



**ABRIL – JULIO**

**2021**

**Vs.10.2**

# **“Introducción al Data Science para el Sector Público” (IDSSP)**

Introducción a técnicas de Automatización, Big Data y Modelos de Clasificación (Predictivos, Supervisados, No Supervisados y Exploratorios)

Ámbito: Data Science

## IDENTIFICACIÓN DEL DIPLOMA

### PRESENTACIÓN DEL DIPLOMA

Hoy en día en las instituciones se recogen grandes volúmenes de datos en diversas escalas y magnitudes, particularmente en aquellos que tienen un fin público, con importantes impactos en la gobernabilidad y en la gobernanza. Hoy en día la creación e innovación de valor público se encuentra, en gran magnitud, en estos yacimientos, y por sobre todo, en el “tratamiento inteligente” de estos datos, los que se transforman en un activo estratégico para efficientar, a través de procesos automatizados, el monitoreo, seguimiento y prospectiva adecuada para la previsión y generación de diversos escenarios en el cumplimiento de los fines y propósitos para los cuales cada organización o institución fue creada.

Para colocar en valor y dominio toda esta data, las instituciones requieren acelerar la incorporación de competencias, nuevas tecnologías y herramientas, así como establecer una nueva cultura basada en los datos que se cristalice en la instalación de nuevas competencias en sus funcionarios, para así responder a la creación o formación de ese nuevo funcionario “Data tecnócrata y/o burócrata” (“Data Technocrat Scientist” or “Data Bureaucrat Scientist”<sup>1</sup>)

En el mundo actual, con economías globalizadas y abiertas al mundo, las habilidades digitales son un elemento fundamental para el éxito y/o sustentabilidad en la provisión inteligente de bienes y servicios en cualquier ámbito. En este escenario, las organizaciones enfrentan diversos y complejos problemas referidos a la necesidad de mejorar la inteligencia (tratamiento) de datos con el fin de tomar mejores y ágiles decisiones.

Atendiendo a estas emergentes necesidades y, con la finalidad de alcanzar un dominio y aplicación adecuada de estas herramientas y metodologías, el Instituto de Asuntos Públicos de la Universidad de Chile ofrece el diplomado de “Introducción al Data Science para el Sector Público” (IDSSP) a los diversos funcionarios de la Administración del Estado, buscando así, satisfacer la creciente demanda de profesionales con habilidades y conocimientos en análisis, automatización y gestión inteligente de datos.

Lo anterior les permitirá no sólo analizar los datos históricos de una institución u organización, sino que además predecir escenarios hipotéticos futuros para la toma de decisiones y optimizadas, contribuyendo de esta forma a la generación de escenarios deseados y preventivos. De esta manera se podrá aportar hacia una transición de desarrollo. Desde una gestión basada en indicadores a una gestión basada en el mejoramiento de la eficiencia por medio de una institucionalidad automatizada, analítica y predictiva.

---

<sup>1</sup> Concepto acuñado por Adrian Araneda Toro.

## 1. OBJETIVO GENERAL

Los estudiantes de este diplomado obtendrán herramientas analíticas y prácticas que les permitirá conocer y ejecutar acciones en todo el proceso de la Inteligencia de Negocios, desde la adquisición y limpieza de los datos hasta la creación de modelos predictivos (cuidando los supuestos estadísticos básicos detrás de cada modelo analítico), incluyendo el aprendizaje e internalización de los conceptos de Big Data, Visualización, e introducción al Machine Learning entre otras, de tal manera que les permita comprender el ciclo completo de la “Analítica” o Análisis de Datos, adquiriendo las habilidades tanto técnicas como de articulación para poder implementarlos de manera efectiva dentro de sus organizaciones para la resolución de problemas. El diplomado en “Introducción al Data Science para el Sector Público” (IDSSP), posee una orientación profesionalizante, dirigida a la resolución de problemas.

## 2. OBJETIVOS ESPECÍFICOS

Se busca que los estudiantes del IDSSP investiguen, apliquen los conocimientos adquiridos y contribuyan a la solución de problemas específicos y complejos, con la capacidad de contextualizarlos a la realidad profesional institucional y de negocio, de cada estudiante del programa o según sus intereses particulares, balanceando aspectos estratégicos, técnicos y de economicidad.

