Alumno: Carlos Andrade

Docente: Ing. Diego Quisi.

Materia: SE

Ciclo: 9no

Fecha: 15/05/2020

Creación de Nodos

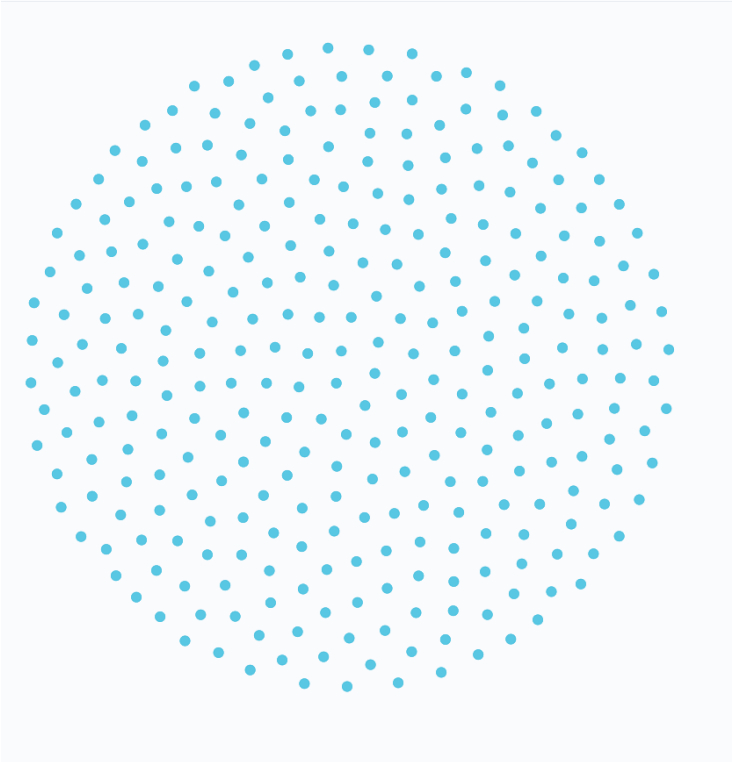
LOAD CSV FROM "http://archive.ics.uci.edu/ml/machine-learning-databases/voting-records/house-votes-84.data" as row

CREATE (p:Person)

SET p.class = row[0],

p.features = row[1..];

resultado



Todos los resultados.

MATCH (n:Person)

WHERE "?" in n.features

RETURN count(n)

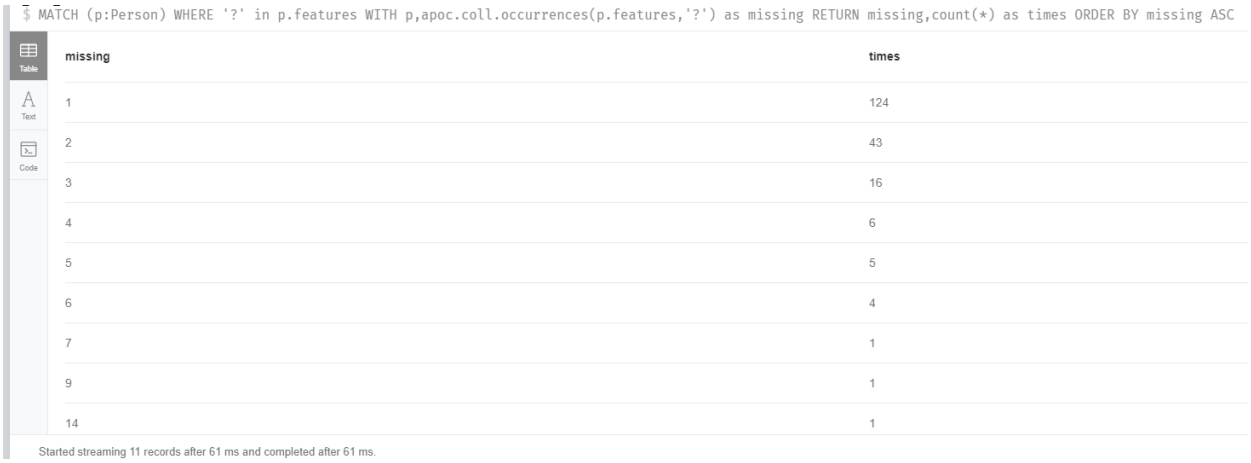
Dentro de los datos hay campos vacíos que tenemos que eliminar.

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



MATCH (p:Person)

WITH p,apoc.coll.occurrences(p.features,'?') as missing

WHERE missing > 6

DELETE p

Dependiendo de los datos de las votaciones hay que mapear estos en tres tipos.

- "y" a 1
- "n" a 0
- "?" a 0,5

Pasar los vectores a entidades

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

Implementación del algoritmo

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

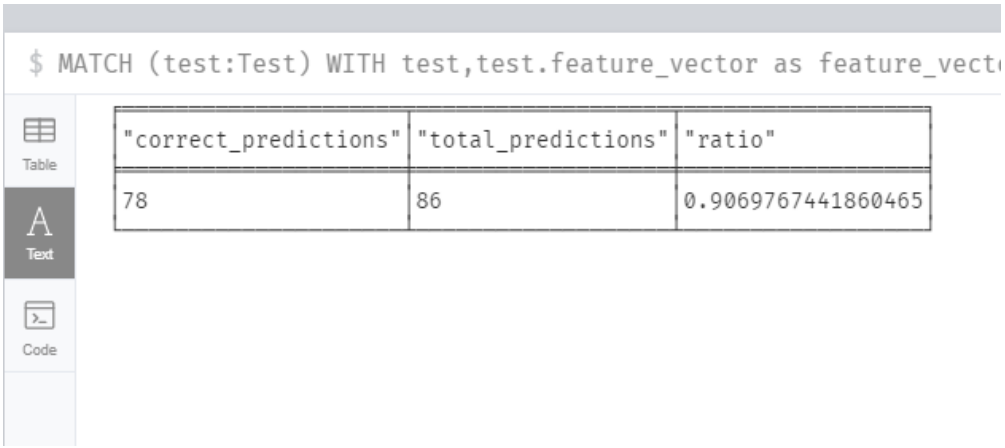
RETURN collect(training.class) as classes',

{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



Conclusión:

Es un método para obtener la similitud entre objetos con propiedades parecidas ayudándonos a clasificar los datos y hacer proyecciones de posibles resultados ya que el algoritmo toma la mayoría de las clases de vecinos y los puede ir puntuando de acuerdo con si similitud.