## INFORME DE PROYECTO

# Catherine Rincón Maya

## Descripción de la estructura de los notebooks entregados

El repositorio de github <a href="https://github.com/cjrincon/Proyecto\_Deep\_Learning">https://github.com/cjrincon/Proyecto\_Deep\_Learning</a> contiene los siguiente:

**Entrega 1 - Catherine Rincon Maya.pdf:** En este archivo de PDF se define el contexto de aplicación del proyecto de Deep learning, el objetivo de machine learning que es la predicción del RUL, el tipo de dataset, cantidad de datos, métricas de desempeño asociadas al proyecto, resultados de algunos trabajos previos por otros autores y referencias.

**Data:** Es una carpeta con 4 archivos .csv que contiene la información de las 4 baterías #5, #6, #7 y #18 con las siguientes columnas: Ciclo, temperatura ambiente, fecha (fecha y hora de inicio del ciclo), voltaje de la batería, corriente de la batería, temperatura de la batería, corriente de carga, voltaje de carga, tiempo (tiempo del ciclo en segundos), capacidad de la batería.

**01. Análisis exploratorio y preprocesado.ipynb:** En este archivo de colab se define el objetivo del proyecto, los datos de entrada y salida, se realiza la lectura de los datos y exploración de los mismos determinando el tipo de dato de cada variable, cantidad de nulos y medidas descriptivas como cantidad de datos, media, desviación estándar, mínimo, cuartiles y máximo. Se hacen boxplot e histograma para cada una de las variables y realizar una exploración gráfica de las variables. Se realiza una matriz de correlaciones y su respectiva gráfica para validar las relaciones lineales directas e inversas entre las variables.

También se realiza un cálculo del estado de salud (SOH) a partir de la capacidad de la batería y se agrega al dataset. Se determina el número de ciclos por batería y la capacidad inicial de cada una, la cual representaría un estado de salud (SOH) del 100%. Así mismo, se valida el valor mínimo del SOH de cada batería.

- **02. Arquitectura de linea base.ipynb:** En este archivo de colab se definen las arquitecturas de los modelos a implementar para hacer la predicción de muchos a uno, y se realiza una comparación de los resultados obtenidos en los modelos de predicción implementando los datos obtenidos a partir del promedio de las variables y la mediana.
- **03. Prediccion del RUL.ipynb:** En este archivo de colab se definen los modelos de predicción de uno a uno, para predecir la capacidad en ciclos futuros y así establecer la predicción en tiempos hacia adelante.

## Descripción de solución y de las iteraciones

Se trabaja con un conjunto de datos de 4 baterías de la NASA, enumeradas así: #5, #6, #7 y #18. Cada una de las baterías tiene información asociada a: Ciclo, temperatura ambiente (°C), fecha (fecha y hora de inicio del ciclo), voltaje de la batería (Voltios), corriente de la batería (Amperios), temperatura de la batería (°C), corriente de carga (Amperios), voltaje de carga (Voltios), tiempo (tiempo del ciclo en segundos), capacidad de la batería (Amperios/h).

Teniendo en cuenta esta información, se puede calcular el estado de salud (SOH) de las baterías a partir de la información del ciclo de carga y descarga como:

$$SOH = \frac{Capacidad\ actual\ en\ el\ ciclo}{Capacidad\ al\ inicio\ de\ la\ vida\ útil}*100\%$$

Con esta variable, se determina el grado de envejecimiento que tiene la batería, lo cual es un indicador de salud para predecir la vida útil. La vida útil restante (RUL) de la batería, se puede determinar a partir del SOH de la batería o a partir de la capacidad de la misma.

Debido a que en cada ciclo se tienen diferentes mediciones para las variables de temperatura, voltaje y corriente, es decir, una serie de datos de cada una de las variables por cada ciclo, se define obtener las medidas descriptivas como mínimo, máximo, promedio, desviación y mediana para cada una de las variables, con el fin de determinar la medida a usar para evaluar los modelos. Con esto, se determina que a partir del promedio se pueden evaluar los modelos.

