
# AI PREDICTIVE MODEL FOR CREDIT UNDERWRITING

# INTRODUCTION TO AI & MACHINE LEARNING

What is Artificial Intelligence (AI)?

✅ AI refers to the simulation of human intelligence in machines that can think, learn, and make decisions.

What is Machine Learning (ML)?

✅ ML is a subset of AI that allows systems to learn from data and improve their performance without being explicitly programmed.

AI & ML in Credit Underwriting

✅ Traditional credit underwriting relies on manual analysis of financial data, which is slow and error-prone.

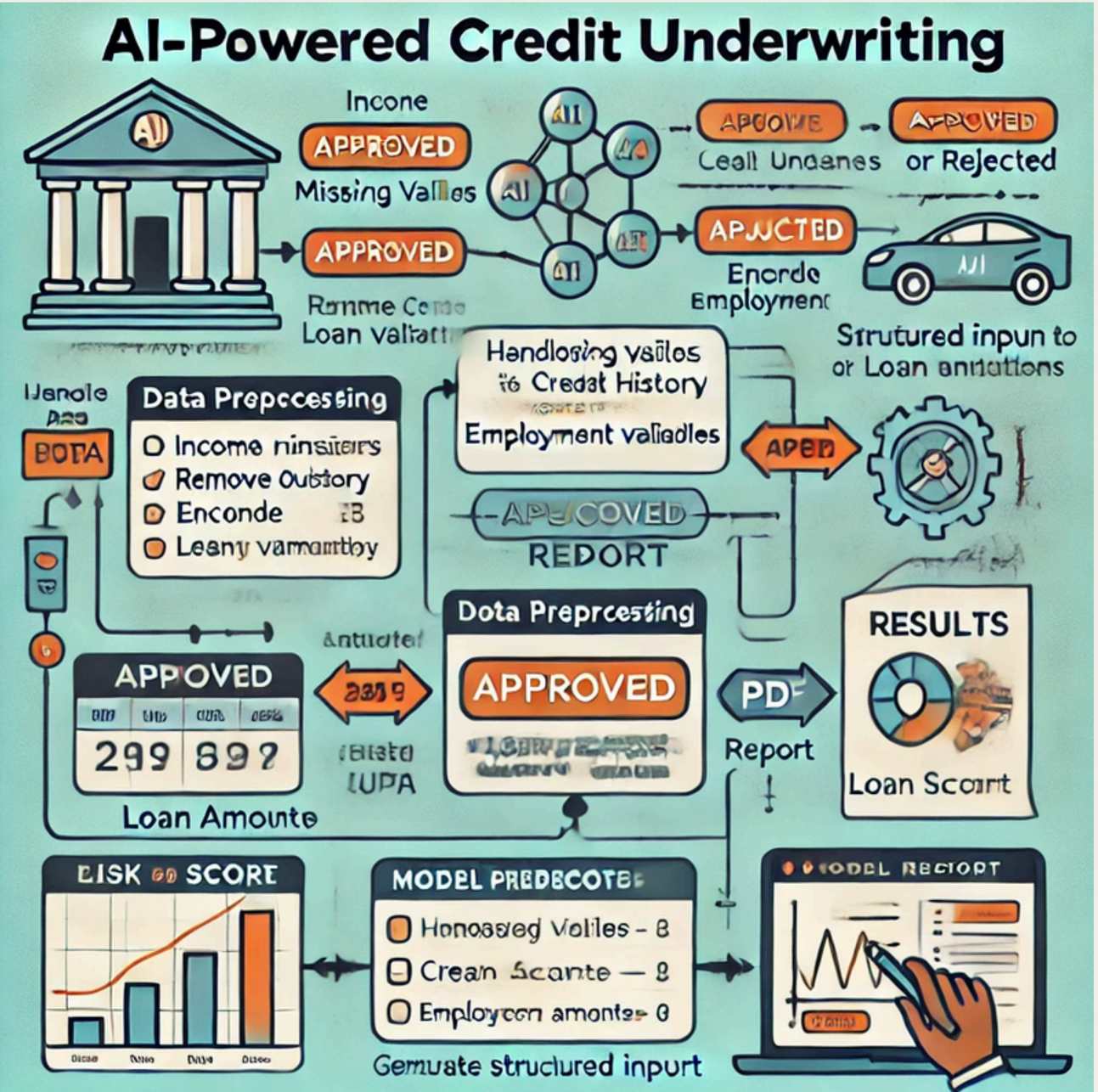
✅ AI/ML automates and enhances the process by evaluating credit risk with speed, accuracy, and fairness.

