

Componente: Reto Integrativo – Desafío BREIT

El Reto Integrativo permite poner en práctica todo lo aprendido durante el Programa, tanto a nivel técnico como de global skills.

En equipos abordarán un problema social y co-diseñarán junto a una organización una solución aplicando analítica de datos.

Key players en el proceso:

- Estudiantes: Aplican todo el conocimiento y habilidades técnicas y no técnicas, para lograr que el proyecto tenga el mayor impacto posible.
- Organizaciones sociales: Facilitan la data e información, así como participar en el diseño y validación de la solución.
- Mentores: Data Scientists con experiencia que acompañan los equipos durante el challenge
- Expertos: Locales de diferentes sectores brindan feedback técnico ad-hoc a los equipos
- Equipo MIT IDSS: Asesoran a los estudiantes para asegurar la solidez técnica del proyecto.
- Equipo Breit: Facilitan la comunicación y cumplimiento de compromisos de los stakeholders y apoyan a los equipos.

Beneficios

A nivel de Organización:

- ✓ Acceder a profesionales especializados que los ayudan a diseñar e implementar una solución relevante.
- ✓ Reconocer la importancia del manejo de datos dentro de una organización de impacto social, de cara a implementar mejoras alineadas al análisis de datos.

A nivel de equipo de estudiantes:

- ✓ Aplicar lo aprendido a nivel técnico y de global skills a proyectos con data sets reales.
- ✓ Aprender sobre la gestión y desarrollo de un proyecto analítico desde su concepción, desarrollo y puesta en marcha.
- ✓ Generar espacios de aprendizaje donde el foco es la gestión de proyectos y el trabajo en equipo desde la innovación y las metodologías ágiles.
- ✓ Potenciar skills de comunicación y negociación con distintos tipos de stakeholders.
- ✓ Lograr un impacto en el desarrollo de país a través del trabajo con organizaciones sociales.

Soporte en el proyecto, presentaciones y entregables

Durante el proyecto tendrán soporte constante para asegurar una experiencia de aprendizaje adecuada y lograr los objetivos del proyecto. Además, tendrán que realizar distintos entregables (actas, reportes, presentaciones) que acompañen el progreso de sus proyectos y den cuenta de sus avances a los grupos de interés: Docentes, Equipo MIT y ONGs.

1. MIT

Tendrán espacios para recibir feedback técnico de los TAs del MIT. Estos espacios darán cuenta del avance del proyecto y asegurarán la solidez técnica del mismo. Serán un total de 5 momentos de feedback:

- First check-in (Presentación oral, no calificado)
- Initial deliverable (Reporte escrito y Presentación oral. Ambos calificados)
- Midway Deliverable (Reporte escrito y Presentación oral. Ambos calificados)

- Second checkin (Presentación oral, no calificado)
 - Final Deliverable (Reporte escrito y Presentación oral. Ambos calificados)
- *El detalle de lo que debe considerar cada entregable y la rúbrica de evaluación se encuentra como anexo en este documento.

2. Organizaciones sociales

Tras las presentaciones de los reportes escritos y orales con el MIT, compartirán con las organizaciones la documentación que dé cuenta de sus avances.

Cabe resaltar, que además de dichos entregables podrán enviar otros según los requerimientos de las organizaciones:

- Presentación First check-in (presentación en castellano de los puntos más relevantes)
- Presentación del Initial deliverable (Presentación en castellano)
- Presentación del Midway deliverable Presentación en castellano)
- Presentación Second check-in (presentación en castellano de los puntos más relevantes)
- Presentación final del Proyecto

Para la etapa de transferencia del proyecto recomendamos contar con actas de reuniones y de todos aquellos entregables que les compartirán, así como manuales y/o videos que se requieran para realizar con éxito la futura ejecución de la solución diseñada por parte de la organización.

3. Soporte BREIT (profesora, mentores, equipo BREIT)

Durante el proyecto tendrán clases estructuradas que los acompañarán a lo largo de cada etapa. Según lo que se trabaje en clase, la docente solicitará trabajos y/o presentaciones, como el MVP o el MMP el cual se deberá cumplir con todas las especificaciones requeridas.

