1. Cuál arquitectura se uso ? -- Tipo, Tamaño y Definición (CNN, RNN, etc)

* Se usó una CNN con dos capas con 64 filtros, con funciones de activación sigmoides, y al final tenemos una capa densa interconectada con únicamente un filtro.

2. Cómo la utilizaron? -- Qué problema solucionaron? Y que tipo Predicción?

Recomendación? Generación? Análisis de intensiones?

* Se utilizó en un escenario de clasificación, en este caso fue el de clasificación de ubicaciones óptimas para la venta de las sucursales.

3. Se necesitó hacer ajustes durante la implementación? -- cambiar learning rate,

arquitectura, tamaño, función de activación, etc.

* Sí, primeramente se había utilizado un modelo de regresión, sin embargo, notamos que no era tan eficiente con los datos que recopilamos para este problema, por lo que hicimos el cambio a un modelo de clasificación y obtuvimos excelentes resultados.

4. ¿Qué otras arquitecturas se usaron y qué resultados obtuvieron?

* Uno de los primeros modelos desarrollados por nuestro equipo fue un modelo de redes neuronales que predecía las coordenadas de las siguientes sucursales Dhl. Tomando como input los valores poblacionales y como salida las coordenadas. Sin embargo, este modelo representaba overfitting en los resultados ya que no parecían tener correlación con la información previa y no se contenían dentro de la región de Nuevo León. Por esta razon, se optó por modificar el modelo y cambiar la base de datos con el objetivo de generar una correlación más fuerte y brindar mejores predicciones.

5. ¿Cuáles entradas (input data) usaron y qué pre-procesamiento se realizó?

Explicar cómo se definieron (manual, automático?)

Los input del modelo son los siguientes campos:

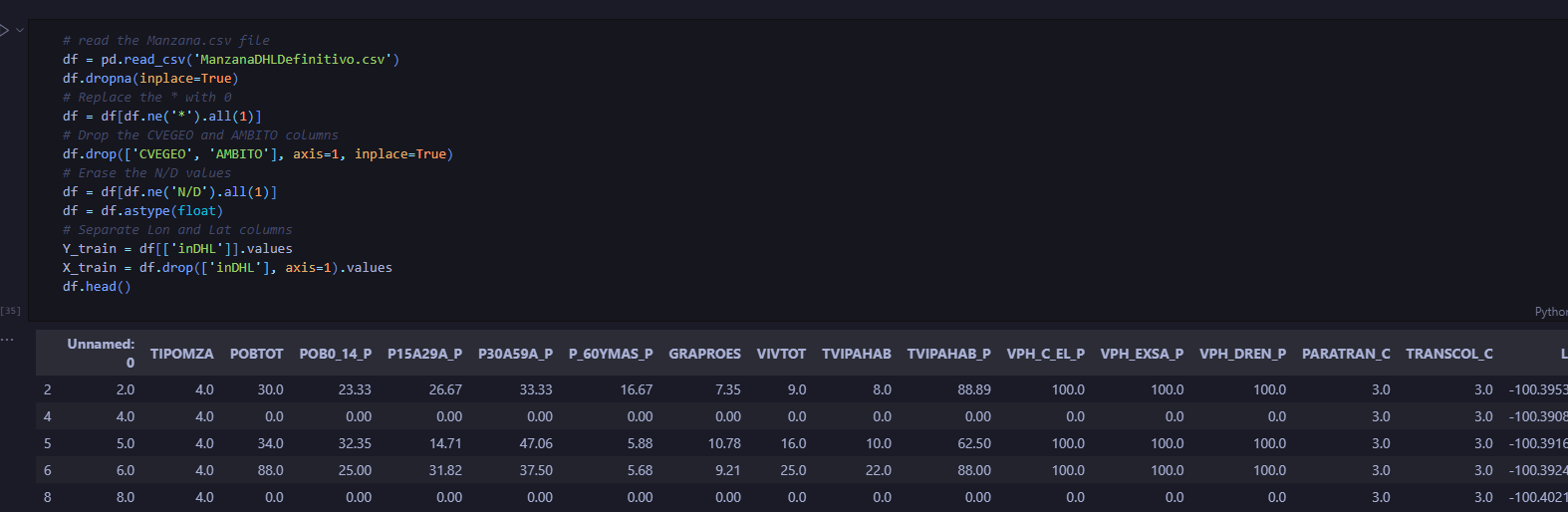
| CaracteristicasPoblacionalesPorManzanaconDHL.csv | |
| --- | --- |
| Variables | Significado |
| POB\_TOT | Poblacion Total |
| POB0\_14\_P | Porcentaje de la Poblacion de 0 a 14 años |
| P15A29A\_P | Porcentaje de la Poblacion de 15 a 29 años |
| P30A59A\_P | Porcentaje de la Poblacion de 30 a 59 años |
| P\_60YMAS\_P | Porcentaje de la Poblacion de 60 años y mas |
| GRAPROES | Grado Promedio de Escolaridad |
| VIVTOT | Total de Viviendas |
| TVIPAHAB | Total de Viviendas Particulares Habitadas |
| TVIPAHAB\_P | Porcentaje de Viviendas Particulares Habitadas |
| VPH\_EXSA\_P | Porcentaje de Viviendas Particulares Habitadas que disponen de Excusado o Sanitario |
| VPH\_DREN\_P | Porcentaje de Viviendas Particulares Habitadas que disponen de Drenaje |
| PARATRAN\_C | Disponibilidad de parada de transporte colectivo en la manzana |
| TRANSCOL\_C | Disponibilidad de transporte colectivo |
| Latitud | Coordenada de Latitud sobre la posicion de la manzana |
| Longitud | Coordenada de Longitud sobre la posicion de la manzana |