Es así que, en el mismo orden de consideraciones, el Diplomado IDSSP tiene por objetivo desarrollar un conjunto de competencias de manera que sus egresados y egresadas sean capaces de:

- a. Abordar problemas complejos en sus organizaciones mediante la exploración adecuada de la data.
- b. Integrar datos de distintas fuentes y gestionar datos de diversas escalas de magnitud.
- c. Aplicar modelos descriptivos, predictivos y prescriptivos para permitir a las organizaciones tomar mejores decisiones, más rápidas e informadas.
- d. Convertir datos en productos y servicios mediante el uso de herramientas analíticas.
- e. Descubrir oportunidades de negocio accionables a través del análisis exploratorio y analítica de datos, contribuyendo de esta manera al desenvolvimiento con equipos multifuncionales, transversales y multidisciplinarios para, a través de todo un Macro Proceso, definir los requisitos de negocio, desarrollo/llar el producto, implementar las soluciones generadas, y evaluarlas a través de medidas estadísticas probabilísticas sus resultados.

### 3. DESTINATARIOS

- Profesionales del Sector Público que manipulen grandes volúmenes de Datos para distintos fines o modelos de negocio.
- Profesionales del Sector Público que realicen analítica para los procesos de toma de decisiones o inteligencia de negocio.
- Profesionales del Sector Público que destinen parte de sus funciones o se desenvuelvan en activadas o en procesos evaluatorios, de programas, iniciativas, intervenciones, etc., de distintas naturalezas, basadas en medidas objetivas, y medidas de optimización (datos y parámetros).
- De esta manera, el programa está dirigido a titulados o graduados, cuyas carreras de origen tengan una duración oficial de a lo menos 4 años de duración:

“Título profesional universitario nacional o extranjero y/o grado equivalente de licenciado en 8 semestres”.

### 4. METODOLOGÍA DE TRABAJO

El Diploma incluye cursos metodológicos, aplicados y talleres prácticos que proveen el conocimiento de prácticas avanzadas de manejo del negocio, con un énfasis cuantitativo del pre procesamiento, análisis exploratorio y modelamiento, incorporando los elementos conceptuales adecuados.

### 5. SISTEMA DE EVALUACIÓN

El Programa contempla principalmente trabajos individuales y/o grupales por modulo. No obstante, contempla la posibilidad de tomar controles de lecturas o pruebas en clases según la naturaleza de los contenidos, y la definición de cada docente.

## Recursos Tecnológicos a Utilizar

- a. El alumno necesitará una conexión estable a Internet sea por wifi o punto de red.
- b. El hardware con el que deberá contar el estudiante es un Notebook o pc ordenador.
- c. La sesión o perfil que utilice el alumno en dicho notebook u ordenador debe estar desbloqueada, o contar con privilegios de “administrador de equipo” o con los permisos correspondientes para no impedir la instalación de softwares nuevos.
- d. En el mismo sentido que lo anterior, los respectivos antivirus que se encuentren instalados en los ordenadores, también deben contar con la configuración correspondiente para no presentar un corta fuego o bloqueo que interrumpa la instalación de un nuevo software y su uso con normalidad.
- e. Así también, los equipos de los alumnos deberán encontrarse sin problemas de rendimiento, velocidad, procesamiento, problemas en la tarjeta de video o gráfica, discos duros copados o llenos, etc.
- f. Los alumnos deberán contar con el programa “zoom” ya instalado en sus notebooks al primer día de clases, y poseer un nivel de usuario básico para su utilización, para así no afectar la entrega normal de los contenidos y ejercicios por zoom.
- g. Se facilitará a los alumnos con anticipación al módulo 4, los manuales y/o tutoriales correspondientes para dos operativos (Windows y Mac), para la instalación de los softwares que ocuparemos y que serán imprescindibles para este diplomado, RStudio.
- h. Los mismos alumnos deberán llegar con el Software citado, ya instalado a la primera clase del módulo 4. A partir de ese Módulo los alumnos comenzarán a hacer uso de este nuevo lenguaje (de programación, RStudio), el cuál será el que se utilizarán por los siguientes 10 Módulos. Dicho software es Open Source o Código Abierto (gratuito).

