

Project on Similarity Search

Ankit Dhall

The query was pre-defined into variables. The categorical variables of this query were then encoded into numeric values to enable distance calculation.

This encoding was done on the basis of the encoding used in the MIF generation code.

Reference points were also defined on the basis of the reference points used in the MIF generation code.

After running the search algorithm, the nearest/most similar data objects from the input file were retrieved and stored in a csv file.

Query 2:

```
1st_Road_Class = "A"
1st_Road_Number = "3218"
Accident_Severity = "Serious"
Day_of_Week = "Tuesday"
Did_Police_Officer_Attend_Scene_of_Accident = "1"
Junction_Control = "Data missing or out of range"
Junction_Detail = "Not at junction or within 20 metres"
Light_Conditions = "Daylight"
Local_Authority_(District) = "Kensington and Chelsea"
Number_of_Casualties = "1"
Number_of_Vehicles = "1"
Pedestrian_Crossing-Human_Control = "0"
Pedestrian_Crossing-Physical_Facilities = "1"
Police_Force = "Metropolitan Police"
Road_Surface_Conditions = "Wet or damp"
Road_Type = "Single carriageway"
Speed_limit = "30"
Urban_or_Rural_Area = "Urban"
Weather_Conditions = "Raining no high winds"
Year = "2005"
```

Execution of the code was done in a Jupyter Notebook and the following query and results were obtained.

localhost:8888/notebooks/similarity_search.ipynb

jupyter similarity_search Last Checkpoint: 3 minutes ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3

```
In [53]: query
```

Out[53]:

	1st_Road_Class	1st_Road_Number	Accident_Severity	Day_of_Week	Did_Police_Officer_Attend_Scene_of_Accident	Junction_Control	Junction_Detail	Light_Condition
0	A	3218	Serious	Tuesday		1	Data missing or out of range	Not at junction or within 20 metres

```
In [54]: search_result.head(10)
```

Out[54]:

	1st_Road_Class	1st_Road_Number	Accident_Severity	Day_of_Week	Did_Police_Officer_Attend_Scene_of_Accident	Junction_Control	Junction_Detail	Light_Condition
15	A	3217.0	Slight	Tuesday	1.0	Give way or uncontrolled	T or staggered junction	Darkness
19	A	3218.0	Slight	Friday	1.0	Give way or uncontrolled	T or staggered junction	Daylight
78	A	3217.0	Slight	Friday	1.0	Give way or uncontrolled	T or staggered junction	Daylight
85	A	3220.0	Slight	Wednesday	1.0	Give way or uncontrolled	T or staggered junction	Darkness
101	A	3218.0	Slight	Monday	1.0	Data missing or out of range	Not at junction or within 20 metres	Daylight
118	A	3217.0	Slight	Monday	1.0	Give way or uncontrolled	T or staggered junction	Daylight
145	A	3217.0	Slight	Friday	1.0	Give way or uncontrolled	T or staggered junction	Daylight
154	A	3220.0	Slight	Sunday	2.0	Give way or uncontrolled	Roundabout	Darkness
155	A	3220.0	Slight	Sunday	1.0	Give way or uncontrolled	Roundabout	Daylight
170	A	3220.0	Slight	Sunday	1.0	Auto traffic signal	T or staggered junction	Daylight

