

Total Points: 100

1

other methods that you know. (2) Communicate and document your approach in a **two-page**, clear, and polished report. The report is not just a formality; the process of articulating what you did and why deepens your own understanding of the methods and sharpens your analytical thinking.

You are encouraged to use AI agents (e.g., Claude, Gemini, GPT, Cursor, Copilot) throughout this assignment. We have posted a [video on Canvas](#) with guidance on working with AI agents for this challenge, and the Appendix includes additional tips on how to get the most out of your AI collaboration.

2 Details and Logistics

The prediction task studied in the Prediction Challenge is on the topic of predicting fraud in phone financing. The Prediction Challenge is hosted on <http://www.kaggle.com> website in the form of a competition among *OIT* 367 students. The competition is not open to the public, so you should check Canvas Prediction Challenge Module for “How to Join the Competition on Kaggle.” You are responsible for carefully reading the competition details: **description, evaluation, rules, and submission instructions** on the Kaggle website and in this document.

You will be able to start working on the Prediction Challenge on **Thursday, February 12** and it is due by **Monday, March 2, at 9:00pm PST**². The Prediction Challenge is to be completed in groups of **two to three students**. You may choose the same groups that you are using for the problem sets. Here are additional important rules:

- You must log in to Kaggle and officially become a member of your team by **Friday, February 13, at 9:00pm PST**. To be added to a team, one member can be the team leader and add the others (see Appendix, Figure 7). All members of the team should accept the competition rules before the team leader can add them to the team. You can accept the competition rules by trying to download a competition data file where Kaggle requires you to accept the rules. Please contact us if you need help.
- After **Friday, February 13, at 9:00pm PST** each team can only make at most 5 submissions per day.
- You may use an arbitrary display name on Kaggle website. However, by the same deadline (**Friday, February 13, at 9:00pm PST**), your team must send us the display name of the team, the display name of your team members as well as your real names in an email to gsb-oit367-staff@lists.stanford.edu.

²Your report is due by **Tuesday, March 3, at 9:00pm PST**, see Reporting score section for details.

- You are allowed to use any software that you prefer. However, you need to include the scripts with your report. Also, if you use a data analysis package, you need to understand what it is doing and demonstrate that in your report.
- You are allowed to use any outside data. However, you need to include the data with your report and explain the reason for its addition.
- You should not rely on your public leaderboard score. For example, if the choices you make depend on the results of your prior submissions, you may overfit, and your public test set score would be higher than your private test set score. See the “Insights from Prior Competitions” section below for historical evidence on this phenomenon.
- **STANFORD UNIVERSITY HONOR CODE:** By accepting the rules of this competition: You acknowledge and accept the Stanford University Honor Code. You may find copies of the honor code and its interpretation in the MBA Handbook.
- **PRIVACY AND CONFIDENTIALITY AGREEMENT:** Under no circumstances should any participant try to identify the individuals in the data set for any reason. You are not allowed to share the competition dataset with anybody except the members of your OIT 367 team.

Grading Criteria *Each group should make at least one submission on Kaggle and hand in exactly one Prediction Challenge report. Note that you should also select up to two of your submissions on Kaggle to calculate your final score (see the “Prediction Score” section below for more details).*

The total grade in the Prediction Challenge is 100 points which consists of a *Prediction Score* and a *Reporting Score* as described below. All Prediction Challenge team members will receive the same grade on the Challenge unless you specify a peer evaluation score. In particular, similar to the other group assignments in this course, you can provide us (on your report) with a peer evaluation grade of your team member, a grade (0 to 10) for each team member. If a team submits an assignment without this peer evaluation grade, we will assume that all team members contributed equally well and will receive a peer evaluation score of 10. Please refer to the Course Policies page on the course website for details on how peer evaluation is applied to your grade on the Prediction Challenge.

3 Prediction Score: Your final AUC (50 points)

The overall structure of the competition is summarized in Figure 1.