Se plantean dos modelos de predicción:

- 1. Un modelo para predecir la capacidad a partir de diferentes variables. Teniendo en cuenta variables asociadas como: Ciclo, voltaje de la batería (Voltios), corriente de la batería (Amperios), temperatura de la batería (°C), corriente de carga (Amperios), voltaje de carga (Voltios). En este caso, este modelo tiene una arquitectura de muchos a uno. Para este modelo de predicción, se plantean 3 tipos de arquitecturas para su solución: se tiene en cuenta un modelo con capas densas, un modelo con capas LSTM y un modelo con capas recurrentes.
- 1.1. Modelo con capas densas: El primer modelo sencillo que se implementó, tiene sólo capas densas y una capa de dropout para evitar el sobreajuste. La arquitectura del modelo inicial se presenta en la Figura 1, consta de 2 capas densas de 8 neuronas, una capa de dropout con una tasa de 0.3 y una capa densa de 1 neurona que es la encargada de entregar la predicción de la capacidad. Este modelo consta de 145 parámetros.

Layer (type)	Output Shape	Param #
dense (Dense)	(None, B)	64
dense_1 (Dense)	(None, 8)	72
dropout (Dropout)	(None, 8)	0
dense_2 (Oense)	(None, 1)	93
Total params: 145 Trainable params: 145 Non-trainable params: 0		************

Figura 1. Arquitectura modelo de capas densas

1.2. Modelo con capas de corto y largo plazo – LSTM: El segundo modelo que se plantea, tiene 3 capas LSTM y 3 capas densas, donde la última capa densa es la encargada de entregar la predicción. Se plantea un modelo con capas LSTM, ya que estas capturan la información de corto y largo plazo, lo que podría generar una mejor predicción. Este modelo se vuelve más complejo que el anterior, ya que cuenta con 144.097 parámetros. En la Figura 2 se presenta la arquitectura del modelo.

Layer (type)	Output Shape	Paran #
********		**********
Istm_29 (LSTM)	(None, 7, 128)	66560
1stm_10 (LSTM)	(tione, 7, 54)	49498
1stm_31 (LSTM)	(None, 32)	12416
dense_54 (Dense)	(None, 96)	3168
dense_55 (Dense)	(None, 128)	12415
dense_56 (Dense)	(None, 1)	129
otal paraws: 144,897	(1000) 27	**********

Figura 2. Arquitectura modelo con capas LSTM

1.3. Modelo con capas densas: En el tercer modelo se plantea una arquitectura más simple conformada por una capa recurrente, una capa de dropout para evitar el sobreajuste, y por último, una capa densa encargada de entregar la información de la predicción. Se plantea esta arquitectura considerando la recurrencia de los datos y la simplicidad de modelo. La arquitectura del modelo se presenta en la Figura 3, donde se observa que el modelo tiene 2.651 parámetros.

Layer (type)	Output Shape	Paron #
simple_rnn_25 (SimpleMMX)	(None, 58)	2600
dropout_40 (Dropout)	(None, 58)	B
dense_57 (Dense)	(Mone, I)	51
Total params: 2,651 Trainable params: 2,651 Mon-trainable params: 0	and the control of the state of	

Figura 3. Arquitectura modelo con capas recurrentes

**2.** Un modelo para predecir la vida útil restante a partir de la capacidad. En este caso, se pretende predecir la capacidad después de varios ciclos de carga/descarga. Es decir, usando 15 datos del pasado, predecir la capacidad que va a tener la batería a los 15 ciclos siguientes. El modelo de predicción se valida con la arquitectura de los modelos usando capas LSTM y usando la capa recurrente para validar los resultados.

## Descripción de los resultados

Análisis exploratorio y preprocesado

Se consolida un nuevo dataset con la información de las 4 baterías, #5, #6, #7 y #18. Se tiene un total de 185.721 filas en el nuevo dataset y no se presentan valores nulos. Se realiza un análisis exploratorio a las variables, donde se identifica que la variable temperatura ambiente es una variable fija con un valor de 24 °C para toda la toma de datos de las 4 baterías, por lo cual, no se tiene en cuenta dentro de los análisis. Además, el voltaje medido tiene una cola pesada hacia la izquierda y la mayoría de los datos para la corriente medida se encuentran alrededor de -2 amperios.

Se grafica la distribución de la variable de interés y se muestra en la Figura 4, esta variable es la capacidad de la batería, y se observa que tiene un comportamiento normal que se distribuye entre los valores de 1 y 2 amperios por hora, con una media de 1.57 y una desviación estándar de 0.19.

