C964: Computer Science Capstone Template

Task 2 parts A, B, C and D

Contents

[Part A: Letter of Transmittal 2](#_Toc148815217)

[Part B: Project Proposal Plan 5](#_Toc148815218)

[Client Overview & Needs 5](#_Toc148815219)

[Data Summary 8](#_Toc148815220)

[Implementation 9](#_Toc148815221)

[Timeline 12](#_Toc148815222)

[Evaluation Plan 12](#_Toc148815223)

[Resources and Costs 15](#_Toc148815224)

[Part C: Application 17](#_Toc148815225)

[Part D: Post-implementation Report 17](#_Toc148815226)

[Solution Summary 17](#_Toc148815227)

[Data Summary 21](#_Toc148815228)

[Machine Learning 22](#_Toc148815229)

[Validation 25](#_Toc148815230)

[Visualizations 28](#_Toc148815231)

[User Guide 32](#_Toc148815232)

[Reference Page 34](#_Toc148815233)

# Part A: Letter of Transmittal

Saturday, October 21, 2023

Senior Leadership

ABC Multinational Bank

1352 Cliffside Drive, Syracuse NY, 13204

Dear Senior Leadership Team,

I am writing to present a solution addressing a prevalent challenge many businesses face today, including ours: customer retention. The specific problem at hand is identifying customers most likely to discontinue using our banking services, which, if not addressed promptly, could have significant financial implications.

**Summary of the Problem:**

Customer churn is the term used when a client stops using a service. For our institution 'ABC Multinational Bank,' losing customers, especially long-standing ones, is not just a loss of a single account but a potential loss of future revenue, word-of-mouth referrals, and increased acquisition costs for new customers. Moreover, “acquiring a new customer can be up to five times more expensive than retaining an existing one."(*Customer retention – what is it and why does it matter for your business* 2017)

**Proposed Solution:**

Using a data-driven approach, I have developed a machine learning model that analyzes our existing customer data and predicts the likelihood of their attrition. This model uses various demographic and transactional factors, including credit score, age, country, gender, and product holdings. Moreover, I've incorporated an interactive application allowing stakeholders to input specific customer data and receive instantaneous predictions on their likelihood to churn.

.

**Benefits to the Organization:**

By identifying high-risk customers in advance, we can initiate proactive strategies, such as personalized offers or outreach programs, to improve retention rates, leading to increased customer satisfaction, reduced marketing costs, and a better understanding of our customers' needs and preferences.

**Implementation Plan:**

Data Collection: Utilize our existing database of over 10,000 customers.

Data Cleaning & Preprocessing: Ensure data accuracy and relevancy.

Model Training: Using the cleaned data to train our predictive algorithm.

Deployment: Integrating the model into our IT infrastructure and making the interactive tool accessible to relevant teams.

Ongoing Maintenance: Regular updates to the model based on new customer data and feedback.

**Concerns & Expertise:**

The primary costs pertain to the computational resources for model training and potential software licenses for deployment. The estimated timeline for full implementation is four weeks. Ethically, all data used respects customer privacy, with no personally identifiable information utilized in the model. My Data Science and Machine Learning background equips me with the necessary skills to lead and implement this project successfully.

In conclusion, this predictive model will be instrumental in addressing our customer retention challenges, and I request your endorsement to proceed further. I am open to discussing this in more detail and answering any questions.

Thank you for considering this proposal.

Sincerely,

Guson Ulysse, Lead Data Science

# Part B: Project Proposal Plan

**Project Proposal for Customer Churn Analysis & Prediction**

**Introduction:**

In the rapidly evolving financial industry landscape, 'ABC Multinational Bank,' like many other institutions, is grappling with the challenge of customer retention. As the banking sector becomes increasingly competitive, the imperative to understand, predict, and act on potential customer attrition has always remained the same.

Client Overview & Needs**:**

'ABC Multinational Bank' is a progressive financial institution aiming to provide unparalleled customer service. However, recent internal assessments suggest that the bank is witnessing a higher-than-average customer attrition rate. While the bank has a wealth of customer data, it needs advanced analytical tools to extract meaningful insights. Middle management needs actionable insights to devise targeted retention strategies, especially in the client relations and IT sectors.

**Problem Statement:**

The principal challenge is the early identification and prediction of customers likely to discontinue their association with 'ABC Multinational Bank.' Addressing this issue head-on will safeguard revenue streams and empower the bank to refine its service offerings based on tangible data-driven feedback.

