AI PREDICTIVE MODEL FOR CREDIT UNDERWRITING

SYNOPSIS

- **Objective:** Develop an AI-predictive model based credit underwriting system using machine learning to enhance loan approval predictions.
- **Methodology:** Train and compare multiple models (Logistic Regression, Decision Tree, Random Forest, AdaBoost, XGBoost, Gradient Boosting), selecting Gradient Boosting Classifier (GBC) for its superior performance.
- **Results & Impact:** GBC achieves higher accuracy, better risk assessment, and improved decision-making, reducing false approvals & rejections in financial institutions.
- **Outcome:** A Streamlit-based AI application that provides loan approval predictions, EMI calculations, AI-driven financial advice, and automated report generation.

AI PREDICTIVE CREDIT UNDERWRITING VS. TRADITIONAL CREDIT UNDERWRITING

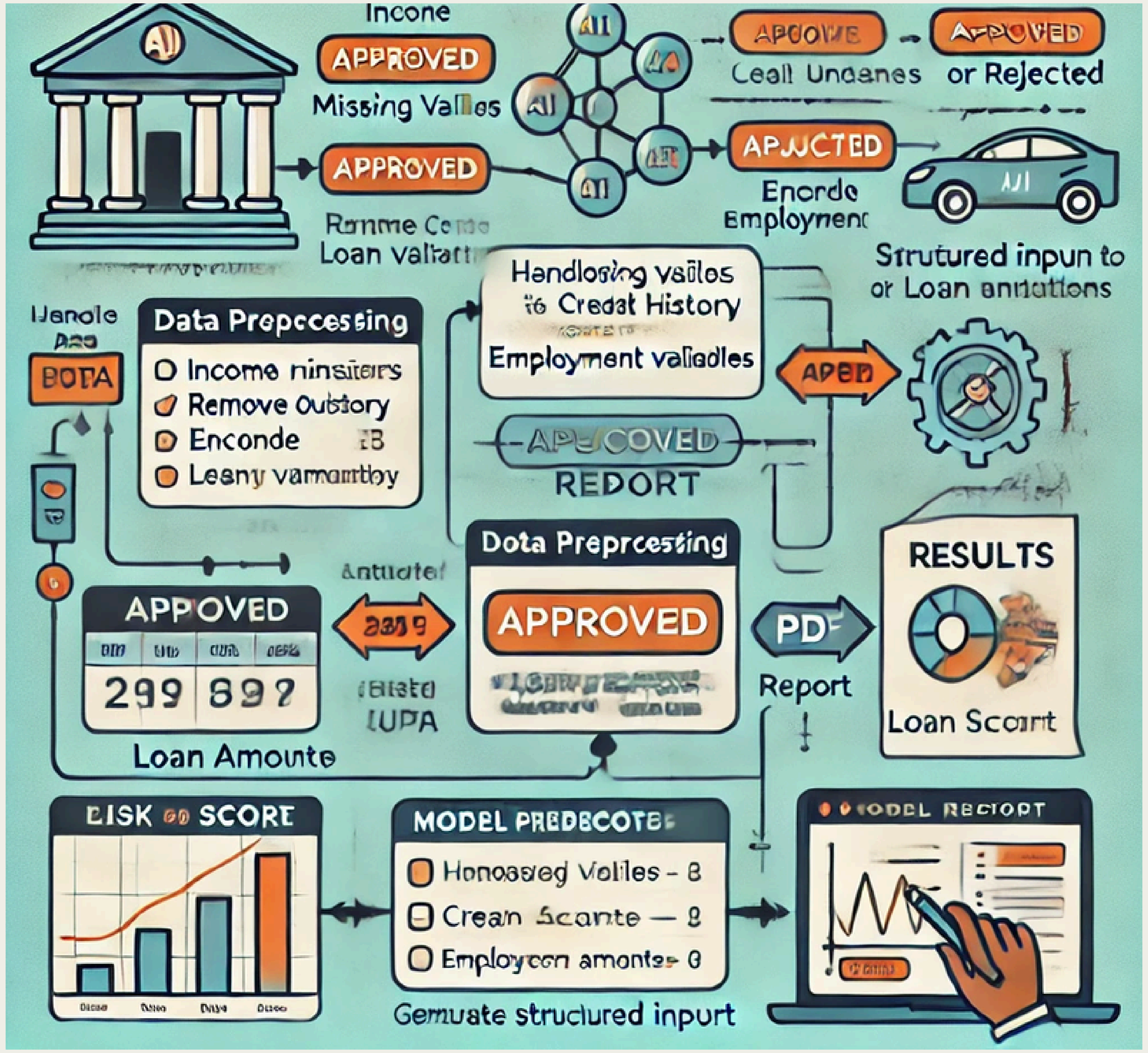
AI Predictive Credit Underwriting

- Automated & Data-Driven
- Real-Time Processing
- Advanced Risk Analysis
- Reduced Bias
- Self-Learning Models

