Michael Cummings

11/21/2022

CS-453 Data Mining

Data Mining Questions for Course Project

Using NOAA Sea Level Trends dataset, with 142 instances of 13 attributes:

* ~~station\_id (nominal)~~
* station\_name (nominal)
* ~~first\_year (time)~~
* ~~last\_year (time)~~
* year\_range (time)
* complete (numerical)
* msl\_trends\_mm\_yr (numerical)
* X95\_ci\_mm\_yr (numerical)
* msl\_trends\_ft\_century (numerical)
* X95\_ci\_ft\_century (numerical)
* Latitude (Geospatial)
* Longitude (Geospatial)
* ~~Location (Geospatial)~~

Questions:

1. Can we associate a geographical region with a rising or falling trend in sea level?
2. Can we cluster different stations together based on how severe their change in sea level is (rising or falling)?
3. Can we associate how many years a station has been in operation with its percentage of complete data?
4. Can we predict how trends in sea level will continue? Can we answer that question specific to geographical region?
5. Can we classify an instance with a certain value of sea level change to a certain geographical region? For example, if we had an instance that was missing its location data, but had the numerical attributes present, to what degree of accuracy could we classify that instance to a particular region of the world?
6. Can we cluster together geographical regions based on how many stations there are within a given area? For example, Hawaii / The West Coast could be a cluster, since there at least 5 stations within the state of Hawaii.