UNIVERSIDAD ADOLFO IBÁÑEZ

Capstone Project
Magíster en Data Science

Rolando de la Cruz, Ph.D.
rolando.delacruz@uai.cl

Luis Aburto, Ph.D.
Luis.aburto@uai.cl

Qué es el Capstone Project

- Curso de Titulación del Master en data science
- Aprendizaje para interconectar todos los cursos aprendidos durante el programa
- Elementos para desarrollar un proyecto en ciencia de datos
- El problema a elegir es de completa elección de los alumnos: oportunidad para profundizar alguna temática o industria que más les interesa.

Trabajo de Graduación: Capstone Project

- El capstone project brinda la oportunidad de integrar todas las habilidades y herramientas aprendidas a lo largo del programa para diseñar y ejecutar un proyecto completo de ciencia de datos.

- El Capstone Project es un proyecto aplicado que cada grupo debe realizar:
  - El caso a desarrollar puede ser de una empresa de donde provenga un integrante del grupo.
  - Si el grupo no tiene un caso a desarrollar, debe indicar al Director del Magíster para ver opciones.

Trabajo de Graduación: Capstone Project

- El proyecto es grupal de 3 a 4 integrantes. No hay excepciones.
- La dedicación es de 8 horas semanales por persona. (32 horas en total en el grupo por semana)
- El proyecto debe resolver un problema práctico de data science de la empresa en un área específica o de alcance corporativo.
  - El alcance debe ser acotado al tiempo que dure el curso de Capstone Project
- El problema debe ser resuelto utilizando los conocimientos aprendidos en los cursos del Magíster.
  - El grupo puede profundizar en un tema específico de data science visto o mencionado en el Magíster para desarrollar su proyecto.

Exigencias

- Asistencia obligatoria
- Participación con preguntas (grupal) en las presentaciones
- Evaluación:
  - Presentaciones parciales (PP): 40%
  - Presentación defensa final (PF): 60%
  - Nota final Capstone Project = PP*0,4 + PF*0,6
- Nota Egreso:
  - 80% Promedio notas cursos
  - 20% Nota final capstone project

CAPSTONE PROJECT

- 10 de Noviembre - CLASE INTRODUCTORIA (CLASE 0 – esta clase)
- 22 y 23 de Marzo - 1ERA PRESENTACIÓN: DEFINICIÓN DEL PROYECTO Y METODOLOGÍA TENTATIVA
- 14 y 15 de Junio- 2DA PRESENTACIÓN: INGENIERÍA DE ATRIBUTOS, VISUALIZACIÓN Y PRIMER MODELO PREDICTIVO
- 20 y 21 de diciembre 2024 3RA PRESENTACIÓN: ANALÍSIS DE MODELOS Y VINCULACIÓN CON DECISIÓN DE NEGOCIO
- DEFENSA DE PROYECTO (por definir con integrantes de comisión)
- **Las fechas no se pueden cambiar**
- **No asistir o no realizar una presentación es nota mínima (1,0)**

1ra presentación del proyecto.

- Los grupos deben presentar su idea de proyecto según las indicaciones dadas.
- Los grupos al comienzo de la jornada deberán subir su presentación (ppt) a Webcursos.
- Los grupos recibirán feedback sobre el alcance del proyecto y si se ajusta como proyecto capstone.
- Cada grupo tendrá 15 minutos para presentar y 10 min de feedback de los profesores.
- La asistencia es obligatoria para que todos puedan aprender del feedback dado a otros grupos.

Alcances

- Es responsabilidad de los alumnos buscar y encontrar tema. Ni los profesores ni la Universidad entrega temas.
- Mientras antes comience mejor. El primer paso es definir un problema factible de resolver en el semestre.

Recomendaciones para definir su proyecto

- Para tener un aporte innovador y novedoso el grupo debe combinar:
  - tecnología + metodología + buenas prácticas + innovación

- El problema debe ser relevante no sólo para la empresa u organización sino que también que pueda ser repetible o implementado a otras industrias.