✅ This approach helps banks and financial institutions make better loan approval decisions, reducing default risks and improving customer experience.

# SYNOPSIS

- Credit underwriting is the process of evaluating an applicant's creditworthiness to determine whether they qualify for a loan or credit.
- Traditionally, it involves reviewing financial documents, credit scores, and other relevant factors to assess risk.
- AI reduces bias, leading to fairer evaluations and wider access to credit.
- AI streamlines underwriting, enabling lenders to make more informed decisions.

# Flowchart Representation of AI-Powered Credit Underwriting




## 📌 Step-by-Step Breakdown

1. User Inputs Data 📥
2. Data Preprocessing 🔄
3. Model Prediction 🤖
4. Display Results (Approved/Rejected) ✅❌
5. Generate PDF Report 📄

# DATASET FINALIZATION

**Key Features Selected:**

- Loan Status, CIBIL Score
- Annual Income
- Loan Amount & Loan Term
- Number of Active Loans
- Employment Status , Residence Type

**Preprocessing:**

- Handled missing
- Feature Encoding
- Scaling & Normalization
- Outlier Detection

# REASON FOR CHOOSING DATASET

Real-World Financial Data

Suitable for Predictive Modeling

Comprehensive Feature Set

Opportunity for Feature Engineering

Imbalance in Loan Approvals

Scalable & Deployable

# METHODOLOGIES

- Data Collection and Preprocessing
- Feature Engineering
- Model Selection
- Model Training and Validation
- Deployment

# ML MODELS

1. Logistic Regression
2. Decision Tree
3. Random Forest
4. AdaBoost
5. Naive Bayes
6. XGBoost
7. Gradient Boosting Classifier

# TESTING AND EVALUATION

## Comparison of Machine Learning Models

| Model | Training Accuracy | Testing Accuracy | Overall Accuracy |
|---|---|---|---|
| Logistic Regression | 55.87196987526477 | 57.38476011288805 | 56.62836494076406 |
| Decision Tree | 100.0 | 98.77704609595484 | 99.38852304797743 |
| Random Forest | 100.0 | 99.15333960489181 | 99.5766698024459 |
| AdaBoost | 97.36408566721582 | 97.3659454374412 | 97.36501555232852 |
| Naive Bayes | 86.93810308307836 | 86.82972718720602 | 86.88391513514219 |
| XGBoost | 99.67051070840198 | 99.15333960489181 | 99.411925156646 9 |

# GRADIENT BOOSTING

**Why Gradient Boosting is the Best Choice?**

**We chose Gadient Boosting for:**

- Higher Accuracy & Performance
- Handles Complex Relationships
- Reduces Overfitting
- Better Precision & Recall
- Feature Importance