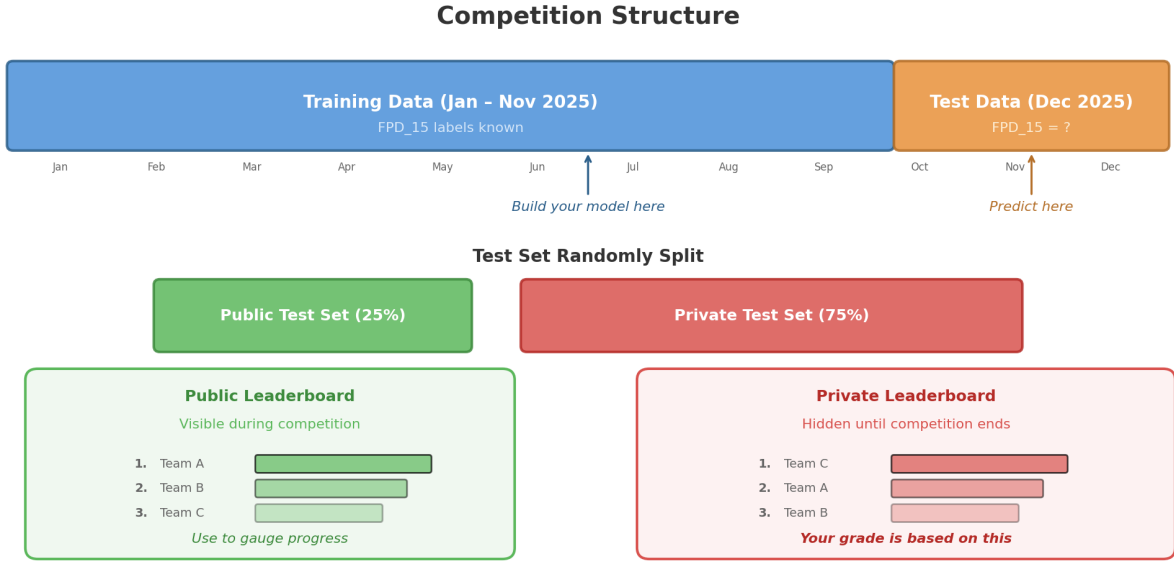


Figure 1: Competition structure: training data with known labels (Jan–Nov 2025), test data with hidden labels (Dec 2025), and how the test set is randomly split into public and private leaderboards.

The leaderboard score on Kaggle will be evaluated based solely on the AUC generated on approximately 25% of the test set, which is called the *public test set*. However, your final submission will be evaluated using the remaining 75% of the test set (called the *private test set*). This separation is performed randomly and will not be disclosed to you to prevent leakage of information from the results of daily submissions. But the same separation is used for all teams. Therefore, your submissions for all rows of the `Test_OrderIDs.csv` file (provided to you) include your predictions for both public and private test sets. Your score on both sets will be very close due to large sample sizes and random partitioning.

According to the competition rules, you can select up to 2 submissions that will be used to calculate your final leaderboard AUC. If you do not select them, up to 2 entries will be chosen for you based on Kaggle’s internal guidelines. Your **final AUC** will be based on the private test data. Your public AUC is only a rough indication of what your final AUC might be. You should choose entries that will most likely be best overall and not necessarily just on the public subset. A snapshot of the portion of the competition website for selecting the submissions is shown in Figure 2. We allocate 50 points to your final AUC according to the curve in Figure 3. The formula used to generate this score is

$$\text{Prediction Score} = \min \left[50, \max \left(0, 50 \times \frac{e^{10 \times (\text{Final AUC} - 0.50287)} - 1}{e^{10 \times (0.61 - 0.50287)} - 1} \right) \right].$$

Overview
Data
Discussion
Leaderboard
Rules
Team
Submissions
Settings

Submissions

Select up to 2 submissions that will count towards your final leaderboard score. If less than 2 are selected, Kaggle will automatically select from your best scoring submissions. [Learn More](#)

Auto-selection candidates

All
Successful
Selected
Errors

Recent

Submission and Description	Public Score	Select
<div> <div>✓</div> <div> benchmark_submission.csv Complete · 6d ago · Benchmark Submission </div> </div>	0.50287	<input type="checkbox"/>

Figure 2: Select up to two submissions on Kaggle website for the calculation of the final score.

Here is some information on this formula:

1. The prediction score generated by the formula above will be rounded to an integer.
2. The final AUC equal to or above 0.61 will receive the full 50 points.
3. The final AUC equal to or below 0.50287 will receive 0 points.
4. The reason for the use of an exponential function rather than a linear function is to put more value on higher values of AUC. Improving the AUC from 0.60 to 0.61 is harder than improving it from 0.52 to 0.53.

Important note: In your report (see below for more details), you should summarize the data-analytic techniques you’ve applied to the dataset so that we can repeat the same steps and reproduce your results. You are required to accompany your codes and scripts that generated the submission file. If we cannot reproduce your result, your score in the Prediction Challenge will be severely affected.

4 Reporting Score: Prediction Challenge report (50 points)

You must submit a clear and polished report describing your data analysis methods by **Tuesday, March 3, at 9:00pm PST**. The report should be at most two pages in length, single-spaced, excluding tables, figures, and code. Additionally, include a one-page cover sheet listing the names of your team members. The main body of the report must contain a detailed description of your methodology and justification for your analytical choices.

