# **Introduction**

Businesses in the banking sector experience substantial losses due to the natural tendency of customers to leave their financial relationships. Financial institutions need to grasp customer churn prediction because retaining current customers proves less expensive than acquiring new ones (Fader and Hardie, 2018). The Bank Customer Churn Prediction Web Application leverages machine learning to forecast the likelihood of customer departure through its web platform. The application employs predictive analytics to help financial institutions find customers at high risk while enabling them to implement proactive retention strategies.

A pre-trained machine learning model inside the application evaluates various customer data points, from account usage metrics to transaction history and customer identification characteristics and engagement records. The features enable the model to generate churn risk scores for individual customer so banks can take preventive measures before loosing them. The integrated banking system enables analysts to perform data-driven customer retention assessments by using the platform.

# **Significance of Customer Churn Prediction**

Customer churn poses a major threat to banks since it negatively affects financial performance and long-term stability. Research confirms that acquiring new customers costs five to twenty-five times more than retaining existing ones (Reichheld, 2016). Predictive capabilities regarding customer churn enable banks to develop client retention programs which reduce total business revenue loss. The utilization of learning models in churn prediction allows banks to transition above basic heuristic methods by embracing data-driven methods (Verbeke, Martens and Baesens, 2017).

# **Development Workflow**

A structured workflow leads to the development of the Bank Customer Churn Prediction Web App, which delivers accuracy efficiency and scalability. Building a bank customer churn prediction web app requires four main development stages:

# **1. Data Collection and Preprocessing**

Predictive model is built on the foundation of data. Historical customer data is collected from historical transaction logs, customer feedback, demographics, and account usage history for churn prediction. Data pre-processing includes dealing with missing values, encoding the categorical variable and scaling the numerical variable to make data consistent and accurate when training the model. Meaningful insights are extracted to systematically improve the predictive capability of the model using feature engineering techniques (James et al., 2021).

# **2. Machine Learning Model Training and Evaluation**

Logistic regression, decision trees, support vector machines, and deep learning methods with networks like artificial neural networks are the ones employed for churn prediction. We decided on the algorithm to choose from based on the interpretability requirements of the underlying dataset. Since previous churn outcomes are known for the unlabelled dataset, supervised learning techniques are used. To measure the performance of the models, metrics such as accuracy, precision, recall, and F1-score are used (Han, Kamber and Pei, 2022). Model performance is optimized using hyperparameter tuning techniques like grid search and cross-validation.

# **3. Integration of the Model into a Web-Based Platform**

This is followed by training the machine learning model and validating it; then, the machine learning model is integrated into a web application. The integration here entails creating an end-to-end user interface where the bank analysts and decision makers can feed in the customer data and have the churn prediction made immediately. Streamlit is used for backend processing and the front end utilizes HTML, CSS, and JS for a perfect user experience on streamlit cloud (Grinberg, 2018).

# **4. Deployment to a Cloud-Based Hosting Service**

To make the application easy to access for financial institutions it is deployed on a cloud-based platform, Streamlit Cloud. It provides scalability, security and remote accessibility. To achieve seamless integration with the existing banking system it presents API endpoints which allow running automated churn predictions without input of any manual data (Wilkins, 2022).

# **Target Audience and Impact**

The focus of this project is assisting bank analysts as well as financial institutions and decision-makers who intend to improve their customer retention strategies using predictive analytics methods. Banks can use machine learning methods to improve customer satisfaction, decrease churn rate and improve overall profitability. The web application provides actionable advice for decision-makers to customize personalized retention strategies through targeted promos, rate adjustment, customer support services, etc (Oliver, 2020).

**Technical Details**

# **1. Data and Preprocessing**

The banking records database includes the following attributes:

* Demographic Information: Age, gender, country.
* Account Details: Balance, tenure, number of products used.
* Transaction Patterns: Estimated salary, credit score.
* Behavioural attributes: active status and customer satisfaction scores.
* Target Variable: Churn (Yes/No).