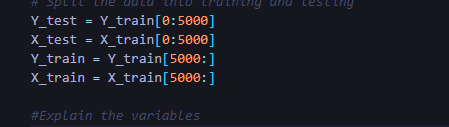
Los output del modelo que se buscan predecir:

| inDhl | Determina de forma booleana si una manzana se encuentra dentro de un radio de 1km cerca de un dhl. |
| --- | --- |

6. Explicación de la división de los datos (training-testing sets)

Los datos fueron divididos en 2 sets siendo los datos de entrenamiento y los datos de testeo para comprobar el funcionamiento y las predicciones del modelo. Para nuestro modelo los datos “Y” contienen el conocimiento sobre si una manzana poblacional se encuentra dentro de un radio de un kilómetro de una sucursal de Dhl. Los datos “X” representan los input que fueron mencionados anteriormente. El test se dividió en todos los datos después de 5000 y los datos de entrenamiento en todos los que estaban antes de 5000.





7. Los detalles de entrenamiento -- Epochs, Learning rate, etc.

Epochs → 5

Función de activación ---> Función sigmoide

Loss function ----> Binary cross-entropy

Se hizo uso de Adam optimization.

8. Los resultados (outputs) obtenidos y cuál es su función final (paso intermedio del

proyecto ?

* Se obtuvo como resultado un conjunto de coordenadas asociadas a las ubicaciones ideales, en base a la densidad de la población, tomando en cuenta un radio de 1 km. Con estas coordenadas, graficamos y obtuvimos una forma más visual de contrastar resultados.

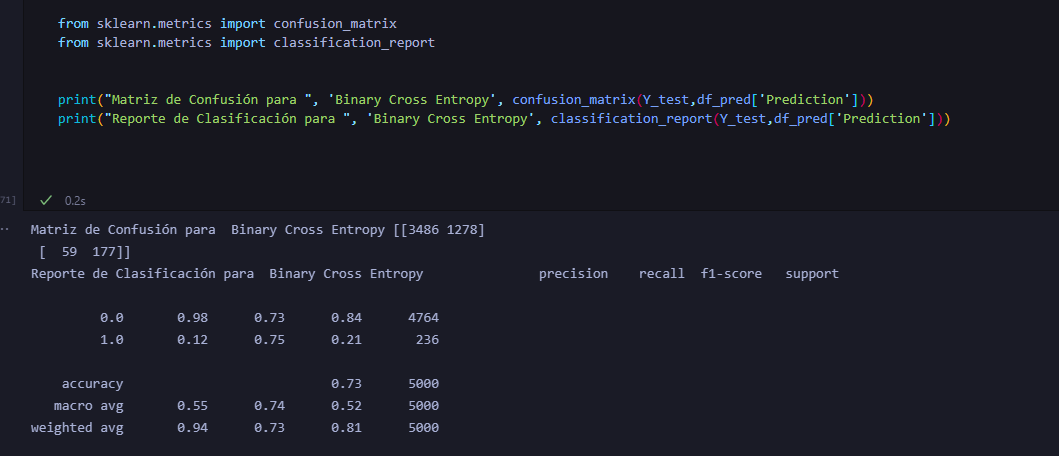
9. Análisis durante el proceso de entrenamiento (training set-- accuracy & loss) y

ejecución (testing set-- accuracy & loss)

La precisión aumento y se tuvo uno pérdida del loss, conforme iban avanzando las épocas.

10. Incluir gráficas de los análisis mencionados arriba

El rendimiento del modelo fue el siguiente



11. Incluir visualizaciones o explicaciones de los datos de entrada utilizados y de los

resultados

Se obtuvo como resultado las siguientes ubicaciones:

