S1 - HOPE PCA_v2

November 15, 2020

0.1 Import data CSV.

[1]:
```
# pip install pymysql
from sqlalchemy import create_engine
import pymysql
import pandas as pd
import numpy as np
```

[2]:
```
dfPCA = pd.read_csv('hope_dataset_cleaned.csv')
```

[3]:
```
dfPCA.head(10)
```

[3]:

	pedido.data.attributes.age	pedido.data.attributes.diagnostic_main	pedido.data.attributes.gender	articulo	respuesta.articlesRevisedYear	respuesta.articlesRevisedMonth
0	75.0	FISTULA PERITONEAL	male	27395425	2018	1
1	75.0	FISTULA PERITONEAL	male	28560554	2018	
2	75.0	FISTULA PERITONEAL	male	28641726	2017	
3	75.0	FISTULA PERITONEAL	male	26245344	2016	
4	75.0	FISTULA PERITONEAL	male	28942543	2018	
5	75.0	FISTULA PERITONEAL	male	24782153	2014	
6	75.0	FISTULA PERITONEAL	male	28002229	2018	
7	75.0	FISTULA PERITONEAL	male	27505109	2017	
8	75.0	FISTULA PERITONEAL	male	24850546	2015	
9	75.0	FISTULA PERITONEAL	male	29371050	2019	

```
1                                4
2                               12
3                               12
4                                6
5                                6
6                                9
7                                4
8                                1
9                                4
```

```
                                 respuesta.pubmed_keys  utilidad
0  Abdomen,Adenocarcinoma,Antiemetics,Blood Cultu…       1.0
1  Abdomen,Adenocarcinoma,Antiemetics,Blood Cultu…       NaN
2  Abdomen,Adenocarcinoma,Antiemetics,Blood Cultu…       NaN
3  Abdomen,Adenocarcinoma,Antiemetics,Blood Cultu…       NaN
4  Abdomen,Adenocarcinoma,Antiemetics,Blood Cultu…       NaN
5  Abdomen,Adenocarcinoma,Antiemetics,Blood Cultu…       NaN
6  Abdomen,Adenocarcinoma,Antiemetics,Blood Cultu…       NaN
7  Abdomen,Adenocarcinoma,Antiemetics,Blood Cultu…       NaN
8  Abdomen,Adenocarcinoma,Antiemetics,Blood Cultu…       NaN
9  Abdomen,Adenocarcinoma,Antiemetics,Blood Cultu…       NaN
```

Get only data with 'utilidad' value defined

[4]:
```
dfPCA = dfPCA[pd.notnull(dfPCA['utilidad'])]

print(dfPCA.shape[0])
```

```
51
```

[5]:
```
dfPCA.head(10)
```

[5]:
```
     pedido.data.attributes.age pedido.data.attributes.diagnostic_main  \
0                          75.0                     FISTULA PERITONEAL
32                         75.0                     FISTULA PERITONEAL
230                        36.0             INSUFICIENCIA RESPIRATORIA
290                        51.0                        POLITRAUMATISMO
299                        51.0                        POLITRAUMATISMO
300                        18.0                          ABDOMEN AGUDO
303                        18.0                          ABDOMEN AGUDO
304                        18.0                          ABDOMEN AGUDO
305                        18.0                          ABDOMEN AGUDO
311                        76.0                            TORACOTOMIA

    pedido.data.attributes.gender  articulo  respuesta.articlesRevisedYear  \
0                            male  27395425                           2018
32                           male  28694230                           2017
```

```
230                          male  28805236                          2011
290                          male  27537587                          2011
299                          male  28148670                          2019
300                          male  25055513                          2019
303                          male  29279563                          2017
304                          male  29279563                          2017
305                          male  28065368                          2017
311                          male  30762794                          2019

     respuesta.articlesRevisedMonth  \
0                                  1
32                                12
230                                3
290                                3
299                                3
300                                3
303                                2
304                                2
305                               11
311                                3

                                    respuesta.pubmed_keys  utilidad
0    Abdomen,Adenocarcinoma,Antiemetics,Blood Cultu…       1.0
32   Abdomen,Blood Culture,Catharsis,Diuresis,Drug …       1.0
230  Abdomen,Analgesics,Antitubercular Agents,Cipro…       0.0
290  Abdomen,Analgesics,Bone,Catharsis,Electroconvu…       0.0
299  Abdomen,Analgesics,Bone,Catharsis,Electroconvu…       1.0
300  Abdomen,Anti-Bacterial Agents,Diuresis,Operati…       1.0
303  Abdomen,Anti-Bacterial Agents,Diuresis,Operati…       0.0
304  Abdomen,Anti-Bacterial Agents,Diuresis,Operati…       0.0
305  Abdomen,Anti-Bacterial Agents,Diuresis,Operati…       1.0
311  Abdomen,Amiodarone,Analgesia,Angiodysplasia,Hy…       1.0
```

