

## Centro de Innovación Tecnológica del Instituto Tecnológico e Durango (ITD)

**Curso:** Machine Learning con R

**Instructor:** Rubén Pizarro Gurrola

<b>Nombre del curso</b>	<u>Programación R con R Studio</u>
<b>Objetivo</b>	Construir programas en lenguaje de programación R en entorno R Studio para analizar, interpretar y comunicar datos con enfoque estadístico y probabilístico presentados mediante documentos <i>markdown</i> y publicados en la nube con el servicio <i>RPubs</i> .
<b>Dirigido</b>	<ul style="list-style-type: none"><li>• Estudiantes: Reforzar aprendizajes de probabilidad y estadística</li><li>• Académicos: Desarrollar habilidades en el uso de herramientas estadística, probabilidad y lenguaje de programación como alternativa de enseñanza.</li><li>• Investigadores. Alternativa para difundir trabajos con análisis estadístico de datos.</li><li>• Empresa: Aprovechar las ventajas de utilizar R en el ámbito estadístico y probabilístico para el desarrollo de análisis de datos.</li><li>• A la comunidad. Personas interesadas en R y R Studio.</li></ul>
<b>Duración</b>	10 horas
<b>Requisitos</b>	<ul style="list-style-type: none"><li>• Laptop,</li><li>• Instalar R y R Studio,</li><li>• Conectividad a internet</li><li>• Conocimientos generales de programación y HTML</li></ul>

<b>Nombre del curso</b>	<u>Machine Learning – Aprendizaje Automático con R y R Studio</u>
<b>Objetivo</b>	Construir programas en lenguaje de programación R en entorno R Studio para analizar, soluciones tareas, interpretar y comunicar datos con enfoque basado en algoritmos de <i>machine learning</i> (aprendizaje automático) presentados mediante documentos <i>markdown</i> y publicados en la nube con el servicio <i>RPubs</i> .
<b>Dirigido</b>	<ul style="list-style-type: none"> <li>• Estudiantes: Reforzar aprendizajes de probabilidad y estadística y aprendizaje automático</li> <li>• Académicos: Desarrollar habilidades en el uso de herramientas para aplicar <i>machine learning</i> y lenguaje de programación como alternativa de enseñanza.</li> <li>• Investigadores. Alternativa para difundir trabajos con análisis estadístico haciendo uso de algoritmos de <i>machine learning</i>.</li> <li>• Empresa: Aprovechar las ventajas de utilizar R en el ámbito estadístico, probabilístico y aprendizaje automático para el desarrollo de análisis de datos.</li> <li>• Científicos de datos y comunidad en general. Personas interesadas en desarrollo de competencias haciendo uso de machine learning.</li> </ul>
<b>Duración</b>	30 horas
<b>Requisitos</b>	<ul style="list-style-type: none"> <li>• Laptop,</li> <li>• Instalar R y R Studio,</li> <li>• Conocimientos generales de probabilidad y estadística</li> <li>• Conocimientos de programación y HTML</li> <li>• Conectividad a internet</li> </ul>

# Machine Learning con R

## Temario:

Viernes 11 de marzo 2022 Programación y Entorno R	
Temas	Tiempo
<b>Configuración y puesta en marcha de R y R Studio</b> <ul style="list-style-type: none"><li>• Instalar R</li><li>• Instalar R Studio</li><li>• Entorno de trabajo</li><li>• Organización de carpetas</li><li>• Scripts</li><li>• Instalando paquetes</li><li>• R Markdown</li><li>• R Pubs</li><li>• Reconocimiento de github</li></ul>	1 hrs.
<b>Programación R</b> <ul style="list-style-type: none"><li>• Constantes y variables en R</li><li>• Expresiones</li><li>• Condicionales</li><li>• Ciclos</li><li>• Funciones de paquetes base en R<ul style="list-style-type: none"><li>○ Funciones matemáticas, char, varias ... ..</li></ul></li><li>• Creación de funciones</li><li>• Llamando funciones locales de la propia PC</li><li>• Llamando funciones externas desde WEB</li></ul>	1 hr

<p><b>Tipos de datos y estructuras en R</b></p> <ul style="list-style-type: none"> <li>• Datos numéricos, lógicos o booleanos</li> <li>• Datos char</li> <li>• Datos categóricos tipo factor</li> <li>• Vectores</li> <li>• Conjuntos de datos data.frame</li> <li>• Acceso a elementos de un vector y/o data.frame</li> </ul>	<p>1 hr.</p>
<p><b>Ejercicios ...</b></p> <p>Crear un data frame a partir de tres vectores que contienen datos de nombres, edades géneros [Femenino Masculino]. El data.frame debe contener 10 observaciones y las 3 variables.</p> <p>Importar data.frame y reconocer sus variables y estructura</p>	