Elementos generales para la definición

- No basta con aplicar un modelo de machine learning. Es requisito algún elemento innovador (novedoso) en el desarrollo del proyecto:
  - Un modelo ensamblado
  - Una pregunta de negocios nueva
  - Evaluar una nueva fuente de información para resolver un problema
  - Una aplicación a una industria nueva
  - Una ingeniería de atributos novedosa que aporte a la solución del problema
  - Una nueva forma de implementar
  - Mezclar el modelo de machine learning como parte de una solución más completa

Elementos generales para la definición

- Trabajar con dos clientes (muchas veces quieren cosas distintas)
  - Cliente Profesores: que sea novedoso, robusto metodológicamente
  - Cliente Empresa: que sirva, que funcione
- La forma y el fondo de la presentación:
  - Convincente
  - Visualización – explicabilidad del modelo
  - Capacidad para explicar algo complejo (tomarse el tiempo)
  - Vender una idea
  - Gráficos vs Tablas

Recomendaciones para definir su proyecto

- Evite un proyecto que:
  - busque el testeo o prueba.
  - sólo busque implementar una práctica sin que haya una innovación en la metodología y herramientas utilizadas.
  - no tenga datos.

- En su proyecto debe haber un claro aporte a la organización. El proyecto debe ser innovador y novedoso para resolver el problema que se está abordando.
  - Y por lo tanto sus resultados deben ser medibles dentro del proyecto y no en una etapa posterior.

1ra presentación de Proyecto: Descripción del Problema de Negocio y metodología tentativa

- El grupo debe estructurar su presentación con lo siguiente:

1) **Contexto de la empresa**
   - Describir el caso/problema de negocio a resolver con datos

2) **Oportunidad**
   - Un problema de escala mayor o varios problemas.
   - Indicar cómo el problema está relacionado o alineado con la estrategia de la empresa u organización.

3) **Problema**
   - Selección del problema acotado que será abordado en el proyecto y su cuantificación.

4) **Objetivos**
   - Definir un Objetivo General
   - Definir 3-4 objetivos específicos que deben ser medibles, pero no indicar con números, por ej. Aumentar los niveles de ventas de xxxx, reducir costos de xxxx, mejorar calidad de xxxx, etc.

1ra presentación de Proyecto

5) **Metodologías para resolver el proyecto**

Debe indicar herramientas técnicas (qué métodos, indicadores, qué información hay), buenas prácticas, metodología de solución con pasos claros, resultados por cada paso y resultado final esperado. Debe indicar claramente cuál es su aporte o innovación.

6) **Métricas**

Definir métricas relacionadas a los objetivos específicos para medir el impacto de la solución y los resultados esperados. Una métrica es la fórmula (contar, ratio, porcentaje, $, etc.) con la cual se hace el cálculo con los datos obtenidos de la medición.

7) **KPI's**

Definir KPI's por cada métrica, es decir, un número, porcentaje o rango para comparar el número o porcentaje obtenido al calcular la métrica.

8) **Plan**

Definir las principales etapas del proyecto y agregar una carta Gantt para seguimiento personal y demostrar avances al profesor guía y en las sesiones de presentación de avance.

2da presentación Análisis Descriptivo y Primer Modelo

- La presentación debe contener avances en el desarrollo del proyecto.
- Al comienzo de la jornada los grupos deberán subir a Webcursos la presentación de avance.
- La asistencia es obligatoria para que todos puedan aprender del feedback dado a otros grupos.
- Elementos a evaluar:
  - Análisis descriptivo de los datos y Visualización
  - Ingeniería de Atributos
  - Un primer modelo predictivo resolviendo el problema

3ra presentación: Proyecto Final

- 3ra presentación
  - La presentación debe contener avances casi finales en el desarrollo de la solución del proyecto.
    - Modelo Final
    - Comparación de modelos y Optimización de parámetros
    - Vinculación del modelo con decisión de negocio y evaluación económica

Defensa de Proyectos

- Cada grupo tendrá 30 minutos para presentar y 10 minutos de preguntas del comité de defensa.
- La asistencia es obligatoria para todos los grupos.
- La presentación final del capstone Project es en grupo, pero la comisión de defensa evaluará individualmente a cada estudiante.
- Un integrante del comité de defensa debe ser de la empresa para la cual se desarrolla el capstone.
- Durante el semestre tienen el apoyo de un profesor Tutor de la Facultad para apoyar el problema específico que van a resolver.
- **En vez de tesis, se entrega un documento corto formato paper que resuma el problema y el trabajo realizado durante el semestre.**

Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

line 1: 1st Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 2nd Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 3rd Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 4th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

***Abstract*—This electronic document is a "live" template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract. (*Abstract*)**

***Keywords*—component, formatting, style, styling, insert (key words)**

I. Introduction (*Heading 1*)

This template, modified in MS Word 2007 and saved as a "Word 97-2003 Document" for the PC, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout a conference proceedings. Ease of Use

A. Selecting a Template (Heading 2)

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the US-letter paper size. If you are using A4-sized paper, please close this file and download the file "MSW_A4_format".

