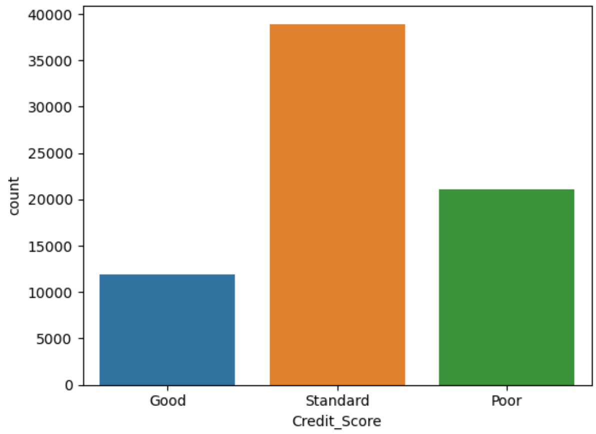
```
In [18]: # List of columns to remove outliers from
columns_to_remove_outliers = ['Delay_from_due_date', 'Total_EMI_per_month', 'Amount_invested_monthly', 'Monthly_Balance',
                'Outstanding_Debt', 'Monthly_Inhand_Salary']

# Function to remove outliers using IQR method
def remove_outliers_iqr(data, col):
    Q1 = data[col].quantile(0.25)
    Q3 = data[col].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR

    initial_rows = data.shape[0]
    data = data[(data[col] >= lower_bound) & (data[col] <= upper_bound)]
    removed_rows = initial_rows - data.shape[0]
    print(f"Removed {removed_rows} rows from '{col}' column.")
    return data

# Remove outliers from the specified columns
for col in columns_to_remove_outliers:
    Score_df = remove_outliers_iqr(Score_df, col)
```

```
Removed 4002 rows from 'Delay_from_due_date' column.
Removed 4919 rows from 'Total_EMI_per_month' column.
Removed 7393 rows from 'Amount_invested_monthly' column.
Removed 6805 rows from 'Monthly_Balance' column.
Removed 3661 rows from 'Outstanding_Debt' column.
Removed 1339 rows from 'Monthly_Inhand_Salary' column.
```

