1. Regresión

1.1. Descripción del conjunto de datos: Datos personales de costos médicos, datos de dominio publico tomados de repositorio de datos Kaggle (https://www.kaggle.com/samirkaggle2000/a-regresion-lineal-predictivo-eduar-samir-01/data?select=datos_personales_de_costos_mdicos.csv). Se pretende predecir el costo medico por beneficiario.

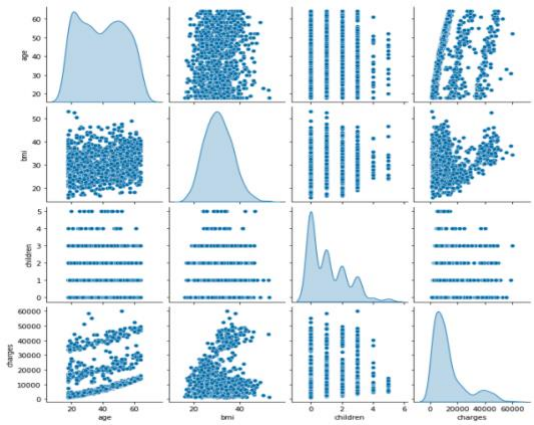
1.2. Caracterización del conjunto de datos

Tabla 1. Muestra tipos de datos originales y transformados a partir del Dataset.

Atributo	Descripción	Tipo original	Tipo conversión	Instancias	Mean	Std	Min	25%	50%	75%	Max	Val Original	Val Nuevo
age	Edad del beneficiario principal	cuantitativa (discreta)	cuantitativa (discreta)	1070	39,04	14,14	18,00	26,00	39,000	51,00	64,00		
bmi	Indice de masa corporal (kg/m^2) relación altura-peso. Idealmente 18,5 a 24,9	cuantitativa (continua)	cuantitativa (continua)	1070	30,74	6,06	15,96	26,32	30,50	34,80	53,13		
children	Número de dependientes	cuantitativa (discreta)	cuantitativa (discreta)	1070	10,93	1,21	0,0	0,0	1,000	2.0	5,00		
female	Genero del Contratista de seguros femenino, masculino	cualitativa (nominal)	cuantitativa (continua)	1070	0,502	0,50	0,0	0,0	1,000	1.0	1,00	'female'	1.0
male	Genero del Contratista de seguros femenino, masculino	cualitativa (nominal)	cuantitativa (continua)	1070	0,498	0,50	0,0	0,0	0,00	1.0	1,00	'male'	
yes	Indica si es o no fumador	cualitativa (nominal)	cuantitativa (continua)	1070	0,199	0,40	0,0	0,0	0,00	0.0	1,00	'yes'	
no	Indica si es o no fumador	cualitativa (nominal)	cuantitativa (continua)	1070	0,801	0,40	0,0	1,0	1,00	1.0	1,00	'no'	
southwest	Area residencial del beneficiario en EE.UU	cualitativa (nominal)	cuantitativa (continua)	1070	0,235	0,42	0,0	0,0	0,00	0.0	1,00	'southwest'	
southeast	Area residencial del beneficiario en EE.UU	cualitativa (nominal)	cuantitativa (continua)	1070	0,281	0,45	0,0	0,0	0,00	1.0	1,00	'southeast'	
northwest	Area residencial del beneficiario en EE.UU	cualitativa (nominal)	cuantitativa (continua)	1070	0,236	0,42	0,0	0,0	0,00	0.0	1,00	'northwest'	
northeast	Area residencial del beneficiario en EE.UU	cualitativa (nominal)	cuantitativa (continua)	1070	0,249	0,43	0,0	0,0	0,00	0.0	1,00	'northeast'	
charges	costos médicos individuales. variable de estudio	cuantitativa (continua)	cuantitativa (continua)	1070	13056,550	11994,26	1121,9	4566,0	9289,08	15826,11	60021,39		

- **Número de instancias en total**: 1338.
- **Número de atributos**: 7 incluyendo la clase originalmente, 12 luego de conversión de variables categóricas.