Además, durante todo el proyecto tendrán el soporte de mentores con quienes tendrán sesiones periódicas y de quien se podrán apoyar para resolver desafíos durante el proyecto o validar aspectos técnicos de su solución.

Finalmente, tendremos espacios de checkpoints con el Equipo BREIT en los compartirán con el equipo su avance según lo que se les solicite anticipadamente.

Fases:

✓ Ideación

Punto de partida del proyecto, a partir de reuniones (levantamiento de información) y revisión de la data, el equipo propondrá posibles soluciones para la organización.

Es importante tener en cuenta que previo al feedback de la organización, el equipo organizará espacios de manera interna para proponer ideas, imaginar soluciones, realizar benchmarks sobre proyectos similares y discutir juntos que podrían desarrollar bajo la visión de agregar valor a la organización y llevar sus aprendizajes a la cancha.

Esta fase va desde el 16 de marzo hasta el 19 de agosto

Hitos:

Fecha	Actividad/Grupo de interés	Propósito
Semana del 18 de marzo	Reunión inicial ONG	<p>Objetivo:</p> <ul style="list-style-type: none"> ▪ Presentación del equipo de estudiantes/equipo ONG ▪ Entender el propósito y propósito de la organización ▪ Discutir la necesidad para el proyecto y objetivo ▪ Solicitar la información relevante para realizar el proyecto

Semana del 25 de marzo	Reunión de consultas ONG	Objetivo: <ul style="list-style-type: none"> Resolver consultas sobre la información revisada
Semana del 08 de abril	Checkpoint Equipo BREIT	Objetivo: <ul style="list-style-type: none"> Presentar el avance del equipo según lo solicitado por BREIT
Semana del 15 de abril	Reunión de Aprobación de Project Charter ONG	Objetivo: <ul style="list-style-type: none"> Presentar las alternativas de solución previstas por el equipo de estudiantes Recibir feedback por parte de la ONG sobre las propuestas
Semana del 29 de abril	First checkin Equipo MIT IDSS	Presentación del avance del Proyecto según los criterios especificados en el anexo.
Semana del 06 de mayo	Presentación de avance ONG	Presentación del avance del Proyecto según el Feedback recibido por el equipo del MIT IDSS, la docente y los mentores a la ONG.

✓ *Desarrollo*

Tras haber priorizado con la organización la solución a realizar, el equipo inicia la ejecución. De manera interna, el equipo planificará su Gantt, definiendo fechas, roles y entregas para el avance adecuado del proyecto.

Considerar que el Capstone Exam tiene como fecha de inicio el 14 de octubre, por lo que las fechas a partir de setiembre serán confirmadas en los siguientes meses.

Esta fase va desde el 20 de mayo hasta noviembre.

Hitos:

Fecha	Actividad/Grupo de interés	Propósito
Semana del 13 de mayo	Reunión de Lanzamiento del desarrollo de la solución ONG	Objetivo: Presentar de manera oficial el proyecto y su alcance: <ul style="list-style-type: none"> ✓ El proyecto por realizar y su alcance ✓ Cronograma ✓ Información y herramientas que se requiere por parte de la ONG ✓ Persona de contacto por parte de la ONG Tras la reunión, el equipo debe elaborar un acta escrita con lo visto en la reunión y los acuerdos dentro de la misma.
Semana del 03 de junio	Checkpoint Equipo BREIT	Objetivo: Presentar el avance del equipo según lo solicitado por BREIT
Semana del 17 de junio	Presentación del Initial Deliverable (reporte escrito) Equipo MIT IDSS	Presentación del reporte escrito según los criterios especificados en el anexo.
Semana del 24 de junio	Presentación oral del Initial Deliverable Equipo MIT IDSS	Presentación oral del Initial Deliverable, contar con apoyo visual.
Semana del 08 de julio	Checkpoint Equipo BREIT	Objetivo: <ul style="list-style-type: none"> Presentar el avance del equipo según lo solicitado por BREIT