Traditional Credit Underwriting

- Manual & Rule-Based
- Slower Processing
- Limited Data Usage
- Higher Risk of Bias
- Fixed Risk Assessment

FLOWCHART REPRESENTATION OF AI-POWERED CREDIT UNDERWRITING



📌 Step-by-Step Breakdown

1️⃣ User Inputs Data 📥
2️⃣ Data Preprocessing 🔄
3️⃣ Model Prediction 🤖
4️⃣ Display Results (Approved/Rejected) ✅❌
5️⃣ 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. Gradient Boosting Classifier

TESTING AND EVALUATION

Comparison of Machine Learning Models

#	Model	Training Accuracy	Testing Accuracy	Overall Accuracy
1	Logistic Regression	92.28	92.28	92.28
2	Decision Tree	100.0	98.68	99.38
3	Random Forest	100.0	98.77	99.57
4	AdaBoost	97.36	97.27	97.36
5	Naive Bayes	86.93	93.69	86.88
6	Gradient Boosting	99.67	98.77	99.57

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

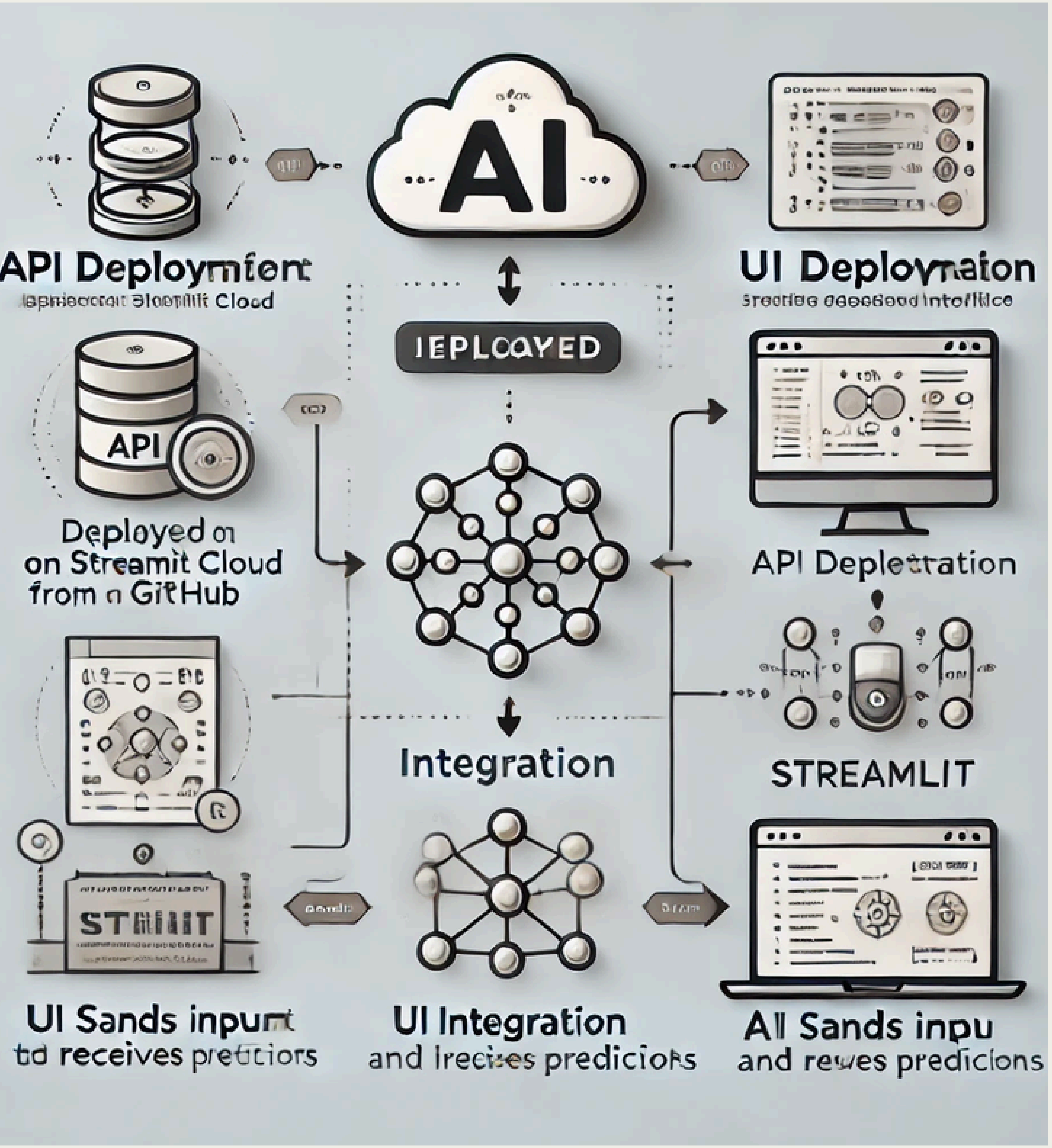
The Gradient Boosting model achieved the following metrics:

