UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR

NOMBRES:

NICOLÁS AÑAZCO

CARRERA:

INGENIERÍA EN SISTEMAS

MATERIA:

SISTEMAS EXPERTOS

FECHA:

22/05/2020

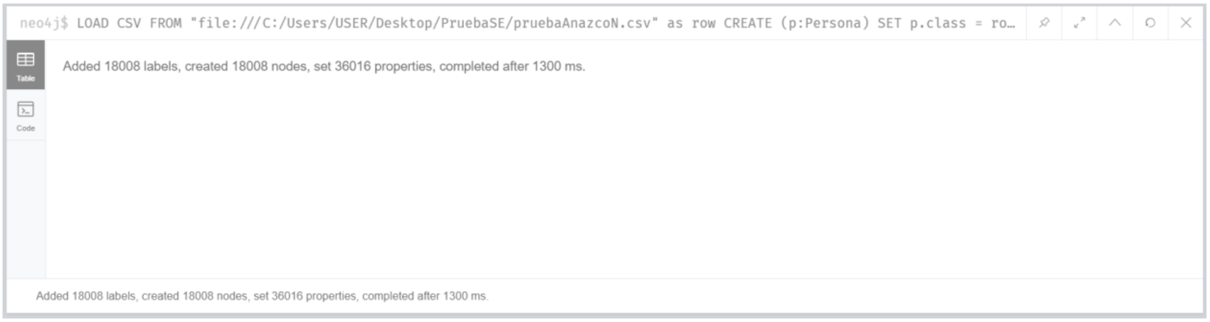
PRUEBA SE 2

- Diseñe y desarrolle un algoritmo Knn en Neo4j para:
  - Fila A - 0: Obtener si le asignamos una garantía en base al perfil de la persona, para ello se debe descargar los datos del siguiente link: https://drive.google.com/file/d/0B21nDwg3DpmWNHU0TC1uOXlGV3c/view.

Cargar el archivo descargado de internet y proceder a la Creación de Nodos:

LOAD CSV WITH HEADERS FROM 'file:///C:/Users/USER/Desktop/PruebaSE/insurance.csv' AS line

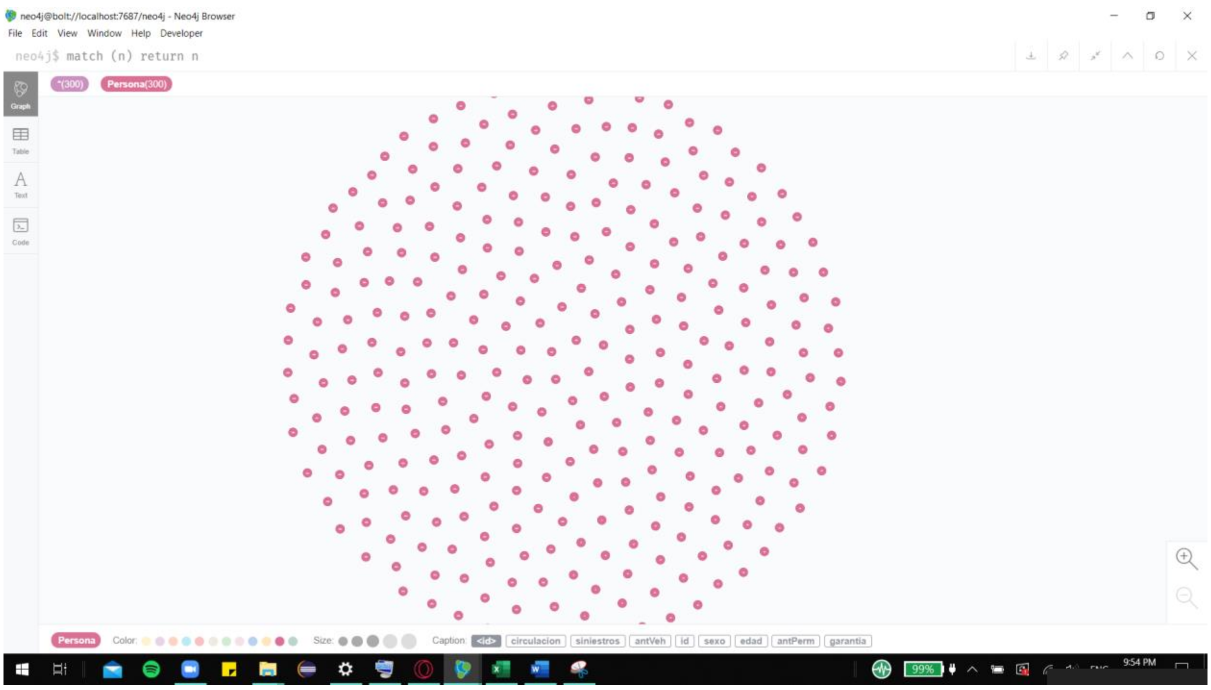
```
LOAD CSV FROM "file:///C:/Users/USER/Desktop/PruebaSE/pruebaAnazcoN.csv" as row

CREATE (p:Persona)

SET p.class = row[3],

   p.features = row[4..];
```



```
neo4j$ LOAD CSV FROM "file:///C:/Users/USER/Desktop/PruebaSE/pruebaAnazcoN.csv" as row CREATE (p:Persona) SET p.class = ro…
```

```
Added 18008 labels, created 18008 nodes, set 36016 properties, completed after 1300 ms.
```

```
Added 18008 labels, created 18008 nodes, set 36016 properties, completed after 1300 ms.
```



