Maharshi Karve Stree Shikshan Samstha's
Cummins College of Engineering for Women, Pune
(An Autonomous Institute affiliated to Savitribai Phule Pune University)

ADD COMPANY NAME

Bank Customer Churn Prediction System

AI-Powered Customer Retention Analysis

By: Deepthi Kambham - UCE2023432
Mayuri Dandekar - UCE2023444
TY COMP A1 (2023-2027)

Understanding Customer Churn in Banking

Customer churn is when customers discontinue their relationship with a bank

Business Impact:
Average churn rate in banking: 20.4%
Cost per lost customer: ~$1,500 (lifetime value)
New customer acquisition: 5x more expensive than retention
Proactive retention is 3x cheaper than reactive approaches

Banks need efficient, data-driven strategies to identify at-risk customers early and implement personalized retention measures, as current manual monitoring is costly and inefficient.

Our Solution

A comprehensive machine learning system that :

- Predicts which customers are likely to leave
- Identifies WHY they might leave (pattern discovery)
- Segments customers into actionable groups
- Recommends personalized retention strategies
- Provides interactive dashboard for business users

Bank Customer Churn Dataset (Kaggle)

Dataset Statistics :

- Total Records: **10,002** customers
- Features:**13** original features → 21 engineered features
- Target Variable: Exited (Churned = 1, Stayed = 0)
- Churn Rate: **20.4%** (class imbalance handled with SMOTE)

Feature Categories :

Demographics - Age, Gender, Geography

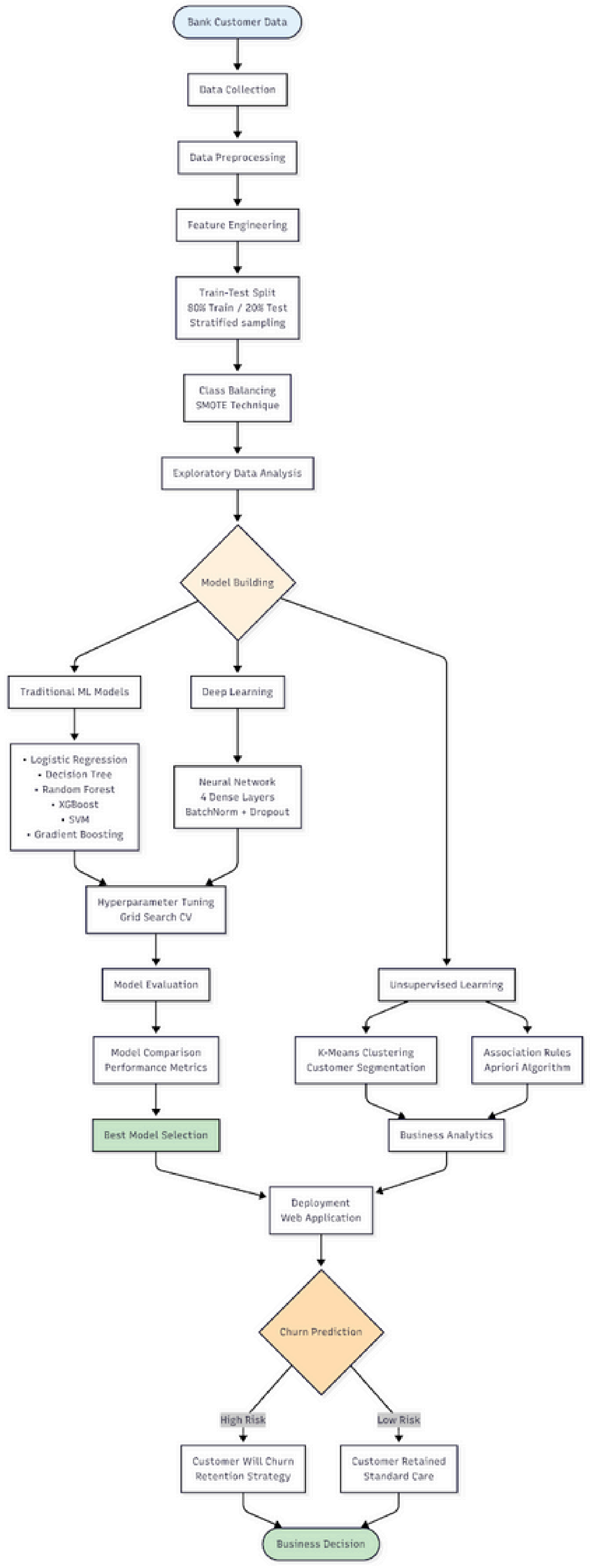
Financial - Credit Score, Account Balance, Estimated Salary

