




Modelo de  
Mejora Continua

LÍNEA TECNOLÓGICA DEL PROGRAMA  
TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES GESTIÓN DE LA INFORMACIÓN  
RED TECNOLÓGICA  
TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES

DISEÑO DE ACCIONES DE FORMACION COMPLEMENTARIA

CÓDIGO:	DENOMINACIÓN DEL PROGRAMA
21710113	LIMPIEZA Y TRANSFORMACION DE DATOS CON PYTHON.
VERSIÓN: 1	SECTOR DEL PROGRAMA: SERVICIOS
Vigencia del Programa	Fecha inicio Programa: 19/12/2023 Fecha Fin Programa: El programa aún se encuentra vigente
DURACIÓN MÁXIMA	48 horas
JUSTIFICACIÓN:	<p>Los algoritmos tradicionales generalmente tienen elementos estáticos que a partir de los mismos datos de entrada ofrecen los mismos resultados de salida. La analítica de datos como parte de la inteligencia artificial por su parte implementa nuevas formas de trabajo en búsqueda de patrones que permiten descubrir en los datos características nuevas a medida que hay cambios en el tiempo adaptándose a diferentes situaciones. Lo anterior parecería magia, sin embargo, todo este tipo de procesos en los que algoritmos aprenden solo tienen sentido a partir de los datos de entrada. Los algoritmos por más evolucionados que estén no tienen la capacidad de intuición de los humanos y el éxito de estos depende exclusivamente de los datos que se les proporcionan. Es por esta razón que es muy importante seleccionar los datos correctos, limpiarlos y transformarlos de forma que los algoritmos puedan descubrir fácilmente los patrones.</p> <p>MinCIT, a través de iNNpulsa y en alianza con MinTIC, el SENA y Colciencias, liderarán el programa manufactura avanzada que busca desarrollar las tecnologías asociadas a las industrias 4.0 (entre las que se incluye la inteligencia artificial y analítica de datos) y que brindará servicios a empresas de todos los sectores.</p> <p>Para dar respuesta al reto anterior, el SENA propone el curso complementario en limpieza de datos que hace parte del módulo enfocado en la ciencia de datos (Big Data) que es una de las disciplinas de la inteligencia artificial que corresponde al proceso de buscar, recopilar, preparar y analizar los datos que permitan la obtención de conocimiento útil y necesario para determinar las necesidades de problemas complejos y ayudar en la toma de decisiones, en donde se trabajara con grandes volúmenes de datos que por su dificultad de análisis se requiere técnicas actuales como Python que ayuda a la extracción de la información requerida.</p>

 Modelo de Mejora Continua	LÍNEA TECNOLÓGICA DEL PROGRAMA TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES GESTIÓN DE LA INFORMACIÓN RED TECNOLÓGICA TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES
<b>REQUISITOS DE INGRESO:</b>	Conocimientos en: Manejo en la manipulación de grandes volúmenes de datos Big Data Manejo de hojas de calculo Manejo de lenguaje de programación en Python (variables, estructuras básicas) Nivel básico de inglés Recopilación de información de acuerdo a requerimientos técnicos Identificación de objetivos del objeto de análisis Identificación de fuentes de datos Académicos: Grado 9 aprobado Requisito adicional: Superar prueba de aptitud, motivación, interés y competencias mínimas de ingreso Requisito de ley: Edad mínima definida en la ley (14 años).
<b>ESTRATEGIA METODOLÓGICA:</b>	Centrada en la construcción de autonomía para garantizar la calidad de la formación en el marco de la formación por competencias, el aprendizaje por proyectos y el uso de técnicas didácticas activas que estimulan el pensamiento para la resolución de problemas simulados y reales; soportadas en el utilización de las tecnologías de la información y la comunicación, integradas, en ambientes abiertos y pluritecnológicos, que en todo caso recrean el contexto productivo y vinculan al aprendiz con la realidad cotidiana y el desarrollo de las competencias. Igualmente, debe estimular de manera permanente la autocrítica y la reflexión del aprendiz sobre el que hacer y los resultados de aprendizaje que logra a través de la vinculación activa de las cuatro fuentes de información para la construcción de conocimiento: -El instructor - Tutor -El entorno -Las TIC -El trabajo colaborativo

COMPETENCIA	
CÓDIGO:	DENOMINACIÓN
220501114	Sistematizar datos masivos de acuerdo con métodos de analítica y herramientas tecnológicas

ELEMENTO(S) DE LA COMPETENCIA
DENOMINACIÓN
Preparar el proceso
Tratar datos
Comprobar resultados

2. RESULTADOS DE APRENDIZAJE
DESCRIPCIÓN



Modelo de  
Mejora Continua

LÍNEA TECNOLÓGICA DEL PROGRAMA  
TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES GESTIÓN DE LA INFORMACIÓN  
RED TECNOLÓGICA  
TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES

RA2: PREPARAR LOS DATOS PARA EL ANÁLISIS DE ACUERDO CON LOS REQUERIMIENTOS TÉCNICOS.

