ΜΙΧΑΗΛ ΜΠΙΖΙΜΗΣ ΑΜ: **1115201500102**

ΚΩΝΣΤΑΝΤΙΝΤΟΣ ΤΣΙΑΡΑΣ ΑΜ: **1115201500165**

# ΤΕΧΝΙΚΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ

# ΕΡΓΑΣΙΑ 2η

# Classification report

**Κώδικας**

* To αρχείο visualization.py το οποίο πραγματοποιεί την οπτικοποίηση των 5 πρώτων διαδρομών (με διαφορετικά journey\_pattern\_id) του train\_set.csv όπως ζητείται από το Ερώτημα 1.
* Το αρχείο dtw\_neighbours.py με την υλοποίηση του Ερωτήματος 2 (Α-1).
* Το αρχείο LCSS.py με την δικιά μας υλοποίηση του αλγορίθμου LCSS και το αρχείο lcss\_neighbours.py που κάνει χρήση αυτού για την υλοποίηση του Ερωτήματος 2 (Α-2).
* Το αρχείο KNN.py με την δικιά μας υλοποίηση του KNN.
* Τα αρχεία knn\_classification.py και knn\_crossval.py τα οποία παράγουν το ζητούμενο αρχείο “testSet\_JourneyPatternIDs.csv” και πραγματοποιούν 10-fold Cross Validation αντίστοιχα, όπως περιγράφεται στο Ερώτημα 3.
* Το αρχείο util.py όπου υλοποιούνται συναρτήσεις που χρησιμοποιούνται από πολλαπλά αρχεία όπως η *harvesineDist()* και η *plotMap()*.

Για όλα τα ερωτήματα, τα datasets θεωρούμε πως υπάρχουν σε έναν φάκελο “./datasets/” στον φάκελο του παραδοτέου.

<https://pypi.org/project/fastdtw/>

**Ερώτημα 1**

**Οπτικοποίηση των Δεδομένων**

|  |  |  |
| --- | --- | --- |
| JP\_ID: 015B0001 | JP\_ID: 015B1001 | JP\_ID: 083A1001 |
| JP\_ID: 00160001 | JP\_ID: 00161001 |  |

**Ερώτημα 2 (Α-1)**

**Εύρεση κοντινότερων γειτόνων**

|  |  |  |
| --- | --- | --- |
| Test Trip 1  Δt = 67 sec | *Neighbor 1*  JP\_ID: 01501001  DTW: 0.0km | *Neighbor 2*  JP\_ID: 01501001  DTW: 3.5km |
| *Neighbor 3*  JP\_ID: 01501001  DTW: 3.8km | *Neighbor 4*  JP\_ID: 01501001  DTW: 4km | *Neighbor 5*  JP\_ID: 01501001  DTW: 4.1km |

|  |  |  |
| --- | --- | --- |
| Test Trip 2  Δt = 41 sec | *Neighbor 1*  JP\_ID: 01200001  DTW: 0.0km | *Neighbor 2*  JP\_ID: 01200001  DTW: 2.8km |
| *Neighbor 3*  JP\_ID: 01200001  DTW: 3.4km | *Neighbor 4*  JP\_ID: 01200001  DTW: 3.4km | *Neighbor 5*  JP\_ID: 01200001  DTW: 3.5km |

|  |  |  |
| --- | --- | --- |
| Test Trip 3  Δt = 51 sec | *Neighbor 1*  JP\_ID: 00791001  DTW: 0.0km | *Neighbor 2*  JP\_ID: 00791001  DTW: 4.7km |
| *Neighbor 3*  JP\_ID: 00791001  DTW: 4.8km | *Neighbor 4*  JP\_ID: 00791001  DTW: 4.8km | *Neighbor 5*  JP\_ID: 00791001  DTW: 4.9km |

|  |  |  |
| --- | --- | --- |
| Test Trip 4  Δt = 42 sec | *Neighbor 1*  JP\_ID: 00010002  DTW: 0.0km | *Neighbor 2*  JP\_ID: 00010002  DTW: 2.5km |
| *Neighbor 3*  JP\_ID: 00010002  DTW: 2.8km | *Neighbor 4*  JP\_ID: 00010002  DTW: 3.2km | *Neighbor 5*  JP\_ID: 00010002  DTW: 3.5km |

|  |  |  |
| --- | --- | --- |
| Test Trip 5  Δt = 40 sec | *Neighbor 1*  JP\_ID: 01300001  DTW: 0.0km | *Neighbor 2*  JP\_ID: 01300001  DTW: 4.3km |
| *Neighbor 3*  JP\_ID: 01300001  DTW: 4.5km | *Neighbor 4*  JP\_ID: 01300001  DTW: 4.6km | *Neighbor 5*  JP\_ID: 01300001  DTW: 4.7km |

**Ερώτημα 2 (Α-2)**

**Εύρεση κοντινότερων υποδιαδρομών**

**Ερώτημα 3**

**Κατηγοριοποίηση**

Accuracy: 0.985411