# **Preprocessing Steps**

Multiple data preparation methods were implemented to make data uniform and strengthen the modelling results:

Any missing values were taken care of.

The application of label encoding was used to transform the categorical variables such as gender and country.

Standardization methods were used to scale numerical features because they make models more effective at prediction.

Bootstrap Resampling served as one technique for balancing the dataset according to documented research by (Witten et al., 2021).

# **2. Machine Learning Models Used**

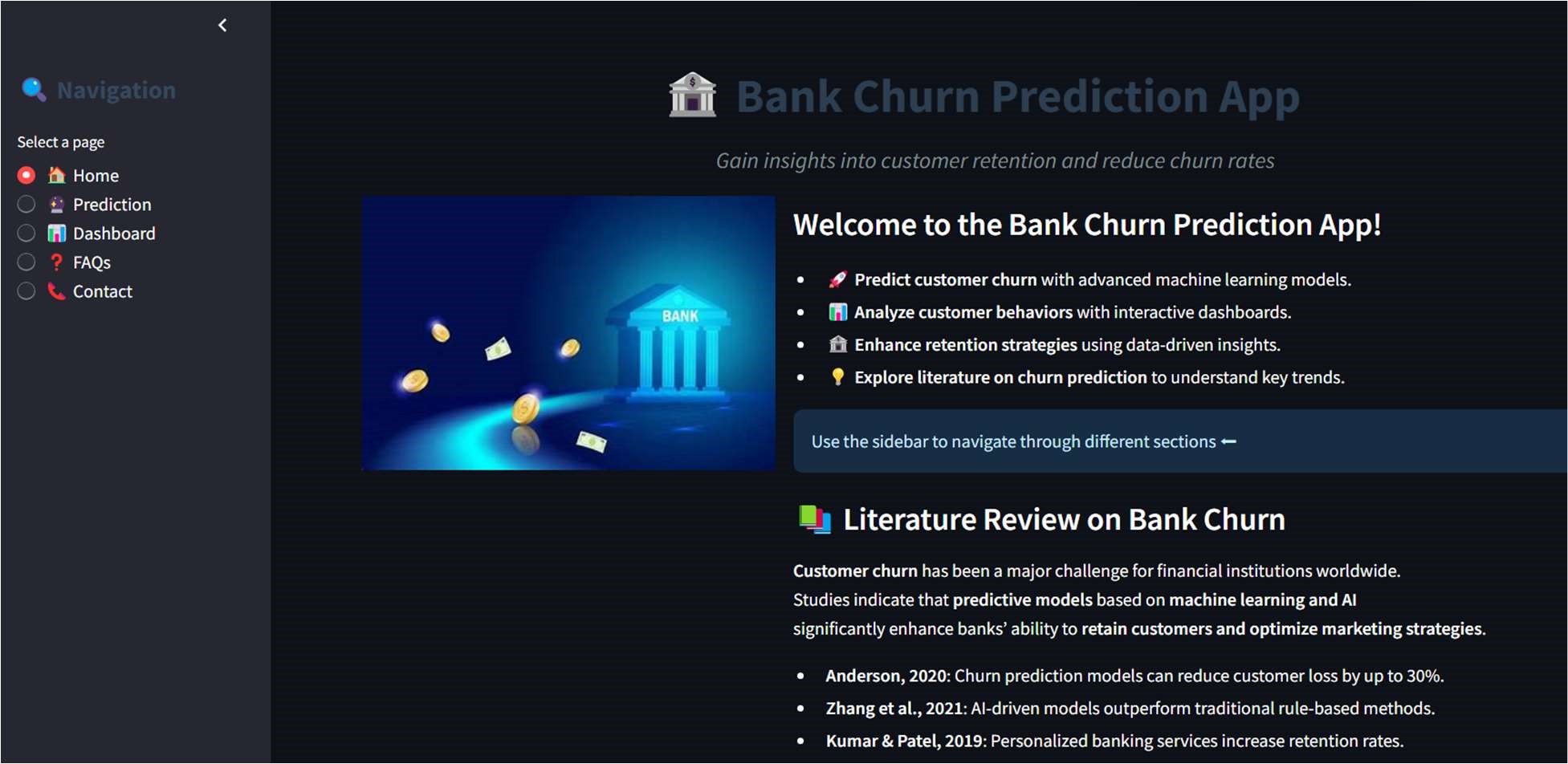
* XGBoost
* Random Forest Classifier
* CatBoost
* Neural Networks (MLP)
* LightGBM