## 6. Inicio y días de clases.

Las clases en Modalidad Virtual se iniciarán el miércoles **14 de Abril 2021**

Las clases serán:

**Miércoles 2.5 hrs. cronológicas** en cada clase de día hábil (18:30 – 21:15 hrs. con 15 minutos de descanso). 4 miércoles en el mes. (15 miércoles)

**Jueves 2.5 hrs. cronológicas** en cada clase de día hábil (18:30 – 21:15 hrs. con 15 minutos de descanso). 4 viernes en el mes. (15 viernes)

**Sábado 4.0 hrs. cronológicas** en cada sesión completa de los días sábados (09:00 – 13:30 hrs. con 30 minutos de descanso) 14 sábados

## 7. CONTENIDOS Y CANTIDAD DE HORAS.

N° DE MÓDULO	NOMBRE MÓDULO	TITULO ABREVIADO	NOMBRE PROFESOR	N° de hrs
1	LA GESTIÓN PÚBLICA Y ÉTICA EN LOS DATOS	LA GESTIÓN PÚBLICA Y ÉTICA EN LOS DATOS	Justo Eduardo Araya Moreno	7
2	BASE DE DATOS E INTRODUCCIÓN AL MACHINE LEARNING	BASE DE DATOS E INTRODUCCIÓN AL MACHINE LEARNING	Rodrigo Lagos	9
3	NIVELACIÓN ESTADÍSTICA: DESDE ESTADÍSTICA DESCRIPTIVA A TEST DE HIPÓTESIS	ESTADÍSTICA DESCRIPTIVA E INFERENCIAL	Juan Carlos Herrera Ordenes	29
4	INTRODUCCIÓN A LA MANIPULACIÓN Y PROCESAMIENTO DE BASE DE DATOS PARA EL ANÁLISIS DE DATOS CON RSTUDIO.	MANIPULACIÓN DE BASE DE DATOS E INTRODUCCIÓN A RSTUDIO	Adrian Armando Araneda Toro	8
5	FUNDAMENTOS DE PROGRAMACIÓN PARA EL PROCESAMIENTO DE BASE DE DATOS. RSTUDIO (CICLOS)	FUNDAMENTOS DE PROGRAMACIÓN Y AUTOMATIZACIÓN	Francisco Acuña	7,5
6	ANÁLISIS EXPLORATORIO DE DATOS (TIPOS DE DISTRIBUCIONES)	ANÁLISIS EXPLORATORIO DE DATOS	José Alexander Yáñez Albornoz	12
7	VISUALIZACIÓN	VISUALIZACIÓN O PLOTEO	Francisco Acuña	15
8	INTRODUCCIÓN A LA ECONOMETRÍA PARA EL CONOCIMIENTO DEL FUNCIONAMIENTO ALGEBRAICO Y MATRICIAL DE LAS TÉCNICAS O MODELOS DE REGRESIÓN, ESPECIALMENTE LOGIT BINOMIAL Y GENERALIZADO	INTRODUCCIÓN A LA ECONOMETRÍA Y REGRESIONES	Rodrigo Pavez Orellana Alex Meléndez Suazo	22
9	MODULO APLICADO DE MINERÍA DE DATOS: 6 TÉCNICAS O MODELOS (ALGORITMOS) DE CLASIFICACIÓN PROBABILÍSTICO Y PREDICTIVOS (LOGIT BINOMIAL Y GENERALIZADO, KNN - K-NEAREST NEIGHBOURS-, NNET – REDES NEURONALES-, TREE – ÁRBOL DE DECISIÓN-, RF – RANDOM FOREST-, SVM – SUPPORT VECTOR MACHINE-). Y CUATRO ALGORITMOS DE CLUSTERING (K-MEANS, PAM, HIERARCHICAL Y NAIVE BAYES).	MODELAMIENTO PREDICTIVO	Juan Carlos Herrera Ordenes José Alexander Yáñez Albornoz	10,5
10	TÉCNICA DE ANÁLISIS EXPLORATORIO Y NO SUPERVISADO "CLUSTERING"	CLUSTERING	Alex Meléndez Suazo Adrian Armando Araneda Toro	7,5
11	QUE ES LA BIG DATA E INTRODUCCIÓN A LA CAPTACIÓN Y ALMACENAMIENTO DE DATOS	BIG DATA E INTRODUCCIÓN A LA CAPTACIÓN Y ALMACENAMIENTO DE DATOS	José Alexander Albornoz Juan Carlos Herrera Ordenes	20
12	INTRODUCCIÓN A LAS TÉCNICAS MINERÍA DE TEXTO Y ANÁLISIS DE SENTIMIENTOS. RSTUDIO	INTRODUCCIÓN A LA MINERÍA DE TEXTO Y ANÁLISIS DE SENTIMIENTOS	Rodrigo Pavez Orellana	7,5