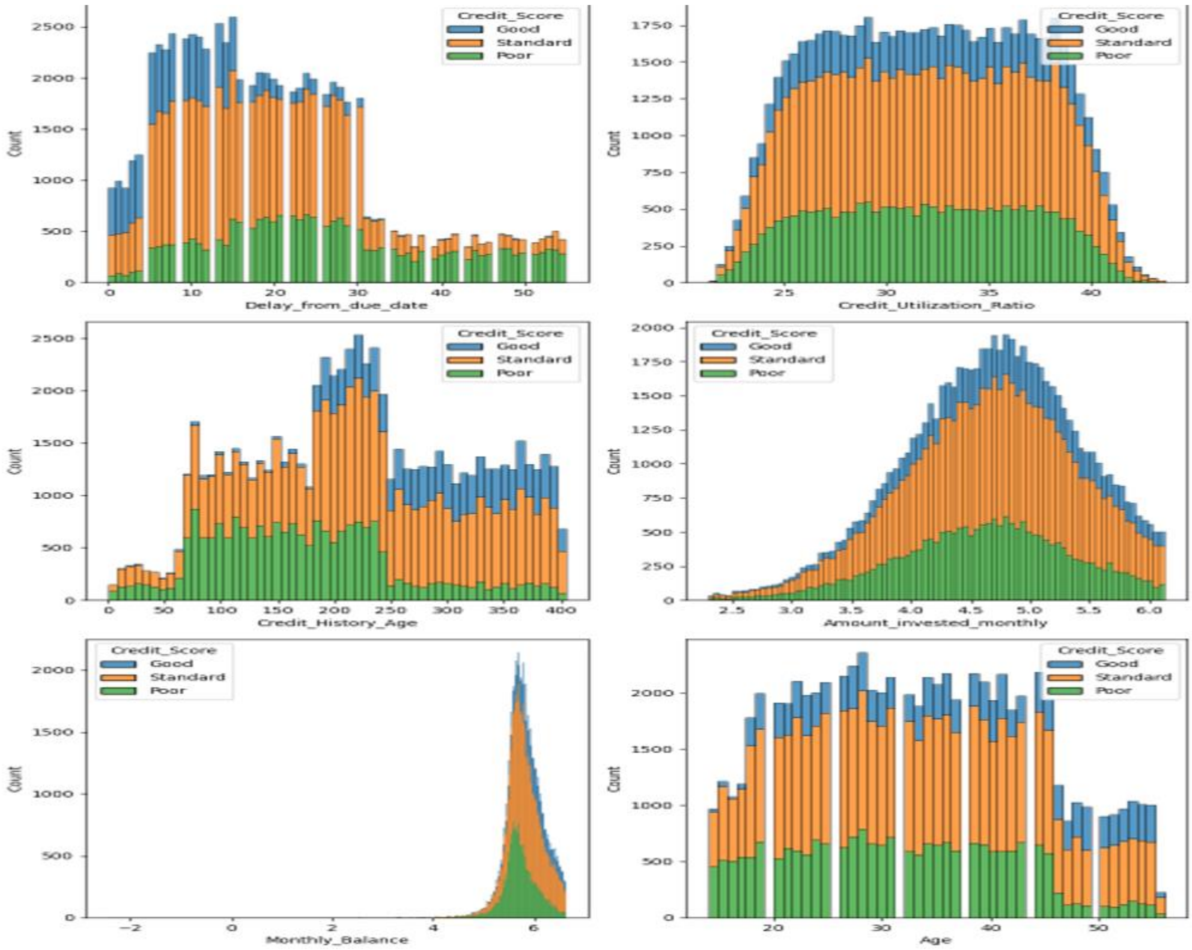
- **Exploratory Data Analysis**

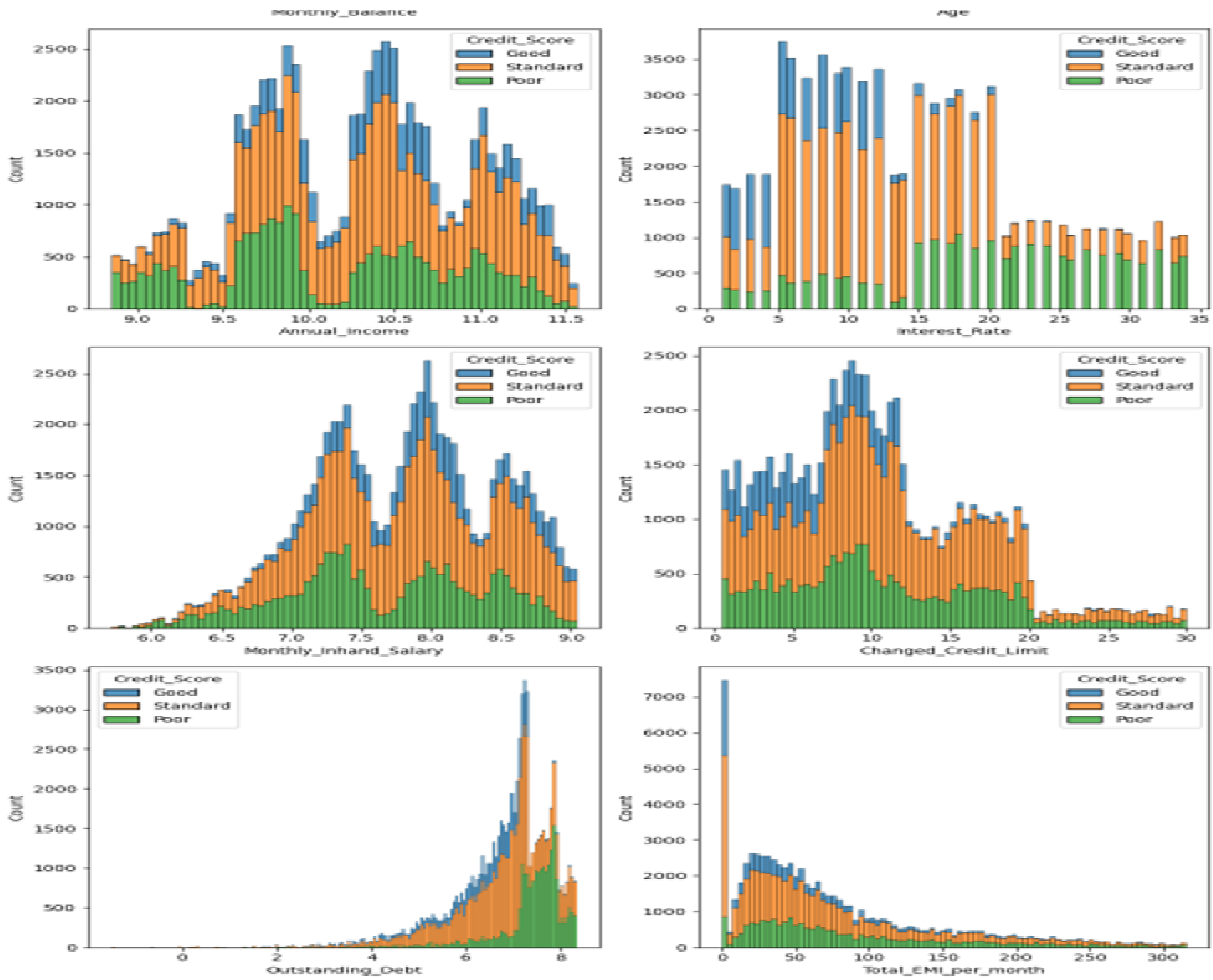
  Target variable – Credit score classes