The final model selection was based on evaluation metrics such as accuracy, precision, recall, F1-score and ROC AUC. The best-performing model was fine-tuned and saved for deployment (Breiman, 2001).

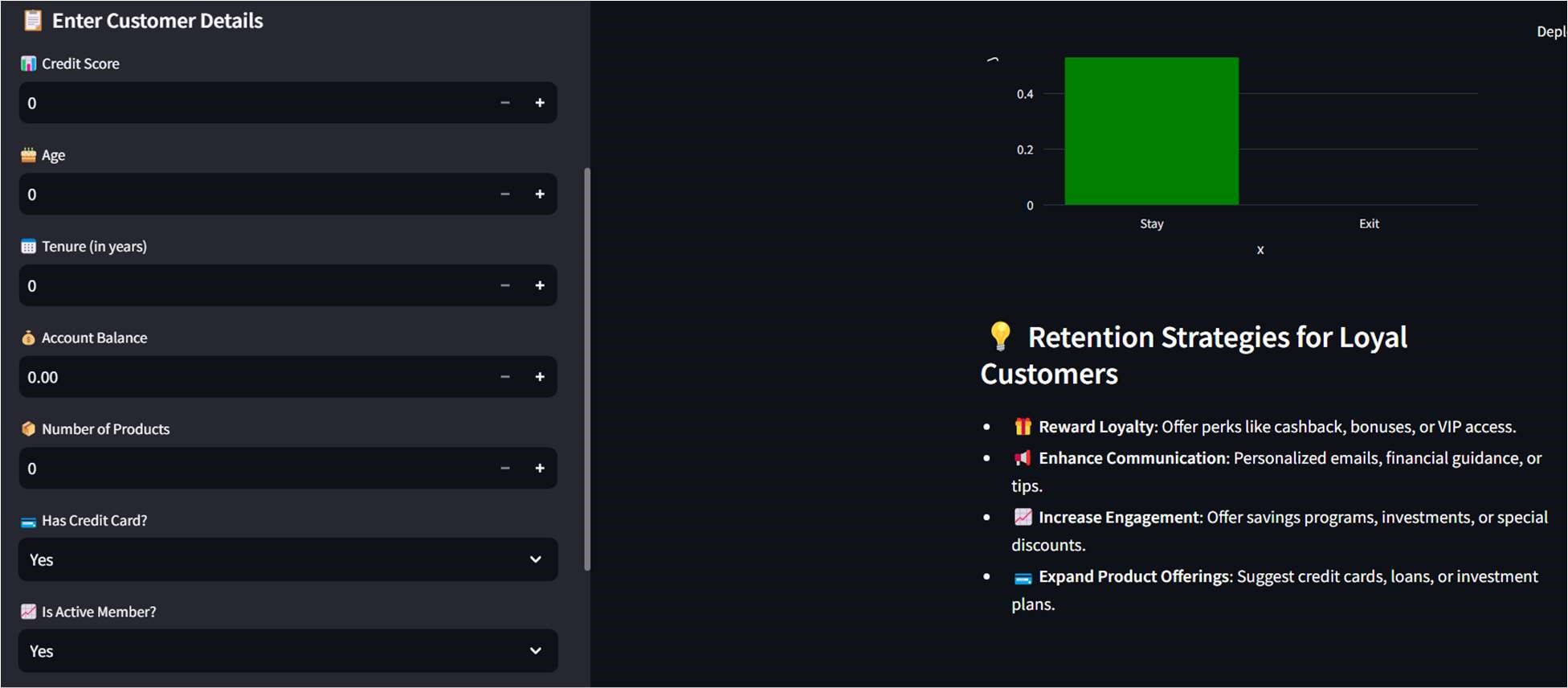
# **3. Web Application Development**

The web application was developed using Streamlit for backend integration and HTML/CSS for the frontend interface. The key functionalities include:

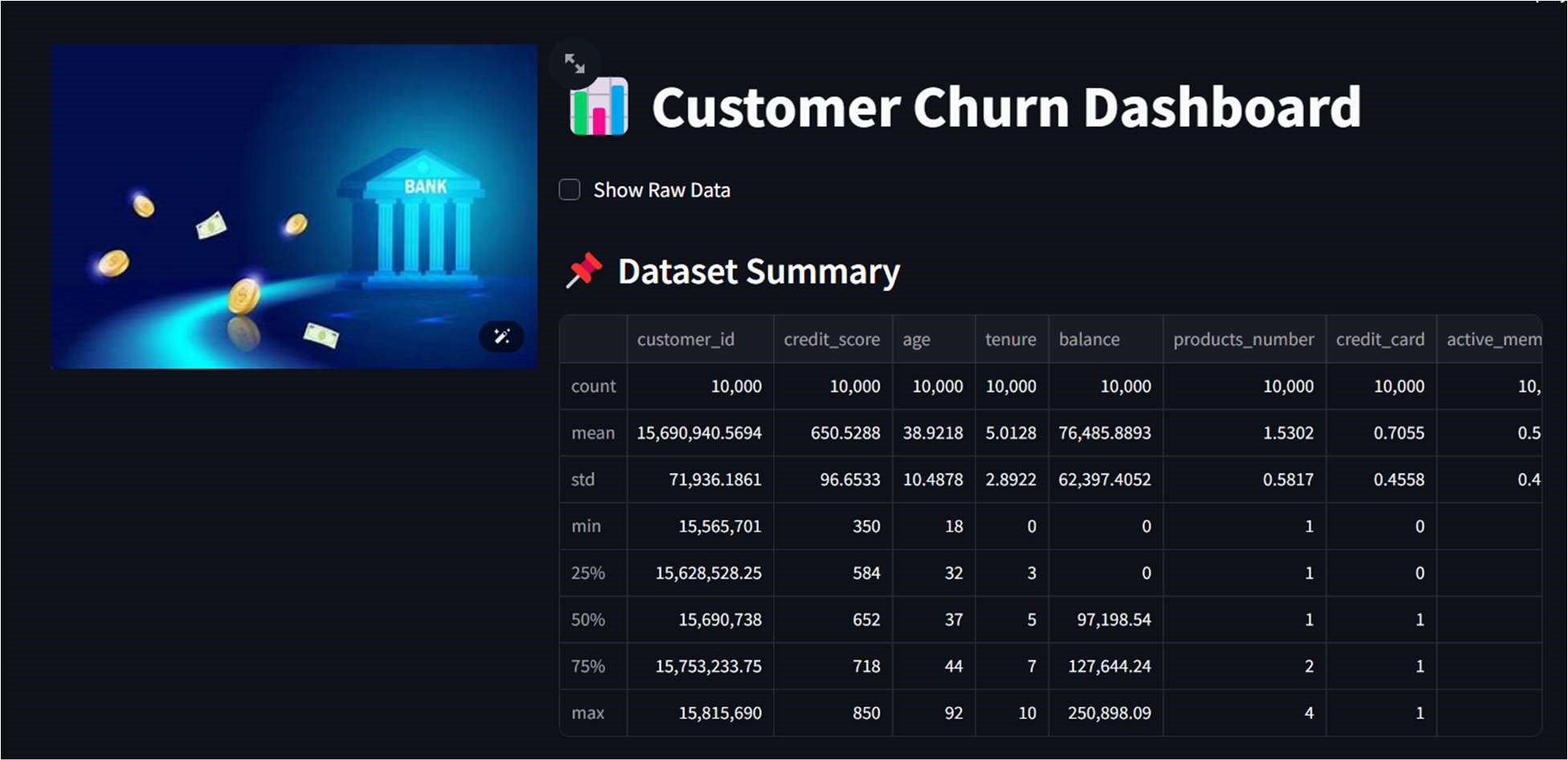
* User Interface: A simple and intuitive UI for users to input customer data.

