

This document describes the outline for the homogeneous multiplex generator from given comma separated format of a dataset.

INPUT 1 (Data File): Structure of the Dataset with N records and K features

- The data file should be in a csv (comma separated) format
- **Line 1:** The first line of the file must specify the **feature names**
 - **Format:** <featureName₁, featureName₂, ..., featureName_K>
- **Line 2:** The second line specifies the **feature type**
 - **Format:** <featureType₁, featureType₂, ..., featureType_K>
 - What are the different types of features that must be handled?
 - **Identifier/Primary Key**
 - The unique identifier corresponding to the record
 - Example – *University ID*: asd3421
 - **Nominal/Categorical**
 - The attributes that have a fixed domain to which the values can belong
 - Example – *Political Belief of a person*: Democratic
 - **Numeric**
 - It can be integer or floating point
 - Example – *Speed of a Vehicle in mph*: 75
 - **Geographical Location**
 - **Latitude**
 - **Longitude**
 - **Time**
 - Format: 24 HRS
 - Example – *Time of Accident*: 1330
 - **Date**
 - Format: MM-DD-YYYY
 - Example – *Date of Movie Release*: 09-19-2018
 - **Set**
 - Format for a set with m values: {value₁, value₂, ... value_m}
 - Example – *Genre of a Movie*: {ACTION, COMEDY, DRAMA}
 - **Text**
 - Format: [<text>]
 - Example1 - *Tweet*: [Congratulations, France #FIFA2018 #Winners]
 - Example2 – *IMDB Review*: [The movie was crap]
- **(N lines) Line 3 to Line N+2:** Each line corresponds to the comma separated data record

INPUT 2 (Configuration File)

- **Line 1: # Dataset Path**
- **Line 2:** The path to the dataset must be specified here
- **Line 3: # Number of Nodes**
- **Line 4:** Specify the number of nodes per layer i.e. the number of records in the file