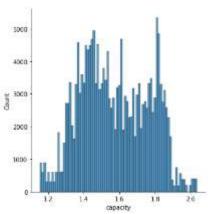


Figura 4. Histograma de la variable respuesta capacidad de la batería

Para explorar las relaciones lineales entre las variables, se realiza un gráfico de correlación que permita determinar las relaciones entre cada una de ellas. La matriz de correlaciones se presenta en la Figura 5, en esta se observan relaciones fuertes directas e inversas. Una de las relaciones fuerte y directa es entre el tiempo y la temperatura, lo que indica que a medida que pasa el tiempo, la temperatura de la batería aumenta, lo que indica degradación en la batería. También se observan relaciones fuertes indirectas entre: el ciclo y la capacidad, temperatura y voltaje, tiempo y voltaje, y voltaje de carga y corriente. Lo que indica que a medida que aumenta una de las variables, la otra disminuye. Cada que aumentan los ciclos de carga/descarga, la capacidad de la batería disminuye, lo cual, está fuertemente relacionado con la degradación de la misma.

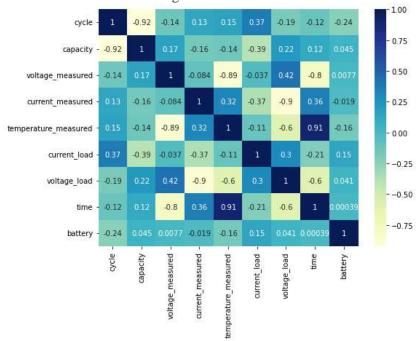


Figura 5. Matriz de correlaciones entre las variables

La relación entre el ciclo y la capacidad para cada una de las baterías, se puede ver gráficamente en la Figura 6, donde se observa que la capacidad tiene un comportamiento decreciente a lo largo de los ciclos, con algunos picos de aumento de la capacidad que representan la regeneración de la batería, sin embargo, la tendencia en general de la capacidad de la batería es a disminuir.

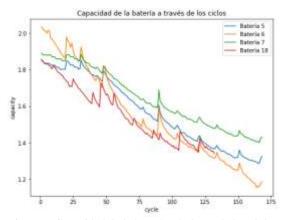


Figura 6. Capacidad de la batería a lo largo de los ciclos

Para comprender el rendimiento de la batería, es importante determinar el SOH a partir de la capacidad de la batería. La relación entre los ciclos de carga/descarga y el SOH de la batería para cada una de las baterías implementadas en los modelos de predicción, se presenta en la Figura 7. Según revisión de literatura, un SOH entre el 70% y el 80%, determina el fin de la vida útil de la batería, por lo cual se grafican ambas líneas. Se observa que en general, las baterías #5, #6 y #7, se dejaron en operación hasta el ciclo de carga/descarga número 168, y la batería #18, sólo se dejó en funcionamiento por 132 ciclos, donde una causa probable fue una mayor degradación en la batería.

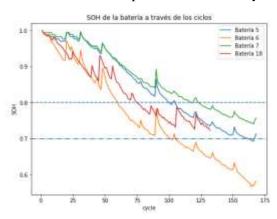


Figura 7. Estado de salud a través de los ciclos de carga/descarga

Así mismo, se presenta el valor de la capacidad inicial para cada una de las baterías en amperios/hora en la Tabla 1. En esta se observa que la batería #6 inició con una capacidad mayor al resto, seguido por la batería #7, luego #5 y por último la batería #18. Así mismo, los valores de capacidad inicial de las baterías #5, #7 y #18 se ven muy homogéneas.

Tabla 1. Capacidad inicial de cada batería

Batería	Capacidad (Ah)		
#5	1.86		
#6	2.03		
#7	1.89		
#18	1.85		

Para realizar los modelos de predicción, se genera un nuevo dataset con medidas descriptivas como mínimo, máximo, media, desviación estándar y mediana para las variables: temperature\_measured, voltage\_measured, current\_measured, voltage\_load, current\_load. Con este nuevo dataset se realizan algunas gráficas de exploración del comportamiento de la corriente, el voltaje y la temperatura a lo largo de los ciclos. En estos, se resalta la media del voltaje medido a lo largo de los ciclos para cada batería, ya que pareciera que el comportamiento de la variable sigue el mismo comportamiento decreciente que la capacidad de la batería a lo largo de los ciclos. Este comportamiento se puede visualizar en la Figura 8.