Semana del 15 de julio	Presentación del Midway Deliverable (reporte escrito) Equipo MIT IDSS	Presentación del reporte escrito según los criterios especificados en el anexo.
Semana del 22 de julio	Presentación oral del Midway Deliverable Equipo MIT IDSS	Presentación oral del Midway Deliverable, contar con apoyo visual.
Del 05 al 13 de agosto	Reunión de presentación del MVP a la ONG	Presentación del MVP a la ONG a fin de recoger feedback
Semana del 19 de agosto	Reunión de presentación del MVP Equipo BREIT	<p>Objetivo: Realizar la presentación considerando los siguientes puntos:</p> <ul style="list-style-type: none"> Breve resumen de la ONG y propósito Pains identificados y definición de la solución a desarrollar Valor que traerá esta solución a la organización (¿Con este MVP la ONG ya podría pilotear alguna acción para ir midiendo el valor? Razones por las que se ha elegido el algoritmo /técnica para la solución Limitaciones encontradas hasta el momento en relación con la data, entendimiento de negocio, procesamiento, etc. Consideraciones y/o dilemas éticos encontrados en el camino. Lecciones aprendidas de los retos, encontrados en la organización interna del equipo <p>En caso el equipo lo decida puede tener como backup sus Notebooks con EDA y experimentos de corridas de modelos, no necesariamente para presentarlos sino como material de consulta en caso alguno de los invitados les haga consultas.</p>
Semana del 26 de agosto	Checkpoint Equipo BREIT	<p>Objetivo:</p> <ul style="list-style-type: none"> Presentar el avance del equipo según lo solicitado por BREIT
Por confirmar	Second check-in Equipo MIT IDSS	Presentación del avance del Proyecto según los criterios especificados en el anexo.
Por confirmar	Entrega del MMP del proyecto ONG	<p>Objetivo: Entregar un MMP del proyecto para recibir feedback al regreso del Capstone Exam</p> <p>*MMP: Un producto comercializable mínimo, aborda todas las necesidades del cliente, siendo la mejor versión del MVP, después de las validaciones, pruebas y errores, y está lista para comercializarse y vender.</p>
Por confirmar	Reunión de feedback preliminar del proyecto ONG	Reunión con la ONG para recibir el feedback sobre la entrega preliminar
Por confirmar	Checkpoint Equipo BREIT	<p>Objetivo:</p> <ul style="list-style-type: none"> Presentar el avance del equipo según lo solicitado por BREIT
Por confirmar	Presentación del Final Deliverable, (reporte escrito) Equipo MIT IDSS	Presentación del Final Deliverable (reporte escrito) según los criterios especificados en el anexo.

Por confirmar	Presentación oral del Final Deliverable Equipo MIT IDSS	Presentación del Final Deliverable (reporte escrito) según los criterios especificados en el anexo.
Por confirmar	Incorporación del Feedback dado por la ONG	El equipo incorpora el feedback de la ONG y realiza las modificaciones pertinentes de cara a finalizar con el proyecto.

✓ *Transferencia*

Proceso final del proyecto en el que se realiza el traspaso de la solución a la organización. Para ello se deberá establecer quién será el o los owners del proyecto por parte de la ONG, el número de capacitaciones que se realizarán y la entrega de todos los documentos/archivos requeridos para el funcionamiento óptimo del proyecto a futuro.

Esta fase se realizará de noviembre hasta 15 de diciembre

Hitos:

Fecha	Actividad/Grupo de interés	Propósito
Semana del 11 de noviembre	Checkpoint Equipo BREIT	Objetivo: <ul style="list-style-type: none"> Presentar el avance del equipo según lo solicitado por BREIT
Semana del 15 de noviembre al 10 de diciembre	Reuniones de Transferencia ONG	El equipo realizará el traspaso del proyecto según los acuerdos alcanzados con la ONG dentro del número definido de reuniones o capacitaciones. Al realizar el traspaso al 100%, cada equipo deberá enviar un acta con las siguientes secciones: <ul style="list-style-type: none"> ✓ Participantes de las reuniones ✓ Breve resumen del proyecto entregado y archivos/documentos asociados ✓ Instrucciones de uso ✓ Recomendaciones *Además de las secciones que, según el proyecto, vean conveniente.
Del 12 al 19 de diciembre	Compartir los entregables del proyecto al equipo BREIT	Envío de los documentos de cierre al Equipo BREIT: <ul style="list-style-type: none"> Presentaciones Documentos/archivo y códigos

Participar activamente del Reto Integrativo es un requisito indispensable para graduarte del Programa, dado que tu participación es pieza fundamental para el desarrollo de este.