Bivariate analysis: Target variable Vs Count of variables





- ## Feature Engineering

Target variable label encoding

In [30]:
```
# Label encoding for target variable
from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
col_list = Score_df.select_dtypes(include="object").columns

for col in col_list:
    Score_df["Credit_Score"] = le.fit_transform(Score_df["Credit_Score"].astype(str))
```

In [31]:
```
Score_df.head()
```

Out[31]:

Amount_invested_monthly	Monthly_Balance	Annual_Income	Monthly_Inhand_Salary	Changed_Credit_Limit	Outstanding_Debt	Total_EMI_per_month	Credit_Score
4.387204	5.744586	9.858183	7.509249	11.27	6.69701	49.574949	0
4.773057	5.651187	9.858183	7.509249	11.27	6.69701	49.574949	0
4.403048	5.802752	9.858183	7.509249	11.27	6.69701	49.574949	0
5.295604	5.409194	9.858183	7.509249	11.27	6.69701	49.574949	0
3.723768	5.833316	9.858183	7.509249	11.27	6.69701	49.574949	0

Scaling features - Standard Scaler

In [36]:
```
from sklearn.preprocessing import StandardScaler

cols = list(Score_df.columns)[:-1]
scaler = StandardScaler()
Score_df[cols] = scaler.fit_transform(Score_df[cols])
```

In [37]:
```
y = Score_df["Credit_Score"]
X = Score_df[cols]
```

In [38]:
```
features = list(Score_df.columns)[:-1]
print("features:", features, sep="\n")
```

```
features:
['Credit_Utilization_Ratio', 'Credit_History_Age', 'Amount_invested_monthly', 'Monthly_Balance', 'Annual_Income', 'Monthly_Inha
nd_Salary', 'Changed_Credit_Limit', 'Outstanding_Debt', 'Total_EMI_per_month']
```

Feature Selection

In [39]:
```
from sklearn.feature_selection import SelectKBest, f_classif

selector = SelectKBest(f_classif, k=5)
selector.fit(X, y)

X_new = selector.transform(X)
#print(data.columns[selector.get_support(indices=True)].tolist())
final_features = Score_df.columns[selector.get_support(indices=True)].tolist()
print(final_features)
```

```
['Credit_History_Age', 'Monthly_Balance', 'Annual_Income', 'Changed_Credit_Limit', 'Outstanding_Debt']
```