**Proposed Deliverables:**

**Predictive Analysis Application:** A machine learning model will be developed to analyze and predict customer churn based on demographic and transactional attributes. This application will provide batch predictions (for large datasets) and real-time predictions for individual queries.

**Interactive Interface:** An intuitive user interface will be built on the Google Colab. This tool will allow stakeholders to input specific customer details and receive instant churn predictions, facilitating immediate decision-making.

**User Guide:** A comprehensive guide detailing the operation of the application and interface and its features.

**Periodic Reports:** Generated insights, predictions, and analyses will be compiled into monthly or quarterly reports, offering a holistic view of the customer base's behavior and tendencies.

**Benefits to the Client:**

The proposed application will be a linchpin in the bank's customer retention strategy. By leveraging the predictive power of machine learning:

**Proactive Intervention:** High-risk customers can be identified in advance, enabling targeted outreach and personalized service offers.

**Financial Savings:** With enhanced retention, the bank can reduce the costly process of new customer acquisition.

**Strategic Planning:** Insights from the data can guide the bank's future offerings, promotions, and services, ensuring they are closely aligned with customer needs and preferences.

In conclusion, the future of banking lies in harnessing the power of data to serve customers better. This project, once implemented, will position 'ABC Multinational Bank' at the forefront of customer-centric innovation in the banking sector.

## Data Summary

**Data Source and Collection Method:**

The raw data for this project has been sourced from Kaggle, a reputable platform for machine learning and data analytics datasets. The dataset can be accessed through the following link: [Kaggle Dataset](https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset/data).

**Data Management and Processing:**

Throughout the application development lifecycle, data will be managed and processed as follows:

**Design Phase:** At this juncture, an in-depth data analysis will be conducted to identify relevant attributes and potential preprocessing requirements. Anomalies, such as outliers and missing data, will be addressed using appropriate statistical and data imputation techniques.

**Exploration Phase with Visualization: The focus will be effectively communicating the data story. Through various types of visualizations, the data will be used to uncover insights that guide the subsequent stages of data preprocessing and model building while attempting to tailor the visualizations to the intended audience.**

**Development Phase:** The data will be split into training and test sets. Various machine learning algorithms will be trained on the training dataset and validated on the test dataset. This division ensures the model's generalization to unseen data.

**Maintenance Phase:** Continuous monitoring of the model's performance will be undertaken. The model will be retrained to ensure its accuracy and relevance as new data becomes available.

**Justification for Data Suitability:**

The chosen dataset comprises attributes highly pertinent to the banking sector, such as credit score, country, gender, and estimated salary. Its comprehensive nature ensures that the models can capture intricate patterns and relationships. While there's a noticeable class imbalance in customer churn, with approximately 79.63% of customers staying with the bank and 20.37% churning, this mirrors real-world scenarios where churn rates can be significantly lower than retention rates.

**Handling of Data Anomalies:**

Given the inherent class imbalance, we will deploy techniques like oversampling, undersampling, or Synthetic Minority Over-sampling Technique (SMOTE) to ensure our models are not biased towards the majority class. Additionally, data outliers will be managed using robust statistical methods, and any incomplete data will either be imputed using sophisticated algorithms or omitted, depending on its significance.

**Ethical and Legal Concerns:**

We place the highest emphasis on data ethics. As a testament to this commitment, we have excluded the Customer ID attribute to preserve the privacy and confidentiality of bank customers. This practice aligns with our ethical standards and ensures that no individual identities or sensitive information can be deduced. Additionally, the data has been sourced from a public platform, negating any proprietary rights or legal impediments. By adhering to these principles, we ensure responsible data usage that benefits the organization without infringing individual rights.