RA1: SELECCIONAR LOS DATOS PARA EL ANÁLISIS DE ACUERDO CON LOS REQUERIMIENTOS TÉCNICOS

### 3. CONOCIMIENTOS

#### 3.1. CONOCIMIENTOS DE CONCEPTOS Y PRINCIPIOS

RA1:  
MANEJO DE DATOS: DEFINICIONES, CONNOTACIONES, DATA SET.  
DENOMINACIONES DE PREPARACIÓN DE DATOS: DATA WRANGLING, DATA PREPROCESSING, DATA PREPARATION, DATA STAGING.  
ESQUEMA DE PROCESOS EN: CRISP-DM, KDD.  
FUENTES DE INFORMACIÓN: PÚBLICAS Y PRIVADAS  
PASOS PARA LA PREPARACIÓN DE DATOS: SELECCIONAR, LIMPIAR, CONSTRUIR, INTEGRAR.  
SELECCIÓN DE DATOS: CALIDAD DE LOS DATOS, COMPLETITUD, CORRECCIONES, LIMITACIONES.  
TAMAÑO DEL DATA SET: PEQUEÑO, GRANDE, HERRAMIENTAS A UTILIZAR (EXCEL), DATOS A TRABAJAR, ESPACIOS DE TRABAJO, DATA STAGING.  
APLANAR ESQUEMA RELACIONAL: CONSULTAS, ATRIBUTOS, TABLAS, EXPORTAR DATOS.  
APLANAR MÚLTIPLES ESQUEMAS RELACIONALES: PREPARA LOS ESQUEMAS, CONSTRUIR CONSULTAS Y SCRIPTS, EXPORTAR LOS ARCHIVOS, CARGAR LOS DATA SET.

RA2:  
LIMPIEZA DE DATOS: DATOS FALTANTES, DATOS ERRÓNEOS, DATOS SUCIOS.  
DATOS SUCIOS: DATOS PERDIDOS, DATOS RUIDOSOS.  
DATOS RUIDOSOS: VALORES PERDIDOS Y VALORES RUIDOSOS.  
TRATAMIENTO DE VALORES FALTANTES: IMPUTACIÓN MEDIA, MEDIANA, MODA.  
DETECCIÓN Y CORRECCIÓN DE ERRORES: ANÁLISIS DE EXTREMOS, MÉTODO DE FACTOR, PRUEBA DE TUKEY.  
TRANSFORMACIÓN DE DATOS: FÓRMULAS MATEMÁTICAS, ATRIBUTOS GENERADOS, REGISTROS GENERADOS, DUMIFICACIÓN.


#### 3.2. CONOCIMIENTOS DE PROCESO

RA1:  
SELECCIONAR LAS FUENTES DE LOS DATOS PÚBLICAS O PRIVADAS.  
DETERMINAR EL TAMAÑO DE LOS DATOS GRANDE O PEQUEÑO.  
REALIZAR PRUEBAS Y ALMACENAMIENTO DE DATOS PARA SU PROCESAMIENTO.  
APLANAR LOS DATOS MEDIANTE LA EXTRACCIÓN DE DATOS Y CONSULTAS.  
CARGAR LOS DATOS PARA SU ANÁLISIS.

RA2:  
DEPURAR LOS DATOS FALTANTES, ERRÓNEOS, DATOS SUCIOS Y DATOS RUIDOSOS DE DATA SET.  
TRANSFORMAR LOS DATOS CREANDO NUEVOS ATRIBUTOS.  
INTEGRAR LOS DATOS A PARTIR DEL ANÁLISIS DE ELEMENTOS REDUNDANTES CON DETECCIÓN DE TUPLAS Y ESCALADO DE VARIABLES.  
ESCALAR VARIABLES TRANSFORMANDO VALORES CONFINADOS EN RANGOS DE VALORES.

### 4. CRITERIOS DE EVALUACIÓN

RA1:  
RECOPILA LOS DATOS DE FUENTES PÚBLICAS O DE LA MISMA ORGANIZACIÓN SEGÚN LOS CRITERIOS DEFINIDOS.

