Diagram

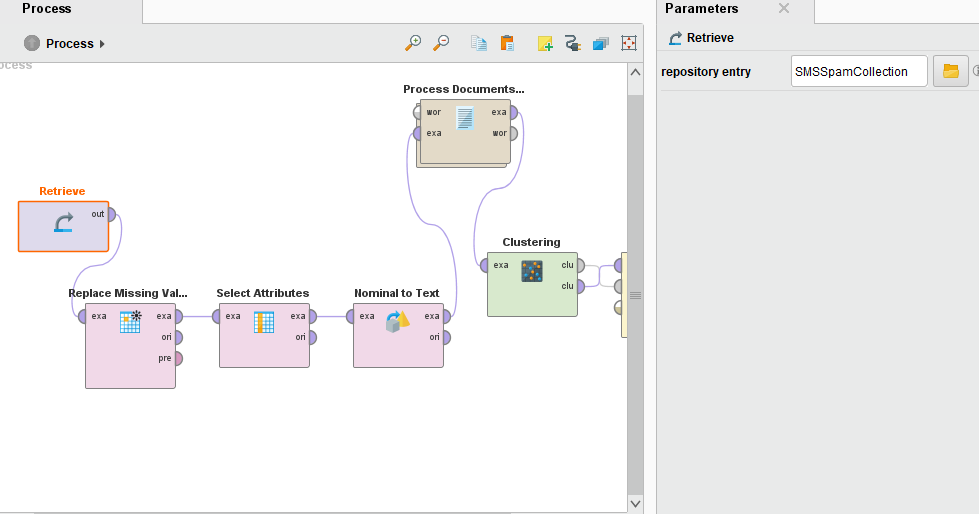
Description automatically generated**SCREENSHOT OF IMPLEMENATION OF K-MEANS CLUSTERING ON SMSSpamCollection Dataset on RapidMiner**