## Implementation

1. **Industry-Standard Methodology: CRISP-DM (Cross Industry Standard Process for Data Mining)**
   1. **Business Understanding**
      * Define the project objectives.
      * Assess the current situation by gathering information about available resources, constraints, and other factors.
      * Determine business success criteria.
   2. **Data Understanding**
      * Collect initial data from Kaggle.
      * Describe the data to get familiar with it.
      * Explore data quality issues and address them.
      * Verify that the data is sufficient for the project goals.
   3. **Data Preparation**
      * Clean the data (removing the customer IDs)
      * Format the data into a suitable form for modeling.
   4. **Modeling**
      * Select appropriate modeling techniques.
      * Generate test design for model assessments.
      * Build the models.
      * Assess the models’ quality.
   5. **Evaluation**
      * Review the process and assess if it can be improved.
   6. **Deployment**
      * Deploy from Jupyter Notebook to Google Colab
      * Install necessary libraries to ensure the Application runs as intended
2. **Implementation Plan Outline**
   1. **Project Initialization**
      * **Objective Definition:** Clearly articulate the problem to be solved.
      * **Resource Allocation:** Assign roles for data scientists, data engineers, and other essential personnel.
   2. **Data Collection and Understanding**
      * **Data Acquisition:** Retrieve data from Kaggle
      * **Exploratory Data Analysis (EDA):** Visualize and summarize the data to gain insights.
   3. **Model Development**
      * **Model Selection:** Choose the appropriate algorithms based on the problem type (classification in this case.).
      * **Training:** Use the training data to train the models.
      * **Validation:** Hypertune the models.
   4. **Model Evaluation**
      * **Testing:** Evaluate model performance on unseen test data.
      * **Performance Metrics:** Use metrics like accuracy, precision, recall, F1 score, etc., to assess model quality.
      * Choose the highest-performing model to proceed with.
   5. **Deployment**
      * **Integration:** Integrate the model into Google Colab to allow its use across the company.

By following the CRISP-DM methodology and the outlined implementation plan, a systematic and thorough approach can be taken to develop and implement machine learning solutions in a business context.

## Timeline

|  |  |  |  |
| --- | --- | --- | --- |
| Milestone or deliverable | Duration  (hours or days) | Projected start date | Anticipated end date |
| Project Initialization | Three days | 10/21/23 | 10/23/23 |
| Data Collection and Understanding | Seven days | 10/24/23 | 10/31/23 |
| Model Development | Fourteen days | 11/1/23 | 11/15/23 |
| Model Evaluation | Seven days | 11/16/23 | 11/23/23 |
| Deployment | Two days | 11/24/23 | 11/26/23 |

## Evaluation Plan

Verification Methods at Each Stage of Development:

1. **Data Pre-processing Stage**:
   1. **Data Integrity Checks**: Ensure no data corruption occurs during collection or loading. This was done by checking for null values.
2. **Model Development Stage**:
   1. **Formatting the data**: The country and gender attributes were converted to boolean values (True or False).
3. **Model Tuning Stage**:
   1. **Hyperparameter Sensitivity**: Ensure that the model is not overly sensitive to minor changes in hyperparameters. It shouldn't drastically change performance with minor tweaks.
   2. **Consistency Checks**: The model's performance should be consistent across different subsets of the data.
4. **Model Evaluation and Selection:**
   1. **Evaluation**: After preprocessing and formatting the data, we tested it across four machine-learning models to determine the best performer for our specific use case. The models evaluated were K-Nearest Neighbors (KNN), Random Forest, XGBoost (XGB), and Logistic Regression.
   2. **Selection**: Each model was assessed based on accuracy, precision, recall, and other relevant performance metrics. The aim was to choose a model that fits the data well and aligns with our business goals of accurately identifying customers likely to churn.
5. **Deployment Stage**:
   1. **Integration Testing**: Ensure that the model integrates seamlessly into Google Colab.
6. **Validation Method Upon Completion of the Project:**
   1. **Test Validation**: This involves setting aside a portion of the dataset (e.g., 20%) that the model has yet to see during training. After training on the remaining data, the model's performance is evaluated on this held-out set. This gives an idea of how the model might perform in real-world scenarios.
   2. **Performance Metrics**: Specific metrics will be prioritized depending on the business objective. For churn prediction, metrics like Accuracy, Precision, Recall, and F1-score
   3. Upon analysis, the RandomForest model emerged as the top performer among the four. To dive deeper into its performance, we used a confusion matrix. This matrix provided a comprehensive breakdown of the true positives, true negatives, false positives, and false negatives predicted by the model.
   4. **Real-world use**:Post-deployment, gather feedback from end-users or stakeholders. This real-world feedback is invaluable for validating the model's predictions in business contexts.

The evaluation plan is structured to ensure the model's robustness, consistency, and relevance in real-world scenarios. By employing a combination of verification and validation methods, we can confidently assess the model's readiness for deployment and its expected impact on the production environment.

