

# Truist 2023 Modeling Competition

**September 2023**

**Disclaimer:** The dataset used in the modeling competition is derived from a public repository. It does not contain any information related to Truist or Truist customers.

Please contact Aoran Wang ([aoran.wang@truist.com](mailto:aoran.wang@truist.com)) if there is any question

# Background and Problem Statement

## Competition Objective

In this machine learning competition, you will be challenged with a compelling binary classification problem: predicting whether a client will enroll in a Certificate of Deposit (CD) or not. Your task is to create an algorithm that can make this prediction with high accuracy and efficiency, which in turn will optimize our marketing efforts and streamline our approach towards potential clients. The target variable is binary, classified as '1' if the client has enrolled in a CD, and '0' otherwise.

## Background

A Certificate of Deposit is a critical financial product offered by Truist that many consumers substitute for a traditional savings account. CDs are risk-free assets with fixed maturities which pay a fixed interest rate over the life of the CD. Most CDs are short-term assets which mature within two years of purchase. The interest rate on CD's is higher than on savings accounts, making them attractive to highly risk-averse investors who want to earn a little extra money on their savings. Truist uses the proceeds of CDs to fund day-to-day operations and loans.

Accordingly, Truist spends considerable resources to promote and sell CD's to existing and potential clients. In fact, Truist has a call center which is dedicated to reaching out to potential clients about CDs. It is an effective, but expensive advertising method. To enhance the efficiency of our team's efforts and increase CD enrollment rates while minimizing call center costs, we need a precise understanding of the factors that influence a client's decision to enroll in a CD.

## Why We Need Your Expertise

Truist is currently navigating a period of large demographic shifts. It is projected that by 2030 millennials will inherit over \$68 trillion from their parents. The collective wealth of millennials will increase by a factor of five and the generational wealth gap between baby boomers and millennials will shrink. As a result, our existing and potential clients' financial behaviors and saving habits are rapidly changing. By accurately predicting who is likely to purchase CD's, we can tailor our services and marketing strategies more effectively to the preferences of different generations.

Your mission is to develop a machine learning model that can help us predict CD enrollments. Your insights will play a pivotal role in our quest to understand generational financial trends better and optimize our marketing efforts.

# Expectations and Detailed Tasks

## EDA

- Explore the dataset and get meaningful insights that can help with the business.

## Classification Model

- Create a classification algorithm to predict whether a client will enroll in a CD or not based on several features.
- Clean the data and avoid data leakage while building your classification model.
- We encourage you to research external sources to understand and consider other factors that may influence the prediction.
- You can create new features from the existing features or drop features which do not have much impact but remember to validate your actions.

## Evaluation Metric

- All submissions will be evaluated on Kaggle using the AUC-ROC (area under the curve) approach.
- For more details on ROC AUC scoring see our Kaggle competition portal.

# Registration and Submission

## Registration instructions

To register for the competition, each team is required to submit their team information through [https://truistex.sic1.qualtrics.com/jfe/form/SV\\_3f51Yw36M6UZOUC](https://truistex.sic1.qualtrics.com/jfe/form/SV_3f51Yw36M6UZOUC) (only one registration is needed per team). Please note that each team member is required to sign an NDA to participate, and each team must upload separate NDAs for all team members through the link. Additionally, each team must have one Kaggle account and register on our dedicated Kaggle competition portal using their team's name. Please be advised that teams are only permitted to submit their prediction twice per day and that duplicate registrations are not allowed. The deadline for registration is **September 24<sup>th</sup>, 2023**, while the deadline for prediction submission is **September 29<sup>th</sup>, 2023**.

## Teams

There is no limitation on the number of teams that can participate in each institute. However, each team should only have **a minimum of 2 and a maximum of 4** members.

## Kaggle

- The modeling competition is hosted on Kaggle platform, and all participants in will need a specific access link to participate.
  - Morehouse College: <https://www.kaggle.com/t/d0a2c1e5dd4d4171a02523132c005d65>
  - Clark Atlanta University: <https://www.kaggle.com/t/855f4c5532114f97bbab54ddd72e8fc9>
  - Spelman College: <https://www.kaggle.com/t/4bfd53c715b04336bc7ebae17fc9dbad>
  - GSU: <https://www.kaggle.com/t/971117e505a5446b88964b26249dcd94>
  - EMORY: <https://www.kaggle.com/t/a302341c60ed991e5856220d6af60f9f>
  - UNCC: <https://www.kaggle.com/t/32952aacd70a4478ac9464cbe5cdcacb>
- Only one member from each team will need to have a Kaggle account registered in the competition; please use the same team's name and include your teammates' Kaggle account under the team if they have one. Do not register more than once for a team. If you don't have a Kaggle account, create an account using the below link, free of cost. <[Kaggle: Your Home for Data Science](#)>
- Only two prediction submissions are allowed a day for each team, until the deadline.

## Scoring

Your submissions will be scored based on ROC AUC. Teams will see their scores and rankings on the leaderboard section once they submit their predictions. Scores and rankings might change after a team upload a new prediction.

## Prediction Evaluation and Presentation

- Once the submission deadline passes (**September 29<sup>th</sup>, 2023**), the **top 2** ranked teams in each participating institute based on their model performance (teams can find their rankings on Kaggle leaderboard) will be reached out and given an opportunity to construct their findings and insights and present to a broader Truist analytical community.
- The presentation should be geared towards a business audience. Some technical details of your submission should be included in your presentation, but we also want to evaluate your ability to extract and explain useful insights to a non-technical audience.
- The judges for the presentation will consist of a mix of data scientists and managers.
- The presentation day will be **October 6<sup>th</sup>, 2023** (Exact time TBA)
- The winners (Top three teams) will be announced on **October 10<sup>th</sup>, 2023**.

# Data Dictionary

Field Name	Data Type	Description
age	int	client age
generation	string	generation
job	string	client job type
marital	string	client's marital status
education	string	client's education
default	string	client has default in credit
balance	int	average of yearly balance
housing	string	client has housing loan
loan	string	client has personal loan
contact	string	communication type
day	int	most recent contact day of the month
month	string	most recent contact month of the year
duration	int	duration of the most recent contact (in seconds)
campaign	int	number of contacts performed during this campaign for this client
pdays	int	number of days since client was last contacted from previous campaign (-1: client not previously contacted)
previous	int	number of contacts performed before this campaign for this client
poutcome	string	outcome of the previous marketing campaign
state	string	state
zip code	string	zip code
cd	string	client has enrolled in a Certificate of Deposit

# Schedule / Timeline

Date	Event
9/18/2023	Email Kick-off
9/24/2023 by EOD	Team Registration Deadline
9/29/2023 by EOD	Kaggle Prediction Submission Deadline
10/06/2023	Finalists Presentation Day (Exact Time TBA)
10/10/2023	Winners Anounced