<b>Sábado 12 de marzo 2022</b> <b>Estadística descriptiva</b>	
Temas	Tiempo
<b>Importar datos en R</b> <ul style="list-style-type: none"> <li>• Importar datos de un csv</li> <li>• Importar datos de un archivo excel</li> <li>• Importar datos de un archivo texto tabulado</li> <li>• Importar datos de una tabla html de una WEB</li> </ul>	1 hrs.
<b>Estadística descriptiva</b> <ul style="list-style-type: none"> <li>• Población y Muestra sample()</li> <li>• Medidas de tendencia central (media, mediana, cuartiles)</li> <li>• Medida de dispersión (varianza y desviación estándar)</li> <li>• Resumen de los datos</li> </ul>	1 hr
<b>Agrupación de datos</b> <ul style="list-style-type: none"> <li>• Frecuencia fdth()</li> <li>• Frecuencia relativa</li> <li>• Frecuencia porcentual</li> <li>• Frecuencia acumulada</li> <li>• Tablas de contingencia table()</li> </ul>	1 hr.
<b>Visualización de datos con plot() de paquetes base</b> <ul style="list-style-type: none"> <li>• Diagrama de puntos</li> <li>• Diagrama de dispersión</li> <li>• Diagrama de barra</li> <li>• Diagrama de barra</li> <li>• Histograma</li> <li>• Diagrama de caja o bigotes</li> <li>• Diagrama de línea y puntos</li> <li>• Diagrama con intersecciones</li> </ul>	1 hrs

<ul style="list-style-type: none"> <li>• Diagrama con títulos, subtítulos y etiquetas de los ejes</li> </ul>	
<p><b>Visualización de datos con ggplot del paquete ggplot2</b></p> <ul style="list-style-type: none"> <li>• Diagrama de puntos</li> <li>• Diagrama de dispersión</li> <li>• Diagrama de barra</li> <li>• Diagrama de barra</li> <li>• Histograma</li> <li>• Diagrama de caja o bigotes</li> <li>• Diagrama de línea y puntos</li> <li>• Diagrama con intersecciones</li> <li>• Diagrama con títulos, subtítulos y etiquetas de los ejes</li> </ul>	1 hrs
<p><b>Ejercicios ...</b> Cargar un conjunto de datos en un data.frame y realizar los siguientes diagramas:</p> <ul style="list-style-type: none"> <li>• Diagrama de dispersión de dos variables</li> <li>• Diagrama de barra de dos variables</li> <li>• Histograma de datos</li> <li>• Diagrama de caja o bigotes</li> </ul>	

Sábado 18 y 19 de marzo 2022	
Machine Learning Algoritmos supervisados Regresión	
Temas	Tiempo
Covarianza	1
Correlación	
<b>Regresión Lineal Simple</b> <ul style="list-style-type: none"> <li>• Cargar librerías</li> <li>• Cargar datos</li> <li>• Explorar datos y variables de interés</li> <li>• Limpiar datos</li> <li>• Diagrama de dispersión</li> <li>• Partir datos</li> <li>• Crear modelo con datos de entrenamiento</li> <li>• Estadísticos y evaluación del modelo</li> <li>• Predicciones con datos de validación</li> <li>• Predicciones con nuevos datos</li> <li>• Evaluación de las predicciones <ul style="list-style-type: none"> <li>○ R square</li> <li>○ rmse</li> </ul> </li> </ul>	1
<b>Regresión Polinomial de segundo nivel</b> <ul style="list-style-type: none"> <li>• Cargar librerías</li> <li>• Cargar datos</li> <li>• Explorar datos y variables de interés</li> <li>• Limpiar datos</li> <li>• Diagrama de dispersión</li> <li>• Partir datos</li> <li>• Crear modelo con datos de entrenamiento</li> <li>• Estadísticos y evaluación del modelo</li> <li>• Predicciones con datos de validación</li> <li>• Predicciones con nuevos datos</li> </ul>	1

