

Car Evaluation using AI

AUTORES	AFILIACIONES	REFERENCIAS
Julio Gudiel 18001201 Jorge Cuevas 22000138	Universidad Galileo FISICC - AN Seminario Profesional 1	UCI Machine Learning Repository. (n.d.). Archive.ics.uci.edu. https://archive.ics.uci.edu/dataset/19/car+evaluation



INTRODUCCIÓN

En Colab se presenta el desarrollo e implemetación completa del proyecto de Inteligencia Artificial asignado en el curso de Seminario Profesional I, el objetivo del proyecto fue aplicar conocimientos de IA que fueron aprendidos en el curso para poder resolver el problema de clasificación asignado, utilizando el dataset "Car Evaluation" de **UCI Machine Learning Repository** y en base a los resultados obtenidos se realizó este póster científico.

DESCRIPCIÓN DEL DATASET

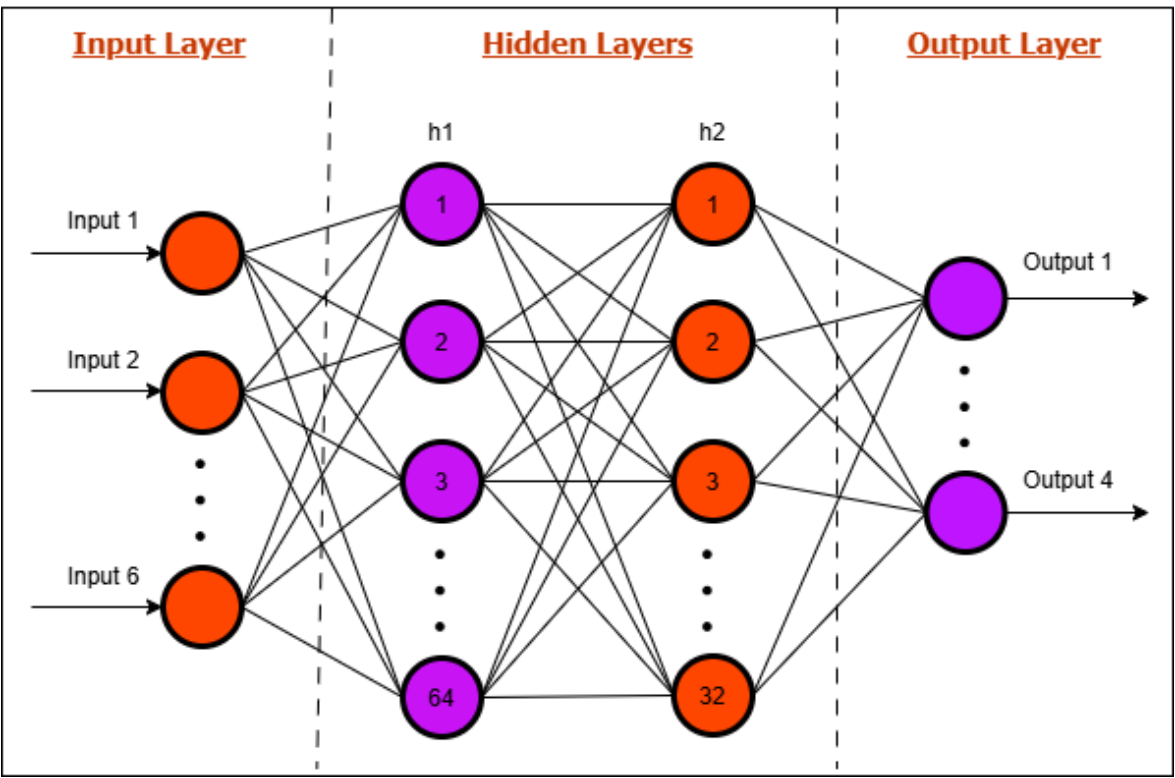
El dataset "Car Evaluation" contiene 1728 instancias con 6 atributos categóricos (precio de compra, costo de manteminiendo, número de puertas, capacidad de personas, tamaño del maletero y seguridad) y una variable objetivo con 4 clases (inaceptable, aceptable, bueno, muy bueno). Entonces, el fin del modelo que se desarrolló es poder clasificar correctamente el nivel de aceptabilidad de un automovil, también manejando el desbalanceo de clases que existe, ya que el 70% de las instancias se encuentran en la clase 'Inaceptable'.

