

Diplomatura en Ciencia de Datos

Escuela de Ciencia y Tecnología
Universidad Nacional de San Martín (UNSAM)

Contenido

1. Identificación de la formación
 - 1.1. Denominación
 - 1.2. Ubicación
 - 1.3. Coordinación Académica
 - 1.4. Modalidad
2. Fundamentación
3. Objetivos de la Diplomatura
4. Certificación que otorga la formación
5. Destinatarios
6. Requisitos de ingreso
7. Diseño y organización curricular
 - 7.1. Estructura del plan de estudios
8. Modalidad de evaluación
9. Contenidos mínimos de las asignaturas

1. Identificación de la formación

1.1. Denominación

Diplomatura en Ciencia de Datos

1.2. Ubicación

Escuela de Ciencia y Tecnología

1.3. Coordinación Académica

Dr. Rodrigo Díaz

1.4. Modalidad

Presencial (33% de carga virtual)

2. Fundamentación

Nos encontramos en una época de transformación digital a nivel global. La disponibilidad de grandes volúmenes de datos (Big Data) y la aparición de técnicas disruptivas para analizarlos y extraer conocimiento, información y valor de ellos (principalmente la Inteligencia Artificial), están cambiando la forma en que se desarrollan las actividades productivas y sociales.

En este contexto la Ciencia de Datos, entendida como el conjunto principios, definiciones de problemas, algoritmos y procesos necesarios para extraer patrones no obvios y útiles a partir de grandes conjuntos de datos, se ha transformado en un campo disciplinar de gran atracción que se espera provea soluciones para las demandas del mundo actual y del venidero. Por tal motivo se hace necesario proveer una adecuada formación en los conceptos asociados a este campo disciplinar.

Esta diplomatura tiene como objetivo adentrar a los y las estudiantes en el mundo del análisis de datos a través de la práctica. Se busca proveer una visión global del proceso de la ciencia de datos y permitir a quienes cursan experimentar todos los pasos del proceso, formando un espíritu crítico y una capacidad de análisis de los datos basados en el contacto directo con el objeto de estudio desde el primer módulo. Además, se introducen herramientas y técnicas de un nivel básico, pero que son suficientes para la realización de una amplia gama de tareas del científico de datos. De esta manera, al finalizar la diplomatura, se contará con una visión global del proceso de la ciencia de datos, que permitirá una profundización ulterior.

3. Objetivos de la Diplomatura

- Proporcionar un marco formativo adecuado referido al objeto de estudio de la ciencia de datos.
- Capacitar para la resolución de problemas novedosos que se plantean en este campo.
- Generar recursos humanos capacitados para un área de fuerte desarrollo y relevancia socioeconómica.

4. Certificación que otorga la formación

Quienes aprueben las actividades previstas recibirán un Diploma en Ciencia de Datos.

Adicionalmente, en caso de contar con un título oficial de escuela secundaria y querer cursar la carrera de Licenciatura en Ciencia de Datos de la UNSAM, se les reconocerá la aprobación del Curso de Preparación Universitaria (CPU) y la asignatura Introducción a la Ciencia de Datos de la carrera.

5. Destinatarios

La Diplomatura en Ciencia de Datos está orientada a todas aquellas personas que deseen desarrollarse en la temática de Ciencia de Datos, o que poseen conocimientos focalizados en temáticas relacionadas con los Sistemas de Información y desean incorporar la visión y conocimientos técnicos específicos de la Ciencia de Datos.

6. Requisitos de ingreso

Para el ingreso a la Diplomatura, los aspirantes deberán presentar la documentación requerida por la normativa vigente.

7. Diseño y organización curricular

El plan de estudios de la Diplomatura se desarrolla a lo largo de 32 semanas (1 año) y consta de un total de 10 (diez) módulos, con una carga horaria semanal de 4 horas presenciales, 2 horas virtuales sincrónicas, y 2 horas asincrónicas. La carga horaria de la Diplomatura es de 192 horas sincrónicas, con una carga adicional de 64 horas de trabajo personal.

