



Programa de capacitación

Ciencias de los Datos

Escuela de Ingeniería en Computación del Tecnológico de Costa Rica.

Contacto:

Para mayor información del programa ingrese al siguiente link:

<https://www.tec.ac.cr/fundatec/programa-ciencias-datos-escuela-computacion>

Coordinadora del Programa:

Máster Alicia Salazar Hernández, asalazar@itcr.ac.cr

Asistente Administrativa del Programa:

Licda. Sarela Gómez Brenes, sagomez@itcr.ac.cr

<p>Requisitos de Ingreso</p>	<ul style="list-style-type: none"> ▪ Inglés Técnico. ▪ Personas con algún tipo de grado en Estadística, Ingeniería en Computación, Ingeniería en Sistemas de Información, Matemáticas, Física. ▪ Es importante que la persona interesada cuente con amplia experiencia en programación de algoritmos. Debe contar con conocimientos en Python o en su defecto haber programado en algún otro lenguaje y que por cuenta propia pueda aprender Python. ▪ Se solicitan conocimientos básicos de Linux.
<p>Perfil de Ingreso</p>	<ul style="list-style-type: none"> ▪ Programa dirigido a personas egresadas o que se desarrollen en alguna de las siguientes áreas: <ul style="list-style-type: none"> ○ Ingeniería en Computación. ○ Ingeniería en Sistemas. ○ Matemática. ○ Ingeniería.
<p>Perfil de Salida</p>	<ul style="list-style-type: none"> ▪ Una persona especialista en ciencias de los datos, es aquella capaz de crear y utilizar algoritmos de aprendizaje automático, diseñar modelos, determinar el dominio aplicable de cada uno de ellos; de forma tal que pueda realizar análisis de datos y análisis cuantitativos, predicciones e interpretación.
<p>Sobre el programa</p>	<ul style="list-style-type: none"> ▪ Programa de Capacitación compuesto por 5 módulos, impartidos a lo largo de un año. ▪ Cada módulo se impartirá de forma bimestral ▪ En el Desarrollo de cada uno de los módulos, se utilizará como principal y de manera transversal, un lenguaje de alto uso en la industria, el cual se verá complementado con algunos otros lenguajes de prototipado. ▪ Entre las técnicas a desarrollar durante el desarrollo de los cinco módulos, se encuentran: <ul style="list-style-type: none"> ○ Técnicas de procesamiento de lenguaje natural ○ Técnicas de procesamiento de imágenes ○ Técnicas de procesamiento de sonido ○ Temas de minería de opinión ▪ Todos los módulos del programa, estarán desarrollados bajo una metodología basada en proyectos y con un alto contenido en visualización.
<p>Bibliografía</p>	<p>Todo el material a utilizar en el desarrollo del curso, ha sido preparado por profesores de La Escuela de Computación, y encontrará de forma digital, el cual puede ser accedido en todo momento por el estudiante.</p>

<p>Secuencia de Módulos</p>	<ol style="list-style-type: none"> 1. Matemática para ciencias de los datos 2. Aprendizaje automático 3. Estadística para ciencias de los datos 4. Big Data 5. Inteligencia de negocios y minería de datos <p>* Los horarios se van definiendo para cada uno de los grupos, se intenta mantener el mismo horario para todos, sin embargo, se depende de la disponibilidad de horarios de cada profesor y laboratorio.</p>
<p>Aprobación del Programa</p>	<p>Para obtener el certificado del Programa de Ciencias de los Datos, es necesario cumplir con la aprobación de los 5 módulos.</p>
<p>Inversión</p>	<p>El precio en modalidad virtual es de \$500 por curso + IVA, un total de \$2550 por todo el programa.</p> <p>En caso de ser presencial, el precio se ha establecido en un monto de \$600 + IVA por cada uno de los módulos, si realizan el pago completo del programa, obtienen un 10% de descuento.</p> <p>Los costos de graduación son definidos por la institución una vez que finalice el programa, actualmente ronda entre los 30.000 (treinta mil) a 40.000 (cuarenta mil colones).</p>

Módulos del Programa:

Módulo 1. Matemática para Ciencias de los Datos

Nombre del Curso	Matemática para Ciencia de los Datos
Descripción	El curso introducirá conceptos matemáticos fundamentales en las ciencias de los datos, como el cálculo multivariable, el álgebra lineal y el cálculo matricial, sirviendo como base para plantear modelos de clasificación y regresión de datos en cursos posteriores.
Tipo de Curso	Aprovechamiento Tipo teórico/práctico Para obtener el certificado correspondiente, es necesario tener una asistencia efectiva de más del 80% a las lecciones, y sus evaluaciones con un promedio mayor o igual a 70.
Cantidad de Horas Lectivas	32 Horas lectivas presenciales
Objetivos	<p>Objetivo General:</p> <ul style="list-style-type: none"> Al finalizar el curso, el estudiante contará con todas las bases matemáticas necesarias para ingresar al mundo de Ciencias de Datos. <p>Objetivos Específicos:</p> <p>Al finalizar el curso el estudiante será capaz de:</p> <ul style="list-style-type: none"> Modelar problema en términos matemáticos, y múltiples variables. Encontrar la solución óptima a un problema planteado según un modelo matemático, utilizando herramientas básicas del cálculo multivariable. Se abordarán clases magistrales por parte del profesor, como introducción a las actividades y conceptos que se desarrollan en cada sesión.
Metodología de la enseñanza	El curso utilizará una metodología de Aprender Haciendo, basado en el desarrollo de proyectos de forma tal que, mediante el desarrollo de casos de estudio y proyectos en el laboratorio, se pueda afianzar los conocimientos adquiridos durante el transcurso de las diferentes lecciones.

Contenidos del programa	<ol style="list-style-type: none"> 1. Álgebra lineal: <ol style="list-style-type: none"> a. Operaciones en vectores y matrices b. Sistemas lineales e independencia lineal. c. Autovectores y análisis de componentes principales. 2. Cálculo matricial <ol style="list-style-type: none"> a. Introducción al cálculo multivariable b. Funciones multivariable c. Derivadas parciales d. Integrales de superficie e. Introducción al cálculo matricial
--------------------------------	--

Módulo 2. Aprendizaje Automático

Nombre del Curso	Módulo 2. Aprendizaje Automático
Descripción	<p>El curso tiene como objetivo desarrollar las habilidades necesarias para la adecuada representación de problemas de reconocimiento de patrones, así como la implementación de soluciones concretas a problemas reales. El estudiante aprenderá sobre la elección e implementación de diferentes algoritmos de clasificación y regresión, populares en el aprendizaje automático.</p> <p>El curso pretende aportar las bases del reconocimiento de patrones y el aprendizaje automático, para además generar curiosidad en el estudiantado para para participar más adelante en proyectos de investigación científica, así como de valor práctico para la industria, en temas de alta complejidad.</p>
Tipo de Curso	<p>Aprovechamiento Tipo teórico/práctico</p> <p>Para obtener el certificado correspondiente, es necesario tener una asistencia efectiva de más del 80% a las lecciones, y sus evaluaciones con un promedio mayor o igual a 70.</p>
Cantidad de Horas Lectivas	32 Horas lectivas presenciales
Objetivos	<p>Objetivo General:</p> <ul style="list-style-type: none"> ▪ Al finalizar el curso el estudiante habrá adquirido las destrezas para construir un sistema de reconocimiento de patrones en general, a través de la experiencia del reconocimiento de patrones basado en imágenes <p>Objetivos Específicos: Al finalizar el curso el estudiante será capaz de:</p>

	<ul style="list-style-type: none"> ▪ Analizar los datos de entrada de un sistema de aprendizaje automático ▪ Diseñar las etapas de preprocesamiento y extracción de características relevantes, en un sistema de reconocimiento de patrones. ▪ Implementar un clasificador bajo un enfoque de aprendizaje automático, con base en las características extraídas de los datos de entrada del sistema. ▪ Evaluar la efectividad de un sistema de aprendizaje automático usando una métrica cuantitativa adecuada.
<p>Metodología de la enseñanza</p>	<p>Se abordarán clases magistrales por parte del profesor, como introducción a las actividades y conceptos que se desarrollan en cada sesión. El curso utilizará una metodología de Aprender Haciendo, de forma tal que, mediante el desarrollo de casos de estudio y proyectos en el laboratorio, se pueda afianzar los conocimientos adquiridos durante el transcurso de las diferentes lecciones.</p>
<p>Contenidos del programa</p>	<ol style="list-style-type: none"> 1. Introducción al Reconocimiento de Patrones <ol style="list-style-type: none"> a. Ejemplos y aplicaciones. b. Etapas de un sistema de reconocimiento de patrones. c. Tipos de aprendizaje. 1. Etapas de preprocesamiento Categorización y discretización de datos <ol style="list-style-type: none"> a. Normalización b. Eliminación de sesgos, redundancia y ruido 2. Etapas de extracción de características <ol style="list-style-type: none"> a. Descriptores e invariantes 3. Etapas de clasificación: Métodos de clasificación supervisada y no supervisada. <ol style="list-style-type: none"> a. Ajuste polinomial de curvas. <ol style="list-style-type: none"> i. Modelos paramétricos lineales de regresión: mínimos cuadrados y mínimos cuadrados regularizados. ii. Selección del modelo (sobreajuste) y validación cruzada. iii. La maldición de la dimensionalidad. b. Métodos supervisados. <ol style="list-style-type: none"> i. PCA y K-vecinos más cercanos. ii. Mínimos cuadrados. iii. Discriminante lineal de Fisher. iv. Perceptrón. v. Análisis de componentes principales (PCA, por sus siglas en inglés). vi. Redes neuronales de retropropagación y con entrenamiento de descenso de gradiente. vii. Redes convolucionales viii. Máquinas de soporte vectorial. c. Métodos no supervisados. <ol style="list-style-type: none"> i. Algoritmo BSAS