Imagen 1. Distribución de los datos en cada atributo.



La **Imagen 1** muestra la distribución de los datos en cada una de las variables, también muestra la correlación entre los atributos de entrada y el atributo de salida **charges,** la variable **bmi** muestra una distribución de los datos simétrica tipo gaussiana, la variable **age** también muestra cierta simetría en la distribución y la variable **children** es mayormente asimétrica que las dos anteriores. La correlación que existe entre los atributos de entrada se puede observar que **bmi**, **age** y **children**, se correlacionan de mayor a menor con la variable objetivo **charges** en el mismo orden que se mencionan.

1.3. Parámetros relevantes

Tabla 2. Muestra parámetros relevantes, su valor y descripción.

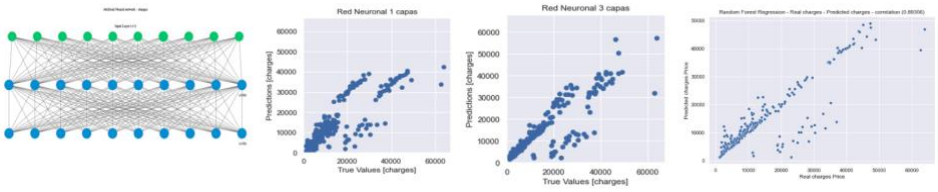
Redes neuronales (1 y 3 capas)		
Parámetro	**Valor**	**Descripción**
Capas (Modelo 1)	1	El modelo se implemento inicialmente con una sola capa y con la función de activación "relu" (Rectified Lineal Unit), la cual nos sirve para entrenar un modelo de regresión, el problema es que cuando una de las iteraciones llega a cero, el modelo deja de aprender.
Capas (Modelo 2)	6	Se crea un modelo de 3 capas, 1 de entrada con 11 atributos, 1 ocultas con función de activación relu y una capa de salia
Neuronas Modelo 2	64,64,1	64 neuronas capa de entrada, capas ocultas y 1 neurona para la capa de salida
Input_shape	11	11 atributos de entrada
Funcion activación	Relu, Linear	La función ReLU transforma los valores introducidos anulando los valores negativos y dejando los positivos tal y como entran.
Training Data	80%	1070 instancias, disminuye el costo de procesamiento
Test data	20%	268, conjunto de datos separados para validar el rendimiento del modelo.
Overfiting	0	Ocurre cuando nuestro modelo aprende demasiado bien nuestros datos de entrenamiento y como resultado sufre al generalizar.
Optimizador.RMSprop(0.001)	0.001	El gradiente estocastico es donde usando derivados, actualizamos todos los pesos en un subconjunto de los datos y brinda muchos beneficios de rendimiento.
mean_squared_error		funcion de costo que definimos en la serie neuronal
epochs	500	Número de veces que se ciclan los datos
verbose	0	muestra a mayor o menor detalle el resultado de entrenamiento del modelo.
validation_split	0.2	muestra aleatoria de datos de entrenamiento, para el caso es de 20% usado para validar la precisión de la red neuronal.
Batch size		Durante cada iteración, el modelo calcula la salida pronosticada para 32 EJ batch size puntos de datos, calcula el costo de cada uno de ellos, promedia este valor y usa derivados para actualizar todos los pesos.
loss		Perdida del conjunto de datos de validacion = 10%
loss_values		Perdida de datos de entrenamiento
R2 (R^2) SCORE		Medida estadística que nos dice que tan cerca están los datos del modelo de regresión. Se extiende de 0 a 100% (0-1) 0: No hay correlación entre la entrada y la salida. 1: Toda la correlación
EarlyStopping		esta función relbe como parametros como la perdida deseada o el costo que desea monitorear, un delta mínimo para establecer un límite de cuanto debe mejorar el modelo en un epoch o ciclo antes de que finalice y el verbose.
Optimizadores de rendimiento		BASH NORMALIZATION, WEIGHT REGULARIZATION, DROPOUT, DATA ARGUMENTATION, OTHERS
criterion='mse'		"squared_error" para el error cuadrático medio, que es igual a la reducción de la varianza como criterio de selección de características y minimiza la pérdida de L2 utilizando la media de cada nodo terminal
Random Forest Regressor		
Parámetro	**Valor**	**Descripción**
criterion	mse / squared_error	"squared_error" para el error cuadrático medio, que es igual a la reducción de la varianza como criterio de selección de características y minimiza la pérdida de L2 utilizando la media de cada nodo terminal
n_estimators	500	Cantidad árboles de estimadores a promediar el resultado mas optimo
random_state	0	Controla la aleatoriedad en la muestra utilizada.