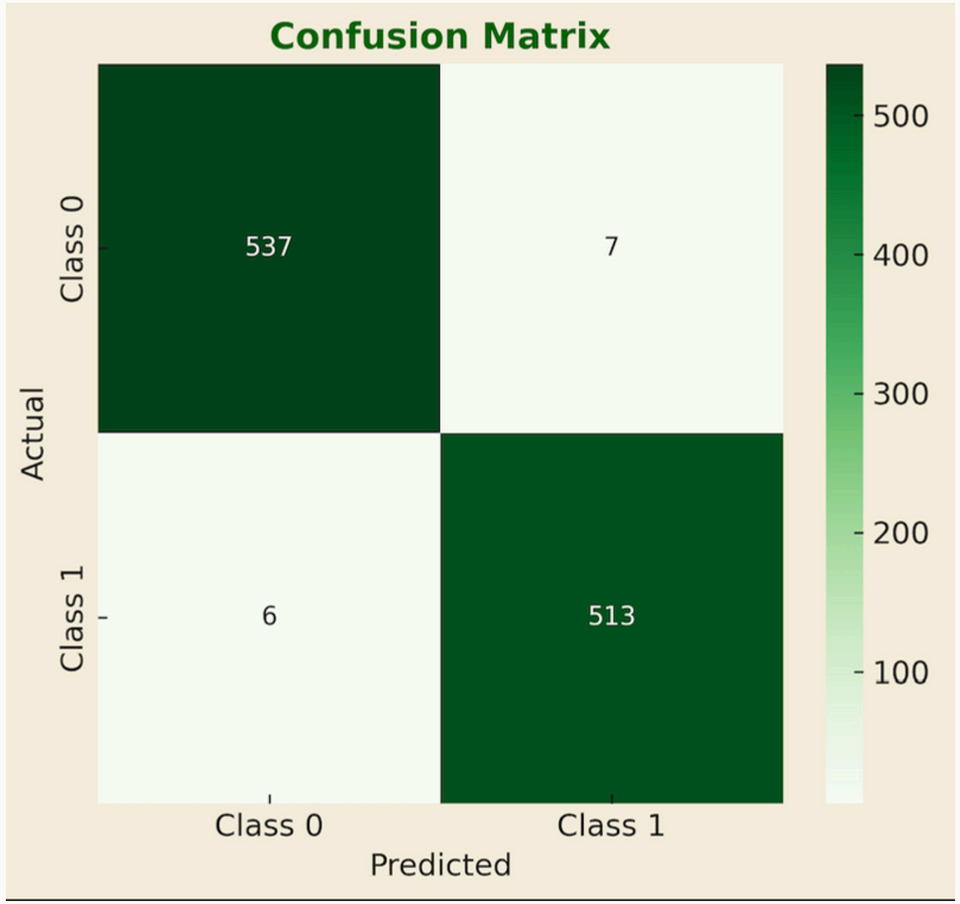
# RESULTS & PERFORMANCE METRICS

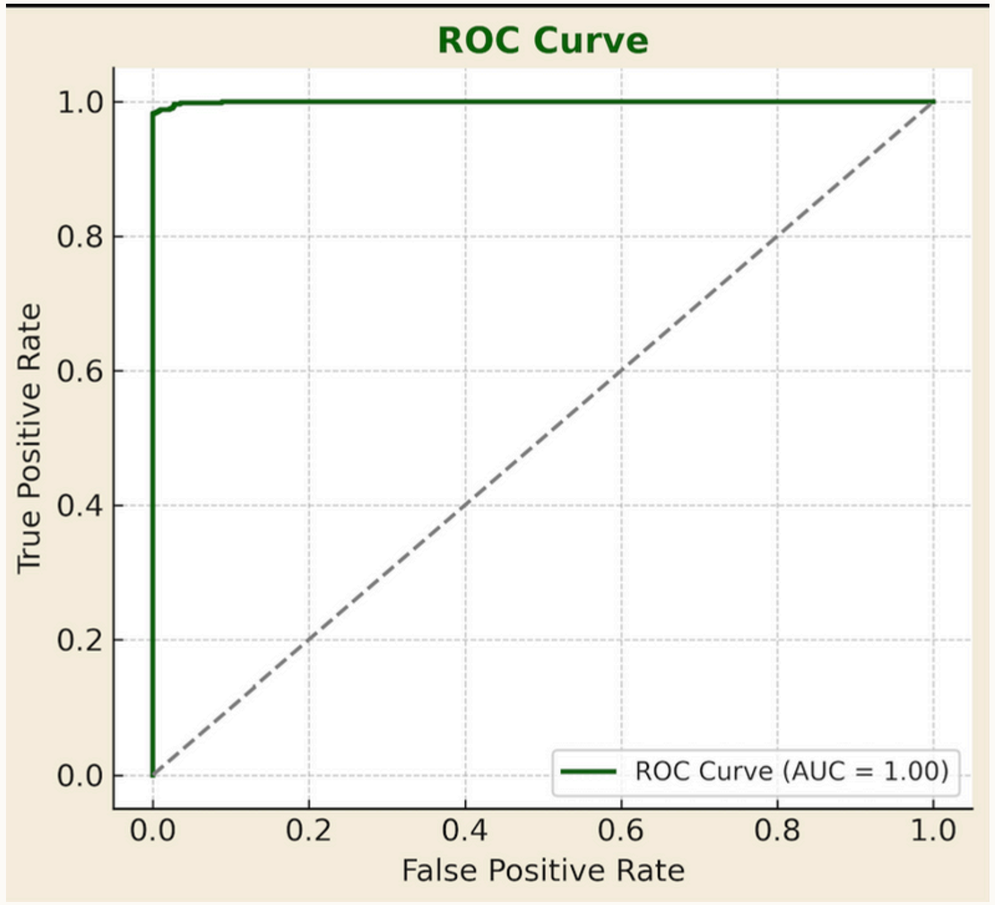
The Gradient Boosting model achieved the following metrics:

- **Accuracy:** 98.78%
- **Precision :** 98.90% (Class 0) , 98.65% (Class 1)
- **Recall :** 98.71% (Class 0) , 98.84% (Class 1)
- **F1-Score :** 98.80% (Class 0) , 98.75% (Class 1)
- **Confusion Matrix:**

  [ [ 537,  7 ] , [6,  513 ] ]

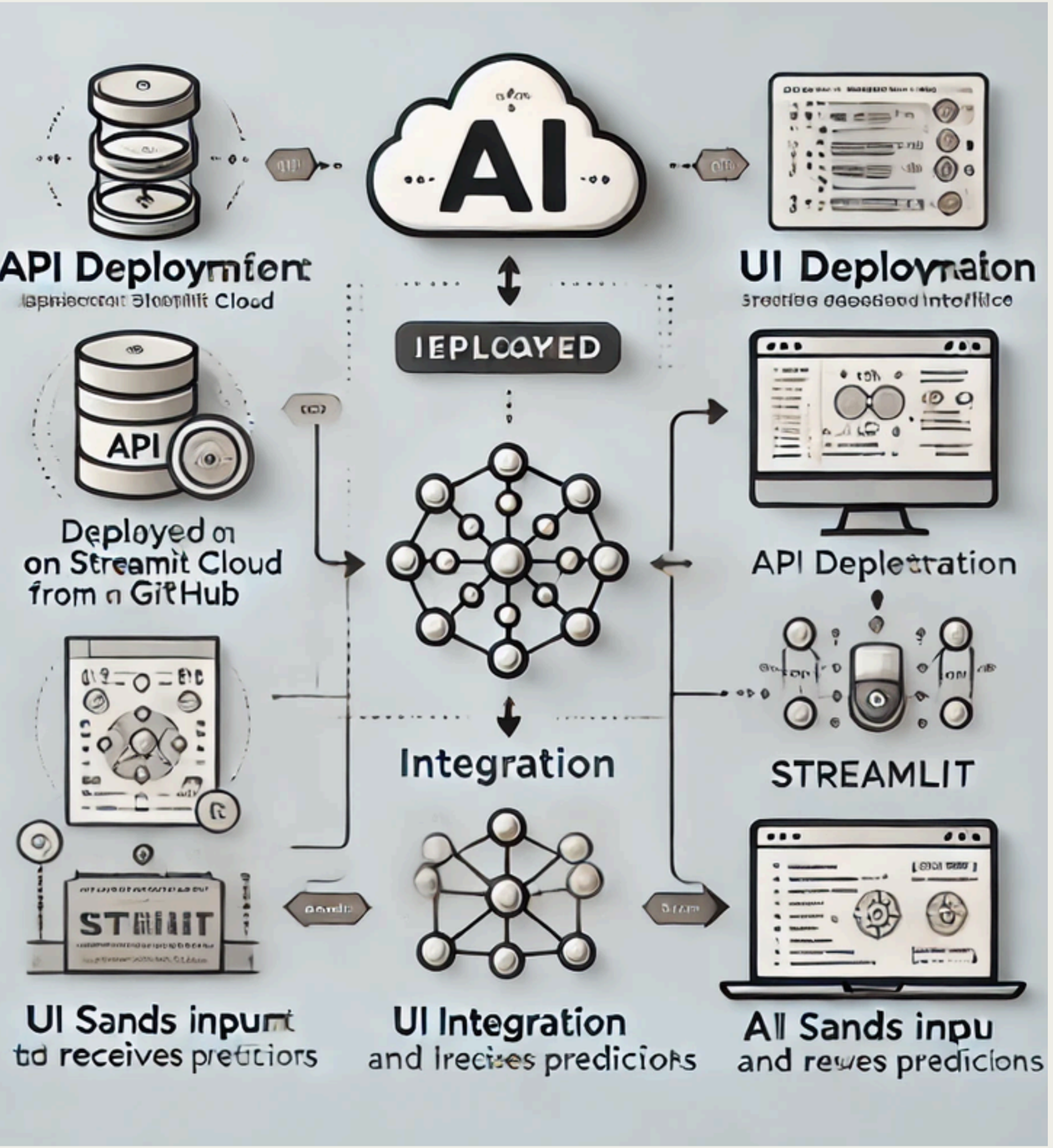
# GRAPHICAL REPRESNTATION



# LIBRARIES

- **Programming Language:** Python
- **Streamlit:** Framework for creating the interactive web app.
- **FPDF:** Library for generating downloadable PDF reports.
- **pandas:** For data preparation and handling user inputs.
- **matplotlib:** For optional visualizations.
- **joblib:** For loading the pre-trained machine learning model.
- **transformers:** For NLP-based chatbot responses (using pre-trained models).
- **langdetect:** For language detection (used in the chatbot).


