



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA ACADÉMICA
DIRECCIÓN DE EDUCACIÓN SUPERIOR



Carmona Bartolome Aldo Armando

1AVI

Introducción a Ciencia de Datos

**Práctica No. 3: Aprendizaje
automático**

Repo del codigo: [Link](#)



Introducción

El Machine learning, o aprendizaje automático, es una forma de la IA que permite a un sistema aprender de los datos en lugar de aprender mediante la programación explícita. Sin embargo, machine learning no es un proceso sencillo. Conforme el algoritmo ingiere datos de entrenamiento, es posible producir modelos más precisos basados en datos. Un modelo de machine learning es la salida de información que se genera cuando entrena su algoritmo de machine learning con datos. Después del entrenamiento, al proporcionar un modelo con una entrada, se le dará una salida. Por ejemplo, un algoritmo predictivo creará un modelo predictivo.

Existen 3 diferentes tipos de algoritmos para la clasificación autónoma de objetos o datos:

- Clasificación supervisada: disponemos de un conjunto de datos. Construimos un modelo en la fase de entrenamiento (training) utilizando dichas etiquetas, que nos dicen si una imagen está clasificada correcta o incorrectamente por el modelo. Una vez construido el modelo podemos utilizarlo para clasificar nuevos datos que, en esta fase, ya no necesitan etiqueta para su clasificación, aunque sí la necesitan para evaluar el porcentaje de objetos bien clasificados.
- Clasificación no supervisada: los datos no tienen etiquetas (o no queremos utilizarlas) y estos se clasifican a partir de su estructura interna (propiedades, características).
- Clasificación semisupervisada: algunos datos de entrenamiento tienen etiquetas, pero no todos. Este último caso es muy típico en clasificación de imágenes, donde es habitual disponer de muchas imágenes mayormente no etiquetadas.

El siguiente programa utiliza un CSV (data1.csv) para la obtención de datos previamente instanciados en el archivo acerca de 50 ofertas de trabajo para los Científicos de datos, en este caso los datos incluidos fueron el salario y los años de experiencia necesarios para ser contratado. El programa utiliza pandas para el manejo de los datos, puesto que esto puede dar una mayor organización si se tienen muchos datos en un solo archivo y por último se usa matplotlib.pyplot para el ploteo de las graficas para demostrar las predicciones que el algoritmo usado "K-means" arroja.



Desarrollo

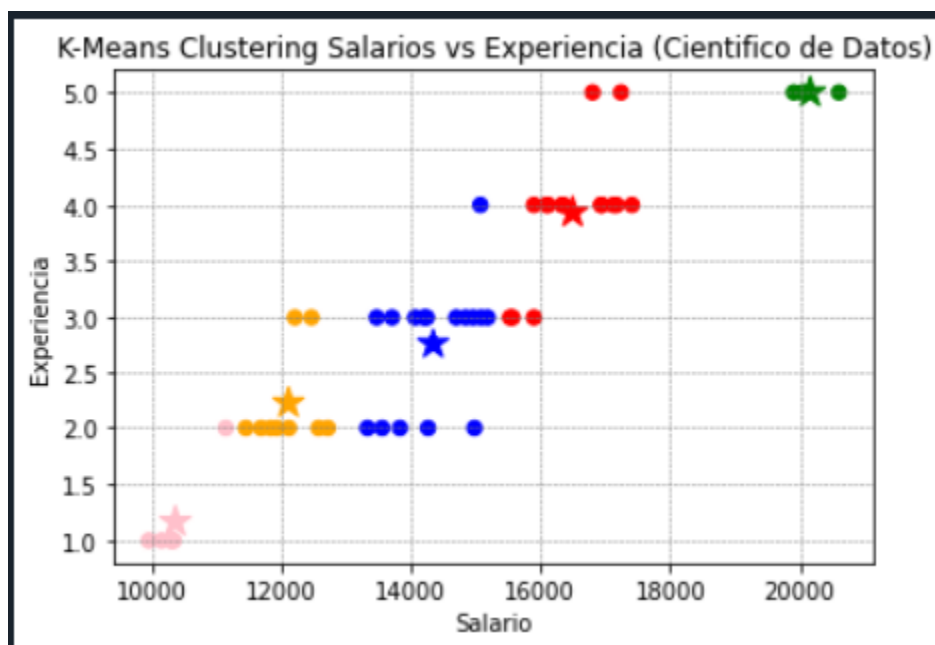
Ejecución

Primeramente, tomemos en cuenta los datos de entrada para el calculo de las estadísticas solicitadas para los salarios de los Científicos de Datos mostrando los datasets referentes a estos. Tomaremos como valor independiente (X) los salarios de estas ofertas y como valor dependiente (Y) tomaremos la cantidad de años que requiere el empleado para obtener el trabajo.

X = Salarios	Y = Experiencia
<pre>27 15073 28 14846 29 12579 30 11840 31 20031 32 11149 33 16944 34 15532 35 15898 36 12469 37 20600 38 13475 39 12214 40 13712 41 14247 42 20092 43 17241 44 16803 45 16350 46 17410 47 15899 48 16118 49 17174 Name: salario, dtype: int64</pre>	<pre>27 4 28 3 29 2 30 2 31 5 32 2 33 4 34 3 35 3 36 3 37 5 38 3 39 3 40 3 41 3 42 5 43 5 44 5 45 4 46 4 47 4 48 4 49 4 Name: experiencia, dtype: int64</pre>



Finalmente se muestra la grafica generada con ayuda del algoritmo seleccionado



Conclusión

Como se puede observar en la grafica se puede deducir de forma evidente que dependiendo del salario que esta dispuesto a pagar el empleador, la cantidad de años de experiencia necesarios va aumentar o disminuir tomando como regla que entre mayor sea la cantidad dispuesta a pagar por el trabajo, mayor será la cantidad de años requerida, siendo esta cantidad menor a la cantidad promedio pagada por 3 años de experiencia. Con esto se puede inferir que ser bien pagado (ms beneficio para el empleado) ronda entre los primeros 3 años de experiencia ya que después de estos la suma que se pretende pagar es “justa” de acuerdo a tus habilidades en el campo.

El aprendizaje automático es de suma vitalidad para el futuro de la industria y en general de la humanidad ya que con ayuda de datos previamente obtenidos podemos tener una pequeña “vista hacia el futuro”, esto siempre tomando en cuenta que métodos se usaron para predecir los próximos datos y por supuesto la integridad y cantidad de los datos que se están tomando como referencia.



Referencias de Apoyo

IBM. (s. f.). *Machine Learning*. machine learning. Recuperado 9 de noviembre de 2021, de

<https://www.ibm.com/mx-es/analytics/machine-learning>

kmeans. (s. f.). K-means. Recuperado 9 de noviembre de 2021, de

https://www.unioviedo.es/compnum/laboratorios_py/kmeans/kmeans.html