	<ul style="list-style-type: none"> ii. Algoritmo K-medias. iii. Algoritmos de aprendizaje competitivo. <p>4. Validación</p> <p>5. Evaluación de modelos</p> <ul style="list-style-type: none"> a. AUC ROC b. Precision / Recall c. Particiones de datos para entrenamiento y pruebas d. Retroalimentación y actualización de modelos
--	--

Módulo 3. Estadística para Ciencias de Datos

Nombre del Curso	Estadística para Ciencias de Datos
Descripción	El curso introducirá a los estudiantes conceptos de estadística descriptiva, pruebas de hipótesis y experimentación usando herramientas estadísticas. Los conceptos estudiados en el curso permitirán proponer preguntas en el ámbito de las ciencias de los datos, y responderlas con rigurosidad estadística.
Tipo de Curso	Aprovechamiento Tipo teórico/práctico Para obtener el certificado correspondiente, es necesario tener una asistencia efectiva de más del 80% a las lecciones, y sus evaluaciones con un promedio mayor o igual a 70.
Cantidad de Horas Lectivas	32 Horas lectivas presenciales
Objetivos	<p>Objetivo General: Al finalizar el curso, el estudiante será capaz de:</p> <ul style="list-style-type: none"> ▪ Utilizar herramientas estadísticas básicas para modelar, resolver e interpretar problemas en ámbitos de aplicación y utilizar herramientas computacionales para realizarlo. <p>Objetivos específicos:</p> <ul style="list-style-type: none"> ▪ Utilizar herramientas de muestreo, diseñar experimentos simples y recolectar datos de manera apropiada. ▪ Analizar datos de forma descriptiva e interpretar los resultados. ▪ Utilizar técnicas inferenciales adecuadas para cada problema específico ▪ Utilizar un lenguaje de programación de alto nivel para implementar modelos estadísticos y analizar los resultados. ▪ Implementar las herramientas estadísticas estudiadas en la resolución de problemas en distintas áreas

<p>Metodología de la enseñanza</p>	<p>Se abordarán clases magistrales como introducción a las actividades que se desarrollan en cada sesión. El curso utilizará una metodología de Aprender Haciendo, de forma tal que se pueda desarrollar el o los proyectos para ejemplificar el aprendizaje</p>
<p>Contenidos del programa</p>	<ol style="list-style-type: none"> 1. Análisis de datos 2. Estadística descriptiva 3. Correlación y causalidad 4. Distribuciones comunes y propiedades <ol style="list-style-type: none"> a. Normal b. Binomial c. Poisson d. Geométrica e. Exponencial f. Beta 5. Pruebas de hipótesis <ol style="list-style-type: none"> a. Abordaje frecuentista b. Intervalos c. p-values 6. Simulación 7. Generación de reportes <ol style="list-style-type: none"> a. Consultas a bajo nivel b. Business Intelligence c. Cubos d. Entities / Facts e. Híbridos 8. Experimentación <ol style="list-style-type: none"> a. Diseño de experimentos b. Poblaciones y muestreo c. A/B Testing d. Prueba de hipótesis en experimentos