# IMPLEMENTATION DEPLOYMENT

## 1. API Deployment:

- Deployed on Streamlit Cloud, pulling from GitHub
- Build command installs dependencies
- Start command runs the ML API

## 2. UI Deployment:

- Deployed on Streamlit, linked to the same repository
- Build command installs dependencies
- Start command runs the Streamlit app

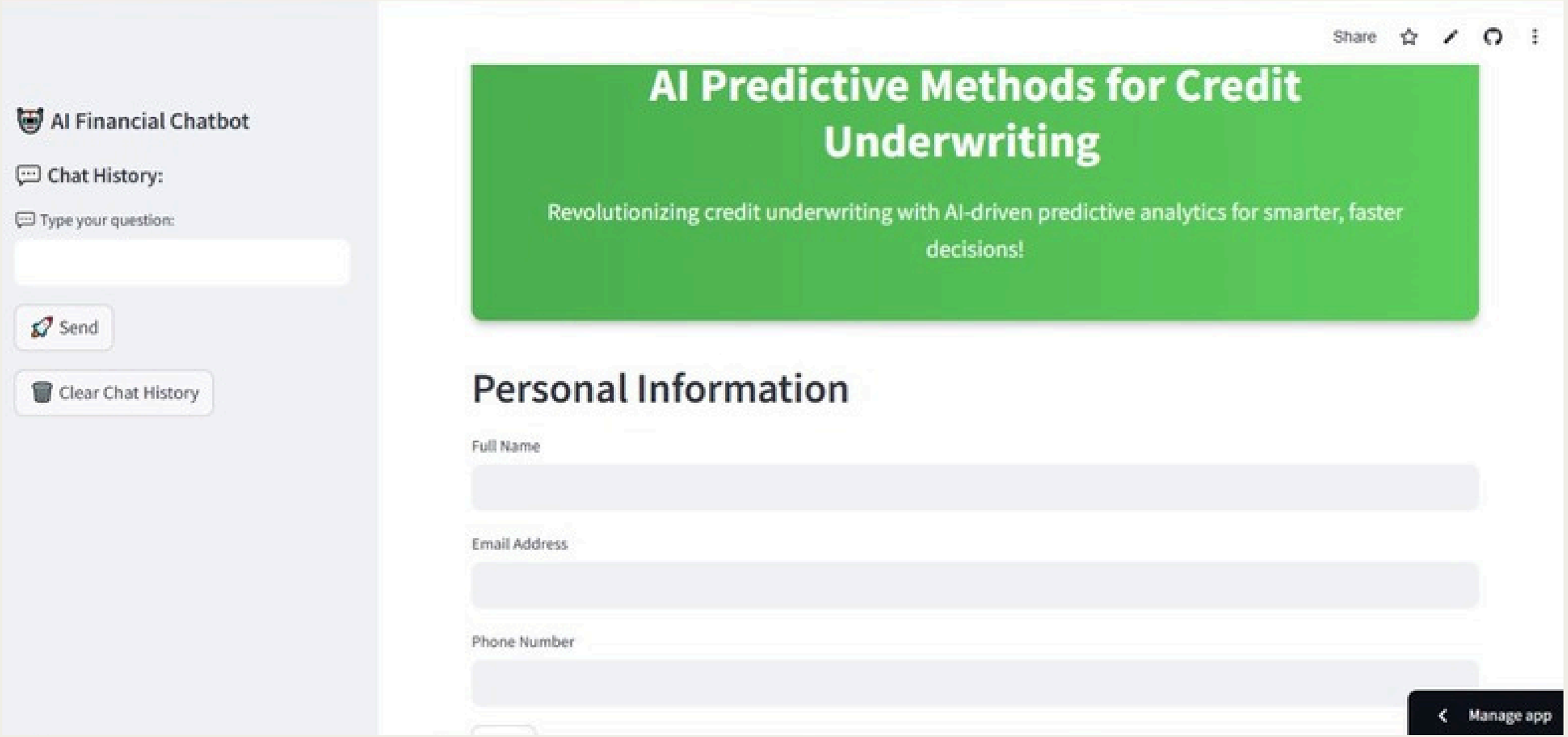
## 3. Integration:

- UI sends input to the API, which processes data and returns predictions

## 4. Testing:

- Both API and UI tested via URLs to ensure proper functionality and smooth integration

# USER INTERFACE



# FEATURES OF UI

- **User-Friendly Interface** – Simple and intuitive design for easy navigation.
- **Loan Application Input** – Users can enter relevant financial details for credit evaluation.
- **AI-Powered Predictions** – Displays instant loan approval/rejection based on the trained model.
- **Real-Time Analysis** – Uses machine learning to assess creditworthiness dynamically.
- **Accessible from Anywhere** – Hosted on Streamlit, making it cloud-based and easily accessible.


# BENEFITS OF MODEL

🚀 Faster Loan Processing

📊 Data-Driven Decision Making

💰 Reduced Risk of Loan Defaults

🎯 High Accuracy in Predictions

📉 Lower Operational Costs

👥 Fair & Transparent Credit Scoring

📈 Continuous Improvement

🔗 Easy Integration with Existing Systems

🌍 Scalable and Adaptable

# CHALLENGES IN TRADITIONAL CREDIT UNDERWRITING

---

⏳ Time-Consuming Process
👥 Human Bias & Subjectivity
📄 Limited Data Utilization
💸 High Operational Costs
⚠️ Higher Risk of Loan Defaults
📈 Low Scalability
📊 Inability to Adapt to Market Trends
📉 Inconsistent Decision-Making
🔗 Lack of Integration with Modern Technologies