B. Maintaining the Integrity of the Specifications

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

II. Prepare Your Paper Before Styling

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as "3.5-inch disk drive".
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: "Wb/m2" or "webers per square meter", not "webers/m2". Spell out units when they appear in text: ". . . a few henries", not ". . . a few H".
- Use a zero before decimal points: "0.25", not ".25". Use "cm3", not "cc". (*bullet list*)

C. Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within

solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \qquad (1)$$

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)", not "Eq. (1)" or "equation (1)", except at the beginning of a sentence: "Equation (1) is . . ."

D. Some Common Mistakes

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o".
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)

E. Figures and Tables

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence.

TABLE I. Table Type Styles

Table Head	Table Column Head		
	Table column subhead	*Subhead*	*Subhead*
copy	More table copy[a]		

[a] Sample of a Table footnote. (*Table footnote*)

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord "Format" pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.

Fig. 1. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

Acknowledgment (*Heading 5*)

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp. 68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.

UNIVERSIDAD ADOLFO IBÁÑEZ

Elementos esenciales para la definición del proyecto Capstone en Data Science

Elementos a considerar

- Introducción (P1)
- Descripción del problema (P1)
- Objetivos (P1)
- Alcances y Entregables (P1)
- Metodología (P1) (P2)
- Revisión bibliográfica (P1)
- Desarrollo Metodológico (P2)
  - Descripción de los datos disponibles (P1)
  - EDA Exploratory data análisis (P2)
  - Ingeniería de Atributos (P2)
  - Construcción del primer modelo (P2)
  - Construcción del mejor modelo (P3)
  - Análisis de sensibilidad y costos (P3)
- Conclusiones (P3)
- Trabajos futuros (P3)
- Carta Gantt del proyecto (P1)
- Bibliografía (P1)

DESCRIPCIÓN DEL PROBLEMA

- ¿Cuál es el problema de negocio a resolver? ¿PARA QUÉ?
- Especificar el o los indicadores de negocio a mejorar
- Evaluar el tamaño del problema.
  - Hacer una evaluación simple respecto del tamaño del problema para la empresa
  - Dimensionar el impacto de la solución analítica a implementar (cliente, proveedores, impacto social)
- Indicar cómo el problema está relacionado o alineado con la estrategia de la empresa u organización.

- ¿Qué otros estudios han resuelto un problema similar? Revisión bibliográfica

Algunos ejemplos de Proyectos

- Desarrollo de un modelo predictivo para detectar casos de fraude interno en una institución bancaria
- Detección temprana de facturas falsas
- Send Time Optimization en cadena retail
- Optimización de concentración de cobre en Minera XXX
- Análisis de reclamos y LDA para Compañía de Seguros
- Deserción académica en educación Instituto Profesional
- Predicción de Fuga y morosidad en WorldVision - México
- Predicción de atrasos de proyectos mineros en Constructora
- Retención de talento en sector financiero
- Clasificador de comentarios usando NPS para XXX
- Clasificador de edad para motor de citas
- Recomendación de Productos para empresa de comida
- Modelo de Rentabilidad para empresa de Retail Financiero
- Modelo de propensión de pagos con deuda en Esval

Más ejemplos de proyectos

- Modelo de default de crédito en retail
- Deserción de docentes en sistema educacional - Mineduc
- Detección de mora temprana para Retail
- Riesgo de Cliente en empresa proveedora de internet
- Modelo Fuga de Clientes para Banco
- Estimación de Venta para Retailer de Ropa
- Modelo de Apertura de crédito para Retail usando Instagram
- Estimación de rotación de repuestos
- Predicción de ocupación de contenedores para TPS
- Estimación de derivaciones médicas en Red de salud