1.4. Resultados y conclusiones

Tabla 3. Muestra los resultados obtenidos de cada algoritmo.

Red Neuronal - 1 capa	Red Neuronal - 3 capa	Random Forest Regressor
Parámtros: * Numero capas: 1 * Input_shape: 11 * Neuronas: C1 = 1 neurona * Función de activación: Relu * epochs: 1000 * Optimización del modelo: RMSprop(learning_rate: 100) Resultados: * loss: 353932224.0000 * mae: 14076.4004 * mse: 353932224.0000 * Puntaje R2 Train set: 0.743 * Puntaje R2 Test set: 0.781 * MAPE DL Score : 0.452	Parámtros: * Numero capas: 3 * Input_shape: 11 * Neuronas: C1 = 64 neurona, C2 = 64 neurona, C3 = 1 neurona * Función de activación: Relu * epochs: 1000 * Optimización del modelo: RMSprop(learning_rate: 0.001) Resultados: * loss: ????? * mse DL : 34798840.000 * mae DL : 3441.627 * R2 DL on the Train set is: 0.862 * R2 DL on the Test set is: 0.782 * MAPE DL Score : 0.301	Parámtros: * criterion: mse / squared_error * n_estimators: 500 * random_state: 0 Resultados: * mse RF : 31995810.052 * mae RF : 2872.022 * R2 RF 0.7960358497640521 * Root Mean Squared Error RF: 54 * MAPE RF Score : 0.274

Imagen 2. Arquitectura red neuronal de 3 capas y tendencias de predicción para cada algoritmo.



La neurona de 1 capa se obtiene un puntaje R2 para los datos de prueba ligeramente mayor al de entrenamiento, lo cual indica que probablemente tengamos un sobre ajuste. La neurona de 3 capas de aproxima mas a una regresión lineal, arrojando como resultado un 86% de precisión (R2). Para las métricas MSE, MAE y MAPE, entre mas cercanas a cero tenemos un pronostico perfecto, para MAPE si el error es negativo el modelo estará subestimado, en caso de ser positivo el pronostico este sobre estimado, para los 2 modelos el valor mas cercano a cero es el del algoritmo RF Regressor, el cual se experimento con un learning_rate de 100 y 0.001, con el de 0.001 se obtuvo un valor R2 negativo, y con el de 100 positivo, lo cual mejoro la calidad de predicción del modelo. Las métricas MSE, MAE, si sus resultados son grandes, el error de pronostico también lo será. La red Neuronal mejoro su R2 adicionando una complejidad mayor de 2 capas adicionales de 64 neuronas y una capa con una neurona de salida, también se realizo la prueba con 10 capas adicionales de 512 neuronas y una de salida, pero se obtuvieron resultados similares a la de 3 capas. Algunas pruebas con la función Early Stopping evitaron que el modelo perdiera la capacidad de continuar aprendiendo en ciclos Epoch posteriores (La detención temprana es una técnica útil para evitar el sobreajuste).

