FORMATO DE INFORME DE PRÁCTICA DE LABORATORIO / TALLERES / CENTROS DE SIMULACIÓN – PARA ESTUDIANTES

CARRERA: Ingeniería en Sistemas		**ASIGNATURA**: Sistemas Expertos
NRO. PRÁCTICA:	8	**TÍTULO PRÁCTICA**: KNN Clasificación de miembros del congreso utilizando algoritmos de similitud en Neo4j

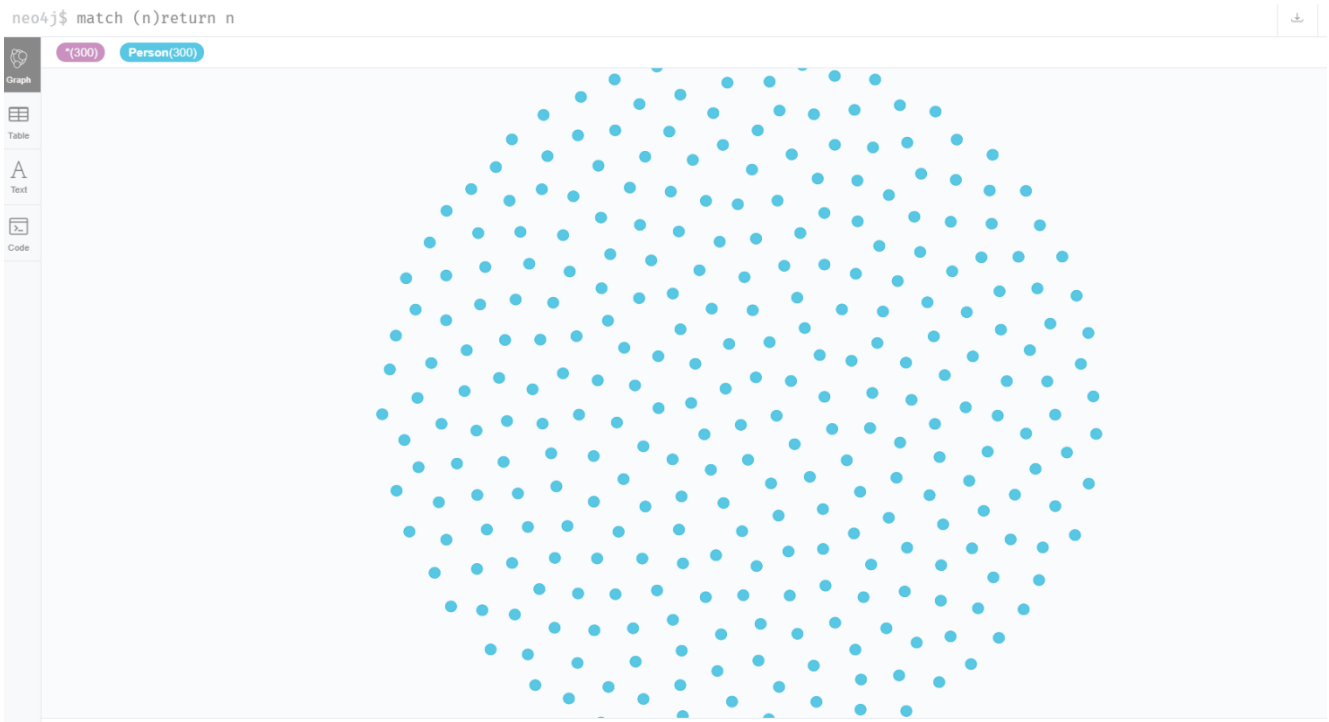
OBJETIVO ALCANZADO:

- Adquirir destrezas en el manejo de algoritmos desarrollados por neo4j.
- Entender la lógica de implementación en inteligencia artificial de los algoritmos de neo4j

ACTIVIDADES DESARROLLADAS

1. La información sobre el conjunto de datos y el conjunto de datos en sí, exportamos los datos.

```
LOAD CSV FROM "http://archive.ics.uci.edu/ml/machine-learning-databases/voting-records/house-votes-84.data" as row
CREATE (p:Person) SET p.class = row[0],
   p.features = row[1..]
```