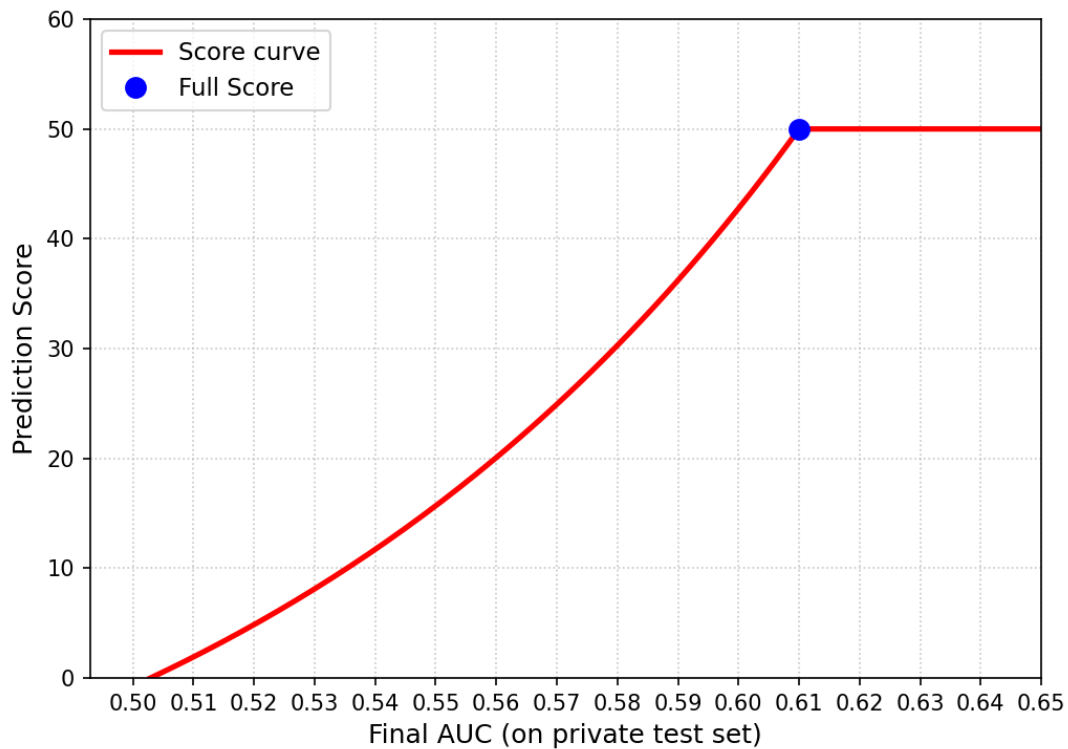


Figure 3: Prediction score as a function of final AUC.

You are also required to submit a technical appendix containing the code and scripts used, as well as any supplementary tables and graphs. The appendix does not count towards the two-page limit. Your report grade, however, will primarily depend on the content of the two-page report. The overall goal is for us to be able to reproduce your leaderboard results from the provided information.

Reports should be submitted on the course website. In grading your reports, we most importantly consider the quality and correctness of your analysis. Important items to keep in mind are:

- Did you do any data cleanup? If you did, how was this performed, and if you did not, what was the reason?
- Did you perform any preprocessing steps such as handling the missing values, binning, or generating new predictors; if it was not done, what was the reason?
- Did you randomize and partition the data correctly (for choosing your model)?
- Did you justify the choices you made? Did you perform a training/validation (or cross-validation) analysis for your variable selection, model selection, or parameter

tuning?

- If you used an AI agent, did you demonstrate that *you* drove the analytical decisions? Your report should reflect your reasoning and judgment, not just the agent's output. What did you ask the agent to do, and why? Did you catch any mistakes or push back on its suggestions?
- Quality of the exposition, writing, and presentation.

5 The formula for calculating the Prediction Challenge score

Total Prediction Challenge score = Prediction Score + Reporting Score.

6 Insights from Prior Competitions

We have analyzed the public vs. private leaderboard data from 7 prior OIT 367 Kaggle competitions (2019–2025), covering 168 teams across 5 AUC-based competitions. The results provide a clear and consistent message: **your private leaderboard score will almost certainly be lower than your public leaderboard score**. This is not because you are doing something wrong; it is the natural consequence of the public leaderboard being computed on only 25% of the test data.

