**PYTHON**

**Final Project Instructions**

1. A dataset (~ 200 rows) will be provided to you with unstructured / partial address information. You will write Python codes that identifies, matches / predicts, standardizes the address to

(a) house/street nbr

(b) locality/colony,

(c) area/ward,

(d) city

(e) pincode

1. Students will need to complete each address based on hierarchy of information, and use heuristics to fill in missing information.  For example, if colony is known, we will typically know the area/city and pincode - even if not specified in the address.  For example, if the original address string only has "jwalamukhi hostel" - the output should resolve to "IIT Delhi, Hauz Khas, New Delhi, 110016".   However, if only house number is known (e.g. house# 1212), and no other fields are known, then we can't determine any subsequent fields (colony / area / city / pincode), so they will be null.

So output that you need to provide is a csv file with the following fields:

* Original input string
* Best predicted full address
* Best predicted address part 1 - House / Street nbr
* Best predicted address part 2 - Locality / Colony
* Best predicted address part 3 - Area / Ward
* Best predicted address part 4 - City
* Best predicted address part 5 - Pincode

1. Here are the metrics that will be used to evaluate the performance of the code:

If any of the address parts 1 to 5 is not null, we will assume that the code / algorithm has determined it can predict something - so it will be a positive prediction.  If all the address parts  1 to 5 are null, we will assume that the code / algorithm has determined it cannot predict - so it will be a negative prediction.  Below are the examples (1 represents positive and 0 represents negative):

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Address string** | **Predicted House / Street nbr (part 1)** | **Predicted locality / colony (part 2)** | **Predicted area / ward (part 3)** | **Predicted City (part 4)** | **Predicted Pincode (part 5)** | **Outcome** |
| jwalamukhi hostel, delhi | Jwalamukhi Hostel | IIT Delhi | Hauz Khas | New Delhi | 110016 | 1 |
| house nbr 1202, delhi | House Nbr 1022 |  |  | New Delhi |  | 1 |
| house nbr 1202, delhi |  |  |  |  |  | 0 |
| hauz nbr 1202, delhi |  |  | Hauz Khas | New Delhi |  | 1 |
| zxdasd |  |  |  |  |  | 0 |

The proportion of positive prediction will be considered as an evaluation criterion for this problem.