- ## Modelling

Train - Test split of dataset

In [42]:
```
from sklearn.model_selection import train_test_split
from imblearn.over_sampling import SMOTE
from collections import Counter
oversample = SMOTE()
X, y = oversample.fit_resample(X, y)

counter = Counter(y)
print(counter)
```

```
Counter({0: 38884, 2: 38884, 1: 38884})
```

In [43]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=42)
```

KNN with Hyperparameter tuning and Grid search

In [47]:
```
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import GridSearchCV, train_test_split
from sklearn.metrics import accuracy_score


knn_classifier = KNeighborsClassifier()

param_grid = {
    'n_neighbors': [3, 5, 7, 9],
    'weights': ['uniform', 'distance'],
    'p': [1, 2]
}

grid_search = GridSearchCV(knn_classifier, param_grid, cv=5, scoring='accuracy')
grid_search.fit(X_train, y_train)

best_params = grid_search.best_params_
best_knn = grid_search.best_estimator_

y_pred = best_knn.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
print(f"Best KNN Model Accuracy: {accuracy:.2f}")
print("Best Hyperparameters:", best_params)
```

```
Best KNN Model Accuracy: 0.85
Best Hyperparameters: {'n_neighbors': 3, 'p': 1, 'weights': 'distance'}
```

Model 2 : Random Forest Algorithm

In [54]:
```
# Import necessary libraries

from sklearn.ensemble import RandomForestClassifier

# Create a Random Forest classifier
rf_classifier = RandomForestClassifier(n_estimators=100, random_state=42)

# Train the classifier
rf_classifier.fit(X_train, y_train)

# Make predictions on the test set
y_pred = rf_classifier.predict(X_test)

# Evaluate the performance
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy:.2f}')

# Print classification report
print("Classification Report:\n", classification_report(y_test, y_pred))

# Generate the confusion matrix
conf_matrix = confusion_matrix(y_test, y_pred)

# Display the confusion matrix using seaborn
plt.figure(figsize=(6, 5))
sns.heatmap(conf_matrix, annot=True, fmt="d", cmap="Blues", xticklabels=['Class 0', 'Class 1'], yticklabels=['Class 0', 'Class 1'
plt.title('Random Forest Confusion Matrix')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.show()
```

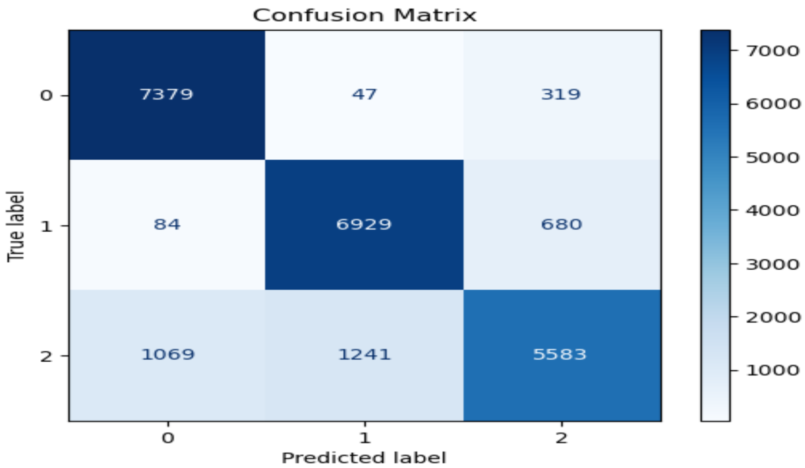
Model 3: Adaboost Classifier Algorithm

In [55]:
```
# Import necessary libraries
from sklearn.ensemble import AdaBoostClassifier