2. Clasificación

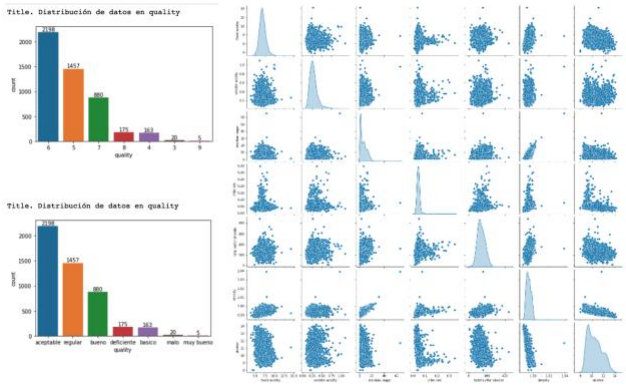
2.1. **Conjunto de datos:** Variantes rojas del vino portugués (Variables Fisiquicoquímicas), datos de dominio publico tomados de repositorio de datos Kaggle https://www.kaggle.com/inzamamsafi/multiclass-classification-wine-quality-beginner/data?select=winequality-red.csv), clasificación calidad del vino.

Tabla 4. Muestra tipos de datos originales y transformados a partir del Dataset.

Atributo	Descripción	Tipo	Instancias	Mean	Std	Min	25%	50%	75%	Max	Diff Max-Min
fixed acidity	la mayoría de los ácidos involucrados con el vino o fijos o no volátiles (no se evaporan fácilmente)	cuantitativa (continua)	4898	6,85	0,84	3,80	6,30	6,80	7,30	14,20	10
volatile acidity	la cantidad de ácido acético en el vino, que en niveles demasiado altos puede provocar un sabor desagradable a vinagre	cuantitativa (continua)	4898	0,28	0,1	0,08	0,21	0,26	0,32	1,1	1
citric acid	encontrado en pequeñas cantidades, el ácido cítrico puede agregar 'frescura' y sabor a los vinos	cuantitativa (continua)	4898	0,33	0,12	0	0,27	0,32	0,39	1,66	2
residual sugar	la cantidad de azúcar que queda después de que se detiene la fermentación, es raro encontrar vinos con menos de 1 gramo/litro y vinos con más de 45 gramos/litro se consideran dulces	cuantitativa (continua)	4898	6,39	5,07	0,6	1,7	5,2	9,9	65,8	65
chlorides	la cantidad de sal en el vino	cuantitativa (continua)	4898	0,05	0,02	0,01	0,04	0,04	0,05	0,35	0
free sufur dioxide	la forma libre de SO2 existe en equilibrio entre el SO2 molecular (como gas disuelto) y el ion bisulfito; previene el crecimiento microbiano y la oxidación del vino	cuantitativa (continua)	4898	35,31	17,01	2	23	34	46	289	287
total sufur dioxide	cantidad de formas libres y ligadas de SO2; en bajas concentraciones, el SO2 es mayormente indetectable en el vino, pero en concentraciones de SO2 libres superiores a 50 ppm, el SO2 se vuelve evidente en la nariz y el sabor del vino	cuantitativa (continua)	4898	138,36	42,5	9	108	134	167	440	431
density	la densidad del agua es cercana a la del agua dependiendo del porcentaje de contenido de alcohol y azúcar	cuantitativa (continua)	4898	0,99	0	0,99	0,99	0,99	1	1,04	0
pH	describe qué tan ácido o básico es un vino en una escala de 0 (muy ácido) a 14 (muy básico); la mayoría de los vinos están entre 3 y 4 en la escala de pH	cuantitativa (continua)	4898	3,19	0,15	2,72	3,09	3,18	3,28	3,82	1
sulphates	un aditivo del vino que puede contribuir a los niveles de dióxido de azufre (SO2), que actúa como antimicrobiano y antioxidante	cuantitativa (continua)	4898	0,49	0,11	0,22	0,41	0,47	0,55	1,08	1
alcohol	el porcentaje de contenido de alcohol del vino	cuantitativa (continua)	4898	10,51	1,23	8	9,5	10,4	11,4	14,2	6
quality	variable de salida (basada en datos sensoriales, puntuación entre 0 y 10)	cuantitativa (discreta)	4898	5,88	0,89	3	5	6	6	9	6

- **Número de instancias en total**: 4898.
- **Número de atributos:** 12 incluyendo la clase originalmente.