Key findings from the historical data:

- **91% of teams** scored worse on the private leaderboard than on public.
- The **median score drop is 0.66%** of the public score (e.g., a public AUC of 0.65 would typically become ~ 0.646 on private).
- The pattern is remarkably consistent across very different prediction tasks and years.

The Leaderboard Shuffle. Figure 4 shows how team rankings changed from the public to private leaderboard across all 7 competitions. Each line connects a team's public rank (left) to their private rank (right). Green lines indicate teams that moved up; red lines indicate teams that moved down. The key takeaway: **rankings are not stable**. A team that is #1 on the public leaderboard is not guaranteed to finish #1 on the private leaderboard.

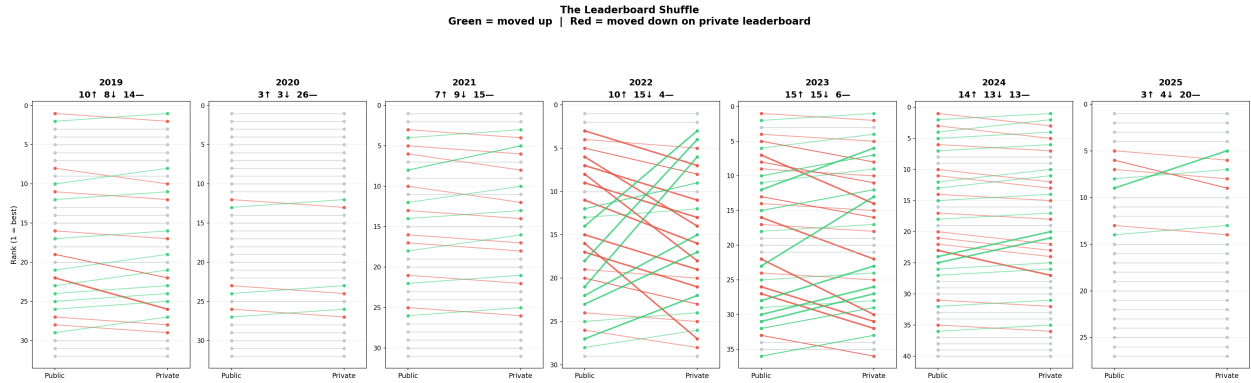


Figure 4: The Leaderboard Shuffle: how team rankings change from public to private leaderboard across 7 competitions (2019–2025). Green = moved up, Red = moved down.

Public vs. Private AUC. Figure 5 plots each team's public AUC against their private AUC for the 5 AUC-based competitions. Points below the diagonal represent teams whose private score was lower than their public score. In every single year, the vast majority of teams fall below the diagonal.

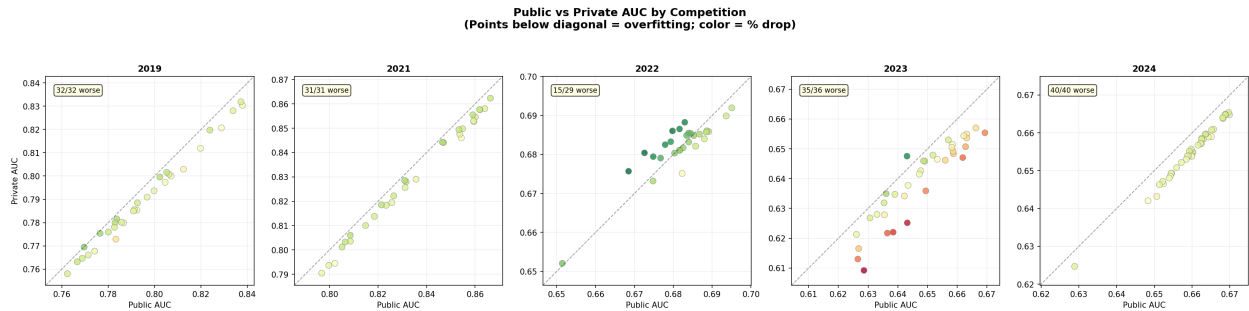


Figure 5: Public vs. Private AUC by year. Points below the diagonal indicate teams whose score dropped from public to private. Color intensity reflects the magnitude of the drop.

Distribution of Score Drops. Figure 6 shows the distribution of the score drop (public AUC minus private AUC) for each year. Most distributions are centered to the right of zero, confirming the systematic tendency for private scores to be lower.