OBJETIVOS: qué queremos hacer

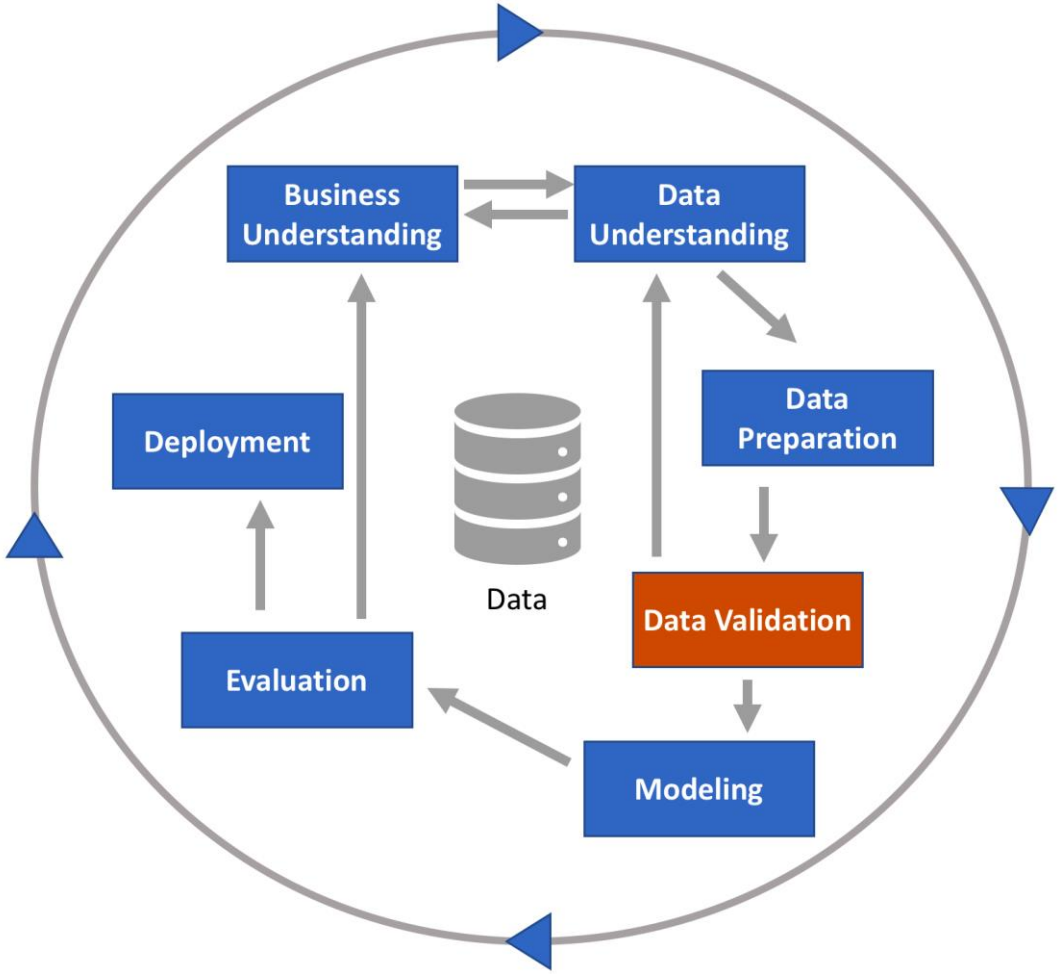
- No son sustantivos, son verbos en infinitivo.
- Son medibles, asociados a un KPI específico
- Al final del proyecto nos preguntaremos si cumplieron dichos objetivos
- Un objetivo general (asociado muchas veces con el título del proyecto) y al menos tres - cuatro objetivos específicos
- Los objetivos específicos no son etapas del proyecto. Están asociados a otros entregables que generan valor para la empresa:
  - Levantar el proceso atención y gestión de reclamos
  - Analizar la calidad de datos del proceso de devolución de productos
  - Recomendar la cantidad óptima de campañas de marketing a realizar
  - Recomendar el número de segmentos a usar por la compañía de seguros

ENTREGABLES Y ALCANCES

- Son sustantivos: informes, modelos, ETL, BBDD, Scripts, evaluaciones, recomendaciones.
- Asociados a cada objetivo del proyecto
- Es lo que recibe la empresa al final del proyecto