## Resources and Costs

|  |  |
| --- | --- |
| Item | Cost (USD) |
| Workstations (5 units) | $5,000 |
| Backup storage system | $1,500 |
| Software licenses: |  |
| - Data management processing suite | $750 |
| Total Hardware and Software Costs | $7,250 |

|  |  |  |  |
| --- | --- | --- | --- |
| Role | Hourly Rate (USD) | Estimated Hours | Total Cost (USD) |
| Data Scientist | $60 | 300 | $18,000 |
| Data Engineer | $55 | 250 | $13,750 |
| Application Developer | $50 | 400 | $20,000 |
| Project Manager | $65 | 150 | $9,750 |
| Quality Assurance Specialist | $45 | 200 | $9,000 |
| Total Labor Costs |  | 1,300 | 70,500 |

|  |  |
| --- | --- |
| Item | Cost (USD) |
| Google Drive (for data storage) | $100 |
| Maintenance software updates | $500 |
| Total Environment Costs (Colab) | $600 |

**Overall Estimated Budget (with Colab): $78,350**

# Part C: Application

Part C is your submitted application. This part of the document can be left blank or used to include a list of any submitted files or links.

Link: [Link to Google Colab](https://colab.research.google.com/drive/1Vx90sdC41tB-qXFVG_svyK_OD5E8dfdJ?usp=sharing)

# Part D: Post-implementation Report

## Solution Summary

**Problem:**

The financial sector, especially banking, is witnessing an increasing rate of customer churn – a term used to describe customers who stop using a bank's services. Identifying potential churners in advance can allow banks to devise retention strategies, saving money and customer relationships.

The challenge was to predict the likelihood of a bank's customers churning using historical data. This data encompassed various attributes of customers, including demographics, account balances, product usage, and other relevant features. A significant aspect of this problem was the class imbalance observed in the dataset: out of 10,000 customers, approximately 79.63% did not churn, while about 20.37% did.

A diagram of a pie chart

Description automatically generated

**Solution:**

To address the problem, we constructed a machine-learning model to predict customer churn. We explored multiple algorithms to ensure the best predictive accuracy and settled on using the RandomForest classifier, which demonstrated the highest performance.

During the initial data processing stage, fields such as country and gender were formatted to Boolean values (True/False) for better compatibility with our models. Given the uneven distribution of churners to non-churners, we employed the SMOTE technique to balance the classes, ensuring our model was not biased toward the majority class.

After testing across four machine learning models, including K-Nearest Neighbors, RandomForest, XGBoost, and Logistic Regression, the RandomForest classifier emerged as the top performer with an accuracy of 89%. Its efficacy was further verified using a confusion matrix, which provided insights into its precision and recall.

A graph of different colored bars

Description automatically generated

The entire solution was hosted on Google Colab, providing a scalable and accessible platform for further iterations and usage. However, we know no model is perfect, and false positives and negatives can occur.  
A yellow and purple squares with numbers

Description automatically generated  
The model performed well in correctly identifying both churning and non-churning customers. However, the number of false negatives (185) indicates room for improvement. Reducing the FN should be a priority, given the business consequences of missing out on potential churners. On the other hand, the model's ability to correctly predict 1368 out of the 1553 actual churners (sum of TP and FN) shows its effectiveness in identifying a significant portion of at-risk customers.

## Data Summary

**Source of Raw Data:**

The raw data for this project was sourced from Kaggle, a platform that hosts datasets contributed by data enthusiasts worldwide. Specifically, the dataset can be accessed via the following link: [Bank Customer Churn Dataset](https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset/data).

**Data Collection:**

The dataset on Kaggle represents a compilation of banking customer details. Each record in the dataset corresponds to a unique customer, encompassing various attributes.

**Data Processing and Management:**

1. **Design Phase:**
   1. **Data Cleaning:** Initially, the dataset underwent a thorough cleaning process. This involved searching, handling null values, and ensuring data consistency across all attributes.
   2. **Data Transformation:** Certain fields, such as 'country' and 'gender,' were transformed from categorical representations to boolean (True/False) format to facilitate more accessible analysis and model training.
2. **Development Phase:**
   1. **Class Imbalance Handling:** Given the observed class imbalance in the 'churn' attribute, the Synthetic Minority Over-sampling Technique (SMOTE) was employed to balance the dataset. This ensured the model would not be biased towards the majority class and could make reliable predictions for both churning and non-churning customers.
   2. **Model Testing:** The processed data was split into training and testing subsets. Four distinct machine learning models were trained on this data, with their performances compared to determine the best-fit model.
3. **Maintenance Phase:**
   1. **Performance Monitoring:** The model's performance metrics were regularly monitored. This enabled timely identification of areas of improvement and subsequent model tweaks via hypertuning.