1 PCA

1.1 Transform (factorice) from Categories to continuous atributes

Transform 'pedido.data.attributes.diagnostic_main' atribute

[6]:
```
categoriesORGDiagnosticMain = dfPCA['pedido.data.attributes.diagnostic_main'].
↪value_counts()

print("total: " + str(categoriesORGDiagnosticMain.size))

categoriesORGDiagnosticMain
```

```
total: 12
```

[6]:
```
DOLOR ABDOMINAL                13
INFECCION DE PARTES BLANDAS     9
INFECCION URINARIA              5
TORACOTOMIA                     4
ABDOMEN AGUDO                   4
HEMORRAGIA DIGESTIVA            3
ACV.ISQUEMICO                   3
CETOACIDOSIS DIABETICA          3
POLITRAUMATISMO                 2
FISTULA PERITONEAL              2
DISNEA                          2
INSUFICIENCIA RESPIRATORIA      1
Name: pedido.data.attributes.diagnostic_main, dtype: int64
```

[7]:
```
dataDiagnosticMain, categoriesDiagnosticMain = pd.factorize(dfPCA['pedido.data.
↪attributes.diagnostic_main'])

categoriesDiagnosticMain
```

[7]:
```
Index(['FISTULA PERITONEAL', 'INSUFICIENCIA RESPIRATORIA', 'POLITRAUMATISMO',
       'ABDOMEN AGUDO', 'TORACOTOMIA', 'INFECCION DE PARTES BLANDAS',
       'DOLOR ABDOMINAL', 'INFECCION URINARIA', 'HEMORRAGIA DIGESTIVA',
       'ACV.ISQUEMICO', 'DISNEA', 'CETOACIDOSIS DIABETICA'],
      dtype='object')
```

0 => first element found => 'FISTULA PERITONEAL'

1 => second element found => 'INSUFICIENCIA RESPIRATORIA'

...

[8]:
```
dfPCA['pedido.data.attributes.diagnostic_main'] = dataDiagnosticMain

dfPCA.head(10)
```

[8]:
```
     pedido.data.attributes.age  pedido.data.attributes.diagnostic_main  \
0                          75.0                                       0
32                         75.0                                       0
230                        36.0                                       1
290                        51.0                                       2
299                        51.0                                       2
300                        18.0                                       3
303                        18.0                                       3
304                        18.0                                       3
305                        18.0                                       3
311                        76.0                                       4

    pedido.data.attributes.gender  articulo  respuesta.articlesRevisedYear  \
```

```
0                              male  27395425                              2018
32                             male  28694230                              2017
230                            male  28805236                              2011
290                            male  27537587                              2011
299                            male  28148670                              2019
300                            male  25055513                              2019
303                            male  29279563                              2017
304                            male  29279563                              2017
305                            male  28065368                              2017
311                            male  30762794                              2019

     respuesta.articlesRevisedMonth  \
0                                  1
32                                12
230                                3
290                                3
299                                3
300                                3
303                                2
304                                2
305                               11
311                                3

                                   respuesta.pubmed_keys  utilidad
0    Abdomen,Adenocarcinoma,Antiemetics,Blood Cultu…       1.0
32   Abdomen,Blood Culture,Catharsis,Diuresis,Drug …       1.0
230  Abdomen,Analgesics,Antitubercular Agents,Cipro…       0.0
290  Abdomen,Analgesics,Bone,Catharsis,Electroconvu…       0.0
299  Abdomen,Analgesics,Bone,Catharsis,Electroconvu…       1.0
300  Abdomen,Anti-Bacterial Agents,Diuresis,Operati…       1.0
303  Abdomen,Anti-Bacterial Agents,Diuresis,Operati…       0.0
304  Abdomen,Anti-Bacterial Agents,Diuresis,Operati…       0.0
305  Abdomen,Anti-Bacterial Agents,Diuresis,Operati…       1.0
311  Abdomen,Amiodarone,Analgesia,Angiodysplasia,Hy…       1.0
```

Transform ‘gender’ atribute

[9]:
```
categoriesORGGender = dfPCA['pedido.data.attributes.gender'].value_counts()

print("total: " + str(categoriesORGGender.size))

categoriesORGGender
```

```
total: 1
```

[9]:
```
male    51
Name: pedido.data.attributes.gender, dtype: int64
```

[10]:
```
dataGender, categoriesGender = pd.factorize(dfPCA['pedido.data.attributes.
↪gender'])

categoriesGender
```