ANEXO I

Guidelines

1. Problem Formulation

- a. Start with a draft problem you aim to solve, and then gradually formulate a concrete problem within the context provided by NGOs by thinking about the following questions.
- b. Why is this problem important? What is the status?
- c. Who are the stakeholders? What actions do the stakeholders decide over for this problem? What information do they need to choose between actions? What are their objectives?

2. Data Collection

- a. Collect the data
 - i. What data do you need to achieve your goal? What information could these data provide for the stakeholders to make decisions? How could those data lead to solutions to your target problem?
 - ii. What sources should you search for those data (e.g., websites, organizations, surveys)? How could the bias in the way of collecting the data affect the feature distribution and your final conclusion? What might make one outcome more likely to occur due to selection issues? Collect the raw data from those sources.
 - iii. Are you using a static dataset or a time series dataset? For example, is your data collected at a particular point in time or collected over time?
- b. Explore the data
 - i. Describe the features within the context of the dataset. Are all features relevant and informative to your target question? How do the features correlate with each other? How might these correlations among features relate to the question you are investigating?
 - ii. How large is the dataset? Do you need the entire dataset or just a subset of it for your project?
- c. Visualize the data
 - i. Visualize the distribution of each feature in the dataset by making appropriate plots (e.g., histograms, scatter plots (averaged over each bin), box-and-whisker plots).
 - ii. Create a summary statistics table by reporting the mean, standard deviation, number of observations, and any other necessary statistics for every variable of interest, in the main population and in any interesting sub-populations.

- iii. Is each feature evenly distributed or skewed towards certain classes/values? Are there different sub-populations that indicate the need for stratification or cluster analysis?
 - iv. Draw insights from these plots and think about how the distribution of each feature could affect your modeling decision and give rise to any potential bias in your conclusion.
 - d. Clean the data
 - i. Standardize numerical features to ensure that each feature is defined based on the same unit of measurement. Convert categorical features via one-hot encoding. Document any transformations you make to your data.
 - ii. Deal with missing entries by either filling in some appropriate values (e.g., mean, median, mode) or removing the corresponding data point. Consider whether the data values are missing at random or due to some selection.
 - iii. Check if there are any outliers in the dataset. Keep them if the outliers make sense or remove them if they are due to some mistakes of measurement.
- 3. Methodology
 - a. Think of your problem type
 - i. Is your target problem descriptive or causal? If you want to predict outcomes for different populations based on certain characteristics, a descriptive model may be fine. If you want to find out any causal effects to guide decision-making, a causal model may be more appropriate.
 - ii. Is your target problem a regression problem or a classification problem?
 - b. Start with something simple
 - i. Think about a minimum viable product. What is the simplest possible version of an answer that can be later expanded into a more complete solution?
 - c. Build your model
 - i. Do you observe/assume any complicated/higher-order interactions among features that are necessary for your prediction?
 - ii. Should you stick to a simple model (e.g., linear regression) or go for a more complicated model (e.g., logistic regression, support vector machine, random forest, neural network)? Simple models are typically more interpretable and statistically testable, while complex models are often more powerful and suitable for pure predictive problems.
 - iii. Do you have sufficient data to ensure that your model is not overfitted?
 - iv. Write codes and fit your model.

4. Experiments

a. Dataset

- i. Split your dataset into a training set, a validation set, and a testing set.
- ii. Prepare for k-fold cross-validation.
- iii. Structure the inputs to ensure that the model is not overfitted. For example, if you are making predictions, make sure to include both positive and negative controls.

b. Evaluation metrics

- i. Think of the concerns/objectives of the stakeholders.
- ii. A single metric may not be enough to evaluate your method in an all-around way. Some potential metrics: accuracy, precision, recall, F1 score, etc.
- iii. Sometimes the time and computational cost of your method also matter.
- iv. Perform necessary diagnostics to see if your model is appropriate (e.g., regression diagnostics, cross-validation)

c. Methodology

- i. Check the existing approaches and think of some intuitive simple baselines.
- ii. What are the hyperparameters of your method? Tune them based on cross-validation.
- iii. Compare all potential methods and shortlist one or a few based on your metrics.