Strict adherence to ethical guidelines was maintained throughout the application development life cycle. All data processing and management activities were executed to preserve data integrity, authenticity, and privacy. With customer trust paramount, sensitive attributes like the Customer ID were deliberately omitted from the dataset, reinforcing our commitment to data privacy and ethical use.

## Machine Learning

*What?*

**Overall Product:**

Our application was developed to predict customer churn, allowing ABC Multinational Banks to identify potential customers who might leave their service. By doing so, companies can take preemptive measures to retain these customers and better understand the underlying factors leading to customer attrition.

**Machine Learning Role in Problem Solving:**

The core of our solution lies in using machine learning to analyze historical customer data and predict future churns. To do this:

1. **Algorithms:** We tested multiple algorithms for our model, including:
   1. K-Nearest Neighbors (KNN): A simple, instance-based learning algorithm.
   2. RandomForest: An ensemble method that builds a 'forest' of decision trees and aggregates their results.
   3. XGBoost: A gradient-boosting algorithm that can work on structured data.
   4. Logistic Regression: A statistical method for modeling binary outcomes.
2. **Libraries and Tools:**
   1. Python was our primary programming language due to its rich ecosystem for data science and machine learning.
   2. Scikit-learn for classic machine learning models and tools.
   3. XGBoost for the gradient boosting algorithm.
   4. Imblearn to handle class imbalance using SMOTE.
   5. Jupyter notebook model training and experimentation.
   6. Google colab to deploy.

*How?*

**Application's Implementation Plan:**

1. **Data Acquisition:** Our initial dataset was sourced from Kaggle, containing various customer attributes.
2. **Data Preprocessing:** We transformed fields like 'country' and 'gender' into binary values. Any missing values or outliers were addressed for model robustness.
3. **Model Development:** We trained multiple models for comparison using the preprocessed data.
4. **Model Evaluation:** We assessed each model's prediction capabilities using performance metrics like accuracy, recall, precision, and f1, then graphed the results.
5. **Deployment:** The best-performing model, RandomForest, was then integrated into our application hosted on Google Colab.
6. **Machine Learning Development:**
7. **Handling Class Imbalance:** Given the skewed distribution of our target variable (churn), we employed SMOTE to generate synthetic samples and create a balanced dataset.
8. **Model Training:** With the balanced dataset, we trained each algorithm. We also performed hyperparameter tuning to optimize their performance.
9. **Model Evaluation and Iteration:** Post-initial training, we used a confusion matrix to evaluate the RandomForest's results, given its superior performance. This iterative process allowed us to refine and retrain our model, improving accuracy.

Why?

1. **Algorithm Choice:** RandomForest emerged as the top-performing algorithm with an 89% accuracy. Its inherent ability to handle large datasets, cater to feature interactions, and reduce overfitting made it suitable for our application. Moreover, ensemble methods like RandomForest perform better than individual models due to their aggregated results.
2. **Training Process Justification:** Using SMOTE for class balancing was essential to avoid model bias towards the majority class. Training on an imbalanced dataset might result in a model that performs well in accuracy but fails to identify the minority class effectively.
3. **Evaluation Metrics:** The confusion matrix was vital to understanding true positive, false positive, true negative, and false negative rates. This holistic view allowed us to ascertain the model's ability to predict churning and non-churning customers accurately.

By aligning our machine learning methodology with the business goal of accurately predicting customer churn, we provided a solution that helps optimize targeted customer retention strategies.

## Validation

We turned to a series of metrics for this assessment, including precision, recall, f1-score, and overall accuracy. These metrics were derived using the sklearn. Metrics library and helped us determine the best model for our application.