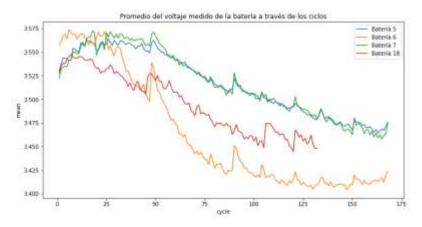


Figura 8. Promedio del voltaje medido de la batería a través de los ciclos

#### Arquitectura de línea base

Se plantean 3 modelos: un modelo con capas densas, un modelo con capas LSTM y un modelo con capa recurrente. Cada uno de estos modelos se prueba con datos obtenidos a partir del promedio de las variables y de la mediana. Para la implementación de estos modelos, se agregan como features las variables: cycle, temperature\_measured, voltage\_measured, current\_measured, voltage\_load, current\_load, SOH, a partir de las cuales se realiza la predicción del RUL. Se realiza la división de los datos teniendo en cuenta 3 baterías para el train y el test, y una batería para la validación.

En la evaluación del primer modelo (capas densas), se obtiene un RMSE inferior al 10% en los datos de test, en los datos de validación, se obtiene un RMSE del 15,6%. Así mismo, se evalúa el modelo para cada uno de los conjuntos de la batería y los resultados se muestran en la Figura 9.

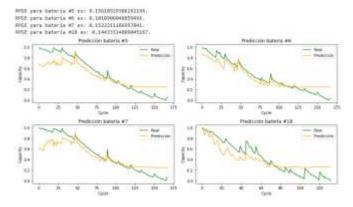


Figura 9. Resultados del modelo de capas densas usando el promedio de las variables

Se realiza la evaluación del segundo modelo (capas LSTM) y se obtiene un RMSE inferior al 8% en los datos de test, en los datos de validación, se obtiene un RMSE del 9,6%. Así mismo, se evalúa el modelo para cada uno de los conjuntos de la batería y los resultados se muestran en la Figura 10. En este caso, se ve una mejoría en los datos de validación en cuanto a la métrica de desempeño.

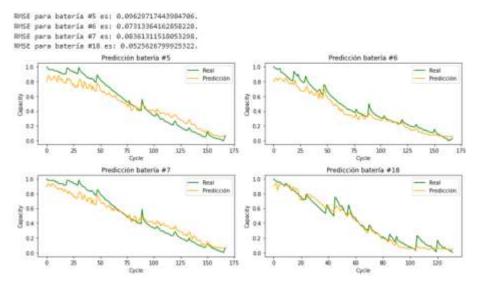


Figura 10. Resultados del modelo de capas LSTM usando el promedio de las variables

Para el tercer modelo, que se realiza a partir de una capa recurrente, se obtiene un RMSE inferior al 5%, y en los datos de validación, se observa que el RMSE puede ser aún menor que en el modelo planteado anteriormente, lo que muestra que un modelo más sencillo en cuanto a número de capas y parámetros, también puede hacer una buena predicción de la capacidad.

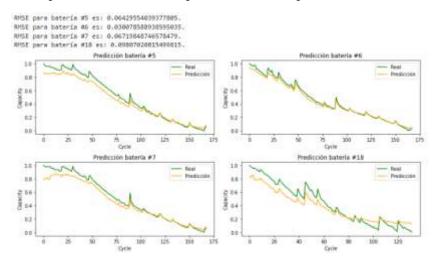


Figura 11. Resultados del modelo de capa recurrente usando el promedio de las variables

Los 3 modelos planteados también se evalúan a partir de los datos obtenidos con la mediana de cada una de las variables, donde se observa una mejoría en la métrica de desempeño al usar los datos obtenidos con la mediana en comparación con los datos del promedio. En la Tabla 2, se muestran los resultados de los 3 modelos para cada el conjunto de test y para cada una de las baterías usando como datos de entrada, los datos obtenidos a partir de la media y la mediana. En general, se observa que se tiene un mejor desempeño con el modelo que implementa la capa recurrente y los datos obtenidos a

partir de la mediana, sin embargo, el modelo usando la mediana de los datos tiende a suavizar más la predicción en comparación con el modelo que usa el promedio de los datos. Así mismo, se observa que el modelo de redes LSTM y datos promedio, se ajustan bien a la predicción de los valores de la batería # 18.