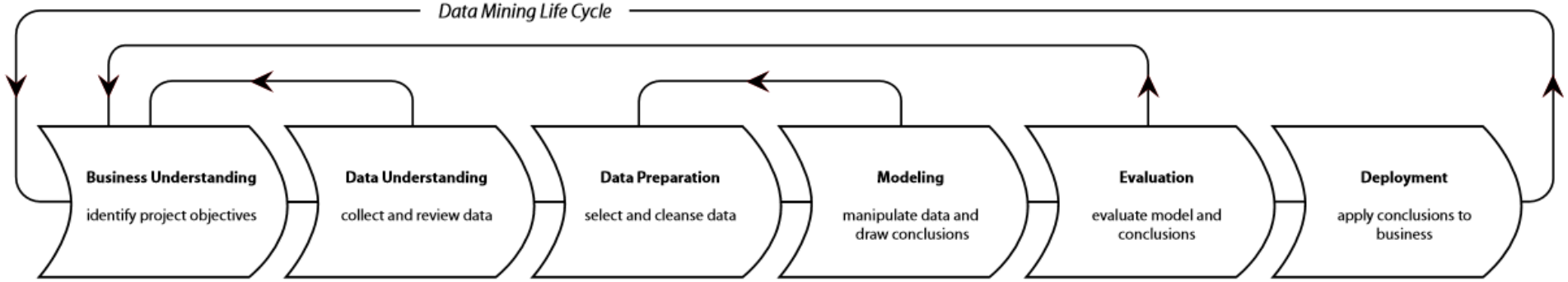
- Alcances: ¿hasta donde llega el proyecto? Especificar POR ESCRITO desde el inicio cual es el fin y que cosas NO incorpora el proyecto. Implementación? Puesta en Marcha? Extensión del modelo a otros problemas?

Metodología Analítica: cómo haremos el proyecto



- **No es copiar CRISP-DM**. Es adaptarlo al problema particular de cada proyecto. QUEREMOS DETALLES:
  - Qué tipo de modelos
  - Qué tipo de errores
  - Qué predecirán
  - Que variables usarán
  - Qué transformaciones usarán
- En general, antes de desarrollar un modelo predictivo se recomienda seguir una metodología analítica.
- La metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) nos orienta mediante un proceso de 6 fases:
  - Entendimiento del negocio
  - Entendimiento de la información
  - Preparación de la información
  - Modelamiento
  - Validación
  - Implementación

Phases



Determine Business Objectives
Background
Business Objectives
Business Success Criteria
(Log and Report Process)

Assess Situation
Inventory of Resources, Requirements, Assumptions, and Constraints
Risks and Contingencies
Terminology
Costs and Benefits
(Log and Report Process)

Determine Data Mining Goals
Data Mining Goals
Data Mining Success Criteria
(Log and Report Process)

Produce Project Plan
Project Plan
Initial Assessment of Tools and Techniques
(Log and Report Process)

Collect Initial Data
Initial Data Collection Report
(Log and Report Process)

Describe Data
Data Description Report
(Log and Report Process)

Explore Data
Data Exploration Report
(Log and Report Process)

Verify Data Quality
Data Quality Report
(Log and Report Process)

Data Set
Data Set Description
(Log and Report Process)

Select Data
Rationale for Inclusion/ Exclusion
(Log and Report Process)

Clean Data
Data Cleaning Report
(Log and Report Process)

Construct Data
Derived Attributes
Generated Records
(Log and Report Process)

Integrate Data
Merged Data
(Log and Report Process)

Format Data
Reformatted Data
(Log and Report Process)

Select Modeling Technique
Modeling Technique
Modeling Assumptions
(Log and Report Process)

Generate Test Design
Test Design
(Log and Report Process)

Build Model Parameter Settings
Models
Model Description
(Log and Report Process)

Assess Model
Model Assessment
Revised Parameter
(Log and Report Process)

Evaluate Results
Align Assessment of Data Mining Results with Business Success Criteria
(Log and Report Process)

Approved Models
Review Process
Review of Process
(Log and Report Process)

Determine Next Steps
List of Possible Actions
Decision
(Log and Report Process)

Plan Deployment
Deployment Plan
(Log and Report Process)

Plan Monitoring and Maintenance
Monitoring and Maintenance Plan
(Log and Report Process)

Produce Final Report
Final Report
Final Presentation
(Log and Report Process)

Review Project
Experience Documentation
(Log and Report Process)

Generic Tasks
Specialized Tasks
(Process Instances)