Banking Relationship - Tenure, Number of Products, Has Credit Card

Behavioral - Is Active Member

System Architecture & Workflow

Data Collection
Data Preprocessing
Feature Engineering
Train–Test Split (80/20, Stratified Sampling)
Class Balancing (SMOTE)
Exploratory Data Analysis
Model Building
Traditional ML Models
Deep Learning
Unsupervised Learning
Hyperparameter Tuning (Grid Search CV)
Model Evaluation
Model Comparison
Best Model Selection
Deployment (Web Application)
Churn Prediction
High Risk
Low Risk
Business Decision



ML Models Trained

Logistic Regression

Baseline model; simple & interpretable — Training time: 0.12s

Decision Tree

Handles non-linear boundaries — Training time: 0.40s

Random Forest

Ensemble of 100 trees; reduces overfitting — Training time: 0.99s

XGBoost

Gradient boosting; high performance — Training time: 0.57s

SVM (RBF Kernel)

Captures non-linear patterns — Training time: 36.95s

Gradient Boosting

Sequential boosting approach — Training time: 9.23s

Neural Network (Deep Learning)

4 dense layers (128 → 64 → 32 → 1), BatchNorm + Dropout — Training time: 443.47s

K- Means Clustering : Understanding Customer groups

Algorithm: K-Means clustering

Features Used: Age, Balance, Tenure, NumOfProducts, CreditScore

Optimal K: 4 clusters (determined by elbow method)

Cluster Profiles

- Cluster 0 – Dormant Accounts
- Cluster 1 – Critical High-Risk
- Cluster 2 – Mass Affluent
- Cluster 3 – Premium High-Value

Association Rule Mining (Apriori)

Methodology

- Algorithm: Apriori Association Rule Mining
- Min Support: 2%
- Min Confidence: 70%