Diagram

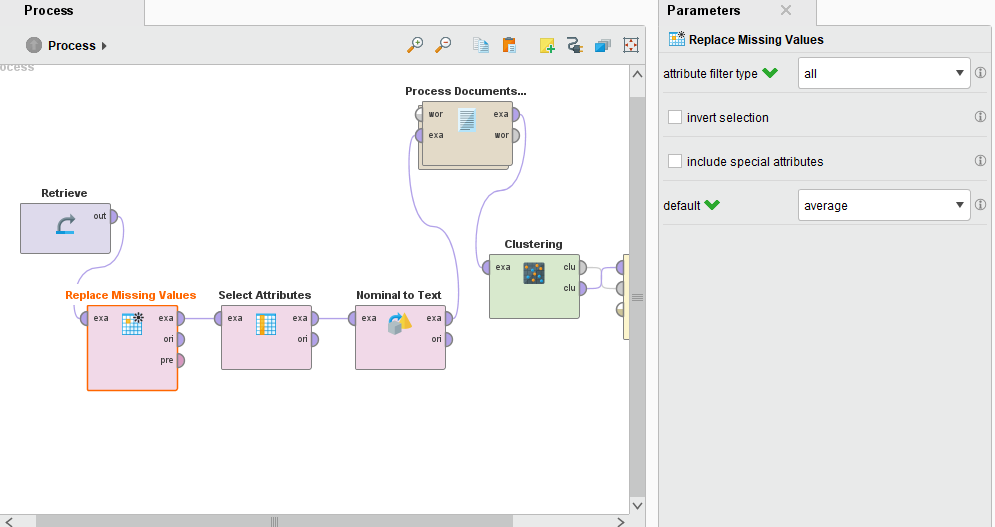
Description automatically generated

**SCREENSHOT OF EACH OPERATOR ARE LISTED BELOW:**

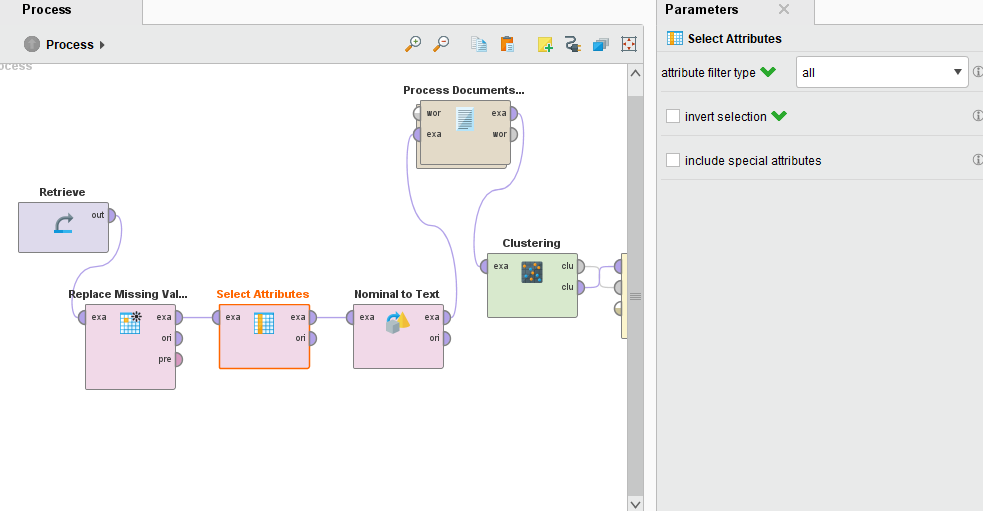
**Retrieve Operator:**



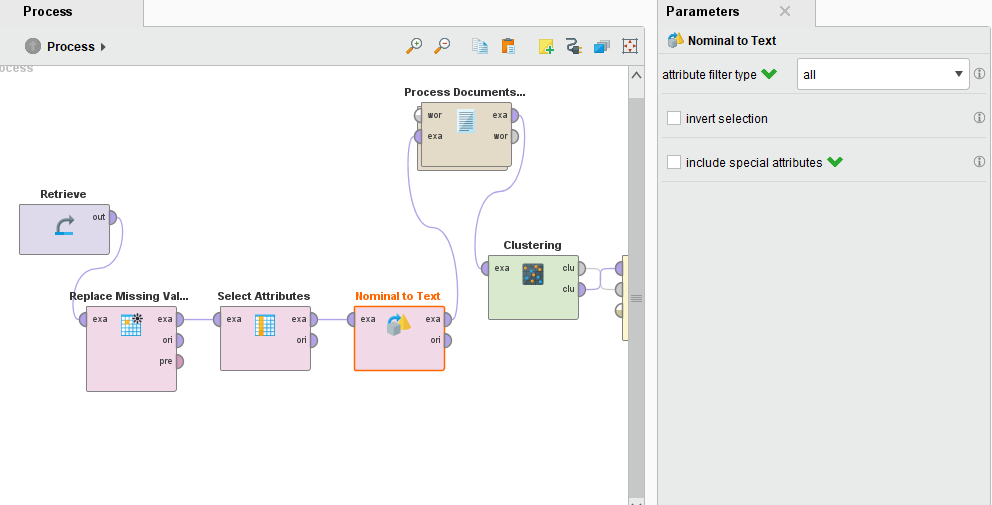
**Replace Missing Value Operator:**



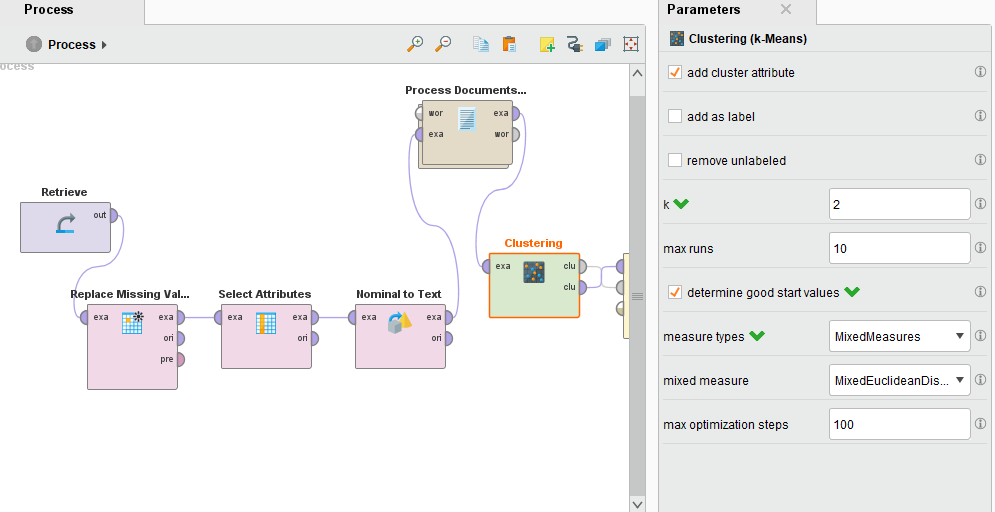