Módulo 4. Big Data

Nombre del Curso	Big Data
Tipo de Curso	Aprovechamiento Tipo teórico/práctico Para obtener el certificado correspondiente, es necesario tener una asistencia efectiva de más del 80% a las lecciones, y sus evaluaciones con un promedio mayor o igual a 70.
Cantidad de Horas Lectivas	32 Horas lectivas presenciales
Objetivos	<p>Objetivo General:</p> <ul style="list-style-type: none"> Entender y aplicar técnicas de análisis de grandes cantidades de datos para la resolución de problemas concretos a través de tecnologías de manipulación, extracción y sintetización estadística. <p>Objetivos Específicos:</p> <ul style="list-style-type: none"> Aplicar bibliotecas para la transformación de datos a gran escala para poder sintetizar el conocimiento para futuro análisis. Aplicar técnicas de análisis de datos para extraer patrones que mejoren el entendimiento de un problema concreto. Aplicar técnicas para aprendizaje automatizado de patrones, basado en datos existentes, para mejorar la certeza de la solución aplicada a problemas concretos.
Metodología de la enseñanza	Se abordarán clases magistrales por parte del profesor, como introducción a las actividades y conceptos que se desarrollan en cada sesión. El curso utilizará una metodología de Aprender Haciendo, de forma tal que, mediante el desarrollo de casos de estudio y proyectos en el laboratorio, se pueda afianzar los conocimientos adquiridos durante el transcurso de las diferentes lecciones.
Contenidos del Programa	<ol style="list-style-type: none"> Fuentes y repositorios de datos Diferencias entre fuentes <ol style="list-style-type: none"> SQL vs NoSQL Archivos simples o distribuidos Schema vs No Schema Escalabilidad Procesamiento de fuentes (data frames) <ol style="list-style-type: none"> Uniones Agregaciones Transformaciones Filtros Procesamiento de atributos <ol style="list-style-type: none"> Selección por importancia y redundancia

	<ul style="list-style-type: none">b. Normalizaciónc. Discretización <p>5. Organización de datos procesados</p> <ul style="list-style-type: none">a. Tablas de estadísticasb. Data warehousingc. Vistas (dinámicas y materializadas)d. Escalabilidad <p>6. Análisis de datos</p> <ul style="list-style-type: none">a. Estadística descriptivab. Correlación y causalidadc. Simulaciónd. Generación de reportes<ul style="list-style-type: none">i. Consultas a bajo nivelii. Business Intelligenceiii. Cubosiv. Entities / Factsv. Híbridos <p>7. Uso de modelos de aprendizaje automático en big data</p> <ul style="list-style-type: none">a. Sin supervisión<ul style="list-style-type: none">i. Clusteringb. Supervisado<ul style="list-style-type: none">i. Regresión linealii. Árboles Aleatorios
--	--

Módulo 5. Minería de Datos e Inteligencia de Negocios

Nombre del Curso	Minería de Datos e Inteligencia de Negocios
Descripción	Este curso pretende estudiar las técnicas de minería de datos y de inteligencia de negocios que persiguen como objetivo esencial el descubrimiento automático del conocimiento contenido en la información almacenada de modo ordenado en conjuntos de datos que pueden tener dimensión elevada. Estas técnicas tienen por objeto descubrir patrones, perfiles, tendencias y otras relaciones presentes en la información, pero ocultas si no se trata adecuadamente. Cuando la dimensión de la información es muy alta, proviene de múltiples orígenes y no está estructurada.
Tipo de Curso	Aprovechamiento Tipo teórico/práctico Para obtener el certificado correspondiente, es necesario tener una asistencia efectiva de más del 80% a las lecciones, y sus evaluaciones con un promedio mayor o igual a 70.
Cantidad de Horas Lectivas	32 horas presenciales
Objetivos	<p>Objetivo General</p> <ul style="list-style-type: none"> Capacitar a los estudiantes para alcanzar una mentalidad crítica y analítica dentro de la empresa, mediante el conocimiento de los diferentes sistemas de información de empresa, los métodos y técnicas de análisis de datos, la formulación de preguntas e hipótesis, la obtención y la visualización de datos para conclusiones útiles en la toma de decisiones. <p>Objetivos Específicos</p> <ul style="list-style-type: none"> Diseñar el proceso de obtención y preparación de datos a alto nivel, en las diferentes fuentes, tipos de datos para métodos cuantitativos y cualitativos de análisis. Conocer las técnicas de análisis multidimensional (estructurado y no estructurado (durante todo el proceso de construcción y mantenimiento) de alto nivel (tableau, powerBI). Conocer técnicas, herramientas y algoritmos de análisis en minería de datos con el fin de adquirir capacidades superiores de comprensión de problemas, formulación de hipótesis, interrogación e interpretación de la información en procesos empresariales como: gestión económico-financiera, marketing y ventas y operaciones y logística. Modelar los procesos de negocio Formulación de preguntas de ciencias de los datos