Pattern 1 – Inactive Senior Females

89% confidence, 4.37× lift

Pattern 2 – Inactive Seniors with 1 Product

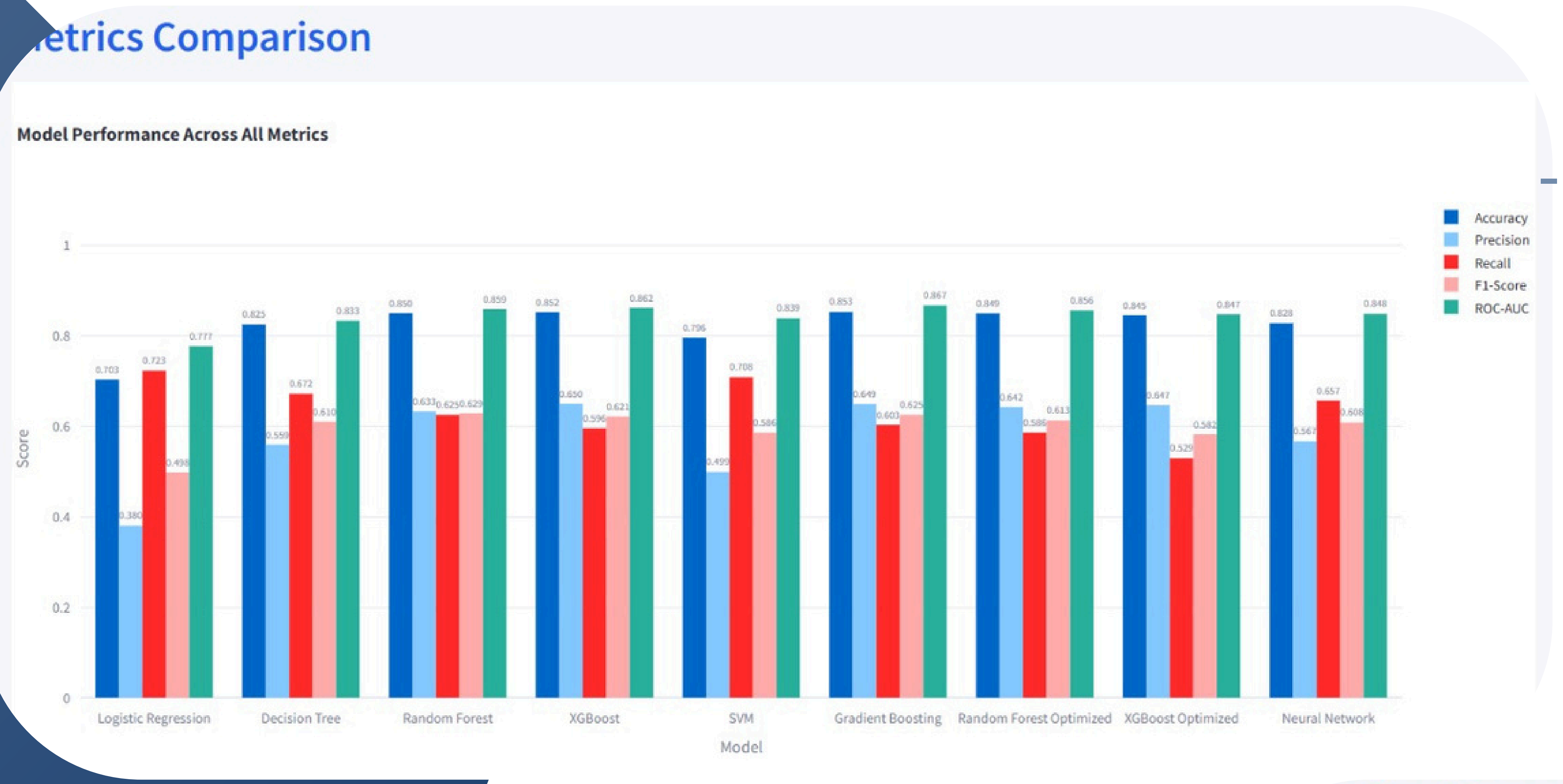
87.7% confidence

Pattern 3 – Customers with 3 Products

82.7% confidence

Surprising: More products ≠ better retention!

Model Performance Results



Best Overall Model: Gradient Boosting

- Highest accuracy (85.3%) and ROC-AUC (86.7%)

- Fast training time (9.23s)

