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 1:

1st_Road_Class = "A"
1st_Road_Number = "23"
Accident_Severity = "Slight"
 Day_of_Week = "Thursday"
Did_Police_Officer_Attend_Scene_of_Accident = "2"
Junction_Control = "Auto traffic signal"
Junction_Detail = "T or staggered junction"
Light_Conditions = "Daylight"
Local_Authority_(District) = "Lambeth"
 Number_of_Casualties = "1"
Number_of_Vehicles = "1"
Pedestrian_Crossing-Human_Control = "0"
Pedestrian_Crossing-Physical_Facilities = "0"
Police_Force = "Metropolitan Police"
Road_Surface_Conditions = "Dry"
Road_Type = "Single carriageway"
Speed_limit = "30"
Urban_or_Rural_Area = "Urban"
Weather_Conditions = "Fine no high winds"
Year = "2005"

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

Jupyter **similarity_search** Last Checkpoint: 7 minutes ago (autosaved)   Logout

File Edit View Insert Cell Kernel Widgets Help   Trusted | Python 3 ○

In [63]: `query`

Out[63]:

	1st_Road_Class	1st_Road_Number	Accident_Severity	Day_of_Week	Did_Police_Officer_Attend_Scene_of_Accident	Junction_Control	Junction_Detail	Light_C
0	A	23	Slight	Thursday	2	Auto traffic signal	T or staggered junction	

In [64]: `search_result.head(10)`

Out[64]:

	1st_Road_Class	1st_Road_Number	Accident_Severity	Day_of_Week	Did_Police_Officer_Attend_Scene_of_Accident	Junction_Control	Junction_Detail	Li
9961	A	23.0	Serious	Thursday	1.0	Data missing or out of range	Not at junction or within 20 metres	
9975	A	23.0	Slight	Saturday	1.0	Give way or uncontrolled	T or staggered junction	
9997	A	23.0	Slight	Monday	1.0	Data missing or out of range	Not at junction or within 20 metres	
10003	A	23.0	Slight	Friday	1.0	Give way or uncontrolled	T or staggered junction	
10021	A	23.0	Slight	Friday	1.0	Give way or uncontrolled	T or staggered junction	
10033	A	23.0	Slight	Friday	1.0	Data missing or out of range	Not at junction or within 20 metres	
10041	A	23.0	Serious	Monday	2.0	Auto traffic signal	T or staggered junction	
10049	A	23.0	Slight	Monday	2.0	Data missing or out of range	Not at junction or within 20 metres	
10054	A	23.0	Slight	Thursday	1.0	Give way or uncontrolled	T or staggered junction	

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

	A	B 1st_Road	C 1st_Road	D Accident_S	E Day_of_W	F Did_Po	G Junction_Co	H Junction_Detai	I Light_Cond	J Local_Auth	K Numbe	L Numbe	M Pedes	N Pedes	O Police_Force	P Roac	Q Road_Type	R Speed	S Urban	T Weather_Conditions	U Year
2	Query	A	23	Slight	Thursday	2	Auto traffic	T or staggered	Daylight	Lambeth	1	1	0	0	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
4	9961	A	23	Serious	Thursday	1	Data missing	Not at junction	Daylight	Lambeth	1	1	0	4	Metropolitan	Wet	Single carriagew	30	Urban	Fine no high winds	2005
5	9975	A	23	Slight	Saturday	1	Give way or	T or staggered	Darkness -	Lambeth	1	1	0	0	Metropolitan	Wet	Single carriagew	30	Urban	Fine no high winds	2005
6	9997	A	23	Slight	Monday	1	Data missing	Not at junction	Daylight	Lambeth	1	1	0	5	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
7	10003	A	23	Slight	Friday	1	Give way or	T or staggered	Daylight	Lambeth	1	2	0	0	Metropolitan	Wet	Single carriagew	30	Urban	Raining no high winds	2005
8	10021	A	23	Slight	Friday	1	Give way or	T or staggered	Daylight	Lambeth	1	1	0	0	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
9	10033	A	23	Slight	Friday	1	Data missing	Not at junction	Darkness -	Lambeth	1	1	0	0	Metropolitan	Wet	Single carriagew	30	Urban	Fine no high winds	2005
10	10041	A	23	Serious	Monday	2	Auto traffic	T or staggered	Daylight	Lambeth	1	1	0	5	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
11	10049	A	23	Slight	Monday	2	Data missing	Not at junction	Daylight	Lambeth	1	1	0	0	Metropolitan	Wet	Single carriagew	30	Urban	Raining no high winds	2005
12	10054	A	23	Slight	Thursday	1	Give way or	T or staggered	Darkness -	Lambeth	1	1	0	0	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
13	10073	A	23	Serious	Sunday	2	Auto traffic	T or staggered	Daylight	Lambeth	1	2	0	1	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
14	10075	A	23	Slight	Tuesday	1	Give way or	T or staggered	Daylight	Lambeth	1	2	0	5	Metropolitan	Wet	Single carriagew	30	Urban	Raining no high winds	2005
15	10095	A	23	Slight	Thursday	1	Give way or	T or staggered	Daylight	Lambeth	1	1	0	0	Metropolitan	Wet	Single carriagew	30	Urban	Raining no high winds	2005
16	10103	A	23	Slight	Friday	1	Auto traffic	T or staggered	Daylight	Lambeth	1	2	0	5	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
17	10172	A	23	Slight	Sunday	1	Data missing	Not at junction	Darkness -	Lambeth	1	2	0	0	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
18	10201	A	23	Slight	Saturday	1	Give way or	T or staggered	Daylight	Lambeth	2	2	0	0	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
19	10211	A	23	Slight	Monday	1	Data missing	Not at junction	Daylight	Lambeth	1	1	0	0	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
20	10227	A	24	Slight	Saturday	1	Auto traffic	More than 4 ar	Daylight	Lambeth	1	1	0	5	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
21	10233	A	23	Slight	Friday	2	Give way or	T or staggered	Daylight	Lambeth	1	1	0	0	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
22	10290	A	23	Slight	Friday	1	Give way or	T or staggered	Daylight	Lambeth	1	2	0	0	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
23	10293	A	23	Slight	Monday	1	Data missing	Not at junction	Daylight	Lambeth	1	2	0	0	Metropolitan	Wet	Single carriagew	30	Urban	Raining no high winds	2005
24	10298	A	23	Slight	Thursday	1	Give way or	T or staggered	Daylight	Lambeth	1	1	0	0	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
25	10302	A	23	Slight	Tuesday	1	Data missing	Not at junction	Daylight	Lambeth	1	2	0	0	Metropolitan	Dry	Dual carriagewa	30	Urban	Fine no high winds	2005
26	10319	A	24	Slight	Monday	1	Give way or	T or staggered	Daylight	Lambeth	1	3	0	0	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
27	10322	A	23	Slight	Friday	2	Auto traffic	More than 4 ar	Daylight	Lambeth	1	1	0	5	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
28	10330	A	23	Slight	Monday	1	Data missing	Not at junction	Daylight	Lambeth	1	1	0	5	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005
29	10350	A	23	Slight	Thursday	1	Data missing	Not at junction	Darkness -	Lambeth	1	1	0	0	Metropolitan	Dry	Single carriagew	30	Urban	Fine no high winds	2005

query1_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.