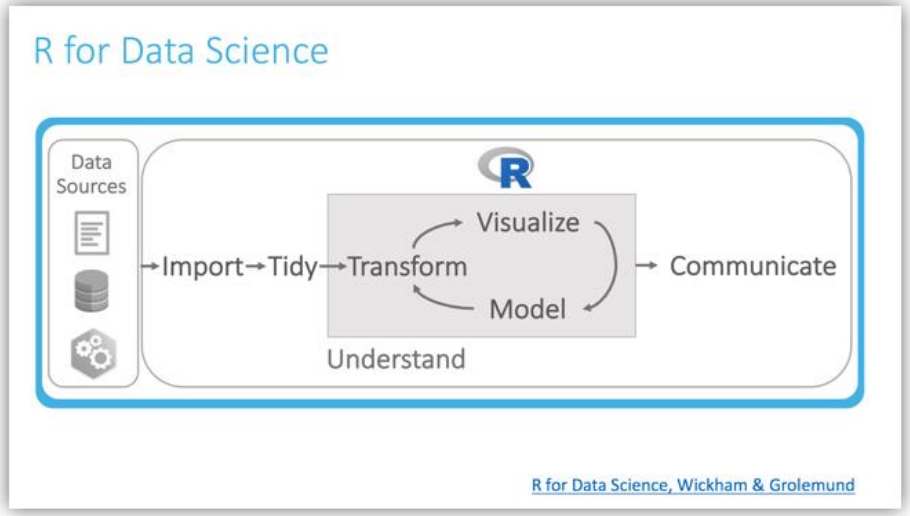
a visual guide to CRISP-DM methodology

SOURCE CRISP-DM 1.0
http://www.crisp-dm.org/download.htm
DESIGN Nicole Leaper
http://www.nicoleleaper.com

cc BY NC ND

Otros caminos: KDD; Semma; Wickham (2016)

- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, *39*(11), 27-34.
- Shafique, U., & Qaiser, H. (2014). A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, *12*(1), 217-222.
- Fernandez, G. (2010). *Statistical data mining using SAS applications*. CRC press.
- Wickham, H., & Grolemund, G. (2016). *R for data science: import, tidy, transform, visualize, and model data*. " O'Reilly Media, Inc.".



DESARROLLO METODOLÓGICO

- Descripción de los datos disponibles (P2)
- EDA Exploratory data análisis (P2)
  - Visualización de la data
  - Herramientas no supervisadas de agregación de información
  - Justificar que con los datos disponibles es posible resolver el problema de negocio planteado
- Ingeniería de Atributos (P2)
  - Creación de atributos relevantes para el problema de negocio
- Construcción del primer modelo (P2)
  - Demostrar que resuelven el problema (no necesariamente de la mejor forma)
- Construcción del mejor modelo (P3)
  - Comparación de modelos
  - Tuning / optimización de parámetros de modelos
  - Combinación de modelos
- Análisis de sensibilidad y costos (P3)
  - Optimización de decisiones según modelo

BIBLIOGRAFÍA

- https://scholar.google.cl/
- Referenciar todo! (sino, es plagio)
  - https://www.websitehostingrating.com/es/plagiarism/
  - https://www.turnitin.com/es/blog/cinco-tipos-plagio-mas-frecuentes
- Mucho mejor la fuente original que una en español

- ASÍ SE HACE: Formato APA:
  - Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, *39*(11), 27-34.

Presentación 1

- Introducción
- Justificación del problema de negocio
  - Destacar novedad/innovación en su proyecto
- Objetivos
- Entregables y Alcances
- Metodología Tentativa
  - Detalles de modelos, datos y variables a predecir
- Análisis Descriptivo de los datos
- Gantt
- Bibliografía

UNIVERSIDAD ADOLFO IBÁÑEZ

Capstone Project
Magíster en Data Science

Rolando de la Cruz, Ph.D.
rolando.delacruz@uai.cl

Luis Aburto, Ph.D.
Luis.aburto@uai.cl

UNIVERSIDAD ADOLFO IBÁÑEZ

Frequent Ask Questions

1. ¿Qué datos mínimos deben ser abordados en la descripción de la empresa? ¿Qué extensión debe tener la descripción?

En el ppt una lámina es suficiente.

2. ¿Qué requisitos debe cumplir el problema a abordar?

El problema debe ser un caso de negocio (puede ser una mejora a lo que ya tienen) donde se usen datos para resolverlo y que haya disponibilidad de datos.

3. ¿Qué requisitos mínimos debe cumplir la solución?

El requisito fundamental es que no requiera una simple aplicación de algún método visto (pueden profundizar/usar otros métodos eventualmente) en el programa. La solución analítica debe generar innovación en la empresa y debe ser cuantificable con los KPIs que se definan.