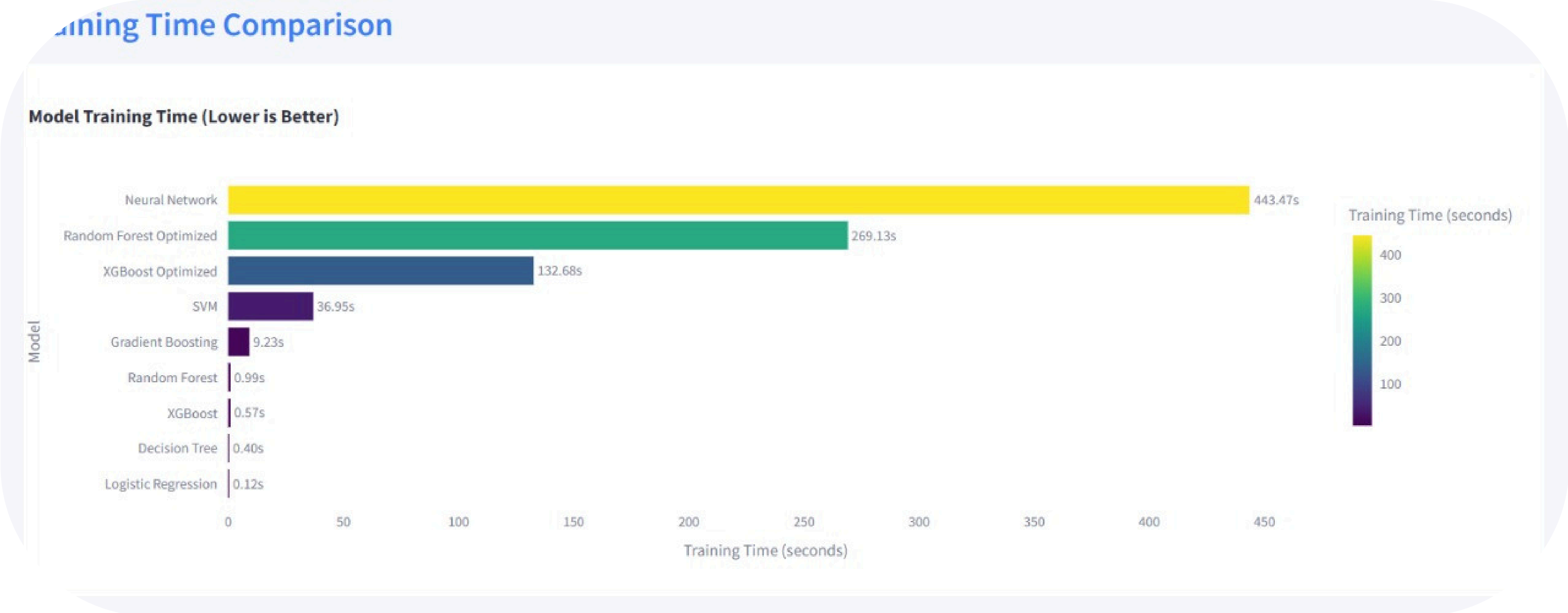
- Balanced precision-recall tradeoff

Best for Recall: SVM (70.8%)

- Catches most churners but many false alarms

Why Recall Matters:

