

|  |
| --- |
| Texto  Descripción generada automáticamente  Machine Learning  Tarea 01  Alumna: Sheila Leyva López.  e-mail: sheileyva29@gmail.com Introducción |
| Page 1 |

# Objetivo

Desarrollar habilidades en el manejo de datos, mediante el lenguaje de programación de python, para su aplicación en Machine Learning.

**Tratamiento de datos**

**Tarea 01**

# Fundamentación teórica

Es importante conocer conceptos básicos dentro de la inteligencia artificial y machine learning que permitan entender el contexto en el que se trabaja:

## Instancias en inteligencia artificial

En general, los sistemas tienen entradas y salidas.  Específicamente los conjuntos de datos entrada y salida pueden ser instancias, atributos y características. Si aterrizamos estos conceptos en una base de datos, las instancias serían las filas ‘y’ [1].

## Atributos en inteligencia artificial

Los atributos o registros, corresponden a una instancia particular ‘y’, en la matriz de la base de datos son las ‘x’ o columnas. Existen distintos tipos de atributos: numéricos, nominales, categóricos ordinales, entre otros [1].

## Estadística de atributos

Cuando se trabaja con datos es importante saber el tipo de operaciones, cálculos y medidas que nos permiten conocer la tendencia de su comportamiento. A continuación, se describen algunas medidas importantes [2][3]:

* Máximo

Corresponde al valor más grande contenido en un conjunto de datos.

* Mínimo

Corresponde al valor más pequeño contenido en un conjunto de datos.

* Desviación estándar

Se define como una medida de la dispersión de un conjunto de datos. A mayor desviación estándar, mayor dispersión de los datos. Su expresión matemática es la siguiente:

* Media

Es una medida de la tendencia central de los datos. Es la suma de los valores de los datos entre la cantidad de datos.

* Moda

Dentro de los atributos corresponde al valor que más se repite dentro de los atributos.

Tipo de distribución

El histograma permite conocer el tipo de distribución de los datos con los que se trabajará, existen distintos tipo de distribución [1]:

Uniforme

Graphical user interface, text, application, email

Description automatically generated