Chemical Entities

Caratterizzazione generale dei composti con lo scopo di codificarne le proprietà in poche *caratteristiche portanti.*

* Reactivity
* Stability
* Polarity
* Steric property
* flexibility/rigidity
* Size
* Structural

**Input**:

**Target molecule** and **Training dataset**.

Il processo è unsupervised dato che si vuole costruire una metodologia per descrivere i composti in poche caratteristiche salienti. Il Training set, infatti, può essere un qualsiasi insieme di molecole a prescindere dalla conoscenza del valore sperimentale.

**Processo**:

Si proietta ogni singolo composto del training in diversi sottospazi. Ogni sottospazio è definito da pochi descrittori selezionati e rappresenta la caratteristica i-esima, esempio **Polarity**:

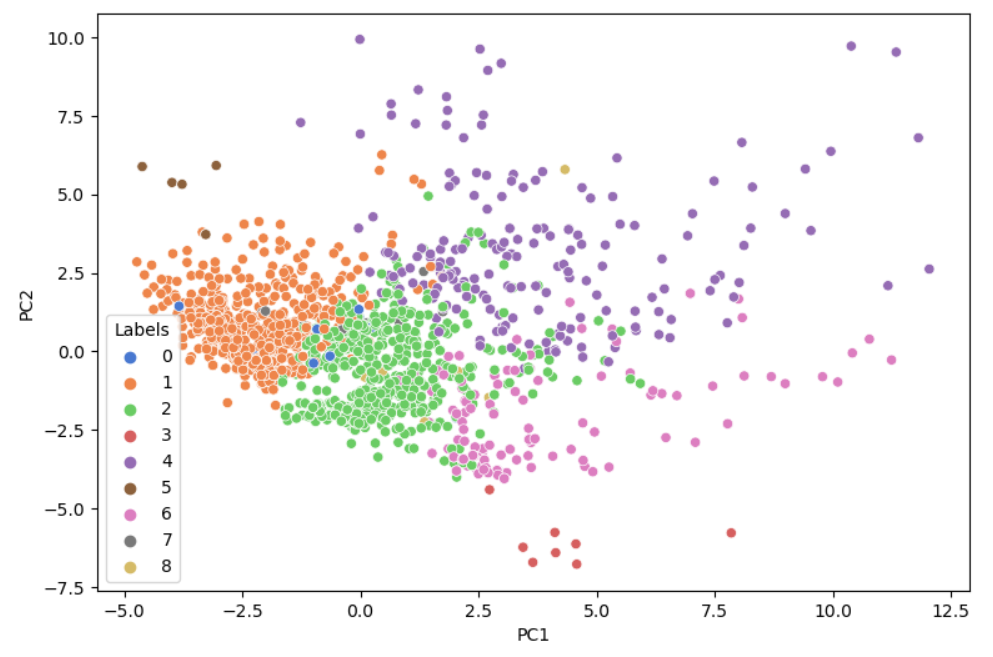
* nBase
* nAcid
* apol
* bpol
* TopoPSA
* WPol

Per ogni sottospazio chimico si effettua una cluster analysis sui dati di training. In questo modo si possono trovare dei “cluster di reazione”, “cluster strutturali”.. etc.

Per il target saranno calcolati gli stessi descrittori e sarà proiettato in ognuno dei sottospazi. Per ogni sottospazio il target potrà appartenere a 1 o più classi in relazione al valore dei descrittori calcolati. L’appartenenza al cluster j-esimo della caratteristica i-esima permette di originare una descrizione binaria 0-1 simile ad una fingerprint.