**Select Attributes Operator:**



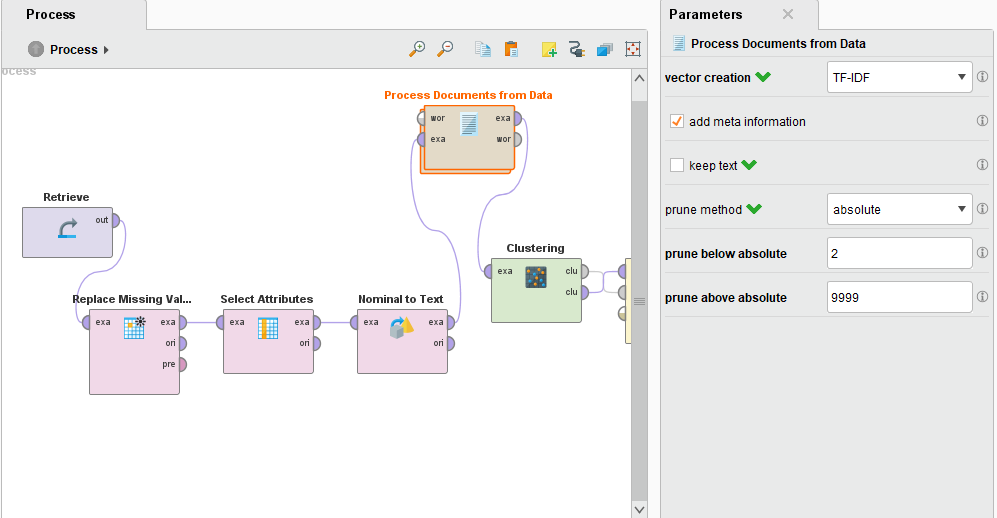
**Nominal to Text Operator:**



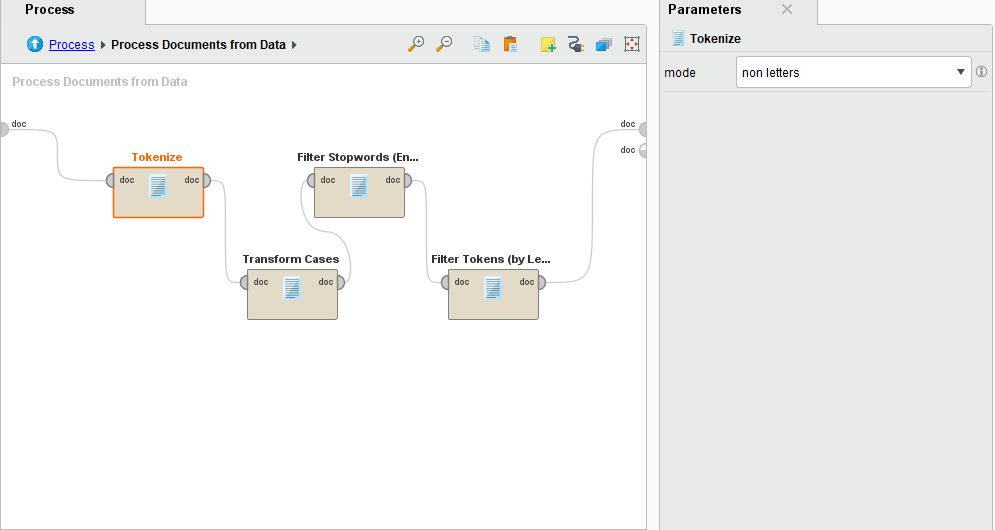
**Clustering (K-MEANS) Operator:**



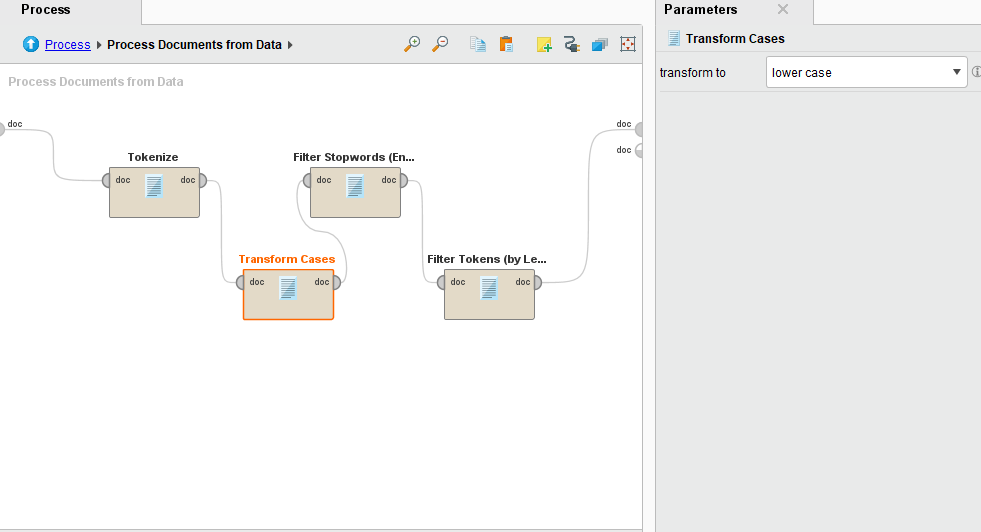
**Process Documents from Data Operator:**



