**Especialización en Inteligencia Artificial**

**Taller semana 6 de Machine Learning**

**Profesor: Juan Carlos Valencia**

**Junio 2025**

**¿Cuándo se aplican modelos de machine learning en problemas reales?**

Estos modelos se aplican cuando se desea detectar patrones ocultos o anomalías, automatizar y buscar algún tipo de predicción basados en la información suministrada. Es necesario tener las variables definidas para poder orientar el modelo hacía una predicción definida, que sin estas, estaríamos buscando algo que no tenemos idea de que es.

Nos puede ayudar a tomar decisiones en tiempo real, permitiendo la reacción rápida. Si tenemos datos históricos podemos hacer uso de machine learning para poner al modelo a aprender con toda esta data, de esta forma tiene entrenamiento robusto y poder predecir acciones futuras.

Con aprendizaje continuo, los modelos pueden mejorar automáticamente a medida que reciben más datos, por ejemplo, los agentes de chatbot.

**¿Cuáles son los sectores en donde se aplican los modelos de machine learning?**

Hay una gran variedad de sectores donde se puede aplicar, por no decir que en todas. Pero teniendo en cuenta las más relevantes, podemos mencionar el sector de la salud, haciendo diagnósticos automáticos, predicciones de enfermedades antes de etapas terminales. Otro sector bastante beneficiado son lo que trabajan con cadenas de producción, ya que se pueden mejorar sustancialmente las etapas por las cuales un producto pasa hasta estar listo; Detectar defectos de fábrica, errores en la cadena de suministro, controles de calidad.

Finanzas, comercio, Educación, transporte, son algunos de los campos de acciones donde se puede explotar de manera exponencial el machine learning, dando resultados rápidos y de fácil compresión.

**Ejercicio práctico 1**

Para el siguiente ejercicio vamos a tener en cuenta los parámetros que se muestran en la imagen para saber los valores de la data:

Interfaz de usuario gráfica, Tabla

El contenido generado por IA puede ser incorrecto.

Para este primer ejercicio vamos a utilizar un dataset que contiene los ingresos, nivel educativo y características generales de personas. Esta data tiene contiene los valores numéricos para su mejor manejo.

Pantalla de computadora

El contenido generado por IA puede ser incorrecto.

En la imagen anterior podemos ver que el **promedio de edad de las personas es de 35 años y que el promedio de ingresos es de 120.954 USD**. Al menos el 50% de los encuestados es universitario pero esta cifra puede llegar hasta el 75%. Algo curioso es que puede que el 25% o menos sean de gran ciudad y tengan un nivel educativo más alto.

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

En este caso vamos a utilizar 4 clusters siguiendo el patrón de “codo”.

Gráfico, Gráfico de dispersión

El contenido generado por IA puede ser incorrecto.

Esta es la segmentación que se realizó y podemos observar que evidentemente el nivel educativo marca una clara tendencia a tener mejores ingresos. Aunque hay algunos puntos que está con ingresos bastante buenos como lo es el cluster 2, podemos observar que el cluster 0 está ubicado dentro de la gráfica con ingresos altos y constantes teniendo un nivel educativo superior, mientras que el cluster 1, que por lo general tiene un nivel de educación menor tiene ingresos menores.

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

Con base a la siguiente gráfica podemos tener lo siguiente:

* En el cluster 0 tenemos un alto nivel educativo, es el que mayor promedio tiene.
* El cluster 1 es el que tiene el promedio más bajo de los 4 en casi todas las áreas.
* Cluster 2 con personas con un nivel educativo no tan alto, pero que prefieren vivir en ciudades más grandes. De hecho según la gráfica de los cluster, tiene ingresos elevados.
* El cluster 3 es el que tiene más promediado sus áreas pero sin mucha relevancia.

**Ejercicio práctico 2**

En este ejercicio tenemos información completa sobre las transacciones bancarias, con especial atención a la identificación de actividades fraudulentas.

En la siguiente gráfica vemos las proporciones entre las dos marcaciones de accidentes según las características en el historial.

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

Por la imagen que se presenta a continuación podemos decir que:

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

* 14.497 accidentes fueron detectados como reales cuando estos si lo fueron.
* 923 accidentes fraudulentos que fueron detectados como real.
* El modelo detecta todos los casos como Accidente, sin importar si realmente lo son o no