Unsupervised Learning Report

**Rishabh Kaushick**

NU ID: 002808996

College of Engineering

Northeastern University

Toronto, ON

kaushick.r@northeastern.edu

# Assignment 3

# Abstract

|  |  |  |
| --- | --- | --- |
| Dataset # | Dataset | Unsupervised Learning Algorithm |
| Dataset #1 | Near Earth Objects | k-Means Clustering |
| Gaussian Mixture Model (GMM) |
| Principal Component Analysis (PCA) |
| Independent Component Analysis (ICA) |
| One more feature selection Algo? |
| Dataset #2 | Heart Attack Risk Prediction | k-Means Clustering |
| Gaussian Mixture Model (GMM) |
| Principal Component Analysis (PCA) |
| Independent Component Analysis (ICA) |
| One more feature selection Algo? |

Table 1: Datasets & algorithms covered.

# Datasets

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dataset | Name | Dataset Characteristics | Attribute Characteristics | Associated Task | Number of Instances | Number of Attributes |
| 1 | NASA Near Earth Objects (NEOs) | Multivariate | Real | Clustering & Dimensionality Reduction | 904 | 27 |
| 2 | Heart Attack Risk Prediction | Multivariate | Real | Clustering & Dimensionality Reduction | 8764 | 25 |

Table 2: Dataset Details

# Exploratory Data Analysis (EDA)

Defining the Problem

Data Characteristics

Diving Into Data