# BUSINESS PROPOSITION

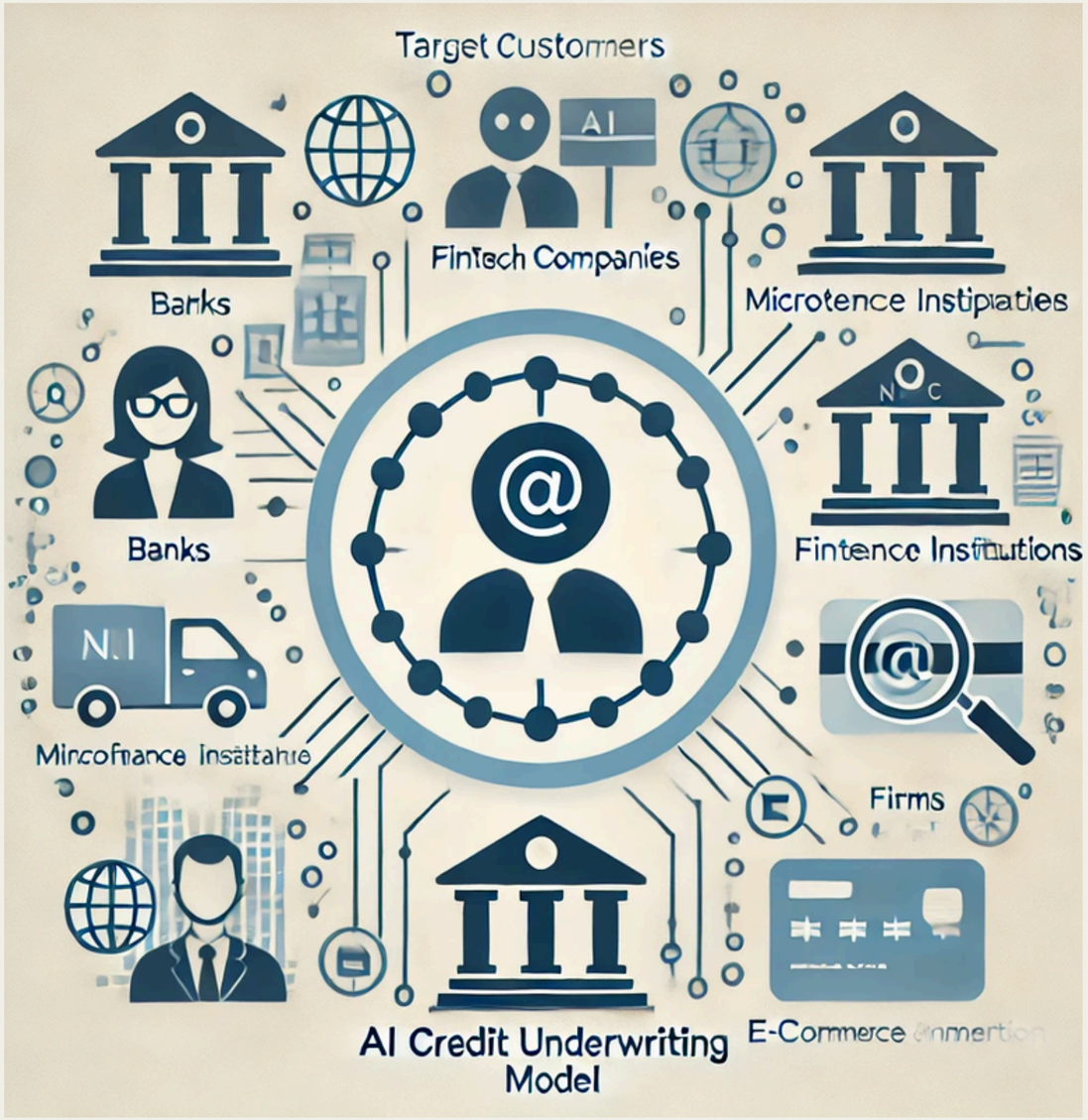

Our AI-powered credit underwriting model offers a fast, accurate, and data-driven approach to loan approvals, helping financial institutions minimize risk, reduce costs, and enhance decision-making.

**Key Value Propositions:-**

- Faster Loan Approvals
- Lower Default Risks
- Cost Efficiency
- Fair & Unbiased Decisions
- Scalability & Adaptability

# TARGET CUSTOMERS

- Banks & Financial Institutions
- NBFCs (Non-Banking Financial Companies)
- Fintech Companies
- Microfinance & Lending Startups
- E-commerce & BNPL (Buy Now, Pay Later) Services
- Insurance & Investment Firms



# FUTURE SCOPE

1. Integration of blockchain for enhanced security and data integrity.
2. Use of advanced deep learning models for even more accurate predictions.
3. Expanding to global financial markets with region-specific risk assessments.
4. Real-Time Credit Scoring System.
5. Continuous Model Optimization

# CASE STUDY

**AI-Powered Credit Underwriting in Action Background:**

A leading NBFC (Non-Banking Financial Company) faced challenges in loan approval delays, high default rates, and biased decision-making due to manual underwriting processes.

**Problem Statement:**

- Lengthy loan approval process (average 7–10 days).
- High default rates due to inefficient risk assessment.
- Manual underwriting led to inconsistent & biased approvals.
- Limited scalability to process a large volume of applications.
- Solution: Implementation of AI-Based Credit Underwriting
- The company deployed an AI-driven predictive model to automate and enhance credit risk assessment.

**Results & Impact:**

📈 Approval Time Reduced by 85% – From 7–10 days to 4–6 hours.
💰 Default Rates Dropped by 30% – More accurate risk prediction prevented bad loans.
🔍 Scalability Increased by 200% – Able to process triple the number of loan applications.
✅ Compliance & Transparency Improved – AI-driven decision-making aligned with fair lending regulations.

# CONCLUSION

After evaluating multiple machine learning models for credit underwriting, we observed significant variations in accuracy. The Decision Tree, Random Forest, and XGBoost models outperformed others, achieving over 99% accuracy. This indicates their robustness in predicting credit risk. Logistic Regression showed subpar results, highlighting its unsuitability for this use case due to complex data patterns. The Streamlit framework was employed to deploy the final solution, ensuring an interactive and user-friendly interface for real-world applications.

# Acknowledgement

We sincerely thank **Mr. Vivek Gautam sir** for his invaluable guidance and support throughout this internship. His constructive feedback was instrumental in the successful completion of our project.


We also extend our gratitude to **Infosys Springboard** for providing this amazing opportunity to apply our academic knowledge to real-world challenges.


This internship allowed us to gain hands-on experience in machine learning, data preprocessing, model deployment, and UI development, contributing significantly to our professional and personal growth.

# REFERENCES


1. **Books:**
   - Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow by
     **Aurélien Géron**

2. **Online Resources:**
   - Scikit-learn Documentation: scikit-learn.org
   - Pandas Documentation: pandas.pydata.org
   - Kaggle: kaggle.com

3. **Articles:**
   "A Comprehensive Guide to Regression in Machine Learning"

4. **Streamlit Documentation:** https://streamlit.io

# THANK YOU!