DESCRIPCIÓN DEL MODELO

El modelo utilizado es una Red Neuronal Pre-alimentada o Feedforward (FNN) que utiliza el API de Keras de TensorFlow.

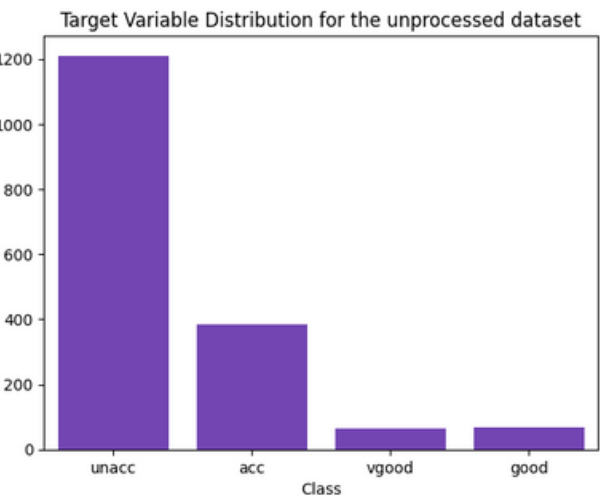
Se utilizó la función de activación **ReLU** para las *Hidden Layers* y **Softmax** para la *Output Layer*.

Para la compilación del modelo se utilizó: **Adam**, un algoritmo popular de optimización. La función *Loss* utilizada fue **SCCE** y la métrica '**Accuracy**' para evaluar el rendimiento. El modelo se entrenó en 10 iteraciones o *Epochs* en muestras de 32 instancias.

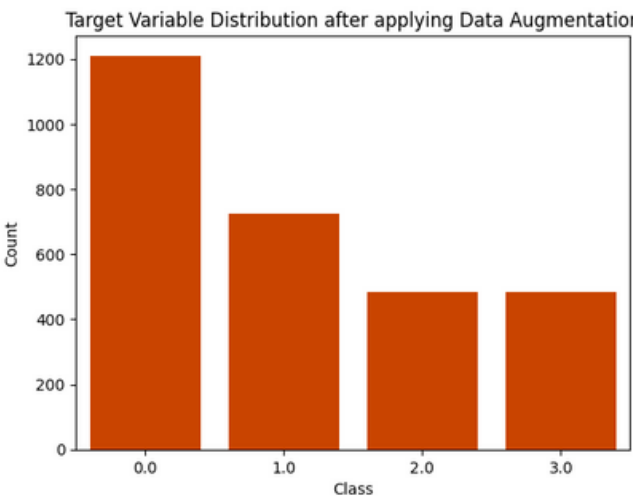


ANÁLISIS VISUAL

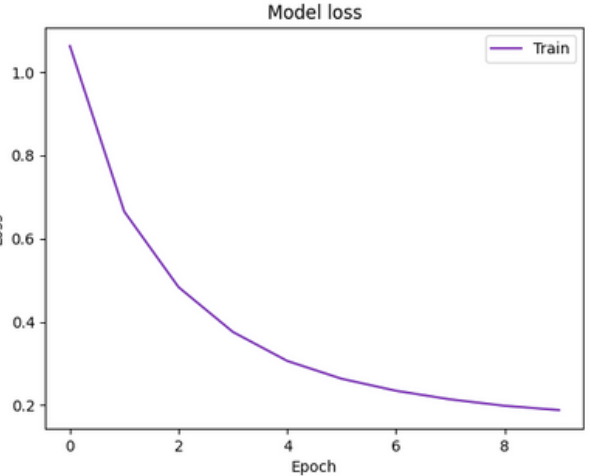
Antes de entrenar el modelo fue necesario hacer un pre-procesamiento del dataset donde se aplicó Oversampling, una técnica de Data Augmentation para lidar con el desbalanceo que existía en la clasificación de la data.