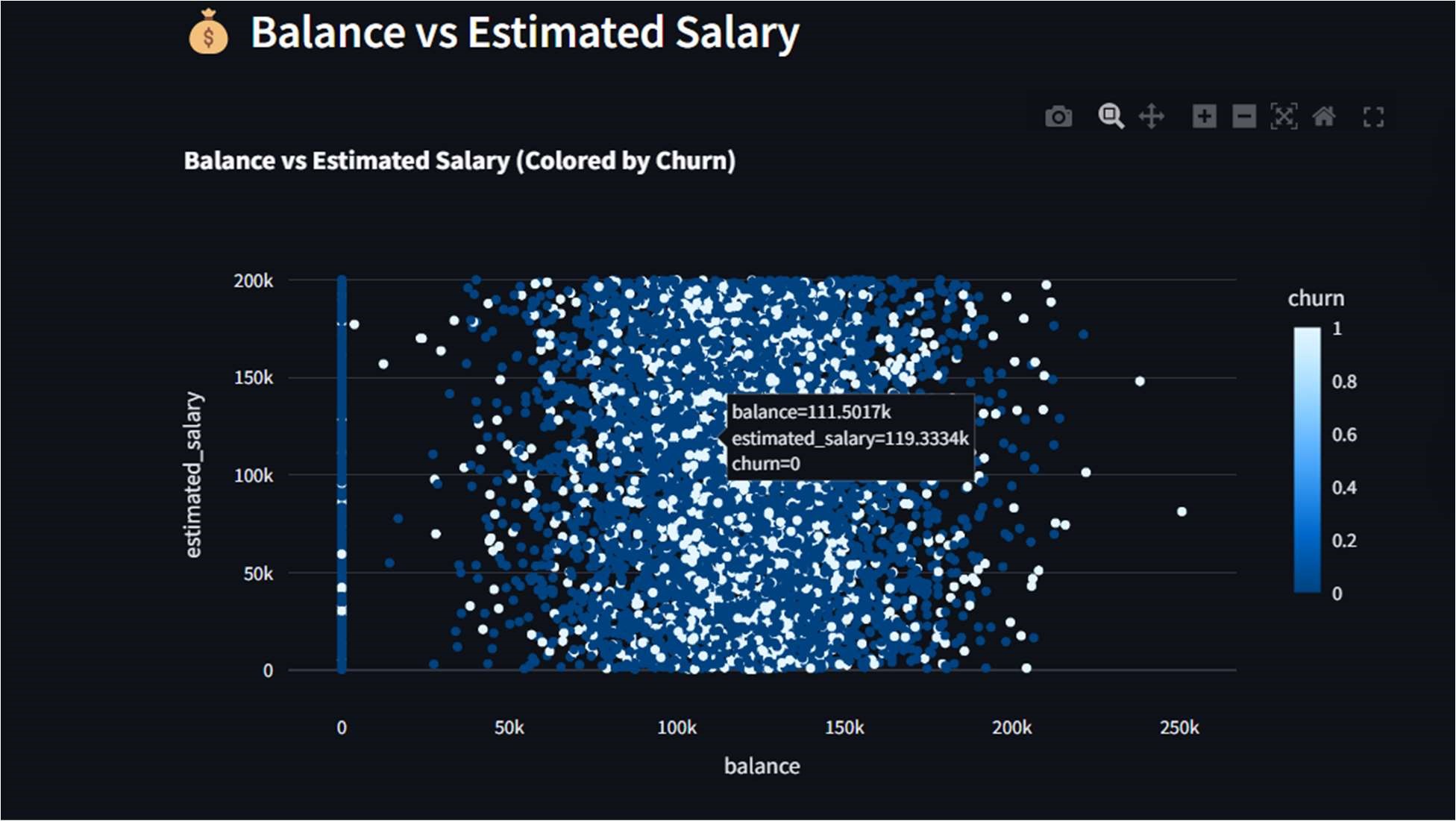


* Model Prediction: The backend processes the input data and returns the probability of churn.