5. Result Interpretation

a. Quantitative evaluation

- i. How large is your improvement against existing methods or simple baselines? Is the gap significant enough so that the stakeholders are willing to adopt your method? Explain the improvement in performance based on the context of the target problem.
- ii. Quantify the uncertainties in the predictions: confidence interval, hypothesis testing, Monte-Carlo simulation
- iii. Test the robustness of your results: how does changing the hyper-parameters or the ways of dealing with outliers and missing values affect your conclusion?

b. Qualitative analysis

- i. What are the insights your method can provide for stakeholders, given the problem context? Any interesting causation/correlation patterns your method has discovered? Describe both the statistical features of your model

- (e.g., p-values, MSE, goodness-of-fit) and the real-world implications of your findings (e.g., the coefficient values, the contributions of certain variables towards an outcome of interest).
- ii. Under what circumstances does your method meet or do not meet your expectations? Are there regions of the population that the model systematically underperforms on? Any idea about the difference in performance?
 - c. Convey your results to the stakeholders.
 - i. To what extent can your method address the concerns/needs of the stakeholders?
 - ii. What are the takeaway messages based on your analysis? How will the results inform the decision-making process of the stakeholders?
 - iii. What are the limitations of your analysis? What extra care do the practitioners need to take when deploying your method in the real world?
6. Engineering Product
- a. Product brainstorming
 - i. What kind of product do you plan to deliver?
 - ii. What features do you plan to include in the product so as to fulfill the needs of NGOs?
 - iii. What features do you plan to include in the product so as to incorporate your analysis results?
 - b. Final product
 - i. Does your overall product convey the analysis results and meet the specific needs of NGOs?
 - ii. Is your product interface easily accessible and user-friendly? Have you grouped different types of results into different pages?
 - iii. Do you provide detailed documentation of your product so that staff from NGOs can understand your design?
 - iv. Do you provide backend access so that staff from NGOs can continue to input new data and update your model/product in the future?

Rubric (Total: 130 pts)

- First Check-in (Not for Grading Purposes)

1. Problem Formulation
 - a. If they have talked to NGOs about their concerns
 - b. If they have an initial idea about potential problems, they are going to investigate
2. Data Collection

- a. If they have obtained all the potential datasets from NGOs
 - b. If they have taken an initial look at those datasets
- 3. Engineering Product
 - a. If they have an initial idea about the kind of product they are going to provide to NGOs
- 4. Timeline and Division of Work
 - a. If they have met their group members and decided on each's role in the group

-Initial Deliverable (end of ideation phase before NGO presentation, e.g. After the first 6 weeks, Total: 17 pts)

- 1. Problem Formulation
 - a. Formulate a concrete problem within the context provided by NGOs (2 pt)
 - b. Explain why the target problem is important (1 pt)
 - c. Describe the current status of this problem (1 pt)
 - d. Identify all the relevant stakeholders (1 pt)
 - e. For each group of stakeholders, identify their objectives (1 pt), the decisions they are trying to make (1 pt), and the necessary information to support their decision-making processes (1 pt)
- 2. Data Collection
 - a. Describe the data collection methods you have tried (1 pt)
 - b. Ensure that you have already collected your dataset(s) (1 pt)
 - c. Explain why you choose your current dataset(s) over the other options (1 pt)
 - d. Explain any potential bias in your dataset(s) (1 pt), how the bias may affect your analysis (1 pt), and how you plan to address the bias (1 pt)
- 3. Methodology:
 - a. Propose your initial idea (1 pt)
- 4. Timeline and Division of Work
 - a. Specify the responsibility of each team member in each month (1 pt)
 - b. Specify the goal of each month so that the project can be completed in time (1 pt)

-Midway Deliverable (Total: 23 pts)