- **Line 5: # Number of Features**
- **Line 6:** Specify the number of features present in the dataset
- **Line 7: # Number of Layers**
- **Line 8:** Specify the number of layers that need to be generated for this layer, say L
- **(L lines) Line 9 to Line $L+8$:** Each line from here specifies the configuration for the layer to be generated. The configuration will depend on feature type and similarity metric. If a layer has to be generated for the i^{th} feature, then depending on the feature type the configuration needs to be provided in one of the following formats
 - **Nominal**
 - **NOMINAL,< i >,METRIC_EQUALITY**
 - **Numeric**
 - **NUMERIC,< i >,<measurement unit of feature i >,METRIC_THRESHOLD,EUCLIDEAN,<threshold value>**
 - **NUMERIC,< i >,<measurement unit of feature i >,METRIC_FIXEDRANGE,<number of ranges, say m >,< r_1 >,< r_2 >,< r_3 >,< r_4 >,< r_5 >,< r_6 >,< r_7 >,< r_8 >,< r_9 >,< r_{10} >,< r_{11} >,< r_{12} >,< r_{13} >,< r_{14} >,< r_{15} >,< r_{16} >,< r_{17} >,< r_{18} >,< r_{19} >,< r_{20} >,< r_{21} >,< r_{22} >,< r_{23} >,< r_{24} >,< r_{25} >,< r_{26} >,< r_{27} >,< r_{28} >,< r_{29} >,< r_{30} >,< r_{31} >,< r_{32} >,< r_{33} >,< r_{34} >,< r_{35} >,< r_{36} >,< r_{37} >,< r_{38} >,< r_{39} >,< r_{40} >,< r_{41} >,< r_{42} >,< r_{43} >,< r_{44} >,< r_{45} >,< r_{46} >,< r_{47} >,< r_{48} >,< r_{49} >,< r_{50} >,< r_{51} >,< r_{52} >,< r_{53} >,< r_{54} >,< r_{55} >,< r_{56} >,< r_{57} >,< r_{58} >,< r_{59} >,< r_{60} >,< r_{61} >,< r_{62} >,< r_{63} >,< r_{64} >,< r_{65} >,< r_{66} >,< r_{67} >,< r_{68} >,< r_{69} >,< r_{70} >,< r_{71} >,< r_{72} >,< r_{73} >,< r_{74} >,< r_{75} >,< r_{76} >,< r_{77} >,< r_{78} >,< r_{79} >,< r_{80} >,< r_{81} >,< r_{82} >,< r_{83} >,< r_{84} >,< r_{85} >,< r_{86} >,< r_{87} >,< r_{88} >,< r_{89} >,< r_{90} >,< r_{91} >,< r_{92} >,< r_{93} >,< r_{94} >,< r_{95} >,< r_{96} >,< r_{97} >,< r_{98} >,< r_{99} >,< r_{100} >,< r_{101} >,< r_{102} >,< r_{103} >,< r_{104} >,< r_{105} >,< r_{106} >,< r_{107} >,< r_{108} >,< r_{109} >,< r_{110} >,< r_{111} >,< r_{112} >,< r_{113} >,< r_{114} >,< r_{115} >,< r_{116} >,< r_{117} >,< r_{118} >,< r_{119} >,< r_{120} >,< r_{121} >,< r_{122} >,< r_{123} >,< r_{124} >,< r_{125} >,< r_{126} >,< r_{127} >,< r_{128} >,< r_{129} >,< r_{130} >,< r_{131} >,< r_{132} >,< r_{133} >,< r_{134} >,< r_{135} >,< r_{136} >,< r_{137} >,< r_{138} >,< r_{139} >,< r_{140} >,< r_{141} >,< r_{142} >,< r_{143} >,< r_{144} >,< r_{145} >,< r_{146} >,< r_{147} >,< r_{148} >,< r_{149} >,< r_{150} >,< r_{151} >,< r_{152} >,< r_{153} >,< r_{154} >,< r_{155} >,< r_{156} >,< r_{157} >,< r_{158} >,< r_{159} >,< r_{160} >,< r_{161} >,< r_{162} >,< r_{163} >,< r_{164} >,< r_{165} >,< r_{166} >,< r_{167} >,< r_{168} >,< r_{169} >,< r_{170} >,< r_{171} >,< r_{172} >,< r_{173} >,< r_{174} >,< r_{175} >,< r_{176} >,< r_{177} >,< r_{178} >,< r_{179} >,< r_{180} >,< r_{181} >,< r_{182} >,< r_{183} >,< r_{184} >,< r_{185} >,< r_{186} >,< r_{187} >,< r_{188} >,< r_{189} >,< r_{190} >,< r_{191} >,< r_{192} >,< r_{193} >,< r_{194} >,< r_{195} >,< r_{196} >,< r_{197} >,< r_{198} >,< r_{199} >,< r_{200} >,< r_{201} >,< r_{202} >,< r_{203} >,< r_{204} >,< r_{205} >,< r_{206} >,< r_{207} >,< r_{208} >,< r_{209} >,< r_{210} >,< r_{211} >,< r_{212} >,< r_{213} >,< r_{214} >,< r_{215} >,< r_{216} >,< r_{217} >,< r_{218} >,< r_{219} >,< r_{220} >,< r_{221} >,< r_{222} >,< r_{223} >,< r_{224} >,< r_{225} >,< r_{226} >,< r_{227} >,< r_{228} >,< r_{229} >,< r_{230} >,< r_{231} >,< r_{232} >,< r_{233} >,< r_{234} >,< r_{235} >,< r_{236} >,< r_{237} >,< r_{238} >,< r_{239} >,< r_{240} >,< r_{241} >,< r_{242} >,< r_{243} >,< r_{244} >,< r_{245} >,< r_{246} >,< r_{247} >,< r_{248} >,< r_{249} >,< r_{250} >,< r_{251} >,< r_{252} >,< r_{253} >,< r_{254} >,< r_{255} >,< r_{256} >,< r_{257} >,< r_{258} >,< r_{259} >,< r_{260} >,< r_{261} >,< r_{262} >,< r_{263} >,< r_{264} >,< r_{265} >,< r_{266} >,< r_{267} >,< r_{268} >,< r_{269} >,< r_{270} >,< r_{271} >,< r_{272} >,< r_{273} >,< r_{274**