# Create an AdaBoost classifier
adaboost_classifier = AdaBoostClassifier(n_estimators=50, random_state=42)

# Train the classifier
adaboost_classifier.fit(X_train, y_train)

# Make predictions on the test set
y_pred = adaboost_classifier.predict(X_test)

# Evaluate the performance
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy:.2f}')

# Print classification report
print("Classification Report:\n", classification_report(y_test, y_pred))

# Generate the confusion matrix
conf_matrix = confusion_matrix(y_test, y_pred)

# Display the confusion matrix using seaborn
plt.figure(figsize=(6, 5))
sns.heatmap(conf_matrix, annot=True, fmt="d", cmap="Blues", xticklabels=['Class 0', 'Class 1'], yticklabels=['Class 0', 'Class 1'
plt.title('Adaboost Confusion Matrix')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.show()
```

- **Evaluation**

K Nearest Neighbors

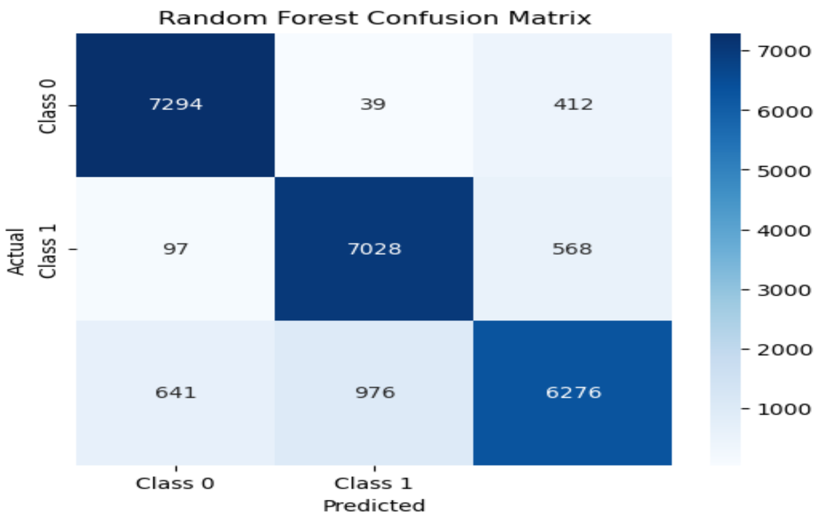
```
Classification Report:
              precision    recall  f1-score   support

           0       0.86      0.95      0.91      7745
           1       0.84      0.90      0.87      7693
           2       0.85      0.71      0.77      7893

    accuracy                           0.85     23331
   macro avg       0.85      0.85      0.85     23331
weighted avg       0.85      0.85      0.85     23331
```




```
F1 Score: 0.85
ROC AUC Score: 0.93
```

Random Forest

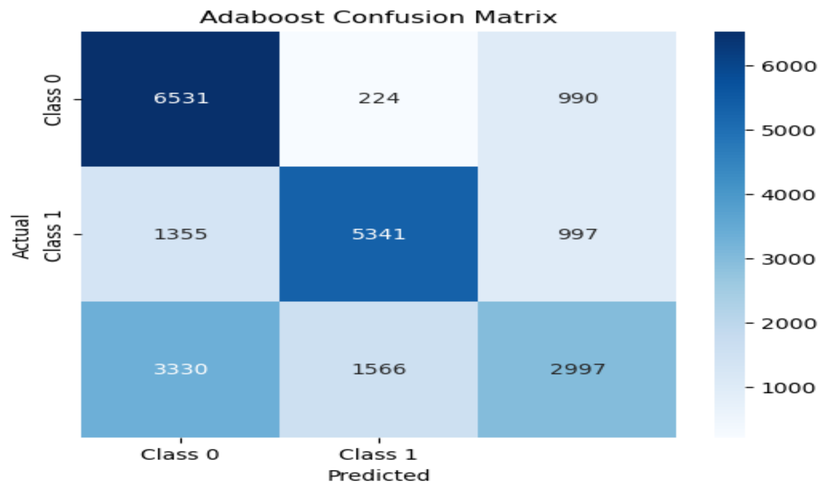
```
Accuracy: 0.88
Classification Report:
              precision    recall  f1-score   support