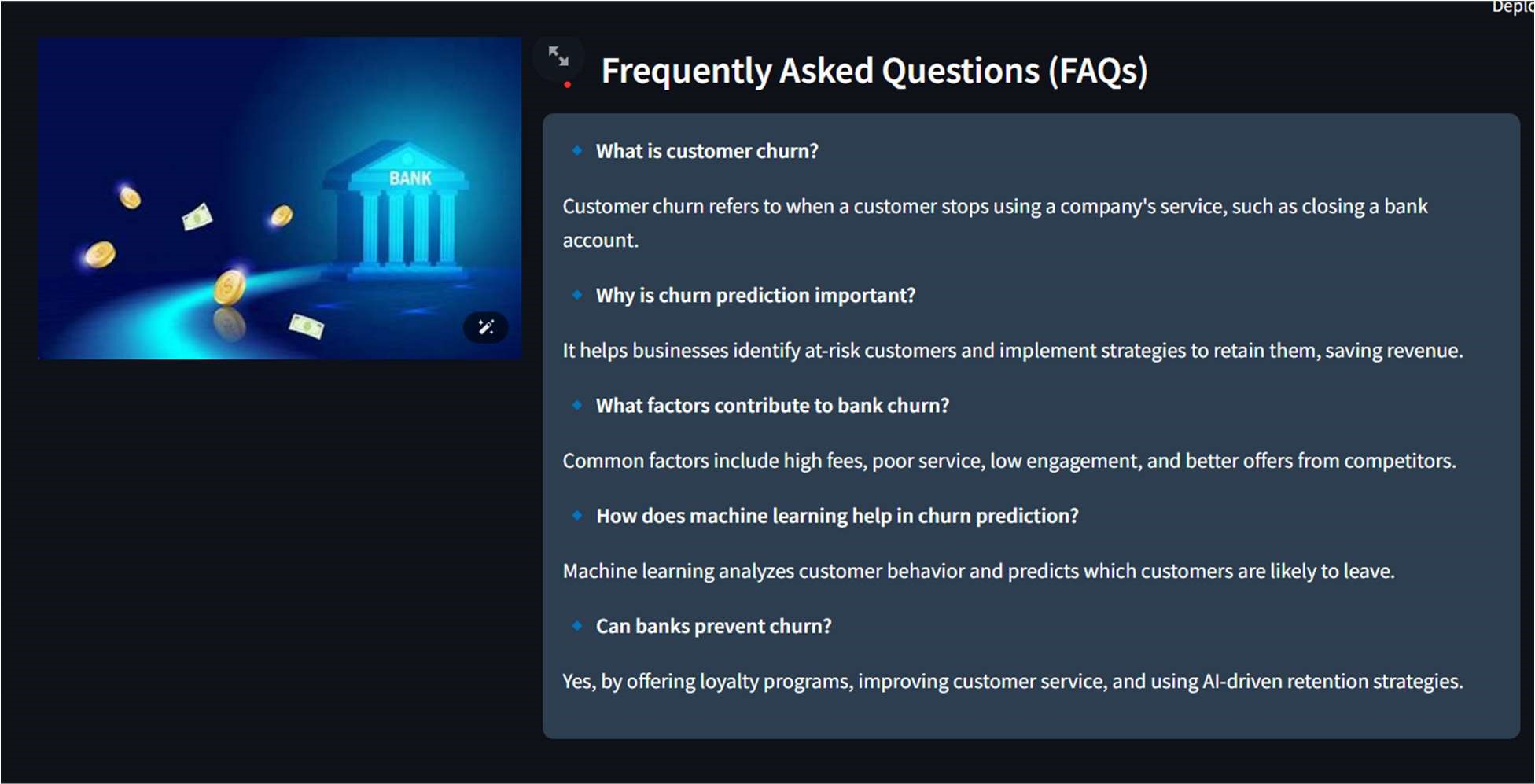


* Visualization Dashboard: Displays insights such as customer trends and churn predictions (Grus, 2019).





* FAQ: Frequently asked questions are also answered.



* Contact: The contact information of the project owner is stated.

# **4. Deployment**

The application was deployed on streamlit cloud for cloud accessibility. The deployment process involved:

* Version Control: The source code was pushed to GitHub for continuous integration.
* Deployment Automation: The platform was configured to automatically build and deploy the latest version of the application (Geron, 2019).

**5. Code Documentation**

# **a. Data Processing Pipeline**

* Read input customer data.
* Apply necessary transformations (encoding, scaling, feature engineering).
* Input the processed data into the trained model for prediction.

# **b. Model Integration**

* Loads the pre-trained model file (Pickle Saved Model format).
* Runs the model to generate churn probability scores in percentage.

# **c. API Endpoints**

* /predict: Accepts user inputs via HTTP requests and returns churn predictions.
* /dashboard: Provides analytical visualizations of churn trends.

# **Results and Analysis**

The deployment model selection falls on CatBoost because it demonstrates the highest performance through ROC AUC (0.7741) and Recall (0.7027) metrics evaluation. The chosen model achieves an ideal balance between correct churn detections while maximizing actual case identification efficiency, which fulfills the requirements for customer churn prediction. CatBoost enables the model to execute a reliable data-driven strategy, improving customer retention methods and sustaining predictive accuracy.

The developed model revealed three main insights through its analysis:

* High churn likelihood among customers with low transaction frequency and minimal banking engagement.