[10]:
```
Index(['male'], dtype='object')
```

[11]:
```
dfPCA['pedido.data.attributes.gender'] = dataGender

dfPCA.head(10)
```

[11]:

	pedido.data.attributes.age	pedido.data.attributes.diagnostic_main	pedido.data.attributes.gender	articulo	respuesta.articlesRevisedYear	respuesta.articlesRevisedMonth
0	75.0	0	0	27395425	2018	1
32	75.0	0	0	28694230	2017	12
230	36.0	1	0	28805236	2011	3
290	51.0	2	0	27537587	2011	3
299	51.0	2	0	28148670	2019	3
300	18.0	3	0	25055513	2019	3
303	18.0	3	0	29279563	2017	2
304	18.0	3	0	29279563	2017	2
305	18.0	3	0	28065368	2017	11
311	76.0	4	0	30762794	2019	3

	respuesta.pubmed_keys	utilidad
0	Abdomen,Adenocarcinoma,Antiemetics,Blood Cultu…	1.0
32	Abdomen,Blood Culture,Catharsis,Diuresis,Drug …	1.0
230	Abdomen,Analgesics,Antitubercular Agents,Cipro…	0.0
290	Abdomen,Analgesics,Bone,Catharsis,Electroconvu…	0.0
299	Abdomen,Analgesics,Bone,Catharsis,Electroconvu…	1.0
300	Abdomen,Anti-Bacterial Agents,Diuresis,Operati…	1.0
303	Abdomen,Anti-Bacterial Agents,Diuresis,Operati…	0.0
304	Abdomen,Anti-Bacterial Agents,Diuresis,Operati…	0.0
305	Abdomen,Anti-Bacterial Agents,Diuresis,Operati…	1.0
311	Abdomen,Amiodarone,Analgesia,Angiodysplasia,Hy…	1.0

Transform 'respuesta.pubmed_keys' atribute

[12]:
```
categoriesORGPubMedKeys = dfPCA['respuesta.pubmed_keys'].value_counts()

print("total: " + str(categoriesORGPubMedKeys.size))

categoriesORGPubMedKeys
```

```
total: 25
```

[12]:
```
Abdomen,Abdominal Pain,Amebiasis,Amebic,Catharsis,Ciprofloxacin,Diarrhea,Diuresi
s,Dysentery,Extremities,Hemorrhage,Hydrocortisone,Intestines,Lower
Extremity,Metronidazole,Pain,Pain Management,Signs and
Symptoms,Tachycardia,Therapeutics,Venous Thrombosis
5
Abdomen,Anti-Bacterial Agents,Diuresis,Operative,Peritonitis,Pregnancy,Signs and
Symptoms,Surgical Procedures,Surgical Wound,Therapeutics,Wounds and Injuries
4
Abdomen,Amiodarone,Analgesia,Angiodysplasia,Hypoventilation,Lung,Operative,Surgi
cal Procedures,Thoracotomy,Tramadol
4
Abdomen,Abdominal
Pain,Amylases,Breast,Catharsis,Cholangiopancreatography,Diuresis,Endoscopic
Retrograde,Extremities,Hyperbilirubinemia,Hypertension,Hypoventilation,Lower
Extremity,Pain,Signs and Symptoms,Skin,Transaminases,Venous Thrombosis
3
Abdomen,Brain
Diseases,Catharsis,Dehydration,Diuresis,Extremities,Fibrosis,Hepatic
Encephalopathy,Lower Extremity,Quarantine,Signs and
Symptoms,Skin,Therapeutics,Venous Thrombosis
3
Abdomen,Anti-Bacterial Agents,Antihypertensive Agents,Catharsis,Diuresis,Hernia,
Hypoventilation,Ileostomy,Intestines,Loperamide,Respiratory Mechanics,Surgical
Wound,Wounds and Injuries
3
```

