About this template document:

1. This document contains a template for your team to fill in for your required AI Studio “Monthly Progress Summary” for October *(due November 1st, 2023 in your AI Studio course in Canvas; and detailing work completed by your team in October 2023 during the “Modeling & Evaluation” phase of your project).*
2. This document should be stored in your team’s Project Folder in Google Drive, for easy access by team members; your AI Studio TA/Tutor/Course Support; and the Break Through Tech AI Program Team. Submission details are available in Canvas.
3. You can review an example Monthly Progress Summary in the Bridge to Studio module of your AI Studio course in Canvas, on the page titled, “Team Breakout: Team Resources”. However, please note that this example corresponds to Data Understanding/Prep, not Modeling & Evaluation.

[View the template below](#_qhyqphsckgi)

# Monthly Progress Summary *(October)*

**I. Modeling Summary**

Question:   
Please provide a summary of your team's Modeling-related accomplishments during the month of October (focusing on the types of tasks detailed in this month’s Progress Evaluation Rubric, in the categories for Algorithm Selection and Model Training and Testing). Remember to include any relevant links to your work (e.g., a Python/Colab notebook showcasing your team’s work).

Student Team Response:

In October, our team focused on the selection and training of the most suitable machine learning algorithm for our dataset, which contains a mix of non-binary categorical variables and numerical features. Our key accomplishments include:

1. Algorithm Evaluation: We assessed the performance of various algorithms, including Random Forest, Gradient Boosting, Support Vector Machine (SVM), and Logistic Regression, to determine the best fit for our dataset.
2. Data Preparation: We employed one-hot encoding to transform categorical variables, making them compatible with the selected algorithms.
3. Model Training and Testing: We split the dataset into training and testing subsets, trained models using the chosen algorithms, and evaluated their performance using metrics like accuracy, precision, recall, and F1-score.
4. Outcome: Our results allowed us to make an informed decision on the most effective algorithm for our dataset. We documented our work in a Python/Colab notebook, showcasing our methodology and findings.

You can review our detailed work by following this link: <https://drive.google.com/drive/folders/1vsKxIcxid-lQHUxXsLrc8HFFwhllMGUF?usp=drive_link>

**II. Evaluation Summary**

Question:   
Please provide a summary of your team's Evaluation-related accomplishments during the month of October (focusing on the types of tasks detailed in this month’s Progress Evaluation Rubric, in the categories for Model Evaluation and Model Improvement). Remember to include any relevant links to your work (e.g., a Python/Colab notebook showcasing your team’s work).

Student Team Response:

During October, our team was actively engaged in model evaluation and improvement tasks. We focused on assessing the performance of different machine learning models and enhancing their effectiveness. Here's a summary of our accomplishments:

Model Evaluation:

1. Random Forest:
   * Accuracy: 0.7
   * Precision: 0.698
   * Recall: 0.7

Link: <https://colab.research.google.com/drive/1OrFl0fZnzXATZgsBrOhdMwLaD7KP2fiJ?usp=sharing>

1. Gradient Boosting:
   * Best Decision Tree Depth: 1
   * Best Decision Tree Accuracy: 0.555
   * Gradient Boosting Accuracy: 0.528

Link: <https://colab.research.google.com/drive/1rEmxm9OTi0Tc_u9RhhFpXpAjaWhC2AIm?usp=sharing>

1. Logistic Regression:
   * Accuracy: 0.543

Link: <https://colab.research.google.com/drive/15p8gdeAnV31taPh2xv14e4MCRQeRPrsO?usp=sharing>

1. SVM (Support Vector Machine):
   * Unfortunately, SVM experienced a prolonged runtime and did not produce a final result.

Link: <https://colab.research.google.com/drive/1GvOx99HnH5QInqIWXoriUcQUwO12PRo9?usp=sharing>

Model Improvement:

We recognized the need for further optimization, fine-tuning, or potentially exploring alternative algorithms. SVM must be modified in order to run results. The provided links will take you to detailed evaluations, including our methodologies and findings. Our ongoing efforts are focused on refining these models and achieving better performance.

**III. Lessons Learned and Challenges**

Question:   
Reflecting on the Modeling and Evaluation phases, what were the key insights or challenges your team encountered? How did you address them? Share any important lessons learned that can help guide future steps in the project.  
  
Student Team Response:  
Key Insights:

1. Data Compatibility: We realized the importance of selecting models that are well-suited to our data types. It became apparent that some models required data type adjustments or preprocessing to achieve optimal results.
2. Model Diversity: Acknowledging that not all models would offer the same levels of accuracy and precision, we conducted experiments with multiple algorithms to gauge their predictive potential. This allowed us to make informed choices.
3. Data Preprocessing: For models like Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR), we employed one-hot encoding to handle categorical features, ensuring seamless integration with the algorithms.

Challenge Faced:

1. Algorithm Expectations: Initially, we had the expectation that Support Vector Machine (SVM) and Gradient Boosting (GB) algorithms would yield the best and most accurate results. However, our findings revealed that Random Forest (RF) outperformed them. Additionally, we encountered a runtime issue with SVM that prevented us from obtaining results.

Lessons Learned:

1. Flexibility in Problem Solving: Our experience taught us the importance of flexibility and adaptability. When confronted with challenges, we needed to explore different approaches to problem-solving rather than rigidly sticking to our initial hypotheses.
2. Data Preprocessing Matters: The success of our models was closely linked to how we preprocessed and transformed the data. Effective data preprocessing is critical to model performance.

These insights and lessons learned have influenced our project's trajectory, emphasizing the need for adaptability, rigorous data preprocessing, and the importance of exploring various approaches. They will guide our future steps to ensure that we make informed choices and continuously improve our models.

**IV. Next Steps (Modeling and Evaluation)**

Question:   
Given your current progress, what additional tasks does your team need to complete in connection with the Modeling and Evaluation phases of your project? What is your plan to complete these tasks?

Student Team Response:  
To further progress in the Modeling and Evaluation phases, our plan includes:

1. SVM Resolution: Addressing runtime issues with SVM by exploring alternative implementation approaches.
2. Model Comparison: Evaluating all four models (Random Forest, Gradient Boosting, Logistic Regression, and SVM) to determine the most accurate one.
3. Integration into Main Codebase: Integrating the chosen model into our main codebase on Google Codelabs for future project use.

**V. Request for AI Studio TA Support**

Question(s):   
What additional support do you need from your AI Studio TA/Tutor/Course Support? Please structure your response as specific questions, related to the Modeling and Evaluation phases of your project. Consider areas where you may require specific guidance, clarifications, suggested approaches, or suggested resources. Your AI Studio TA will review these questions and work through them with you in an upcoming meeting or chat.   
  
Student Team Response:  
Do we have all the result information we need for each model? Is this enough to give concrete results?

Any ideas on how to make SVM quicker and finalize the running process?

What else can we do to further increase our predictive accuracy for Logistic Regression and Gradient Boosting?