13	UTILIZACIÓN DE MINERÍA DE DATOS EN EL SECTOR PÚBLICO. CASOS Y RESULTADOS	ANALÍTICA EN EL SECTOR PÚBLICO	Rodrigo Lagos	6
			<b>Total horas Presenciales</b>	<b>161</b>

Los módulos son los siguientes:

1. LA GESTIÓN PÚBLICA Y ÉTICA EN LOS DATOS. Prof. Justo Eduardo Araya moreno
2. BASE DE DATOS E INTRODUCCIÓN AL MACHINE LEARNING. Prof. Rodrigo Lagos.
3. ESTADÍSTICA DESCRIPTIVA E INFERENCIAL. Prof. Juan Carlos Herrera Ordenes.
4. MANIPULACIÓN DE BASE DE DATOS E INTRODUCCIÓN A RSTUDIO. Prof. Adrian Armando Araneda Toro.
5. FUNDAMENTOS DE PROGRAMACIÓN Y AUTOMATIZACIÓN. Prof. Francisco Acuña.
6. ANÁLISIS EXPLORATORIO DE DATOS. Prof. José Alexander Yáñez Albornoz.
7. VISUALIZACIÓN O PLOTEO. Prof. Francisco Acuña.
8. INTRODUCCIÓN A LA ECONOMETRÍA Y REGRESIONES. Profs. Rodrigo Pavez Orellana y Alex Meléndez Suazo.
9. MODELAMIENTO PREDICTIVO. Profs. Juan Carlos Herrera Ordenes y José Alexander Yáñez Albornoz.
10. CLUSTERING. Profs. Alex Meléndez Suazo y Adrian Armando Araneda Toro.
11. BIG DATA E INTRODUCCIÓN A LA CAPTACIÓN Y ALMACENAMIENTO DE DATOS. Profs. José Alexander Yáñez Albornoz y Juan Carlos Herrera Ordenes.
12. INTRODUCCIÓN A LA MINERÍA DE TEXTO Y ANÁLISIS DE SENTIMIENTOS. Prof. Rodrigo Pavez Orellana.
13. ANALÍTICA EN EL SECTOR PÚBLICO. Prof. Rodrigo Lagos.