Imagen 2. Distribución de los datos en cada atributo y cantidad de instancias por clase



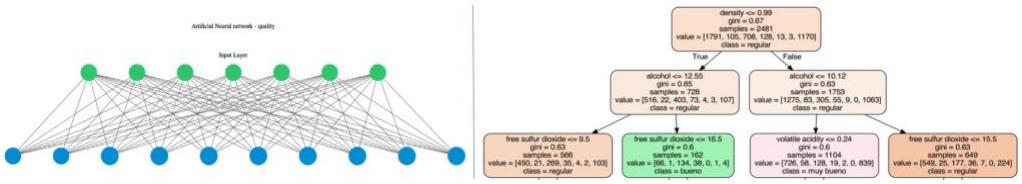
La **Imagen 2** muestra la distribución de los datos en cada una de las variables, también muestra la correlación entre los atributos de entrada y el atributo de salida **quality of wine,** la mayoría de las variables tienen una distribución próxima a la gaussiana, lo cual nos dice que se encuentran distribuidas adecuadamente con pocos Outliers. podemos evidenciar que las clases no están balanceadas en instancias, lo cual puede generar dificultades para predecir los valores correctos.

1.1. Parámetros relevantes

Redes neuronales (2 apas)		
Parámetro	**Valor**	**Descripción**
Capas (Red Neuronal)	2	Se crea un modelo de 2 capas, 1 de entrada con 7 atributos, una capa de salia con función de activación Softmax para multiclase
Neuronas Modelo 2	10, 7	10 para la capa de entrada y 7 que se ajustan a la dimensión de las 7 clasificaciones de calidad del vino (6, 5, 7, 8, 4, 3 y 9)
Input_shape	11	7 atributos de entrada
Training Data	0,8	3918 instancias, disminuye el costo de procesamiento
Test data	0,2	980, conjunto de datos separados para validar el rendimiento del modelo.
Overfiting	0%	Ocurre cuando nuestro modelo aprende demasiado bien nuestros datos de entrenamiento y como resultado sufre al generalizar.
optimizer='adam'	0.001	indican como actualizamos nuestros pesos y Adam es una alternativa al descenso de gradiante estocástico que funciona mejor. El gradiente estocastico es donde usando derivados, actualizamos todos los pesos en un subconjunto de los datos y brinda muchos beneficios de rendimiento.
categorical_crossentropy		funcion de que determina las perdidas del modelo
epochs	20, 100	Número de veces que se ciclan los datos mediante la red neuronal
verbose	0	muestra a mayor o menor detalle el resultado de entrenamiento del modelo.
validation_split	0.2	muestra aleatoria de datos de entrenamiento, para el caso es de 20% usado para validar la precisión de la red neuronal.
num_folds	10	Numero de iteraciones en las que se valida la presición del modelo.
Batch size		Durante cada iteración, el modelo calcula la salida pronosticada para 32 EJ batch size puntos de datos, calcula el costo de cada uno de ellos, promedia este valor y usa derivados para actualizar todos los pesos.
Random Forest Classifier		
Parámetro	**Valor**	**Descripción**
criterion	mse / squared_error	"squared_error" para el error cuadrático medio, que es igual a la reducción de la varianza como criterio de selección de características y minimiza la pérdida de L2 utilizando la media de cada nodo terminal
n_estimators	100	Cantidad árboles de estimadores a promediar el resultado mas optimo
random_state	0	Controla la aleatoriedad en la muestra utilizada.