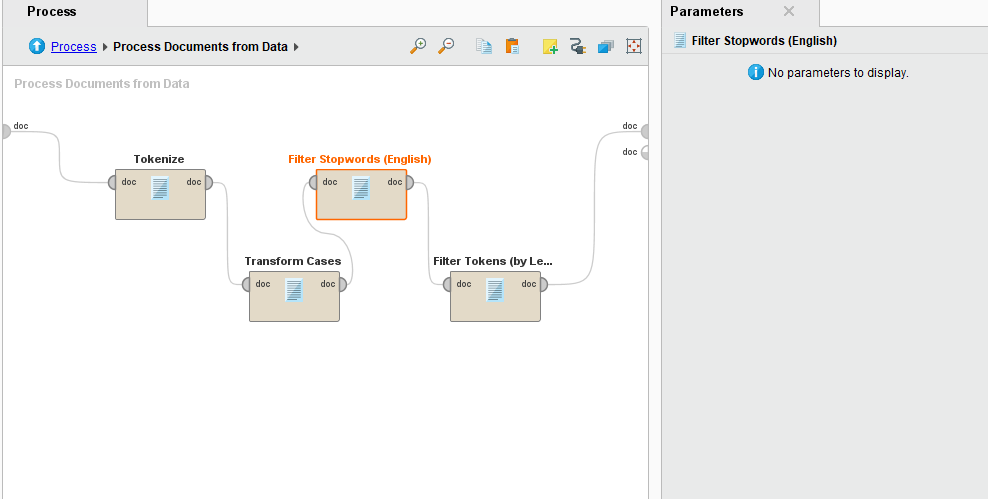
**Process Documents from Data Operator 🡪 Tokenize Operator:**



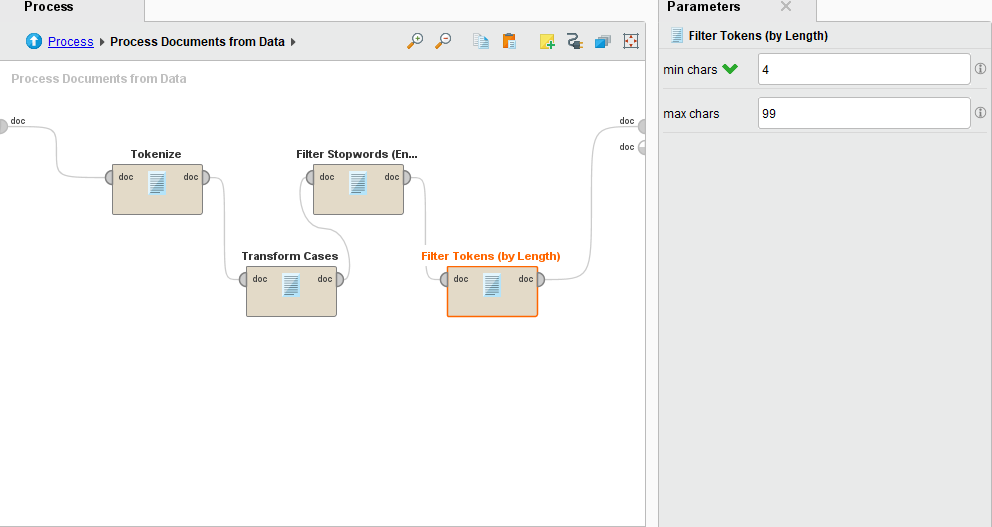
**Process Documents from Data Operator 🡪 Transform Cases Operator:**