2. Verificamos la cantidad de votos.

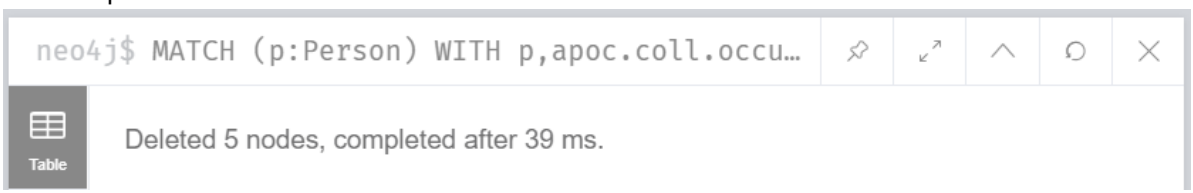
```
MATCH (n:Person)
WHERE "?" in n.features
RETURN count(n)
```

3. **Distribucion de votos faltantes**

```
MATCH (p:Person)
WHERE '?' in p.features
WITH p,apoc.coll.occurrences(p.features,'?') as missing
RETURN missing,count(*) as times ORDER BY missing ASC
```

4. Eliminar nodos que no tienen votos

```
MATCH (p:Person)
WITH p,apoc.coll.occurrences(p.features,'?') as missing
WHERE missing > 6
DELETE p
```



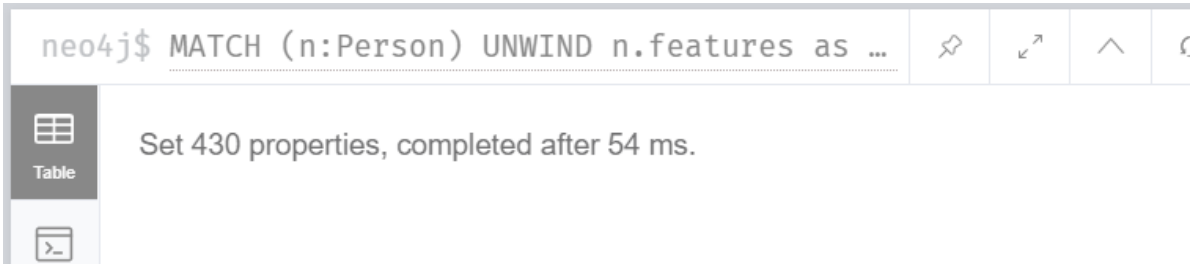
5. Dividimos el conjunto

```
MATCH (p:Person)
WITH p LIMIT 344
SET p:Training;
```


```
MATCH (p:Person)
WITH p SKIP 344
SET p:Test;
```

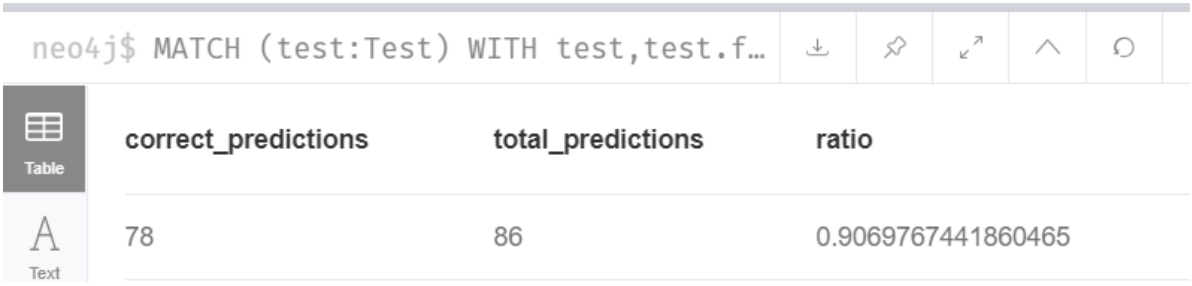
6. Transformar a vector de características

```
MATCH (n:Person)
UNWIND n.features as feature
WITH n,collect(CASE feature WHEN 'y' THEN 1
                    WHEN 'n' THEN 0
                    ELSE 0.5 END) as feature_vector
SET n.feature_vector = feature_vector
```



7. Algoritmo KNN

```
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
```



Nombre de los estudiantes: Fanny Gutama