

The Machine Learning Challenge

20th April 2021 - 08th June 2021

Background and Motivation

Machine Learning challenges are fun. They are full of learning & intense competition, and a lot is packed in a matter of a few days. You get to apply your learning on real-life datasets and also compare yourself against the other students from around the country to see where you stand. The thrill of finding a solution in a competitive environment is addictive. This machine learning challenge, in partnership with our machine learning training partner Analytics Vidhya, is our attempt to inspire the next generation of data scientists to fall in love with machine learning, to learn how to work with real-life data sets, and win exciting prizes.

Project Problem Statement

Your client is a Financial Distribution company. Over the last 10 years, they have created an offline distribution channel across the country. They sell Financial products to consumers by hiring agents in their network. These agents are freelancers and get a commission when they make a product sale.

Overview of your client On-boarding process

Managers at your client are primarily responsible for recruiting agents. Once a manager has identified a potential applicant, they would explain the business opportunity to the agent. Once the agent provides the consent, an application is made to your client to become an agent. This date is known as `application_receipt_date`. In the next 3 months, this potential agent has to undergo a 7 days training at your client's branch (about Sales processes and various products) and clear a subsequent examination in order to become an agent.

The problem - Who are the best agents?

As is obvious in the above process, there is a significant investment which your client makes in identifying, training and recruiting these agents. However, there are a set of agents who do not bring in the expected resultant business after recruitment. Your client is looking for help from data scientists like you to help them provide insights upfront using their past recruitment data. They want to predict whether an agent will source new business in the next 3 months which would help them identify the right agents to hire and thus make the whole process more agile. For the complete problem statement, datasets, data dictionary etc., please access the contest environment as described below.

The Contest Environment:

We have set up a complete contest environment for you on Analytics Vidhya, exclusively for this challenge. To access this contest environment, please follow below steps -

1. Go to <https://datahack.analyticsvidhya.com/contest/internshala-machine-learning-challenge-11/>
2. You will be required to sign up/create an account on Analytics Vidhya (if you don't have one already) **using the same email address with which you registered for this challenge on Internshala.**
If you register with a different email, we will not be able to identify you in the contest environment and evaluate your submission

3. After that, click on 'Register' button on the contest page to complete registration for Internshala Machine Learning Challenge and accept the T&Cs of the challenge
4. Once you are inside the contest environment, you can access the problem statement, datasets, data dictionary and other details relevant to the challenge.
5. Please submit your predictions using the solution checker provided on the challenge page.
6. You can use solution checker to make multiple submissions and see your position on Public Leaderboard before you make final submission
7. Final submission must be selected from all submissions and the code file required to create the submission must be uploaded along with the final predictions.

How will the submissions be evaluated:

- Both train and test combined has data from Apr'2007 to Jan'2009 (For Jan'09 only 99 records for 01-Jan-09)
- Train Data is for the period Apr'2007 to 01-Jul-2008
- Test data (02-Jul-2008 to 01-Jan-2009) is further divided based on time into public and private leaderboard
- The public leaderboard is based on the first 2 months of the test dataset (02-Jul-2008 and 01-Sep-2008) and rest of the test data is used for Private leaderboard
- Leaderboard and scoring will be based on AUC-ROC Score
- Your initial responses will be checked and scored on the Public data.
- The final rankings and winners would be decided based on your private score which will be published once the competition is over and your code review
- The top few contestants may also be interviewed (telephonic or in-person) and the final list of winners will be declared post code review and interviews on result_date.

Rules of Machine Learning Challenge:

1. Your Machine Learning training would begin on (20th April 2021).
2. The contest project problem statement and instructions will be available to you from the progress tracker.
3. The last date to make your final submission (as per the instructions given above) is (08th June 2021). Any late submissions will not be considered.
4. Only Individual participation is allowed. Use of external data is not allowed. Use of 'ID' variable as a predictor is not allowed.
5. You are expected to solve the project using only those algorithms and concepts that have been taught in the training. Use of any other algorithm(s) to make predictions is not allowed.
6. Throughout the challenge, you are expected to respect fellow participants and act with high integrity.
7. The winners will be decided on the basis of private leaderboard and code review post the contest closes on (08th June 2021).
8. The top few contestants may also be interviewed (telephonic or in-person) and the final list of winners will be

declared post code review, and interviews on (15th September 2021).

9. Internshala & Analytics Vidhya hold the right to disqualify any participant at any stage of the competition if the participant(s) are deemed to be acting fraudulently.
10. The decision on the winners and runners-up made by Internshala & Analytics Vidhya will be final and binding.

Important Dates:

- Contest (and training) start date - (20th April 2021)
- Contest end date - (08th June 2021). This is the last date to make your final submission.
- Result declaration - (15th September 2021)

Extra Support for the contestants:

In case of any technical issue or issues related to the understanding of the problem statement, please post your query using the 'forum' provided inside your Internshala Trainings environment.

Here are some additional resources you could refer:

- [Data Science using Python](#)
- [Decision Trees & Tree based Modeling](#)
- [Essentials of Machine Learning Algorithms](#)

All the best!