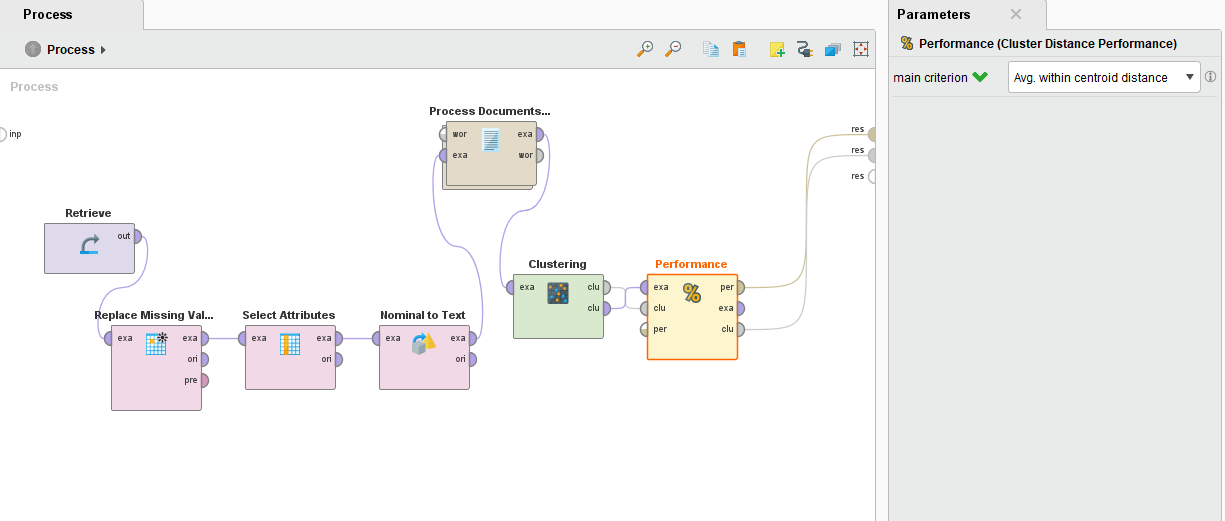
**Process Documents from Data Operator 🡪 Filter Stop words (English) Operator:**