4. ¿La medición de retorno es solo en el ámbito económico o puede ser en HH u otros?

La medición debe ser cuantificable en cómo genera valor para la organización. Allí hay que ver como impacta el uso de la solución analítica con respecto a no usarla.

5. ¿Con cuántas otras soluciones mínimo se debe comparar la solución esperada?

Eso va a depender del problema/reglamentación, etc. Si quieren desarrollar un modelo predictivo binario, lo ideal es comparar resultados con varios métodos (puedo recordar por lo menos 10). También hay que hacer una búsqueda bibliográfica de qué cosas se han hecho en el contexto del problema a resolver.

6. ¿Existe un syllabus con la información detallada del contenido de cada evaluación junto a sus criterios (puntaje por aspecto a evaluar), documentación a exigir a la empresa y fechas de entrega de la documentación?

Para cada presentación futura subiremos una rúbrica. Con respecto a la documentación a exigir a la empresa, eso lo maneja cada grupo con su contraparte (búsqueda del problema, fecha de entrega de datos, retroalimentación de avance, etc).

7. ¿Existe un cobro asociado a la defensa o a la ceremonia de titulación?

Con respecto a la defensa no existe cobro. Para la ceremonia de titulación no manejo esa información. Puedes escribir a Ingrid sobre ese punto.

8. ¿Cuál será la modalidad de defensa y su rúbrica?

La defensa de cada proyecto capstone es ante una comisión integrada por 3-4 profesores. Hay un tiempo de 30 minutos de presentación y posterior tiempo para preguntas. La presentación es grupal pero la nota es individual, y las preguntas pueden ser personalizadas (eso implica que todos los miembros del grupo deben manejar todo sobre el desarrollo del casptone). Publicaremos una rúbrica de defensa.

9. ¿Qué documentos se solicitan al alumno para la titulación?

Para la defensa: todo alumno/a debe estar sin deudas pendientes, no deber libros en la biblioteca, haber completado los cursos obligatorios y electivos. Previo a la defensa (al menos una semana) cada grupo debe enviar titulo y un resumen ejecutivo de a los más 200 palabras del capstone project. Para efectos de titulación, haber completado todos los requisitos antes mencionado para la defensa más la probación de la defensa de capstone.

- ¿La universidad tiene sus propios proyectos capstone (o a empresas asociadas) que estén disponibles para los grupos que lo requieran?

No. La universidad no cuenta con proyectos o propuestas de capstone.

- Relacionado a lo anterior, alguien propuso que se hiciera un listado de distintos proyectos o ideas y luego cada uno eligiera el tema que más le gustara para formar los grupos, ya que quizás algunas personas prefieren hacer su capstone en procesamiento de imágenes y otros en una temática social. Para ello habría que saber qué proyectos pone a disposición la universidad y también proyectos / ideas en las empresas donde trabajamos. ¿Es posible llevar a cabo esta idea?

Rpta: No se dispone de una feria de capstone projects de parte de la U, ya que la U no cuenta con proyectos (algunos académicos tienen proyectos, pero usualmente requieren dedicación exclusiva para su desarrollo). Como les fue informado en el proceso de postulación. Las propuestas de capstone usualmente vienen de las mismas empresas de los alumnos. En el caso que algún grupo no tenga capstone, invitamos a organismos públicos o privados ha presentar propuestas. Pero eso debe ser informado con anticipación como fue requerido por email, para así hacer las gestiones correspondientes.

- Dudas varias para que un proyecto o idea se pueda considerar para el Capstone: ¿Es necesario que uno de los integrantes trabajé en la empresa? ¿Qué tan innovador tiene que ser el proyecto? ¿Que pasa si los KPI no están 100% claros? ¿Debemos nosotros establecer los KPI? Etc.

El problema de capstone puede venir de la empresa de uno de los integrantes del grupo o de otra empresa. Lo importante es que tienen que tener disponiblidad de datos para desarrollar el capstone. El proyecto de capstone no busca una simple aplicación a un caso de negocio, debe ser innovador y generar impacto en la organización donde se desarrolla. Ese impacto tiene que ser cuantificable. Los KPIs tienen que ser claramente definidos en el marco del problema a resolver y que sean cuantificables. Los KPIs deben ser establecidos por el equipo de casptone.