Abdomen,Amiodarone,Anemia,Atrial
Fibrillation,Catharsis,Ceftriaxone,Cephalexin,Colic,Communicable Diseases,Diures
is,Edema,Enzymes,Hematocrit,Hematuria,Hypokalemia,Hypoventilation,Infection,Pulm
onary Atelectasis,Radiography,Renal Colic,Skin,Tachycardia,Therapeutics,Urinary
Tract Infections,Urologic Diseases,Work
3
Abdomen,Catharsis,Diuresis,Kidney
Calculi,Lithiasis,Methods,Nephrolithiasis,Pain,Pain Management,Renal
Colic,Urologic Diseases
3
Abdomen,Acromegaly,Arteries,Breast,Bundle-Branch Block,Catharsis,Chronic
Obstructive,Cough,Craniotomy,Diabetes Mellitus,Diabetes Mellitus,Diuresis,Echoca
rdiography,Edema,Electrocardiography,Extremities,Goiter,Heart
Failure,Hypertension,Hypertension,Hypoventilation,Ischemia,Lower
Extremity,Mastectomy,Myocardial Ischemia,Nose,Oxygen Inhalation
Therapy,Piperacillin,Pulmonary,Pulmonary Artery,Pulmonary Disease,Pulmonary
Embolism,Radiotherapy,Respiratory Insufficiency,Respiratory
Sounds,Segmental,Tazobactam,Therapeutics,Thromboembolism,Thrombosis,Type
2,Venous Thrombosis,Volition 2
Abdomen,Diuresis,Extremities,Hernia,Hernia,Hiatal,Hypoventilation,Intestine,Inte
stines,Lower Extremity,Pain,Signs and Symptoms,Small,Venous Thrombosis,Wounds
and Injuries
2
Abdomen,Adrenal Cortex Hormones,Amebiasis,Amebic,Catharsis,Ciprofloxacin,Colon,C
ysts,Diarrhea,Disease,Diuresis,Dysentery,Edema,Heart
Murmurs,Hematocrit,Inflammatory Bowel Diseases,Intestines,Metronidazole,Signs
and Symptoms,Speech,Therapeutics
2
Abdomen,Analgesics,Bone,Catharsis,Electroconvulsive
Therapy,Extremities,Fractures,Immunologic Memory,Lung,Medical History
Taking,Signs and Symptoms
2
Abdomen,Diuresis,Extremities,Hyperplasia,Hypertension,Hypoventilation,Ileus,Inte
stines,Ischemia,Lower Extremity,Pain,Pain Management,Respiratory Sounds,Signs
and Symptoms,Venous Thrombosis,Work,Wounds and Injuries
2
Abdomen,Aphasia,Aphasia,Atrial Appendage,Broca,Catharsis,Diabetes Mellitus,Diure
sis,Hemiplegia,Hypertension,Neck,Obesity,Palpation,Rehabilitation,Respiratory
Sounds,Stroke
2
Abdomen,Adenocarcinoma,Antiemetics,Blood
Culture,Catharsis,Diuresis,Fistula,Gastrectomy,Incisional
Hernia,Intestines,Muscles,Nausea,Pain,Pain
Threshold,Palpation,Piperacillin,Pleural
Effusion,Pneumonia,Quarantine,Respiratory Sounds,Signs and Symptoms,Surgical
Wound,Tazobactam,Therapeutics,Thorax,Tomography,Wounds and Injuries,X-Ray
Computed

```
1
Abdomen,Blood Culture,Catharsis,Ciprofloxacin,Communicable Diseases,Diuresis,Hyp
erplasia,Infection,Pain,Pelvis,Prostatitis,Therapeutics,Tramadol
1
Abdomen,Clindamycin,Diuresis,Edema,Inflammation,Molar,Therapeutics,Tomography,X-
Ray Computed
1
Abdomen,Alzheimer Disease,Communicable
Diseases,Disease,Infection,Skin,Sleep,Urinary Tract Infections
1
Diabetes Mellitus,Diabetic Ketoacidosis,Diagnosis,Ketosis
1
Abdomen,Analgesics,Antitubercular
Agents,Ciprofloxacin,Defecation,Diuresis,Intention,Intestines,Lupus
Erythematosus,Pain,Pain Management,Parenteral Nutrition,Rifampin,Streptomycin,Sy
ndrome,Systemic,Therapeutics,Tuberculosis,Wounds and Injuries
1
Chronic,Chronic,Diabetes Mellitus,Diabetic Ketoacidosis,Diagnosis,Ketosis,Kidney
Failure,Renal Insufficiency,Renal Insufficiency,Urologic Diseases
1
Abdomen,Arteries,Catharsis,Diabetes
Mellitus,Diuresis,Extremities,Hypertension,Hypoventilation,Lower
Extremity,Oxygen Inhalation Therapy,Signs and
Symptoms,Speech,Stroke,Therapeutics,Venous Thrombosis
1
Diabetes Mellitus,Diabetes Mellitus,Diabetic
Ketoacidosis,Diagnosis,Ketosis,Therapeutics,Type 1
1
Abdomen,Blood Culture,Catharsis,Diuresis,Drug
Therapy,Extremities,Fistula,Hiccup,Intestines,Lower
Extremity,Morphine,Nausea,Pain,Palpation,Piperacillin,Pneumonia,Respiratory
Sounds,Tazobactam,Therapeutics,Wounds and Injuries
1
Abdomen,Adrenal Cortex
Hormones,Catharsis,Ciprofloxacin,Cysts,Disease,Grief,Heart Murmurs,Inflammatory
Bowel Diseases,Infliximab,Intestines,Mesalamine,Metronidazole,Signs and
Symptoms,Syncope,Therapeutics
1
Name: respuesta.pubmed_keys, dtype: int64
```