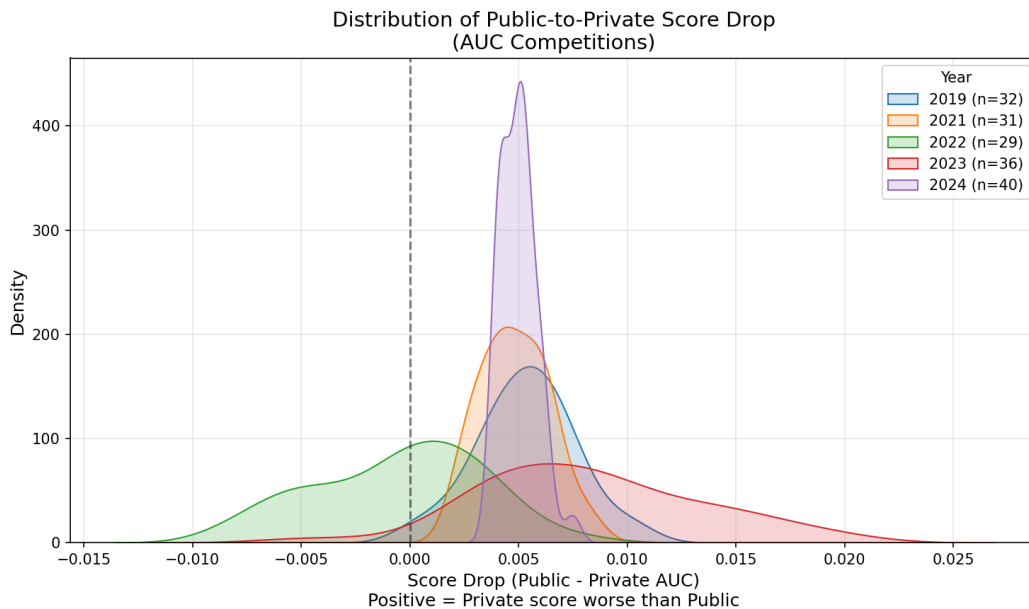


Figure 6: Distribution of score drops (public minus private AUC) across 5 AUC-based competitions. Positive values indicate the private score was lower than the public score.

The Lesson. Use your public leaderboard score to gauge progress, but **do not make modeling decisions based solely on small improvements to the public score**. Instead, focus on building models that generalize well. Use proper cross-validation, avoid overfitting, and select submissions that reflect robust, generalizable performance rather than the highest public score. Your grade will be determined by the **private** leaderboard.

7 Competition Winners

This Prediction Challenge constitutes 20% of your final grade in *OIT 367*. In addition to this 20%, each member of the top 3 teams in the competition will receive a Coupa Cafe gift card! Also, some of the top teams can provide a brief presentation of their approach in class which will help their class participation grade.

A Appendix

B Tips for Working with AI Agents

AI agents are powerful tools that can accelerate your work, but the teams that perform best in this challenge are the ones who **drive the process themselves**. The agent handles implementation. You handle the thinking.

You are in charge. The creative thinking about the problem (what features might matter, what patterns to look for in the data, what modeling approach to take) must come from *you*, not the agent. The agent can build what you design, but it cannot understand the business problem the way you can. Working with an agent is not about asking it to “build the best model.” It is about directing it with specific, well-reasoned requests, evaluating whether the output makes sense, and iterating based on your own understanding of the data and the task.

Bring your own reasoning. The teams that stand out are those who bring creative domain reasoning: they think carefully about what drives first payment default, they inspect the data for patterns the agent would not think to look for, and they make deliberate choices rather than accepting whatever the agent produces.

Check the agent’s work. Treat the agent as a very capable collaborator, not as an answer machine. Question its choices. Push back when something does not look right. Agents can make subtle mistakes (data leakage, wrong evaluation metrics, overfitting to the validation set) that only a human who understands the concepts would catch.

Use it as a learning tool. When the agent produces code or a result you do not fully understand, ask it to explain. This is an opportunity to deepen your understanding of the techniques, not just to get an answer. In your report, you should be able to articulate *why* you made each analytical choice, regardless of who (or what) wrote the code.

The combination of strong AI execution with thoughtful human judgment is what wins.

Your Team

Everyone that competes in a Competition does so as a team - even if you're competing by yourself. [Learn more.](#)

General


TEAM NAME

Stanford OIT367 Test

This name will appear on your team's leaderboard position.

Let others know you're looking for teammates

[See tips about forming teams](#)

☐ Display your status with  on the leaderboard

Team Members

Your team can have a maximum of 3 members.




Stanford OIT367 Test (You)

Team Leader

Send Invitation

You can invite 2 more people to join your team. They will need to have joined the competition.

 Search for the team or user name

Received Invitations

You haven't received invitations from other teams.

Save Changes

Figure 7: Setting up your team on Kaggle website.