1. Data Collection
 - a. Describe the type, features, and size of your dataset(s) (1 pt)
 - b. Describe the features you plan to use as explanatory variables and response variables (1 pt), and explain why they are relevant to your target problem and informative to the stakeholders (1 pt)
 - c. Explain any potential correlations among the features that you plan to exploit with your model (1 pt)
 - d. Specify if you delete any data points in the original dataset(s) and why (1 pt)
 - e. Provide visualization for all the important features in your dataset(s) (2 pt)
 - f. Create a summary statistics table for all the important features in the main population and any interesting sub-groups (1 pt)
 - g. Describe any skewness in feature distributions based on the visualization and the statistics table (1 pt), and explain how the skewness affects your modeling decisions (1 pt)
 - h. Document any transformations you make to your data (e.g., standardize numerical features, encode categorical features) (2 pt)
 - i. Check any missing entries and outliers in your dataset(s), and explain how you deal with them (2 pt)
2. Methodology
 - a. Describe the type of your target problem (e.g., predictive or causal, regression or classification) (1 pt)
 - b. List a set of models you plan to use, and explain your choices (1 pt)
3. Engineering Product
 - a. Describe the type of product you plan to provide to NGOs (1 pt)
 - b. Outline the features you plan to include in your product so as to address NGOs' needs (2 pt)
 - c. Outline the features you plan to include in your product so as to incorporate your analysis results (2 pt)
4. Timeline and Division of Work
 - a. Specify the responsibility of each team member every two weeks (1 pt)
 - b. Specify the goal every two weeks so that the project can be completed in time (1 pt)

-Second Check-in (Not for Grading Purposes)

1. Methodology
 - a. If they have obtained some initial results with a minimum viable approach

- b. If their initial results make sense
 - c. If they have started to code their final approach
 - d. If their final approach is valid
- 2. Experiments
 - a. If they have correctly split the datasets for training/validation/testing
 - b. If they have identified suitable metrics for evaluation
- 3. Engineering Product
 - a. If they have put in considerable effort into developing the engineering product they are going to provide to NGOs
 - b. If their eventual engineering product is going to meet NGOs' needs

-Final Deliverable (Total: 90 pts)

- 1. Problem Formulation
 - a. Clearly state the concrete problem you are trying to solve (2 pt)
 - b. Provide any relevant contextual information for the target problem to illustrate its significance based on the current status of stakeholders (2 pt)
 - c. Highlight your overall contributions in this project (2pt)
- 2. Data Collection
 - a. Describe the features in your dataset(s) that you plan to use as explanatory variables and response variables (2 pt)
 - b. Provide visualization and a summary statistics table for those variables (2 pt)
 - c. Describe any findings based on the visualization and the table, including any skewness in feature distributions and observed correlations among the features (2 pt)
 - d. Document any data transformations, including why you deleted any data points, how you dealt with any outliers and missing entries, and how you standardized the numerical variables and encoded the categorical variables (3 pt)
- 3. Methodology
 - a. State if you are working on a causal/predictive regression/classification problem (2 pt)
 - b. Clearly describe your final model, including the model structure/inputs/outputs/hyperparameters/fitting process, and provide code snippets or model visualization if necessary (8 pt)
- 4. Experiments
 - a. Describe how you perform data splitting into a training set, a validation set, and a testing set (2 pt)

- b. Describe the evaluation metrics, and explain why you choose them (3 pt)
 - c. Describe how you tune the hyperparameters of your final model based on k-fold cross-validation (2 pt)
 - d. Describe at least two simple baselines that you have tried based on the above experimental setting, besides your final model (8 pt)
5. Result Interpretation
- a. Summarize the results of your final model and simple baselines in a table based on the metrics (2 pt)
 - b. Describe any surprising results, and explain why you choose your model given the context of your target problem (3 pt)
 - c. Describe the uncertainties and robustness of your results (2 pt)
 - d. Describe the statistical features of your model and any interesting causation/correlation patterns in your results (3 pt)
 - e. Interpret the real-world implications of these patterns and the insights you can provide for stakeholders (5 pt)
 - f. Identify any change in model performance across the population, and explain any potential reasons leading to the difference in performance (3 pt)
 - g. Explain how your results inform the decision-making process of the stakeholders and to what extent you have solved your target problem (4 pt)
 - h. Describe the limitations of your approach and any extra care the practitioners need to take when deploying your model in the real world (3 pt)
6. Engineering Product
- a. Illustrate your final engineering product (8pt) and explain how it conveys the analysis results (1pt) and meets the specific needs of NGOs (1pt)
 - b. Ensure that the interface of your product is easily accessible and user-friendly (5pt)
 - c. Provide detailed documentation of your product design (5pt)
 - d. Provide backend access so that staff from NGOs can input new data and update your model/product in the future (5pt)