Missing a churner costs $1,500; false alarm costs $50



Financial Impact of churn prediction

Scenario

- Customer base: 10,000 customers

Without Prediction System

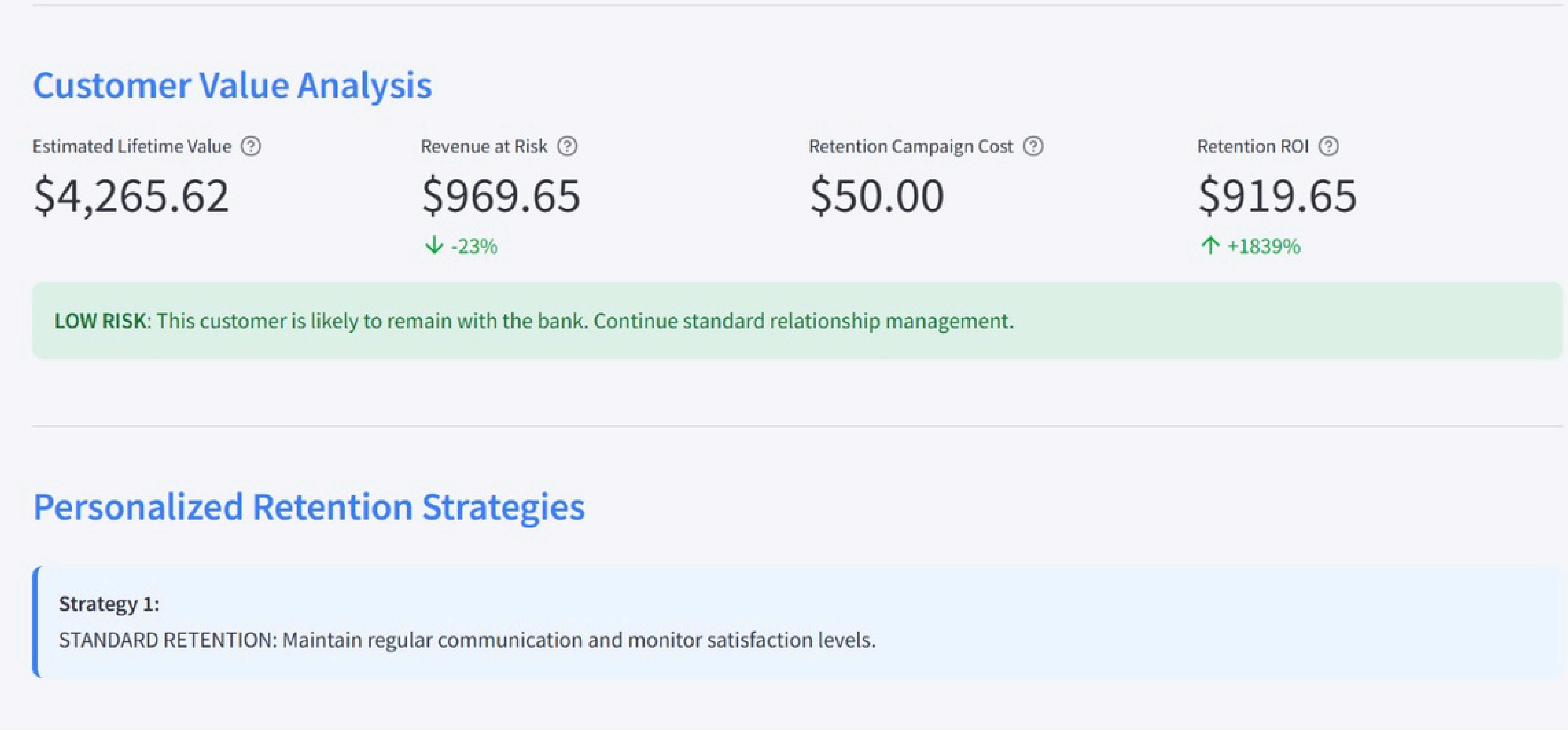
- Customers lost: 2,038 (20.4%)
- Revenue loss: $3,057,000

With Prediction System (85% accuracy)