- **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]]

LIBRARIES

- **Programming Language:** Python
- Streamlit
- FPDF
- pandas
- matplotlib
- joblib
- transformers
- langdetect



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:

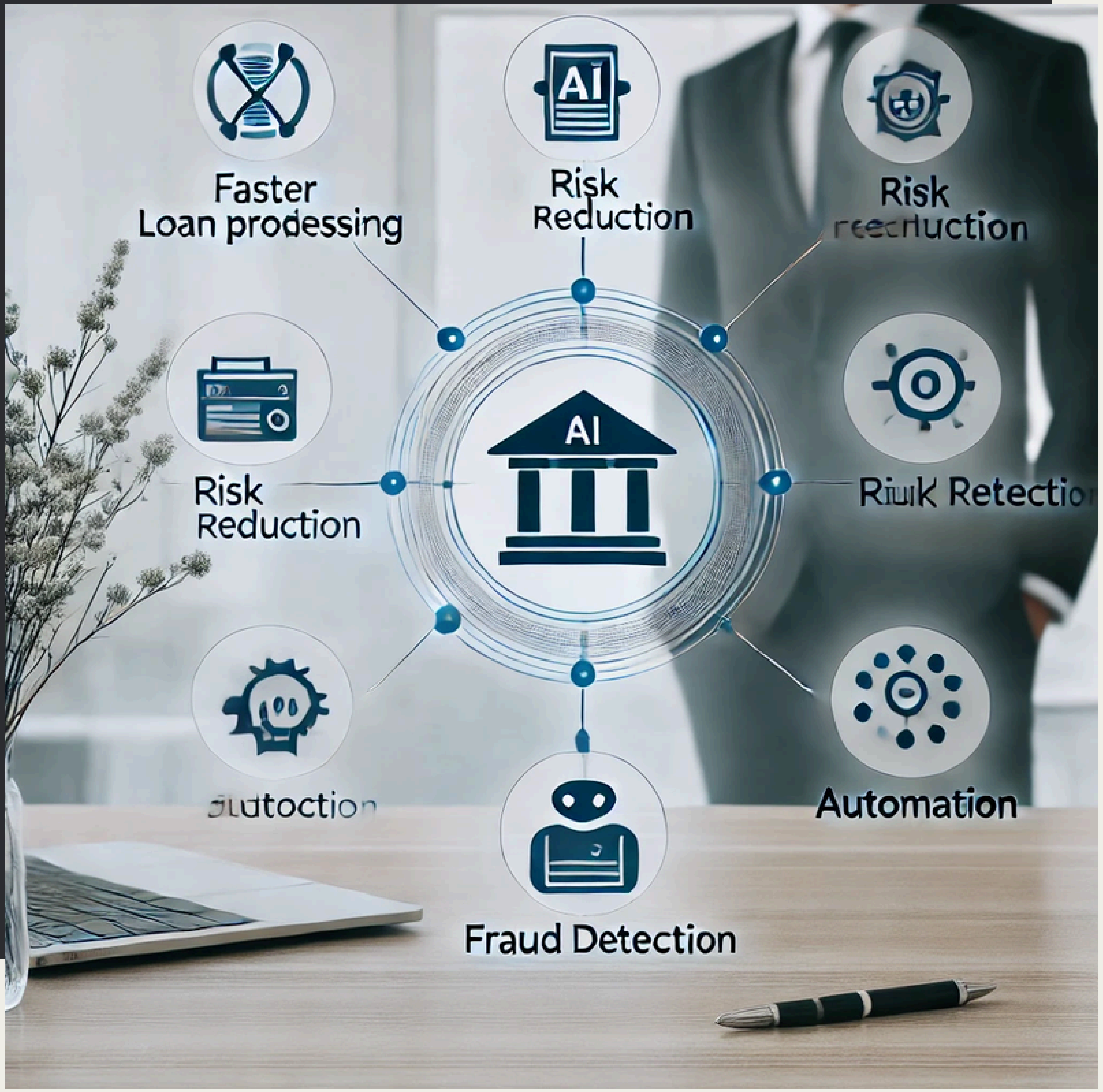
- Deployed on Streamlit, linked to the same repository
- Build command installs dependencies
- Start command runs the Streamlit app

