

PREPARACIÓN AMBIENTE DE TRABAJO – INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL

Proyecto modelo predictivo

Este proyecto tiene como objetivo implementar un modelo de aprendizaje de maquina (machine learning) en un conjunto de datos con la finalidad de crear un algoritmo capaz de predecir una variable de interés del conjunto de datos seleccionado.

Para la realización de este proyecto los estudiantes tendrán dos actividades entregables en la plataforma TESS mediante documento-soporte. Durante el transcurso del curso los estudiantes participaran en 3 practicas orientadas al desarrollo de las actividades de este proyecto.

Clase	Actividad
#4 Deep learning	Primera practica modelos de aprendizaje.
#5 Computer Vision	Entrega de primera practica
#6 NLP	Segunda practica preparación de datos.
#8 Algoritmos IA	Tercera Practica modelos especializados.
#9 robótica	Entrega de segunda actividad
#12 Smart cities	Tercera practica Discusión de modelos

Actividad #1

Ingresar a las paginas de repositorios de datos y escoger un conjunto de información que le llame la atención.

- <https://www.kaggle.com/datasets>
- <https://www.datos.gov.co/browse?limitTo=charts>
- <https://archive.ics.uci.edu/datasets>

Descargar el conjunto de datos y mediante un documento de Word identificar los siguientes elementos del conjunto:

Título: HR Analytics Dataset

Resumen: El conjunto de datos contiene información detallada sobre empleados, programas de capacitación y métricas relacionadas con RR. HH. A continuación, se presenta un resumen:

- Datos de empleados: contiene detalles como nombres, fechas de inicio y salida, puestos de trabajo y supervisores.

- Métricas de desempeño y encuesta: incluye puntajes de compromiso, satisfacción y equilibrio entre vida laboral y personal.
- Información de capacitación: cubre los nombres de los programas, los tipos de capacitación, los resultados, las duraciones, los costos y los detalles del capacitador.
- Detalles de diversidad: incluye género, raza y estado civil.
- Estado y clasificación: indica el estado del empleado (activo/terminado), el tipo y los motivos de la terminación.

Origen: <https://www.kaggle.com/hopesb/hr-analytics-dataset?resource=download>

Número total de variables: 23 (sin tener en cuenta la variable de salida)

Numero de variables cuantitativas: 8

Numero de variables cualitativas: 15 (sin tener en cuenta la variable de salida)

Variable a predecir: Performance Score

Algoritmo de predicción: neural_network MLPClassifier

Actividad #2

1. Instalar librerías de IA en Python (Sklearn, keras)
2. Seleccionar modelo de aprendizaje.
3. Cargar los datos del conjunto seleccionado en la actividad #1.
4. Preparar las variables de entrenamiento.
5. Realizar proceso de normalización y entrenamiento del modelo.
6. Exportar el modelo de aprendizaje.
7. En un documento responder las siguientes preguntas:

- **Precisión de modelo:** 81.72% accuracy_score
- **Variables de entrenamiento:** Antigüedad (variable calculada usando StartDate), Title, BusinessUnit, EmployeeStatus, EmployeeType, PayZone, EmployeeClassificationType, DepartmentType, Division, State, GenderCode, RaceDesc, MaritalDesc, Satisfaction Score, Current Employee Rating, Engagement Score, Work-Life Balance Score, Training Program Name, Training Type, Training Outcome, Training Duration(Days), Training Cost, Age.
- **Variable predicha:** Performance Score
- **Hiper parámetros del modelo:** hidden_layer_sizes=(20,), activation="logistic", max_iter=500, verbose=True, tol=1e-10, random_state=1.

Parametros usados con valores por defecto:

- solver="adam",
- alpha= 0.0001,

- batch_size = "auto",
- learning_rate= "constant",
- learning_rate_init = 0.001,
- power_t= 0.5,
- shuffle= True,
- warm_start = False,
- momentum= 0.9,
- nesterovs_momentum= True,
- early_stopping= False,
- validation_fraction= 0.1,
- beta_1 = 0.9,
- beta_2 = 0.999,
- epsilon = 1e-8,
- n_iter_no_change = 10,
- max_fun = 15000