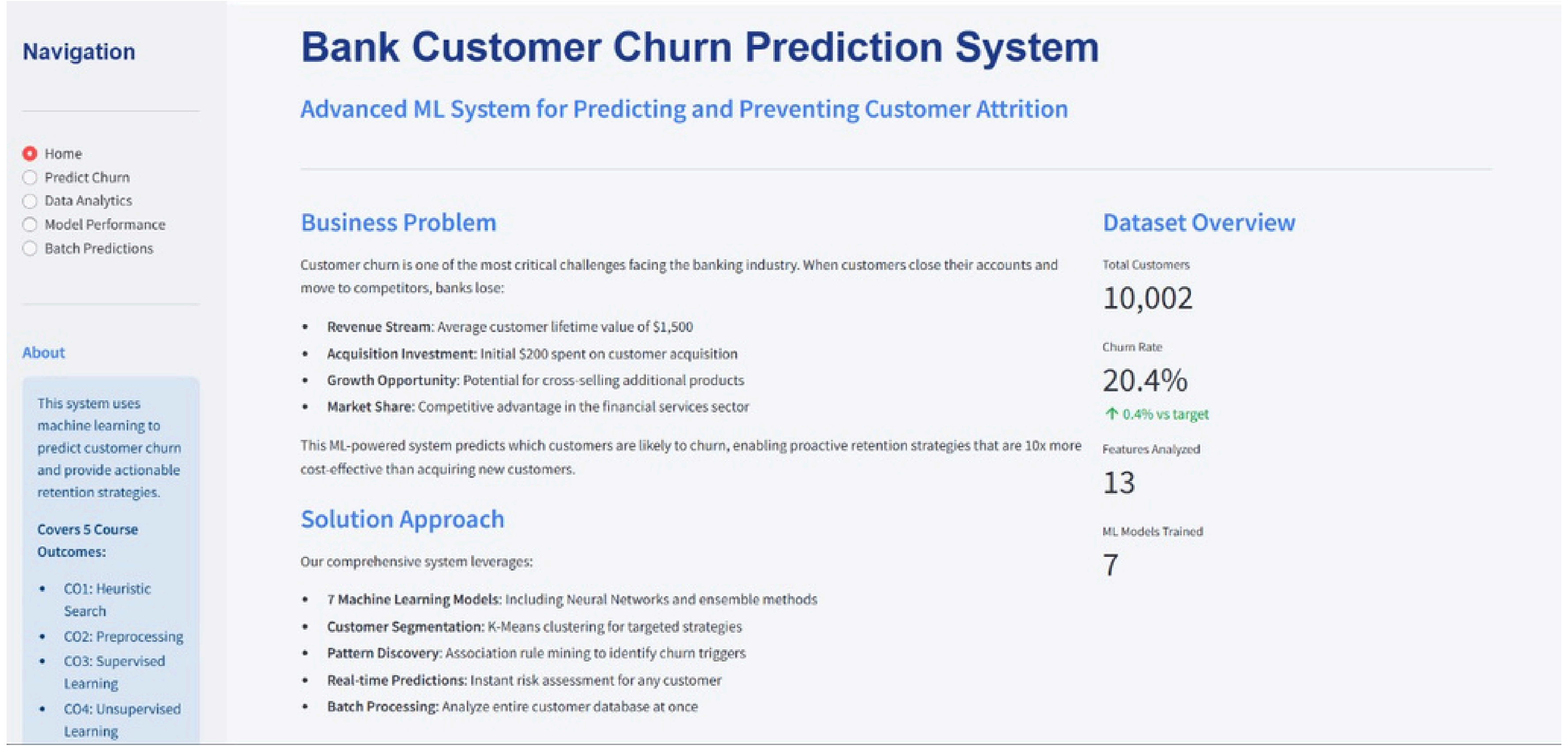
- High-risk customers identified: ~1,700
- Retention success: 50–60%
- Customers saved: 850–1,020