These search results were stored in a CSV file along with the query

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U		
	1st_Roa1st_Road_Accident_SDay_of_Week_Did_Police_Junction_Detail_LiConLocal_Authority_(District)NumbNumberPedestPolice_Force																						
1	Query	A	3218	Serious	Tuesday	1	Data missi	Not at junction	Daylight	Kensington and Chelsea	1	1	0	1	Metropolitan Police	Wet	Single carriageway	30	Urban	Raining	no high	2005	
2																							
3																							
4		15	A	3217	Slight	Tuesday	1	Give way c	T or staggered	Darkness	Kensington and Chelsea	1	2	0	0	Metropolitan Police	Wet	One way street	30	Urban	Raining	no high	2005
5		19	A	3218	Slight	Friday	1	Give way c	T or staggered	Daylight	Kensington and Chelsea	1	2	0	0	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
6		78	A	3217	Slight	Friday	1	Give way c	T or staggered	Daylight	Kensington and Chelsea	1	2	0	1	Metropolitan Police	Wet	Single carriageway	30	Urban	Snowing	no high	2005
7		85	A	3220	Slight	Wednesday	1	Give way c	T or staggered	Darkness	Kensington and Chelsea	1	2	0	0	Metropolitan Police	Wet	Single carriageway	30	Urban	Raining	no high	2005
8		101	A	3218	Slight	Monday	1	Data missi	Not at junction	Daylight	Kensington and Chelsea	1	1	0	0	Metropolitan Police	Wet	Single carriageway	30	Urban	Fine	no high	2005
9		118	A	3217	Slight	Monday	1	Give way c	T or staggered	Daylight	Kensington and Chelsea	1	1	0	1	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
10		145	A	3217	Slight	Friday	1	Give way c	T or staggered	Daylight	Kensington and Chelsea	1	1	0	1	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
11		154	A	3220	Slight	Sunday	2	Give way c	Roundabout	Darkness	Hammersmith and Fulham	1	2	0	0	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
12		155	A	3220	Slight	Sunday	1	Give way c	Roundabout	Daylight	Hammersmith and Fulham	1	2	0	0	Metropolitan Police	Dry	Roundabout	30	Urban	Fine	no high	2005
13		170	A	3220	Slight	Sunday	1	Auto traffic	T or staggered	Daylight	Kensington and Chelsea	1	2	0	0	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
14		225	A	3220	Slight	Friday	1	Give way c	T or staggered	Daylight	Kensington and Chelsea	1	2	0	0	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
15		228	A	3217	Slight	Monday	1	Give way c	T or staggered	Daylight	Kensington and Chelsea	1	2	0	0	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
16		268	A	3220	Serious	Wednesday	1	Auto traffic	T or staggered	Daylight	Kensington and Chelsea	1	2	0	0	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
17		286	A	3220	Slight	Tuesday	1	Give way c	T or staggered	Daylight	Kensington and Chelsea	1	2	0	0	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
18		293	A	3220	Slight	Monday	1	Give way c	T or staggered	Daylight	Kensington and Chelsea	1	3	0	0	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
19		294	A	3220	Serious	Saturday	1	Auto traffic	Roundabout	Daylight	Hammersmith and Fulham	1	2	0	0	Metropolitan Police	Dry	Roundabout	30	Urban	Fine	no high	2005
20		296	A	3220	Slight	Monday	2	Auto traffic	T or staggered	Daylight	Kensington and Chelsea	1	3	0	0	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
21		317	A	3220	Slight	Thursday	1	Give way c	Private drive or	Daylight	Kensington and Chelsea	1	2	0	0	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
22		368	A	3220	Serious	Thursday	1	Auto traffic	Roundabout	Daylight	Hammersmith and Fulham	1	2	0	0	Metropolitan Police	Dry	Roundabout	30	Urban	Fine	no high	2005
23		375	A	3217	Slight	Monday	1	Give way c	T or staggered	Daylight	Kensington and Chelsea	1	1	0	4	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
24		384	A	3220	Slight	Tuesday	1	Give way c	T or staggered	Daylight	Kensington and Chelsea	1	2	0	0	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
25		391	A	3220	Slight	Saturday	1	Give way c	T or staggered	Daylight	Kensington and Chelsea	1	1	0	4	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
26		413	A	3218	Slight	Monday	1	Give way c	T or staggered	Daylight	Kensington and Chelsea	1	2	0	0	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
27		475	A	3220	Slight	Saturday	1	Give way c	T or staggered	Daylight	Hammersmith and Fulham	1	2	0	0	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
28		484	A	3220	Slight	Friday	1	Give way c	Private drive or	Daylight	Kensington and Chelsea	1	2	0	0	Metropolitan Police	Dry	Single carriageway	30	Urban	Fine	no high	2005
29		493	A	3220	Slight	Thursday	1	Give way c	T or staggered	Daylight	Kensington and Chelsea	1	2	0	0	Metropolitan Police	Wet	Single carriageway	30	Urban	Raining	no high	2005
	query2_result																						

In the above screenshot, the query has been highlighted in orange. The indices of the data objects (from the original data file) retrieved in the search results are highlighted in grey.