Below are the results of the four models we trained and tuned:

**Logistic Regression**

Before tuning:

**A screenshot of a computer screen

Description automatically generated**

After tuning:

**A screenshot of a computer

Description automatically generated**

**K-Nearest Neighbors**

Before tuning:

**A screenshot of a computer

Description automatically generated**

After tuning:

**A screenshot of a computer

Description automatically generated**

**XGBoost**

Before tuning:

**A screenshot of a computer

Description automatically generated**

After tuning:

**A screenshot of a graph

Description automatically generated**

**RandomForest**

Before tuning:

**A screenshot of a graph

Description automatically generated**

After tuning:

**A screenshot of a computer screen

Description automatically generated**

## Visualizations

**Confusion Matrices**:

Location: on Google Colab

This visualization illustrates the true positives, true negatives, and false positives. The color-coded matrix was beneficial in quickly identifying model strengths and areas of improvement. We additionally included here.

A yellow and purple squares with numbers

Description automatically generated

**Heatmap**:

Location: on Google Colab

This visualization has lighter or darker shades (or different colors) and can quickly show you high or low-magnitude areas, making it easy to spot patterns or anomalies. The intensity and color of each cell in the heatmap correspond to the correlation coefficient between the variables. This helps in:

1. Identifying multicollinearity.
2. Recognizing features that are most correlated with the target variable.
3. Uncovering relationships among elements.

A screenshot of a graph

Description automatically generated

**Roc Curve:**

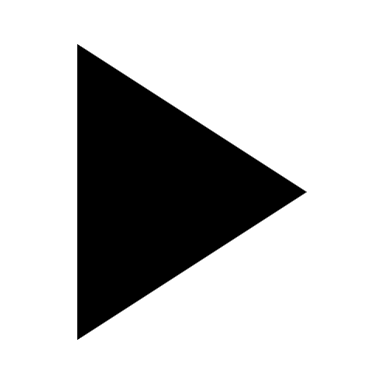
Location: Google Colab

The ROC curve plots the True Positive Rate (sensitivity) against the False Positive Rate (1 - specificity) at various threshold settings. The AUC gives a single value metric that describes the overall performance of a binary classifier; a model with an AUC close to 1 is considered good, whereas a model with an AUC close to 0.5 is no better than random guessing.  
  
A graph of a line

Description automatically generated with medium confidence

## User Guide

Guide to Using the Churn Prediction Application in Google Colab

1. **Access the Application:**
   1. Click on the provided link, redirecting you to Google Colab containing the project.
2. **Navigating Google Colab:**
   1. If you're new to Google Colab, it functions similarly to a Jupyter Notebook.
      * Each cell can be executed by clicking on the play button () that appears when you hover over the cell.
      * Or press Shift + Enter to execute the cell
      * Navigate through the notebook and run all the cells. Ensure that each cell completes its execution before moving to the next one.
        + It’ll show a green checkmark when that cell has run successfully
   2. Otherwise, you can click Runtime on the top left, then hit run all
      * This way, you won’t have to run each cell individually
3. **Using the Interactive Widget:**
   1. The last cell contains the interactive widget that will be used for making churn predictions.
   2. Input Details:
      * Credit Score: Use the slider to select the credit score, ranging from 300 to 850.
      * Age: Adjust the slider to specify the age, ranging from 18 to 100.
      * Tenure: Define the tenure using the slider, which ranges from 0 to 10.
      * Balance: Adjust the slider to specify the account balance, which can be between 0 and 250,000.
      * Products: Select the number of products the customer has, ranging from 1 to 4, using the slider.
      * Salary: Use the slider to set the estimated salary, ranging from 0 to 500,000.
      * Has Card: Use the toggle buttons to indicate whether the customer has a credit card.
      * Active: Use the toggle buttons to indicate if the customer is an active member.
      * Gender: Choose the gender of the customer using the radio buttons.
      * Country: Select the country from the options France, Germany, or Spain using the radio buttons.
      * After inputting all the details, click the Submit button.
4. **Viewing the Prediction:**
   1. Once you click "Submit," the application will process the inputs and predict whether the member will likely churn.
   2. The result will be displayed as "Member likely to churn" or "Member not likely to churn."
5. **Iterative Process:**
   1. You can adjust the sliders, toggle buttons, or radio buttons and click "Submit" again to make a new prediction with different input values.
6. **Closing the Application:**
   1. Once you're done, close the Google Colab tab or browser window.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Note:** No downloading or installation of software is required. The application is web-based and runs directly in your browser using Google Colab.

# Reference Page

*Customer retention – what is it and why does it matter for your business*. SWOT Digital Site. (2017, October 9). https://www.swotdigital.com/customer-retention-what-is-it-and-why-does-it-matter-for-your-business/