## 8. BREVE RESEÑA DE PROFESORES/AS

- **Rodrigo Pavez Orellana.** Ingeniero Comercial, Licenciado en Cs de la Administración de Empresas y Magíster en Gestión mención Finanzas, *Universidad de Valparaíso*. Magíster Economía Financiera, *Universidad de Santiago de Chile*. Profesor ayudante de la cátedra de Econometría II, para alumnos del programa de Magíster en Economía financiera, de la Universidad de Santiago de Chile. Experiencia en Control de Gestión “Banco Consorcio” y en Gerencia estratégica de zona en “Correos de Chile”. Actualmente coordinador o líder del equipo de Análisis de Datos de la Unidad de Análisis de Declaraciones de Intereses y Patrimonio, de Contraloría General de la República.
- **José Alexander Yáñez Albornoz.** Ingeniería Civil Industrial y Magister en Ingeniería Industrial. Universidad de Concepción. Diplomado en Inteligencia de Negocios. Universidad de Chile 2017. Ingeniero Senior Business Analytics - División Factoring en Grupo Security 2019-2018. Business Intelligence en Grupo Gerdau Aza 2018-2017. Primer lugar Concurso Innovación Inchalam: Idea para mejorar tracking logístico. Concepción, 2016. Premio Roberto Ovalle Aguirre: Premio por el mejor proyecto de título relacionado con la instalación o explotación de una industria relevante para el fomento de la economía nacional. Otorgado por Instituto de Ingenieros de Chile, 2015 • Premio Facultad de Ingeniería. Universidad de Concepción, 2015
- **Juan Carlos Herrera Ordenes.** Ingeniero Estadístico y Magíster en Estadística. Universidad de Valparaíso. Profesor titular en Matemática e Introducción a la Matemática en Universidad Andrés Bello. Especialista en Modelos Predictivos. Área de Procesos y Modelos Predictivos, Banco Estado Microempresas - Gerencia de Riesgo. 2019-2015. Analista Estadístico Investigador. Departamento de Investigación y Desarrollo, Instituto Nacional de Estadísticas (INE) 2015-2011.
- **Alex Meléndez Suazo.** Ingeniero Civil Industrial, Licenciado en Ingeniería Aplicadas, *Universidad de Santiago de Chile*. Magíster Economía Financiera, *Universidad de Santiago de Chile*. Profesor Ayudante de Econometría 1 y 2 para el Magister en Economía Financiera de la *Universidad de Santiago de Chile*. Experiencia profesional en Telecomunicaciones “Claro Chile, Mercado Empresarial”, y en consultorías técnicas “Contac Ingenieros” e “Invensys”. Actualmente miembro del equipo de Análisis de Datos de la Unidad de Análisis de Declaraciones de Intereses y Patrimonio de Contraloría General de la República.
- **Francisco Acuña.** Ingeniero Industrial. Universidad Católica de Córdoba. Actualmente alumno del Magister en Data Science. Universidad Adolfo Ibáñez. Profesor de Física Mecánica I, Física II, Análisis Matemático y Álgebra Lineal. Especializado en lenguaje SQL, Power BI para Data Science y ploteo o Visualización en RStudio. Actualmente presta sus servicios profesionales como Analista y Data Science en CRM Walmart Chile.
- **Adrian Araneda Toro.** Administrador Público. Licenciado en Cs Políticas y Gubernamentales con mención en Gestión Pública. Universidad de Chile. Postítulo en “Inteligencia de Negocios”. Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile 2017. Postítulo en “Estadística y Análisis Masivo de Datos”. Universidad Adolfo Ibáñez 2018-2019. Postítulo en “Data Science” y actualmente alumno de Magíster en “Data Science”. Universidad Adolfo Ibáñez. Autor de Modelo “Microeconómico” y

Seleccionado Expositor en BAFI 2020 <https://www.baficonference.cl/20/default/inicio> de la Facultad de ciencias Físicas y Matemáticas de la Universidad de Chile. Experiencia profesional en el Sector Público y Privado: Departamento de Estadísticas e Información de Salud, de Ministerio de Salud. Encargado de Sistemas de Gestión de Calidad, en Red de Empresas Fidegroup S.A. Actualmente miembro del equipo de “Análisis de Datos” de la “Unidad de Análisis de Declaraciones de Intereses y Patrimonio”, de Contraloría General de la República.

- **Rodrigo Lagos Chávez** Bioquímico, Universidad Católica. Magister en Bioestadísticas Universidad de Chile. Diplomado en Filosofía de la Ciencia. Experiencia en Data Science con énfasis en aplicación en proyectos de investigación en ciencias básicas y aplicadas. Métodos estadísticos frecuentistas, métodos Bayesianos y técnicas de machine learning (clustering, arboles de decisiones, entre otros).
  
- **Eduardo Araya Moreno. (Coordinador)** Administrador Público, Magister en Ciencia Política, mención Gobierno Universidad de Chile. Master en Sociedad de la Información y el Conocimiento UOC. Doctor ©en Sociedad de la Información y el Conocimiento UOC. Barcelona.  
Profesor de pregrado y posgrado en Universidades chilenas y extranjeras.  
Líneas de investigación en Políticas públicas, Sociedad digital y Formación profesional para el sector público  
Amplia experiencia académica de docencia e investigación en el área. Varias publicaciones en USA, México, Chile en el área.