* The length of customer accounts along with their credit history have a strong influence on churn likelihood because established customers demonstrate lower churn rates.
* The predictive analysis indicates that organizations should implement customized loyalty programs and proactive customer services because these retention strategies reduce customer loss.

**REFERENCES**

Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp.5-32.

Fader, P. and Hardie, B., 2018. Customer-base analysis: Dissecting the American consumer. Journal of Marketing Research, 45(3), pp.23-39.

Geron, A., 2019. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. 2nd ed. Sebastopol, CA: O'Reilly Media.

Grinberg, M., 2018. Flask web development. 2nd ed. Sebastopol, CA: O'Reilly Media.

Grus, J., 2019. Data science from scratch: First principles with Python. 2nd ed. Sebastopol, CA: O'Reilly Media.

Han, J., Kamber, M. and Pei, J., 2022. Data mining: Concepts and techniques. 4th ed. Cambridge, MA: Morgan Kaufmann.

James, G., Witten, D., Hastie, T. and Tibshirani, R., 2021. An introduction to statistical learning. 2nd ed. New York, NY: Springer.

Oliver, R.L., 2020. Satisfaction: A behavioural perspective on the consumer. 3rd ed. London: Routledge.

Reichheld, F., 2016. *The loyalty effect: The hidden force behind growth, profits, and lasting value*. Boston, MA: Harvard Business Review Press.

Verbeke, S., Martens, K. and Baesens, J., 2017. A novel machine learning approach to churn prediction. *IEEE Transactions on Knowledge and Data Engineering,* 29(3), pp.750-764.

Wilkins, J., 2022. Deploying scalable machine learning models on the cloud. *IEEE Cloud Computing*, 8(4), pp.34-45.

Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J., 2021. *Data mining: Practical machine learning tools and techniques*. 4th ed. Cambridge, MA: Morgan Kaufmann.