<ul style="list-style-type: none"> <li>• Evaluación de las predicciones <ul style="list-style-type: none"> <li>○ R square</li> <li>○ rmse</li> </ul> </li> </ul>	
<b>Regresión Polinomial de tercero, cuarto, quinto nivel</b> <ul style="list-style-type: none"> <li>• Cargar librerías</li> <li>• Cargar datos</li> <li>• Explorar datos y variables de interés</li> <li>• Limpiar datos</li> <li>• Diagrama de dispersión</li> <li>• Partir datos</li> <li>• Crear modelo con datos de entrenamiento</li> <li>• Estadísticos y evaluación del modelo</li> <li>• Predicciones con datos de validación</li> <li>• Predicciones con nuevos datos</li> <li>• Evaluación de las predicciones <ul style="list-style-type: none"> <li>○ R square</li> <li>○ Rmse</li> </ul> </li> </ul>	1
<b>Regresión Lineal Multiple</b> <ul style="list-style-type: none"> <li>• Cargar librerías</li> <li>• Cargar datos</li> <li>• Explorar datos y variables de interés</li> <li>• Limpiar datos</li> <li>• Partir datos</li> <li>• Crear modelo con datos de entrenamiento</li> <li>• Estadísticos y evaluación del modelo</li> <li>• Predicciones con datos de validación</li> <li>• Predicciones con nuevos datos</li> <li>• Evaluación de las predicciones <ul style="list-style-type: none"> <li>○ R square ajustada</li> <li>○ Rmse</li> </ul> </li> </ul>	1



<b>Arboles de regresión</b> <ul style="list-style-type: none"> <li>• Cargar librerías</li> <li>• Cargar datos</li> <li>• Explorar datos y variables de interés</li> <li>• Limpiar datos</li> <li>• Partir datos</li> <li>• Crear modelo con datos de entrenamiento</li> <li>• Estadísticos y evaluación del modelo</li> <li>• Predicciones con datos de validación</li> <li>• Predicciones con nuevos datos</li> <li>• Evaluación de las predicciones <ul style="list-style-type: none"> <li>○ R square ajustada</li> <li>○ Rmse</li> </ul> </li> </ul>	1
<b>Regresión logística</b> <ul style="list-style-type: none"> <li>• Cargar librerías</li> <li>• Cargar datos</li> <li>• Explorar datos y variables de interés</li> <li>• Limpiar datos</li> <li>• Partir datos</li> <li>• Crear modelo con datos de entrenamiento</li> <li>• Estadísticos y evaluación del modelo</li> <li>• Predicciones con datos de validación</li> <li>• Predicciones con nuevos datos</li> <li>• Evaluación de las predicciones <ul style="list-style-type: none"> <li>○ R square ajustada</li> <li>○ Matriz de confusión</li> </ul> </li> </ul>	1
<b>Máquinas de soporte vectorial</b> <ul style="list-style-type: none"> <li>• Cargar librerías</li> <li>• Cargar datos</li> <li>• Explorar datos y variables de interés</li> </ul>	1

<ul style="list-style-type: none"> <li>• Limpiar datos</li> <li>• Partir datos</li> <li>• Crear modelo con datos de entrenamiento</li> <li>• Estadísticos y evaluación del modelo</li> <li>• Predicciones con datos de validación</li> <li>• Predicciones con nuevos datos</li> <li>• Evaluación de las predicciones <ul style="list-style-type: none"> <li>○ R square ajustada</li> <li>○ Rmse o matriz de confusión</li> </ul> </li> </ul>	
<p><b>Ejercicios ...</b></p> <p>Realizar y evaluar predicciones con un conjunto de datos de dos variables y comparar modelos de regresión lineal simple y polinómico</p> <p>Realizar y evaluar predicciones con un conjunto de datos de variables múltiples y comparar modelos de regresión lineal múltiple y árboles de regresión.</p> <p>Realizar y evaluar predicciones con un conjunto de datos de variables múltiples y comparar modelos de regresión logística y árboles de regresión y máquinas de soporte vectorial</p>	

Viernes y Sábado 25 y 26 de marzo 2022	
Machine Learning Algoritmos supervisados Clasificación	
Temas	Tiempo
Árboles de Clasificación	2
Bosque aleatorio	2
SVM Máquinas de Soporte Vectorial	2
Algoritmo Bayesiano	2

Viernes y Sábado 1 y 2 abril 2022	
Machine Learning Algoritmos NO supervisados	
Temas	Tiempo
K Means	2
Vecinos cercanos K Nearest	2
Jerárquico	2

Viernes y Sábado 1 y 2 abril 2022	
<b>BONUS</b> Aprendizaje profundo	
Temas	Tiempo
Deep learning	2 hrs