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



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

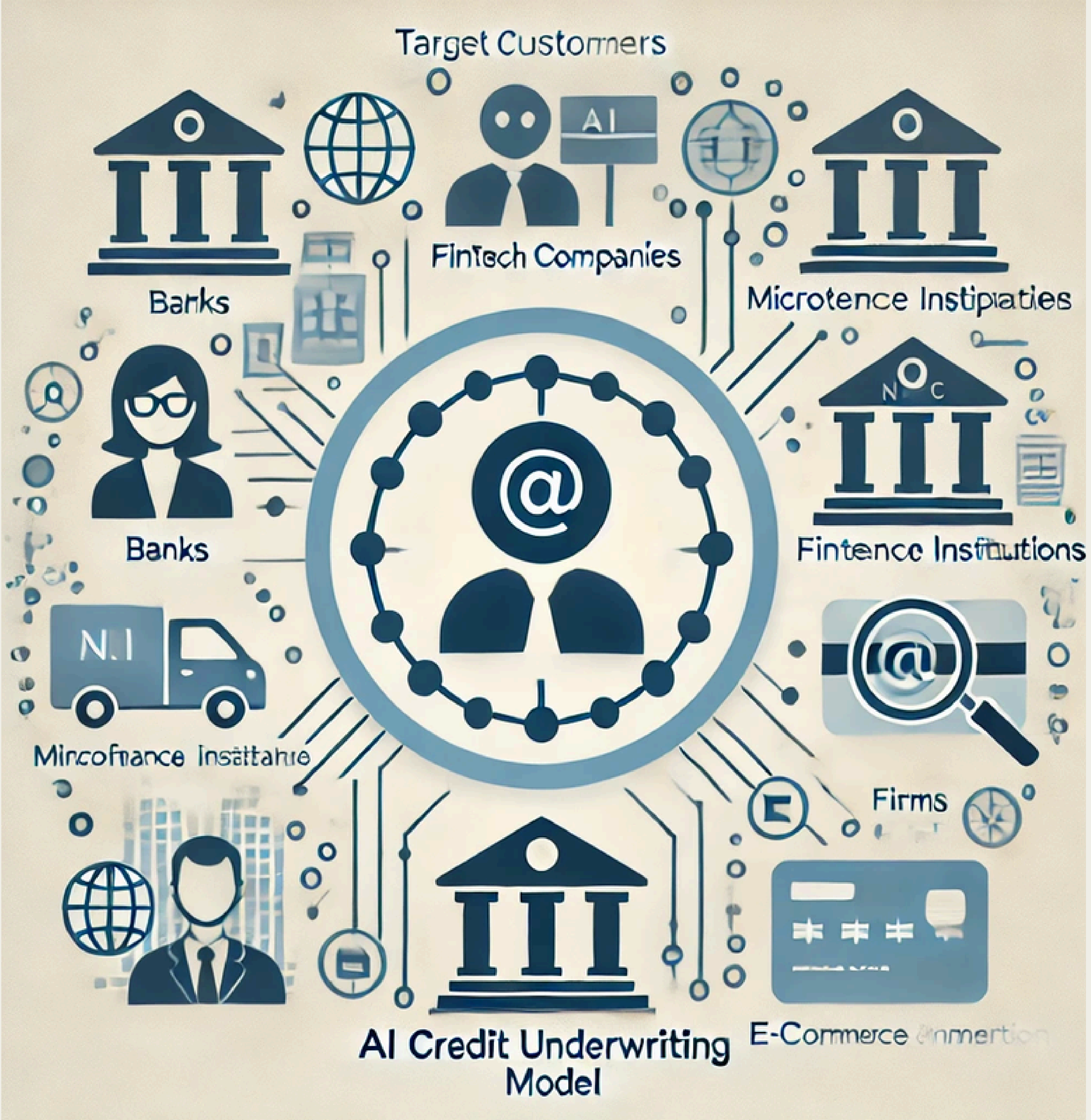
Our AI-powered credit underwriting model offers a fast, accurate, and data-driven approach to loan approvals, helping financial institutions minimize risk, reduce costs, and enhance decision-making.

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

📌 Background

A leading NBFC faced challenges in loan approval delays, high default rates, and biased decision-making due to manual underwriting.

📌 Problem Statement

- Lengthy loan approvals (7–10 days).
- High default rates due to poor risk assessment.
- Manual underwriting caused bias & inconsistency in approvals.
- Limited scalability to process high loan application volumes.

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!