Credit Analytics Case Study

Company overview

Stride Funding is a venture-backed, mission-driven start-up that is transforming access to alternative and higher education by offering outcomes-driven funding products (e.g. income share agreements, deferred tuition agreements, payment plans, etc.). Stride believes in the power of education to fuel economic mobility, and also recognizes the limitations of society's current approach to funding education. At Stride, we seek to increase educational access by leveraging alternative data sets, innovative modeling techniques, and school-supported funding to offer outcomes-driven funding products on the basis of students' future career trajectories, rather than their historical family credit profile. You can play a role in creating meaningful change for our customers, community, and company every day by doing work you truly love, while surrounded by a team that shares your passion! We look forward to hearing from you.

Background / context

Our approach to tackling inequities in education funding and access involves a few key innovations that differ from the traditional cosigner-based student lending system:

* We offer funding options for students pursuing both degree and non-degree programs as we recognize that strong career outcomes can result of programs of all types
* We leverage education and other alternative data to predict, on a cohort-basis, future earnings for program graduates, and thus their ability-to-pay
* We work with credit bureaus to acquire look-a-like data that helps us understand and identify the more specific credit risk of our individual student borrowers, who are typically younger and have thinner files
* We offer innovative funding and repayment structures where our borrowers only owe Stride money when they are earning a living wage

Case study prompt

One of Stride’s primary strategic goals for the remainder of 2022 is developing a unique and innovative credit risk score and associated predictive models to enable us to serve more students without taking on additional risk.

**Overview**

Using the set of provided features, the Stride Funding Credit team would like you to predict which borrowers will default on a financial product (e.g. ISA or Loan).

T**he Challenge**

An applicant that will default is defined by a target variable equaling “1”. You are challenged to construct new variables, employ feature-selection methods, and compare different prediction algorithms to approach this dataset.

More specifically, use python, or other method of choice to develop a model on the “target variable” in the train dataset and score on the associated test set.

**Data**

You will be provided with two similar datasets that include de-identified customer information. One dataset is titled *train.csv* and the other is titled *test.csv*.

* train.csv will contain the details of 16,671 customers and importantly, will reveal whether the customer has defaulted on the financial product (i.e., target).
* test.csv will contain similar information but does not disclose the default/not default. It is your task to predict this outcome.

**Submission**

You will provide us with two files.

1. A CSV file with the test data attached with your prediction named submission.csv. (**a sample submission is provided if a predictive model was applied**). If you are not able to model the data, please provide an analysis of the information that can inform the business to improve credit underwriting strategy.
2. A report with the findings and detail of the model methodology applied (**if a model was applied**). (You could use [this style guide](https://www.dataquest.io/blog/data-science-project-style-guide/) for guidance)

Note: Data Dictionary will not be provided, please use your best judgment as to what a certain variable could mean as it relates to the ‘target variable’.

Related to model methodology information, please provide the following additional commentary:

1. What other data sources would you want to use in addition to what was provided to make your decision on approval and offer terms.
2. What other important variables in the model to determine aforementioned offer terms and what predictive target(s) would you want to look at to make that decision.
3. How would you go about developing a business strategy to portray to business the benefits of the new model and strategy?

Deadline

Once you have reviewed these materials, we ask that you share some proposed times for a 60-minute interview slot over the next two (2) weeks. We ask that you please send your case study materials over before the presentation for us to review in advance. You will then walk through your materials for ~45 of the 60 minutes of your final interview, with intermittent Q&A from the panel.

Please feel free to email us with any questions you have!