2.2. Arquitectura

Imagen 3. Diseño y arquitectura de modelos de clasificación (Random Fores Classifier y Red neuronal)



2.3. Resultados y conclusiones

Imagen 4. Resultados obtenidos de las pruebas de ejecución para ambos modelos (Random Fores Classifier y Red neuronal)

	precision	recall	f1-score	support
aceptable	0.62	0.83	0.71	409
basico	1.00	0.14	0.24	51
bueno	0.70	0.53	0.60	183
deficiente	0.80	0.24	0.37	33
malo	0.00	0.00	0.00	9
regular	0.68	0.64	0.66	295
accuracy			0.65	980
macro avg	0.63	0.40	0.43	980
weighted avg	0.67	0.65	0.63	980

ACC RF : 0.652

```
[[338  50  21   0   0   0]
 [100 189   6   0   0   0]
 [ 78   6  97   2   0   0]
 [ 11   0  14   8   0   0]
 [  0   0   0   0   0   0]
 [  0   0   0   0   0   0]]

exactitud 0.65
presición 0.67
sensibilidad 0.65
f1 0.63
```

La red neuronal tiene una sola capa oculta con 7 entradas (las cuatro 7 de datos de entrada) conectadas a la capa de salida que contiene 7 neuronas.

Usamos softmax como la función de activación de salida en la capa de salida porque significa que todas las neuronas de salida serán iguales a 1, lo que nos permite interpretar la salida del modelo como probabilidades para cada clase. Estamos eligiendo funciones de activación y funciones de pérdida en función de lo que nosotros, como humanos, queremos que genere el modelo.

La función de pérdida para el entrenamiento, la entropía cruzada categórica, es solo una forma elegante de decir que el modelo se entrenará para generar un solo resultado alto (la clase).

Al comparar los 3 modelos RF, Red con 20 y 100 Epoch evidenciamos que RF obtiene un mayor porcentaje de exactitud, pero muy lejano del 100%. A pesar de esto los valores VP Y FP tienen un porcentaje de acierto muy bajo para cada clase, estos modelos se pueden optimizar mediante el incremento de Epoch por ejemplo o numero de estimadores y neuronas. Todas las variables del conjunto de datos aportan un valor significativo, por lo cual solo se omite un 30% en una de las pruebas para optimizar tiempo de aprendizaje y predicción del modelo.

3. Bibliografía

al, K. e. (2018). *https://keras.io/api/optimizers/adam/.* Obtenido de keras.io: https://keras.io/api/optimizers/adam/

GEOTutoriales. (21 de 7 de 2015). Cálculo del Error Porcentual Absoluto Medio o MAPE en un Pronóstico de Demanda.

GEO Tutoriales. (26 de 01 de 2015). *Gestión de Operaciones .* Obtenido de Error Porcentual Absoluto Medio (MAPE) en un Pronóstico de Demanda : https://www.gestiondeoperaciones.net/proyeccion-de-demanda/error-porcentual-absoluto-medio-mape-en-un-pronostico-de-demanda/

Gonzalez, A. c. (12 de 4 de 2019). BOSQUES ALEATORIOS REGRESIÓN - SCIKIT LEARN | #29 Curso Machine Learning con Python.

HD, H. W. (26 de 12 de 2018). Neural Network Regression Model with Keras | Keras #3.

gasolina, R. B. (24 de 2 de 2022). *tensorflow.org.* Obtenido de Regresion Basica: Predecir eficiencia de gasolina: https://www.tensorflow.org/tutorials/keras/regression?hl=es-419

SAFI, I. (1 de 7 de 2021). *Multiclass classification wine quality.* Obtenido de kaggle.com: https://www.kaggle.com/inzamamsafi/multiclass-classification-wine-quality-beginner/data

Analyst, N. D. (9 de 8 de 2021). Easiest Way to Download Kaggle Datasets using Opendataset in Jupyter Notebook.

Rowe, W. (4 de 10 de 2019). *How to Use Keras to Solve Classification Problems with a Neural Network.* Obtenido de bmc.com: https://www.bmc.com/blogs/keras-neural-network-classification/

Brownlee, J. (2 de 6 de 2016). *Multi-Class Classification Tutorial with the Keras Deep Learning Library.* Obtenido de machinelearningmastery: https://machinelearningmastery.com/multi-class-classification-tutorial-keras-deep-learning-library/

Ahmed, J. (5 de 10 de 2018). Keras - Multi Class Classification using a Deep Neural Network with Keras.