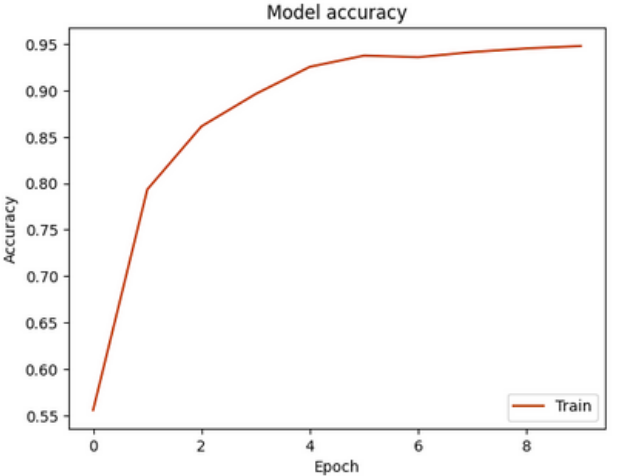
Distribución de la variable explicada antes del Pre-Procesamiento



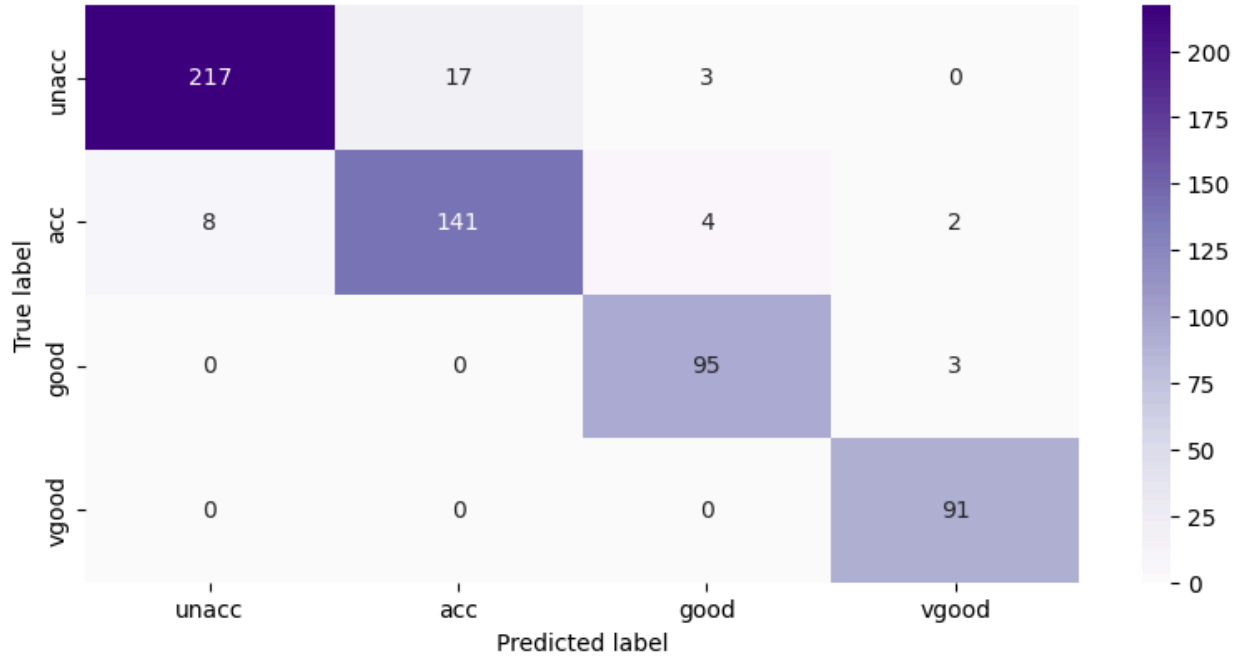
Distribución de la variable explicada después de aplicar Ordinal Encoding y Oversampling