Tabla 2. Comparación de RMSE	para cada uno de los modelos usando el p	promedio y la mediana de los datos

RMSE	Datos con promedio			Datos con mediana		
Datos	Densas	LSTM	Recurrente	Densas	LSTM	Recurrente
Test	6,8%	6,8%	2,8%	8,0%	5,9%	1,6%
Batería #5	15,6%	9,6%	6,4%	15,4%	16,3%	4,3%
Batería #6	10,1%	7,3%	3,0%	11,7%	13,1%	4,4%
Batería #7	15,2%	8,4%	6,7%	13,7%	15,0%	3,3%
Batería #18	14,4%	5,2%	9,8%	14,1%	12,8%	12,4%

#### Predicción del RUL

En este archivo, se valida la predicción de la capacidad que se va a tener a los 15 ciclos de carga/descarga. En este caso, se hace la división de los datos teniendo en cuenta sólo la variable de capacidad, por lo que el modelo de predicción tiene una estructura de uno a uno. Se evalúan dos arquitecturas de modelos de predicción: uno que implementa la capa recurrente y otro que implementa las capas LSTM.

Con el modelo que tiene la capa recurrente y el dropout, se obtiene un RMSE para los datos de entrenamiento del 2%, y un RMSE para los datos de test del 9%. En la Figura 12 se observa el comportamiento de los datos originales, la predicción del entrenamiento y la predicción de test. Se observa al inicio de la predicción en los datos de test, que se tienen algunos picos en la capacidad, que luego se estabilizan y capturan el comportamiento decreciente de la variable a lo largo de los ciclos.

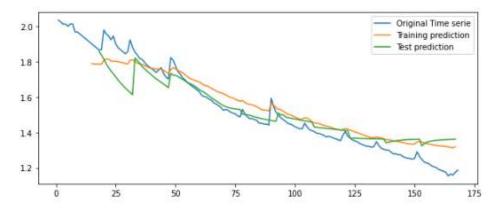


Figura 12. Resultados del modelo de capa recurrente para predecir la capacidad

Así mismo, se evalúa la predicción con el modelo que implementa capas LSTM. En este caso, se obtiene un RMSE para los datos de entrenamiento del 1%, y para los datos de test, se obtiene un RMSE del 5%, lo cual muestra mejores resultados de predicción usando este modelo con capas LSTM

en comparación con el modelo anterior que usa una capa de recurrencia y un dropout. En la Figura 13 se muestran los resultados obtenidos en la predicción implementando los datos de entrenamiento y los datos de test. En esta gráfica, se observa que la predicción es más suavizada en comparación con la predicción obtenida anteriormente. Así mismo, se ve un mejor ajuste en la predicción usando ambos datos de entrenamiento y test.

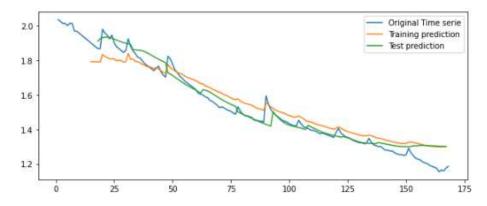


Figura 13. Resultados del modelo de capa recurrente para predecir la capacidad

#### **Conclusiones**

Cada una de las baterías tiene un estado de salud (SOH) del 100% al inicio de la vida, es decir, en el ciclo 1, y este SOH va disminuyendo a lo largo de los ciclos de carga/descarga de la batería, con algunos picos de regeneración en el estado de salud de la batería. Según revisión de literatura, un SOH entre el 70% y 80%, determina el fin de la vida útil de la batería.

La capacidad es una de las variables que se puede predecir a partir de variables como corriente, voltaje y temperatura, y es la variable encargada de determinar la degradación de la batería, lo que permite obtener resultados para predecir el fin de la vida útil. Así mismo, con la variable capacidad, se pueden hacer predicciones varios ciclos a futuro para determinar en qué ciclo ocurrirá el fin de la vida útil de la batería.

Se definen 3 arquitecturas de modelos de predicción usando como datos de entrada, los datos obtenidos a través del promedio y la mediana de las variables corriente, voltaje y temperatura. Se entrenan los modelos con cada uno de los conjuntos de datos y se realiza la predicción, obteniendo un RMSE inferior en la mayoría de los datos en comparación con los otros modelos, implementando el modelo de predicción con la capa de recurrencia y el dropout. Así mismo, se observa que la predicción usando la mediana de los datos, tiende a suavizarse más en comparación con los modelos de predicción que usan el promedio de los datos.

Para la predicción de la capacidad a través de varios ciclos de carga/descarga a futuro, se implementaron dos modelos de predicción, de los cuales, el modelo de predicción que implementa capas LSTM mostró mejores resultados en cuanto al RMSE en comparación con el otro modelo. Así mismo, se observa en las gráficas que la predicción con el modelo de las capas LSTM tiende a capturar mejor el comportamiento de la capacidad que el modelo con la capa de recurrencia.