7.1. Estructura del Plan de Estudios

Módulo	Carga horaria
1. Introducción a los datos	12 hs (2 semanas)
2. Conceptos de programación	36 hs (6 semanas)
3. Visualización de datos	24 hs (4 semanas)
4. Limpieza y preparación de datos (<i>Data Engineering</i>)	18hs (3 semanas)
5. Conceptos de probabilidad y estadística	24 hs (4 semanas)
6. Análisis Exploratorio de datos (<i>Data Analytics</i>)	18 hs (3 semanas)
7. Ciclo de vida de un proyecto de datos (<i>Data Science</i>)	12 hs (2 semanas)
8. Introducción al modelado estadístico de datos	36 hs (6 semanas)
9. Elementos de Inferencia Causal	12 hs (2 semanas)

8. Modalidad de evaluación

Para la obtención del Diploma deberán aprobarse todas las asignaturas, que serán evaluadas de forma independiente en base a la realización de actividades prácticas y exámenes.

9. Contenidos analíticos de las asignaturas/módulos

Módulo 1. Introducción a los datos. Los datos. ¿Qué son los datos? Obtención de los datos. Conceptos de variable, valor, unidad y observación. Población y muestra. Representatividad y sesgos de una muestra de datos. Cuestiones de ética y privacidad. Tipos de datos: datos numéricos (enteros y de punto flotantes), cadenas de caracteres, fechas. Variables categóricas. Exploración de los datos usando herramientas en línea como Google Sheets. Tablas dinámicas.

Módulo 2. Conceptos de programación. Variables y entornos de programación. Estructuras de datos. Plataforma “R” para el análisis de datos. Estructuras de control

Módulo 3. Visualización de datos. Visualización de los datos. Gráficos de barras. Gráficos de barras apiladas. Gráficos de dispersión (scatter plots). Histogramas. Polígonos de frecuencia. Gráficos de densidad. Comparación de distribuciones. Gráficos de cajas (box plots) y de violines. Gráficos de densidad apiladas.

Módulo 4. Limpieza y preparación de datos. La necesidad de la preparación de los datos para el análisis. Evaluación de la calidad de los datos. Datos duplicados. Datos faltantes. Técnicas de imputación de datos faltantes. Compleción por valores medios y medianas segmentada y no-segmentada.

Modulo 5. Conceptos de probabilidad y estadística. Concepto de distribución de parámetros. Métricas descriptivas. Media, mediana y moda. Cuantiles. Cuartiles. Desvío estándar y distancia intercuartil. Sensibilidad a los outliers. Elementos de inferencia estadística. Elementos de pruebas estadísticas. Concepto de p-valor y significancia.

Módulo 6. Análisis exploratorio de los datos. Desarrollo de comprensión sobre los datos. El ciclo del análisis exploratorio. Generación de una pregunta. Búsqueda de respuestas a través de la visualización y el modelado. Revisión de la pregunta original.

Módulo 7. Ciclo de vida de un proyecto de datos. El modelo de proceso CRISP-DM (Cross-industry standard process for data mining). El modelo de proceso ASUM-DM (Analytics Solutions Unified Method for Data Mining/Predictive Analytics). Presentación de un problema o necesidad. Contextualización y formulación de la pregunta (business understanding). Obtención y preparación de los datos apropiados. Generación de una visión del resultado / herramienta / producto final. Visualización de los pasos necesarios para llegar a ese objetivo.

Módulo 8. Modelado de los datos. Concepto de modelo estadístico de los datos. Utilidad de los modelos como herramientas de inferencia y como técnica de resumen de los datos. Conceptos de parámetros e hiperparámetros. Variables predictoras y variable objetivo. Modelo lineal de regresión, simple y múltiple. Obtención de los parámetros por cuadrados mínimos. Evaluación del modelo ajustado. Coeficiente de determinación. Residuos. Dispersión en los residuos. qq-plots. Palanca y observaciones influyentes. Interpretación de los resultados de un modelo lineal. Error en los estimadores.

Módulo 9. Elementos de inferencia causal. Diferencia entre relaciones asociativas y relaciones causales. La importancia de la inferencia causal. La paradoja de Simpson. Variables confundidoras. Variables de control. Variables de control categóricas. Variables de control continuas. Modelos lineales con variables de control. Modelos con términos de interacción.