	<ul style="list-style-type: none"> ▪ Construir informes y cuadros de mando para la toma de decisiones de los empleados y directivos con técnicas y herramientas de visualización de datos.
<p>Metodología de la enseñanza</p>	<p>Se abordarán clases magistrales por parte del profesor, como introducción a las actividades y conceptos que se desarrollan en cada sesión. El curso utilizará una metodología de Aprender Haciendo, de forma tal que mediante el desarrollo de casos de estudio y proyectos en el laboratorio, se pueda afianzar los conocimientos adquiridos durante el transcurso de las diferentes lecciones.</p>
<p>Contenidos del programa</p>	<ol style="list-style-type: none"> 1. Fundamentos de Inteligencia de negocio <ol style="list-style-type: none"> a. Concepto de inteligencia de negocio. b. Tipos de datos explotados. c. Recogida y preparación de datos. d. Metodologías (KDD, CRISP-DM) e. Casos de Estudio 2. Sistemas de inteligencia de negocio: Data Warehouse (DW) <ol style="list-style-type: none"> a. Diseño del almacén de datos b. Integración, limpieza y transformación del almacén de datos c. Explotación y administración de sistemas de DW 3. Sistemas de inteligencia de negocio: Data Lake <ol style="list-style-type: none"> a. Diseño y construcción del lago de datos b. Explotación y administración de sistemas estructurados y no estructurados 4. Análisis y minería de datos <ol style="list-style-type: none"> a. Minería de datos complejos (espacial, temporal, web mining) b. Modelos de clasificación c. Modelos de relación d. Análisis de textos 5. Visualización de datos <ol style="list-style-type: none"> a. Visualización (Gráficos, Mapas, Dashboard) b. Evaluación c. Conclusiones

Políticas de nuestro Programa:

CONGELAMIENTOS:

En caso de que algún estudiante decida congelar el monto de cualquiera de los módulos de los programas de Educación Continua que ofrece la Escuela de Ingeniería en Computación deberá, solicitar el formulario de congelamiento al correo: sagomez@itcr.ac.cr, asalazar@itcr.ac.cr, posteriormente remitirlo a estos dos correos y debe esperar al visto bueno del Coordinador General del Programa.

El congelamiento se hace solamente por **motivos de fuerza mayor**. Este congelamiento tiene una vigencia de un bimestre, es decir, debe de ingresar al siguiente grupo que inicie. En caso de no presentar su reingreso antes de este lapso perderá el monto cancelado, correspondiente al módulo congelado. En caso de reingreso y si hubiese un cambio de precio, se debe cancelar la diferencia en relación a la inversión del módulo cancelado y la inversión del módulo en que indique se incorpora.

DEVOLUCIONES:

Las devoluciones se realizarán debidamente justificadas (**por motivos de fuerza mayor**) y deberán ser aprobadas por la coordinación del programa. Dicha solicitud de devolución puede ser presentada **ÚNICAMENTE** los días anteriores al inicio de las lecciones del módulo correspondiente, si se presenta el primer día de clases o posterior, **NO PROCEDE** dicha devolución, sin excepción. Favor no comprometer al personal administrativo.

Para dicha aprobación el estudiante debe presentar una carta dirigida al Coordinador del Programa (Máster Alicia Salazar Hernández) justificando el motivo de la devolución de dinero, también deberá indicar; nombre, número de cédula, curso al que pertenece, monto que canceló y el número de cuenta para la transferencia en caso de que sea una empresa quien realice el depósito, deberá incluir los mismos datos y la firma del encargado de la inscripción o de quien realice el depósito con número de cédula jurídica, y razón social, en ambos casos **está sujeta a la aprobación de la Coordinación del Programa**, si la misma se aprueba tendrá 8 días para ser efectiva.

Independientemente al motivo de la solicitud de devolución, se procede con el cobro administrativo del 35% aplicado al monto cancelado, quedando disposición de la de la coordinación del programa.

Deberá remitir la carta debidamente firmada a los correos: correo:sagomez@itcr.ac.cr, asalazar@itcr.ac.cr.