**Process Documents from Data Operator 🡪 Filter Tokens (by Length) Operator:**



**Cluster Distance Performance Operator:**



**Results:**

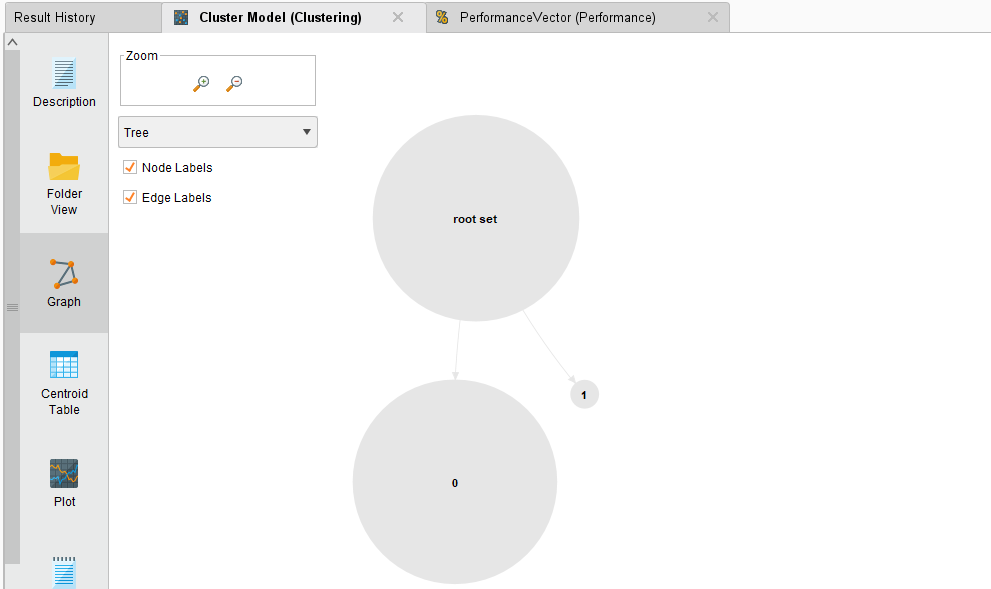
Cluster 0 is ham

Cluster 1 is spam

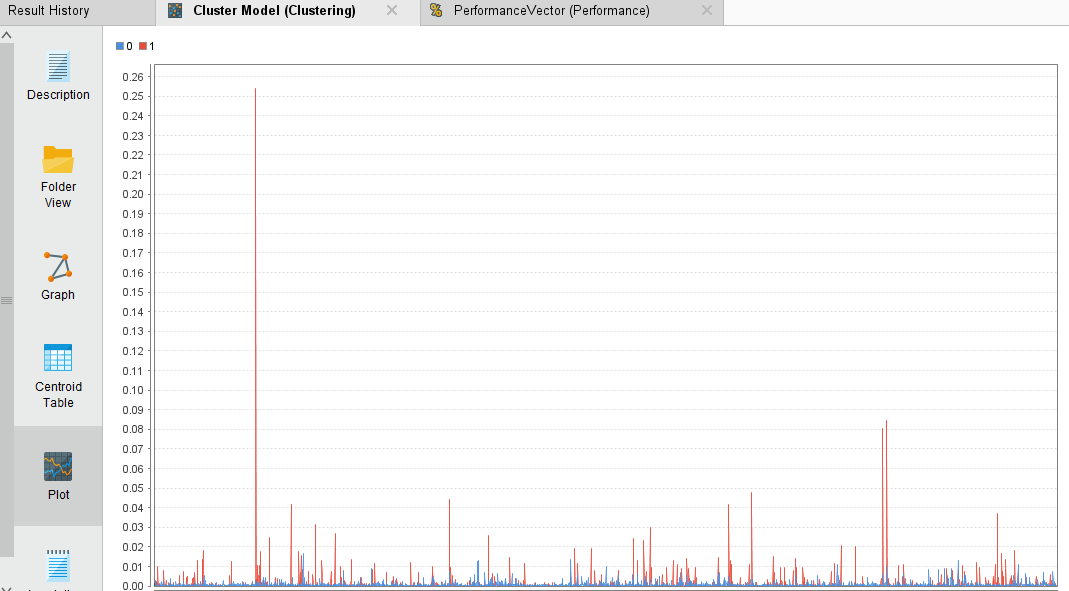
Graphical user interface, text, application

Description automatically generated

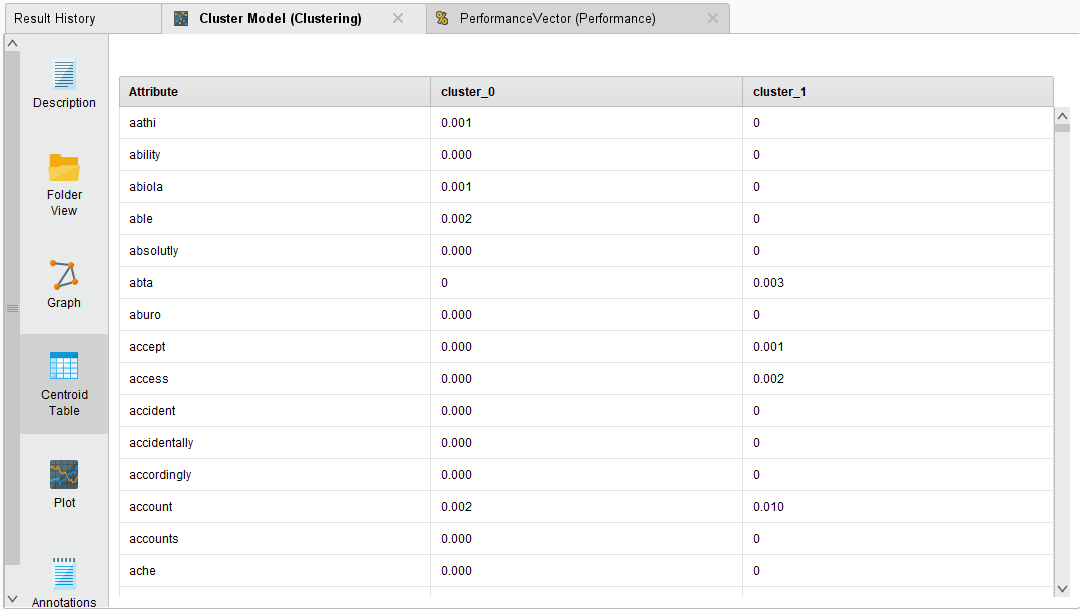
**Graph View:**



**Plot View:**



**Centroid Table (Some of them display in a following picture):**



**Performance Vector by taking average of both clusters distance:**

Graphical user interface, text, application, Word

Description automatically generated

**Performance Vector of cluster 0(ham):**

Graphical user interface, text, application, email

Description automatically generated

**Performance Vector of cluster 1(spam):**

Graphical user interface, text, application, email

Description automatically generated