

# ADENINE — A Data Exploration pIpeliNE

**Samuele Fiorini**  
**Federico Tomasi**  
**Annalisa Barla**

SAMUELE.FIORINI@DIBRIS.UNIGE.IT  
FEDERICO.TOMASI@DIBRIS.UNIGE.IT  
ANNALISA.BARLA@UNIGE.IT

*Department of Informatics, Bioengineering,  
Robotics and System Engineering (DIBRIS)  
University of Genoa  
Genoa, I-16146, Italy*

**Editor:** Editor name

## Abstract

Abstract here.

**Keywords:** Exploratory data analysis, unsupervised learning, dimensionality reduction, clustering

## 1. Introduction

## 2. Implementation

From an algorithmic standpoint, **adenine** is built upon the concept of *pipeline* i.e. a sequence of four fundamental steps: *i*) missing values imputing, *ii*) data preprocessing, *iii*) dimensionality reduction and *iv*) clustering.

## 3. Experiments and results

To assess the quality of the obtained results, we tested **adenine** on a set of synthetic and real dataset.

**{parla qui dei test synth} {TGCA}**

## 4. Conclusions

Table 1: Pipelines building blocks and relative references (which are not reported when the definition is given in Section 2).

Step	Algorithms	Ref.
Imputing	Mean	
	Median	
	KNN	(?)
Preprocessing	Recentring	
	Standardize	
	Normalize	
	MinMax	
Dimensionality reduction	Principal component Analysis (PCA)	(?)
	Incremental PCA	(?)
	Randomized PCA	(?)
	Kernel PCA	(?)
	Isomap	(?)
	Locally linear embedding	(?)
	Spectral embedding	(?)
	Multidimensional scaling	(?)
	t-Distributed Stochastic Neighbor Embedding (t-SNE)	(?)
Clustering	K-means	(?)
	Affinity propagation	(?)
	Mean Shift	(?)
	Spectral	(?)
	Hierarchical	(?)