1. **Analisi Sottospazio: Reattività**

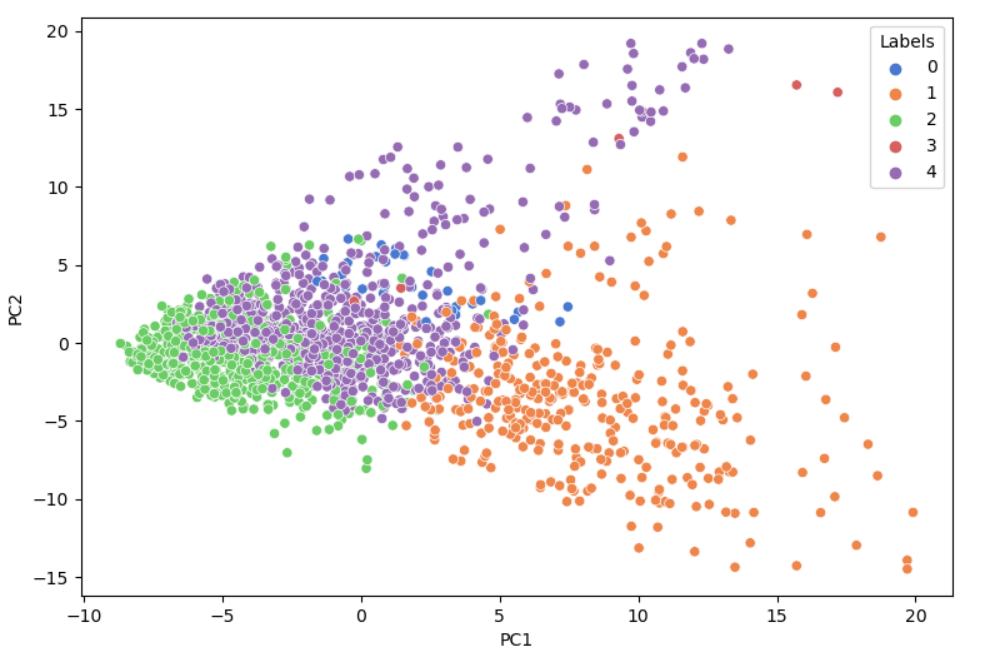
**Test 1: SyGMA smirks**

****

1. Analisi Sottospazio Strutturale

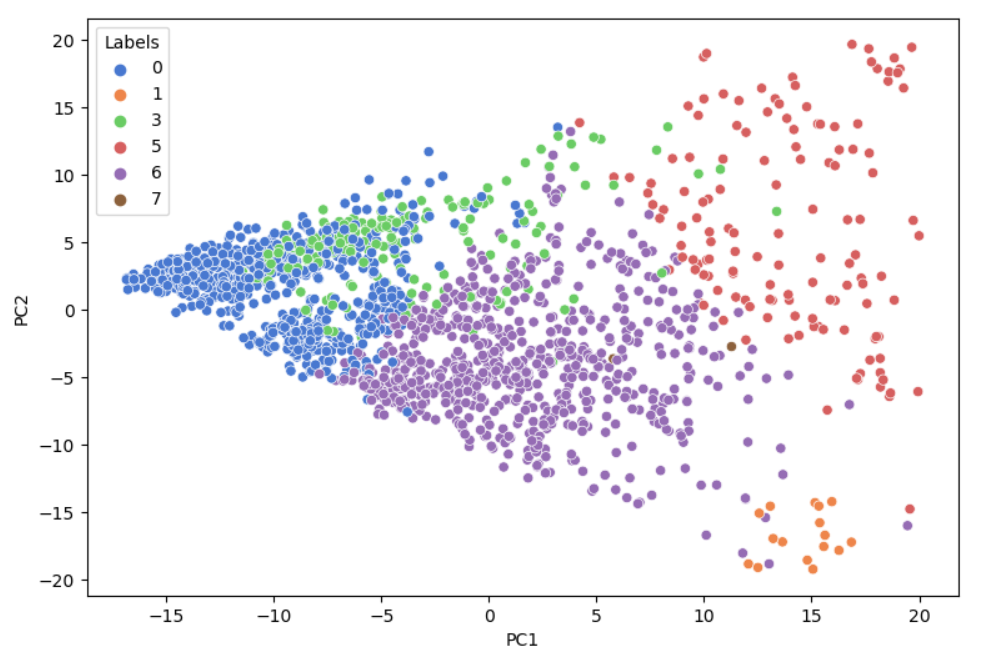
**Test 1:**

**MACCS fingerprint + Circular Fingerprint**

****

**Test 2:**

**MACCS Fingerprint + Circular Fingerprint + Mol2Vec**

****

**Working progress…**