

Informe Proyecto 3 Inteligencia Artificial

Integrantes: Juan Jose Farias
Edinson Troncoso
Cristobal Trujillo

Introducción

El análisis simultáneo de información, en conjunto con el procesamiento estadístico suele ser una de las tareas más importantes a nivel global. En el transcurso de las últimas décadas se ha encontrado la solución a esta ardua tarea de aprendizaje y predicción en una de las cuantas disciplinas procedentes de la *inteligencia artificial*; esta materia o herramienta informática es el ***machine learning***

El machine learning busca mejorar el análisis de datos, en pro de predicción futura, ya sea por la implantación de nuevos sistemas o simplemente el mejoramiento de los ya existentes, mediante el uso de algoritmos basados en información antigua o reciente que permita el funcionamiento óptimo del sistema a trabajar

En este proyecto vamos a usar el dataset “Car Evaluation” del repositorio “UCI Machine Learning”, con el fin de predecir la aceptabilidad del automóvil utilizando una serie de características de este.

Caracterización del dataset

Primero que todo profesora, mencionar que usamos una base de dato que no cumpla con la característica de tener mas de 10 atributos, debido a que no encontramos en la plataforma UCI una base de datos adecuada para cumplir con el objetivo de lograr una clasificación adecuada y en las que probamos tuvimos distintos problemas al preparar para el entrenamiento, es por esto y además del

tiempo, que usamos una que cumpliera similares características, pero que logramos adecuar para usarla en el entrenamiento los algoritmos y con ella pudimos capturar de buena forma el uso de los algoritmos y lo que se esperaba aprender con este proyecto.

- Nombre: Car evaluation
- Número de atributos: 6
 1. Precio: Muy alto, alto, medio, bajo.
 2. Mantenimiento: Muy alto, alto, medio, bajo.
 3. Puertas: 2, 3, 4, 5 o más.
 4. Capacidad (personas): 2, 4 o más.
 5. Maletero: pequeño, mediano, grande
 6. Seguridad: baja, media, alta
- Número de instancias: 1728
- Número de clases: 4
 1. Inaccesible.
 2. Accesible.
 3. Bueno.
 4. Muy bueno.
- Problema a resolver: El modelo evalúa los automóviles de acuerdo al precio, de compra, precio de mantenimiento, características técnicas, comodidad, número de puertas, tamaño del maletero y la seguridad.

Algoritmos seleccionados

Decision Tree: Es un método de aproximación de una función objetivo de valores discretos, en el cual la función objetivo es representada mediante un árbol de decisión.

Regresión lógica: Es una técnica de análisis de datos que predice el valor de datos desconocidos mediante el uso de otro valor de datos relacionado y conocido.

Naive Bayes: El clasificador asume que el efecto de una característica particular de una clase es independiente de otras características. Por ejemplo, un solicitante de préstamo es deseable o no dependiendo de sus ingresos, historial de préstamos y transacciones anteriores, edad y ubicación. Incluso si estas características son interdependientes, estas características se consideran de forma independiente.

Redes neuronales: Una red neuronal es un método de la inteligencia artificial que enseña a las computadoras a procesar datos de una manera que está inspirada en la forma en que lo hace el cerebro humano.

Breve descripción del programa

En el programa que hacemos es leer el archivo y obtener los datos y arreglarlos de manera que podamos utilizarlos en cada tipo de clasificación, transformamos las columnas a datos numéricos para que no se presenten errores al momento de usar los datos, luego de ya tener definidos los datos, definiremos y separemos los datos que usaremos para entrenar y para testear los algoritmos.

El primer algoritmo a describir será el de **regresión lógica** en donde primero creamos el objeto **LogisticRegression** para luego cargar en él los datos que sería X_train y Y_train, luego realizamos la predicción e imprimimos el accuracy obtenido al entrenar el algoritmo.

Para resumir en los demás algoritmos se realiza lo mismo debido a que necesitan los mismos parámetros, tan solo cambiando el tipo de objeto, ya sea

MultinomialNB, **DecisionTreeClassifier**, **MLPClassifier** e imprimiendo en pantalla el accuracy de cada algoritmo.

Performance

Accuracy Logistic Regression: 0.7090558766859345

Accuracy Naive Bayes: 0.6994219653179191

Accuracy Red Neuronal: 0.7495183044315993

Accuracy Decission Tree: 0.720616570327553

Ya teniendo el accuracy de todos los métodos utilizados podemos constatar, que existe una pequeña variación entre los métodos, lo que podemos inferir de estos es que hay métodos más eficientes para realizar la predicción para esta base de datos que otros, por lo que el mejor método en este caso sería la Red Neuronal seguido de el Árbol de Decisión, luego Regresión Lógica y por último Naive Bayes, ya obtenido los datos podemos concluir que los métodos tiene un porcentaje de acierto de un 70% aproximadamente.

Conclusiones

Actualmente la inteligencia artificial está tomando mucha fuerza como forma predilecta para realizar muchas aplicaciones en el área de la informática y la computación, por lo cual nos parece de mucha importancia aprender y saber usarlo. En este proyecto pudimos observar la aplicación de entrenar y predecir a través de una base de datos un “característica” definida para la clasificación de los objetos, junto a esto pudimos diferenciar el cómo se usa y cómo funcionan los distintos algoritmos usados y ver la eficiencia de cada uno de ellos, para concluir que existen mejores para distintos tipos de casos a estudiar.