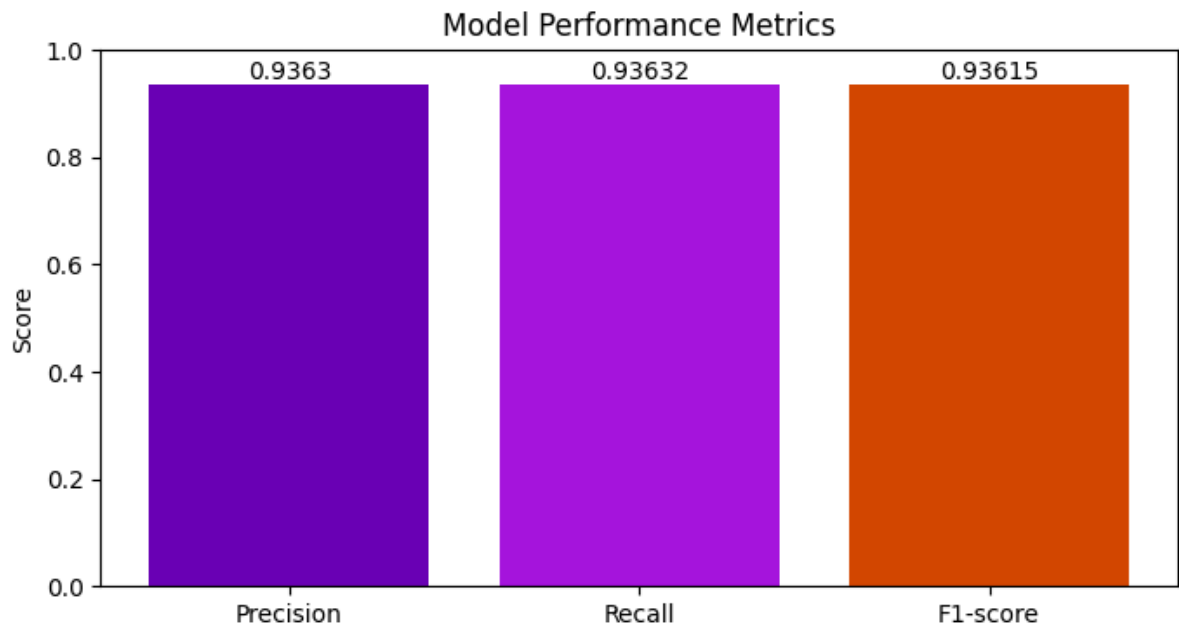
Gráfica de la función Loss durante el entrenamiento



Gráfica de Precisión durante el entrenamiento



Matriz de Confusión de la etapa de evaluación del modelo.



Rendimiento de métricas del modelo

RESULTADOS

El Objetivo del entrenamiento fue minimizar la función Loss y maximizar la métrica de Accuracy. Como se puede observar en las gráficas, tanto la función Loss como el Accuracy muestran una disminución y aumento constante respectivamente. Esto Indica que el entrenamiento fue exitoso.

Durante la fase de prueba, el modelo logró un Accuracy del 93%. Esto significa que el modelo clasificó correctamente el 93% de las muestras en la etapa de prueba.

MEJORAS A FUTURO

Evaluar otros modelos: Comparar con otros algoritmos como *Random Forest*, *SVM* o *XGBoost* para identifcar mejoras en el rendimiento. Ajuste de hiperparametro: Utilizar herramientas como *Keras Tuner* o *GridSearchCV* para optimizar la arquitectura del modelo. Validacion cruzada: implementas cross-validation para obtener una mejor evaluacion. Poder explicar mejor el modelo: Aplicar herramientas como *SHAP* o *LIME* para interpretar el impacto de las variables en las predicciones.

CONCLUSIONES

En el proyecto se logró implementar un sistema efectivo de clasificación de vehículos utilizando técnicas de inteligencia artificial, basado en el conjunto de datos "Car Evaluation" de UCI. Se realizaron procesos adecuados de preprocesamiento, como codificación ordinal y balanceo de clases mediante *SMOTE* y sub muestreo aleatorio, que permitió mejorar la distribución de los datos. Luego se entrenó la red neuronal con *TensorFlow*, obteniendo métricas satisfactorias como Precision, Recall y F1-score. Las visualizaciones empleadas facilitaron el análisis de los resultados y la comprensión del comportamiento del modelo. Se da por concluir que la solución propuesta es apropiada para problemas de clasificación multiclase con variables categóricas.