           0       0.91      0.94      0.92      7745
           1       0.87      0.91      0.89      7693
           2       0.86      0.80      0.83      7893

    accuracy                           0.88     23331
   macro avg       0.88      0.88      0.88     23331
weighted avg       0.88      0.88      0.88     23331
```




```
F1 Score: 0.88
ROC AUC Score: 0.96
```

Adaboost

```
Accuracy: 0.64
Classification Report:
              precision    recall  f1-score   support

           0       0.58      0.84      0.69      7745
           1       0.75      0.69      0.72      7693
           2       0.60      0.38      0.47      7893

    accuracy                           0.64     23331
   macro avg       0.64      0.64      0.62     23331
weighted avg       0.64      0.64      0.62     23331
```




```
F1 Score: 0.62
ROC AUC Score: 0.77
```

- **Performance comparison**

  Performance comparison between all the models,

  - KNN Model : F1 Score: 0.85

    ROC AUC Score: 0.93
  - Random Forest : F1 Score: 0.88

    ROC AUC Score: 0.96
  - AdaBoost : F1 Score: 0.62

    ROC AUC Score: 0.77

- **GitHub Link to code Implementation**

  https://github.com/DharaKhamar/Fintech-Lending---Credit-Score-Classification

15. Conclusion

This Fintech lending application, with the integration of machine learning models, has truly transformed how we evaluate credit. We've successfully categorized applicants into good, standard, and bad credit classes, allowing for more precise decision-making. Model using Random Forest classifier seems to be best match for this process with 0.96 ROC AUC score and 0.88 F1 Score. This approach not only enables personalized lending terms but also ensures a fair and transparent process.

16. References

[1] Baidaa M Alsafy, Zahoor M. Aydam, Wamidh K. Mutlag "Multiclass Classification Methods: A Review". IJAETIS. Volume 5, Issue 3, Page No: 01-10, May-June 2019. ISSN:2455-1651

[2] Yang, Yang, "Analysis and assessment of credit rating model in P2P lending: an instrument to solve information asymmetry between lenders and borrowers". Thesis report. S.M. Massachusetts Institute of Technology, pages 58-67, June 2015

[3] "Indian Fintech Funding & Acquisition 2022-2023" Report

[4] Muhammad Anshari, Masairol Masri, Milan Hrdy"Financial Technology with AI-Enabled and Ethical Challenges" , June 2021, Society 58(4), DOI:10.1007/s12115-021-00592-w

[5] Deniel Felix, Paolo Giudici, "Credit Scoring for Peer-to-Peer Lending", 7 July'23

[6] Ashwani Kumar, Dr D Shanthi, Pronaya Bhattacharya, "Credit Score Prediction System using Deep Learning and K-Means Algorithms", Journal of Physics Conference Series, August 202, DOI:10.1088/1742-6596/1998/1/012027

[7] Ryuichiro Hashimoto, Kakeru Miura, Yasunori Yoshizaki, "Application of Machine Learning to a Credit Rating Classification Model: Techniques for Improving the Explain ability of Machine Learning", Bank of Japan Working Paper Series, April 21, 2023

https://altair.com/newsroom/articles/credit-scoring-series-part-one-introduction-to-credit-scoring?_ga=2.255666696.1468709777.1704982637-432260547.1704982637

https://plaid.com/resources/fintech/how-does-fintech-and-plaid-make-money/

https://yellow.systems/blog/how-to-build-a-fintech-app

https://www.code-brew.com/how-does-p2p-lending-marketplace-work-its-business-revenue-model/

https://www.linkedin.com/pulse/comparison-different-fintech-lending-models-products-msmes-nxddc/?trk=article-ssr-frontend-pulse_more-articles_related-content-card

https://plaid.com/resources/fintech/what-is-fintech/

https://www.leadsquared.com/industries/lending/p2p-lending/

https://www.linkedin.com/pulse/ai-fintech-meet-top-10-startups-powered-techmagic/

https://cloudbankin.com/blog/fintech/top-10-fintech-companies-india-lending-space/