 <p>Modelo de Mejora Continua</p>	<p>LÍNEA TECNOLÓGICA DEL PROGRAMA</p> <p>TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES GESTIÓN DE LA INFORMACIÓN</p> <p>RED TECNOLÓGICA</p> <p>TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES</p>
---	--

DETERMINA EL ALMACENAMIENTO DE LOS DATOS UTILIZADO PARA EL PROCESAMIENTO. (DATAWAREHOUSE).  
EXTRAE ATRIBUTOS DE LAS TABLAS MEDIANTE CONSULTAS PARA APLANAR LOS DATOS DE ACUERDO CON EL ESQUEMA RELACIONAL.  
CARGA LOS DATOS EN LA HERRAMIENTA SELECCIONADA PARA EL ANÁLISIS MATRICIAL DE LA INFORMACIÓN DE ACUERDO CON LOS CRITERIOS DEFINIDOS.


RA2:  
DEPURA LOS DATOS FALTANTES, ERRÓNEOS, DATOS SUCIOS Y DATOS RUIDOSOS DE DATA SET DE ACUERDO CON LA INFORMACIÓN A PROCESAR.  
TRANSFORMA LA INFORMACIÓN DE LOS DATA SET DE ACUERDO CON LOS REQUERIMIENTOS ESTABLECIDOS.  
OBTIENE INFORMACIÓN DERIVADA A PARTIR DE LA INTEGRACIÓN DE DATOS EXISTENTES EN EL DATA SET.

#### 5. PERFIL TÉCNICO DEL INSTRUCTOR

Requisitos Académicos	TÍTULO PROFESIONAL EN EL NÚCLEO BÁSICO DE CONOCIMIENTO EN INGENIERÍA DE SISTEMAS, TELEMÁTICA Y AFINES. TARJETA PROFESIONAL EN LOS CASOS EXIGIDOS POR LA LEY. PREFERIBLEMENTE CON CERTIFICACIÓN EN O RELACIONADAS CON ANALÍTICA DE
Competencias mínimas	<p>GENERALES (PEDAGÓGICAS Y DIDÁCTICAS): FORMULAR, EJECUTAR Y EVALUAR PROYECTOS. TRABAJAR EN EQUIPO. APLICAR TÉCNICAS DIDÁCTICAS ACTIVAS EN LOS PROCESOS FORMATIVOS ORIENTAR FORMACIÓN SIGUIENDO MODELO PEDAGÓGICO SENA ESTABLECER PROCESOS COMUNICATIVOS ASERTIVOS. MANEJAR HERRAMIENTAS INFORMÁTICAS ASOCIADAS AL ÁREA OBJETO DE LA FORMACIÓN COMPETENCIAS ESPECÍFICAS (TÉCNICAS) CONOCIMIENTO EN DISEÑO Y CONSTRUCCIÓN DE BASES DE DATOS RELACIONALES Y NO RELACIONALES MANEJO DE HERRAMIENTAS DE ETL. MANEJO DE HOJAS DE CÁLCULO. CONOCIMIENTO EN DISEÑO Y CONSTRUCCIÓN DE SISTEMAS CON PYTHON</p>
Experiencia laboral y/o especialización	VEINTICUATRO (24) MESES DE EXPERIENCIA RELACIONADA DISTRIBUIDA ASÍ: DOCE (12) MESES DE EXPERIENCIA RELACIONADA CON EL EJERCICIO DEL DESARROLLO DE SOFTWARE

#### CONTROL DEL DOCUMENTO

	NOMBRE	CARGO	DEPENDENCIA / RED	FECHA
--	--------	-------	-------------------	-------

 <p>Modelo de Mejora Continua</p>	<p>LÍNEA TECNOLÓGICA DEL PROGRAMA</p> <p>TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES GESTIÓN DE LA INFORMACIÓN</p> <p>RED TECNOLÓGICA</p> <p>TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES</p>
---	--

Responsable del diseño	YULI SIDNEY GARCES BOLAÑOS	EQUIPO DE DISEÑO CURRICULAR	CENTRO DE COMERCIO Y SERVICIOS. REGIONAL CAUCA	06/03/2022
Responsable del diseño	ANDRES FELIPE VALENCIA PIMIENTA	APROBAR ANALISIS	DIRECCION GENERAL. DIRECCIÓN GENERAL	07/03/2022
Responsable del diseño	CLAUDIA PATRICIA ARISTIZABAL GUTIERREZ	ACTIVAR PROGRAMA	DIRECCION GENERAL. DIRECCIÓN GENERAL	19/12/2023
Aprobación	ANDRES AURELIO ALARCON TIQUE		null. REGIONAL CUNDINAMARCA	08/03/2022