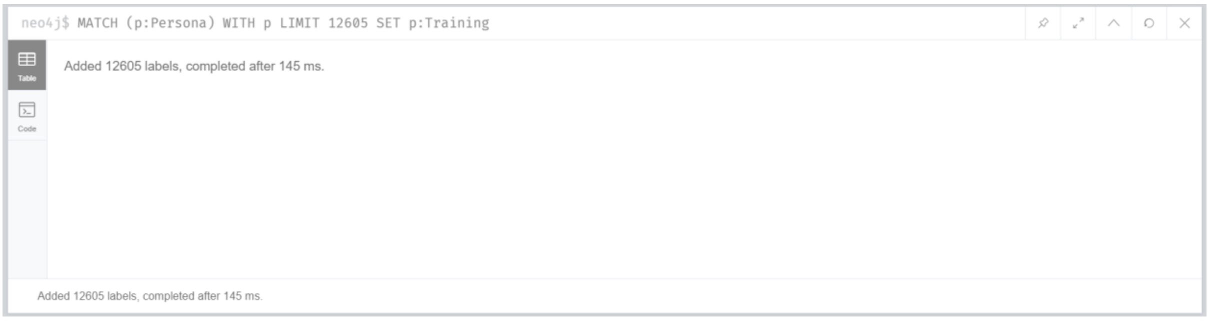
neo4j@bolt://localhost:7687/neo4j - Neo4j Browser

```
neo4j$ match (n) return n
```

File Edit View Window Help Developer
*(300)
Persona(300)
Graph
Table
Text
Code
Persona
Color:
Size:
Caption: <id> circulacion siniestros antVeh id sexo edad antPerm garantia
99%
9:54 PM

Mark training data 70 %:

```
MATCH (p:Persona)
WITH p LIMIT 12605
SET p:Training;
```

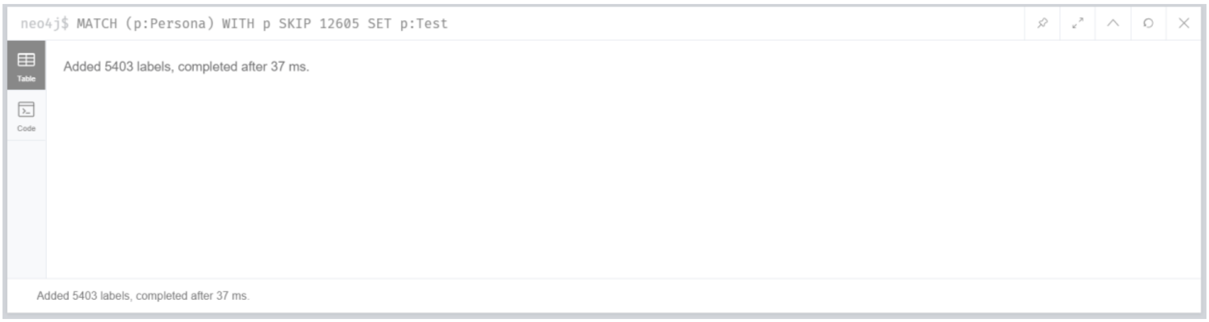


```
neo4j$ MATCH (p:Persona) WITH p LIMIT 12605 SET p:Training
Added 12605 labels, completed after 145 ms.
```

Added 12605 labels, completed after 145 ms.

