EPSILON 2.0

Nucleus

The Analytics Society

Shaheed Sukhdev College of Business Studies

University of Delhi

ABOUT

Welcome to this year's edition of Epsilon – The online data science and analytics case study competition, organized by Nucleus, The Analytics Society of SSCBS.

Last year, participants were tasked with predicting costs of potential policyholders from the perspective of an insurance company wishing to diversify into a new geography.

The sector we would be focusing on for this iteration of Epsilon is the **Banking and Financial Services** industry.

THEME

Banks have diverse sources of revenue. These sources can be classified into non-interest sources and interest sources.

Non-interest income is bank and creditor income derived primarily from fees including deposit and transaction fees, insufficient funds (NSF) fees, annual fees, monthly account service charges, inactivity fees, check and deposit slip fees, and so on.

Traditionally, the most significant method is via charging interest on the capital it lends out to customers. The bank profits from the difference between the level of interest it pays for deposits and other sources of funds, and the level of interest it charges in its lending activities. Historically, profitability from lending activities has been cyclical and dependent on the needs and strengths of loan customers and the stage of the economic cycle.

Of late, one of the major concerns that banks are facing is the increasing amount of non-performing assets or NPAs, often called Bad Debts. When borrowers default on their interest payments to the bank and eventually progress to a situation wherein they can no longer repay the bank, we term the net loss incurred from the borrowing activity as

NPAs. With this as the key objective, banks wish to strengthen their decision-making algorithms to have a better understanding as to the risk they are exposed to with each loan given.

TASK

You work as the **Chief Analytics Officer** in the *retail banking division* at **Royal Stratton Bank**, one of the largest banking institutions in the world.

Recently, many new customers have approached the bank for the purpose of acquiring loans. The bank is interested in finding out whether these customers would be able to repay the amount of loan if provided to them.

Pertaining to this objective, a sample dataset (train set) has been assembled which relates to the previous loan takers of the bank and whether these customers were able to repay the amount of loan taken by them based upon certain factors.

The task at hand for you is to estimate the likelihood of the bank loan default on account of a borrower based on multiple factors like Loan Term, Credit score, Gender, Type of Loan, Employment Status and other such characteristic metrics.

This analysis will then be used further by the bank to form its policies and take decisions regarding any loan queries.

In conclusion, a detailed analysis of the sample population followed by a well-defined predictive model to evaluate the bank loan default risk for each borrower needs to be prepared.

DELIVERABLES

1) You need to analyze the Datasets (train & test set) from a Statistical and Business perspective
2) You need to design a predictive model to predict whether there would be a bank loan default on account of all the new borrowers (in the test dataset)
3) Lastly you need to develop a document entailing the explanation of your analytical strategy. Feel free to include any other qualitative recommendations that you might have pertaining to the data. (more info below)

SUBMISSION GUIDELINES

Each participant needs to submit the following documents:

➜The Excel file labelled as "Sample Submission" after renaming it to "Participant1_Participant2_CollegeName1_CollegeName2" which will include your prediction of whether the bank loan will default or not.

➜All relevant files of Excel/R/Python or any other software used for the analysis.

➜A 3-6 page PDF document entailing the explanation of the analytical strategy undertaken by the team – Kindly explain in detail why and how certain algorithms and/or methodologies were applied.

➜Kindly mention name, college, phone no. & email ID of all participants in the PDF file submitted.

The above files must be submitted through dare2compete before 24th February 2020 11:59 PM.

DISCLAIMER

❖The rules and decisions of the judging panel and organizing team in all matters related to the case study shall be final and binding.

❖ The participants will be judged on the basis of the level of accuracies they have in their model and the depth of their analysis of the case study.

❖The judging panel and the organizing team reserve the right, in their sole discretion, to disqualify any individual they find tampering with the process (in any form) of the competition and/or violating the rules and guidelines.

❖Details and specifications are subject to change, and further changes, if any, would be notified via the Facebook page of Nucleus and/or registered email id.

For any queries contact:

Chirag Gupta - +91 9811550704

Anvit Gupta - +91 9811810610