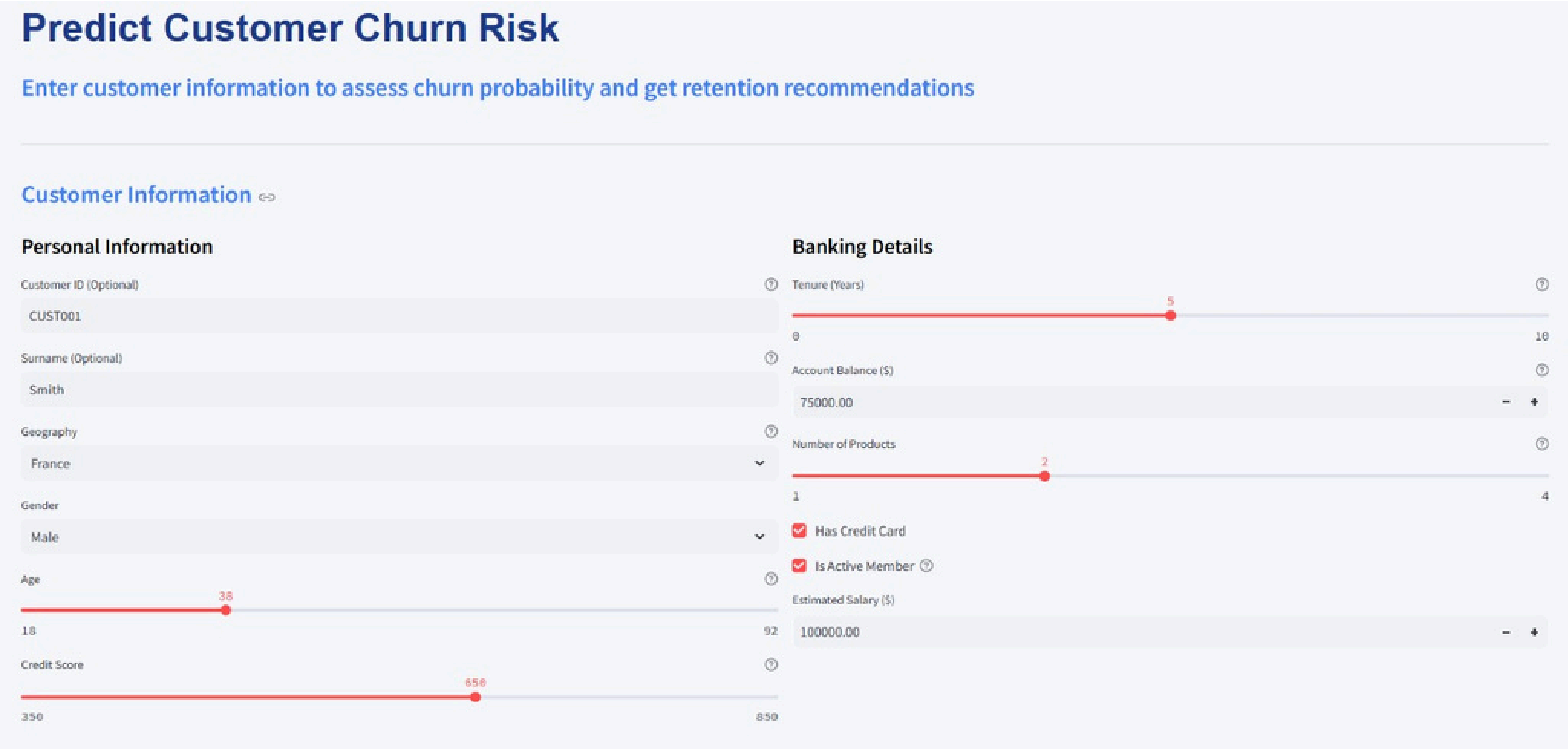
Cost–Benefit Analysis

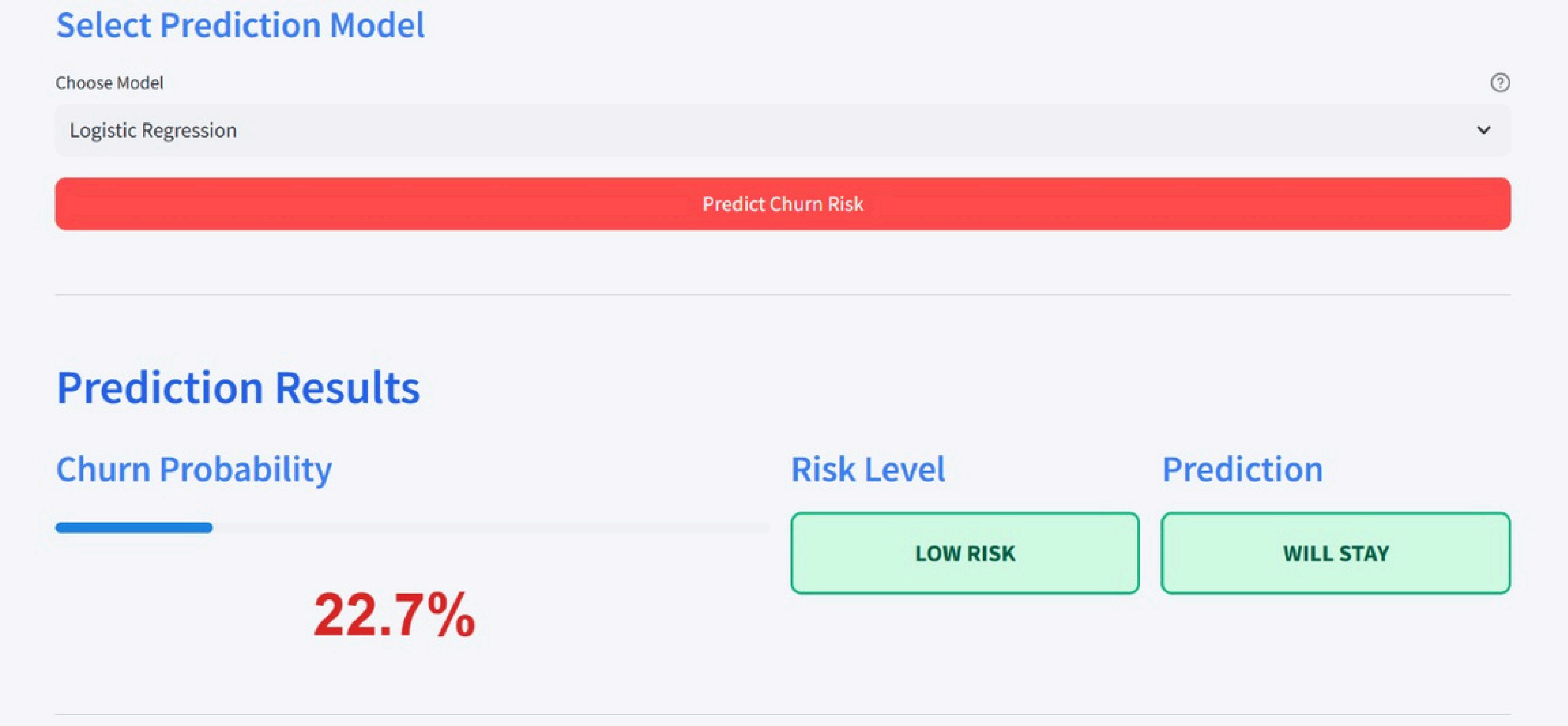
- Revenue Saved: $1,275,000 – $1,530,000
- Campaign Cost: $85,000
- False Positive Cost: ~$25,000
- Net Benefit: $1.16M – $1.42M
- ROI: 1,270% – 1,550%



Interactive Dashboard







Technology Stack & Python Libraries

- Programming & Development: Python 3.11, VS Code, Jupyter Notebook, Git
- Deployment: Streamlit
- Data Storage: CSV, Pickle (.pkl), HDF5 (.h5), JSON
- ML/AI Libraries: scikit-learn, XGBoost, TensorFlow/Keras, imbalanced-learn, mlxtend
- Data Processing: pandas, numpy
- Visualization: Plotly, Matplotlib, Seaborn
- Purpose: End-to-end ML pipeline from data processing to model deployment

Challenges & Solutions :

- Class Imbalance: Only 20% churners → SMOTE improved recall.
- Feature Selection: Identifying key features → Engineered + correlation analysis boosted accuracy.
- Model Selection: Unknown best algorithm → Compared 7 models, Gradient Boosting won.
- Interpretability: Black-box models → Association rules explained predictions.
- Real-time Deployment: Hard for business users → Streamlit dashboard enabled full adoption.

Future Enhancements :

- Enable real-time data integration with banking systems
- Implement automatic model retraining using AutoML
- Predict when customers are likely to churn (time-based forecasting)
- Estimate customer lifetime value for better retention strategies
- Use AI-powered personalized messaging with large language models (LLMs)

Conclusion

- Built end-to-end churn prediction system with 85% accuracy
- Identified 4 customer segments and 6 churn patterns
- Deployed interactive Streamlit dashboard
- **Key Insights**: Inactive members, seniors, single-product customers at highest risk
- **Business Value**: Reduces churn to ~12%, saves $1.2M+, ROI 1,270–1,550%
- **Success Factors**: Explainable AI, business-focused, user-friendly

References

Dataset:

- Bank Customer Churn Prediction Dataset – Kaggle

Documentation:

- Scikit-learn
- TensorFlow/Keras
- XGBoost
- Streamlit
- Imbalanced-learn
- MLxtend

Research Papers:

- "Customer Churn Prediction Using Machine Learning" – IEEE, 2020
- "Deep Learning for Customer Churn Prediction in Banking" – Journal of AI, 2021
- "Association Rule Mining in Banking Analytics" – Data Science Review, 2022

Thank You !