Mark test data 30%:

```
MATCH (p:Persona)
WITH p SKIP 12605
SET p:Test;
```



```
neo4j$ MATCH (p:Persona) WITH p SKIP 12605 SET p:Test
Added 5403 labels, completed after 37 ms.
```

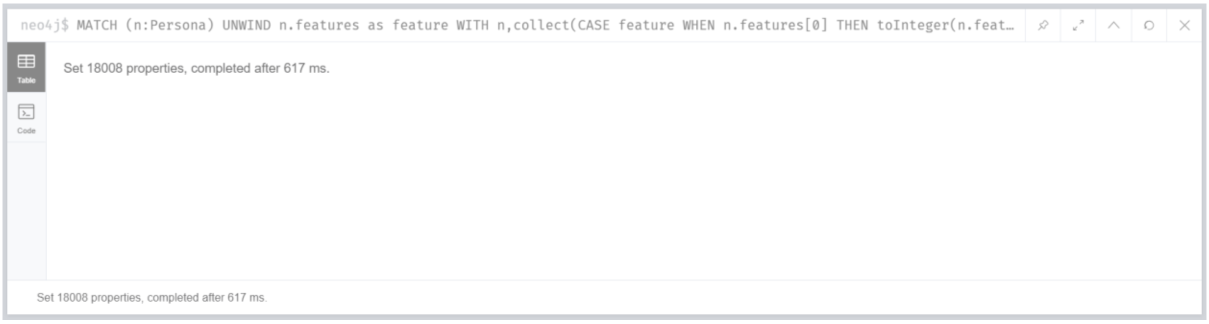
Added 5403 labels, completed after 37 ms.

Convertir a vectores:

```
MATCH (n:Persona)
UNWIND n.features as feature
WITH n,collect(CASE feature
            WHEN n.features[0] THEN toInteger(n.features[0])
            WHEN n.features[1] THEN toInteger(n.features[1])
            WHEN n.features[2] THEN toInteger(n.features[2])
            WHEN n.features[3] THEN toInteger(n.features[3])
            WHEN n.features[4] THEN toInteger(n.features[4])
```

END) as feature_vector

SET n.feature_vector=feature_vector



```
neo4j$ MATCH (n:Persona) UNWIND n.features as feature WITH n,collect(CASE feature WHEN n.features[0] THEN toInteger(n.feat…
```

Set 18008 properties, completed after 617 ms.

Set 18008 properties, completed after 617 ms.

Consulta:

MATCH (test:Test)

WITH test,test.feature_vector as feature_vector

CALL apoc.cypher.run('MATCH (training:Training)

WITH training,gds.alpha.similarity.euclideanDistance($feature_vector, training.feature_vector) AS similarity

ORDER BY similarity ASC LIMIT 3

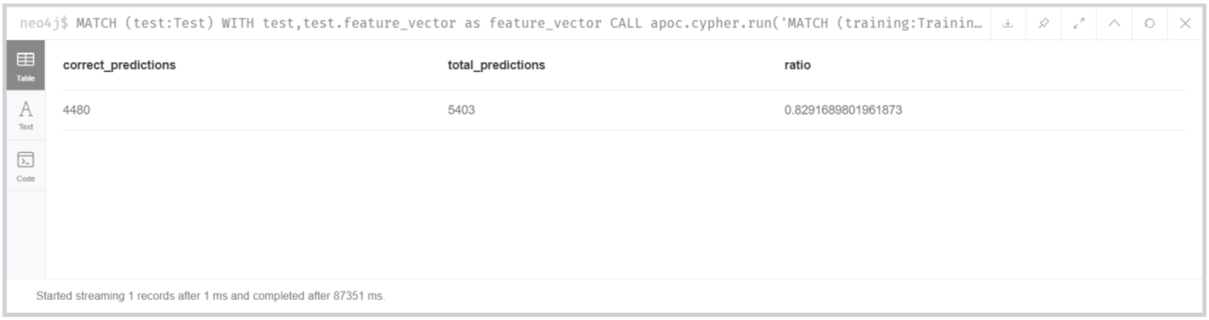
RETURN collect(training.class) as classes',

{feature_vector:feature_vector}) YIELD value

WITH test.class as class, apoc.coll.sortMaps(apoc.coll.frequencies(value.classes), '^count')[-1].item as predicted_class

WITH sum(CASE when class = predicted_class THEN 1 ELSE 0 END) as correct_predictions, count(*) as total_predictions

RETURN correct_predictions,total_predictions, correct_predictions / toFloat(total_predictions) as ratio;



```
neo4j$ MATCH (test:Test) WITH test,test.feature_vector as feature_vector CALL apoc.cypher.run('MATCH (training:Trainin…
```

correct_predictions	total_predictions	ratio
4480	5403	0.8291689801961873

Started streaming 1 records after 1 ms and completed after 87351 ms.