[13]:
```
dataPubMedKeys, categoriesPubMedKeys = pd.factorize(dfPCA['respuesta.
↪pubmed_keys'])
```

[14]:
```
dfPCA['respuesta.pubmed_keys'] = dataPubMedKeys
```

[15]:
```
dfPCA.head(10)
```

[15]:

	pedido.data.attributes.age	pedido.data.attributes.diagnostic_main	pedido.data.attributes.gender	articulo	respuesta.articlesRevisedYear	respuesta.articlesRevisedMonth	respuesta.pubmed_keys	utilidad
0	75.0	0	0	27395425	2018	1	0	1.0
32	75.0	0	0	28694230	2017	12	1	1.0
230	36.0	1	0	28805236	2011	3	2	0.0
290	51.0	2	0	27537587	2011	3	3	0.0
299	51.0	2	0	28148670	2019	3	3	1.0
300	18.0	3	0	25055513	2019	3	4	1.0
303	18.0	3	0	29279563	2017	2	4	0.0
304	18.0	3	0	29279563	2017	2	4	0.0
305	18.0	3	0	28065368	2017	11	4	1.0
311	76.0	4	0	30762794	2019	3	5	1.0

1.2 Standardize the Data

[16]:
```
from sklearn.preprocessing import StandardScaler

features = ['pedido.data.attributes.age',
            'pedido.data.attributes.diagnostic_main',
            'pedido.data.attributes.gender',
            'respuesta.articlesRevisedYear',
            'respuesta.articlesRevisedMonth',
            'respuesta.pubmed_keys',
            'articulo']
```

```
# Separating out the features
x = dfPCA.loc[:, features].values# Separating out the target
dfPCA['utilidad']=pd.Categorical(dfPCA['utilidad'])
my_color=dfPCA['utilidad'].cat.codes

featuresTransformed = StandardScaler().fit_transform(x)

featuresTransformed
```

[16]:
```
array([[ 0.91709628, -2.24479066,  0.        ,  0.48066047, -1.27483261,
        -1.81819247, -0.19317962],
       [ 0.91709628, -2.24479066,  0.        ,  0.11478459,  1.60896212,
        -1.65608091,  0.31397864],
       [-0.8747549 , -1.85670821,  0.        , -2.08047067, -0.75050629,
        -1.49396934,  0.35732434],
       [-0.18558137, -1.46862576,  0.        , -2.08047067, -0.75050629,
        -1.33185777, -0.13766811],
       [-0.18558137, -1.46862576,  0.        ,  0.84653634, -0.75050629,
        -1.33185777,  0.100948  ],
       [-1.70176313, -1.0805433 ,  0.        ,  0.84653634, -0.75050629,
        -1.16974621, -1.10687007],
       [-1.70176313, -1.0805433 ,  0.        ,  0.11478459, -1.01266945,
        -1.16974621,  0.54253988],
       [-1.70176313, -1.0805433 ,  0.        ,  0.11478459, -1.01266945,
        -1.16974621,  0.54253988],
       [-1.70176313, -1.0805433 ,  0.        ,  0.11478459,  1.34679897,
        -1.16974621,  0.06842018],
       [ 0.96304118, -0.69246085,  0.        ,  0.84653634, -0.75050629,
        -1.00763464,  1.12171293],
       [ 0.96304118, -0.69246085,  0.        ,  0.84653634, -0.75050629,
        -1.00763464,  1.12171293],
       [ 0.96304118, -0.69246085,  0.        ,  0.11478459,  1.60896212,
        -1.00763464,  0.66295943],
       [ 0.96304118, -0.69246085,  0.        ,  0.11478459, -0.22617998,
        -1.00763464, -1.74384989],
       [-1.19636921, -0.30437839,  0.        ,  0.84653634, -0.75050629,
        -0.84552307,  0.41640915],
       [-1.19636921, -0.30437839,  0.        ,  0.11478459,  1.34679897,
        -0.84552307,  0.32427757],
       [ 1.10087588,  0.08370406,  0.        ,  0.84653634, -0.22617998,
        -0.68341151, -0.59661724],
       [ 1.10087588,  0.08370406,  0.        ,  0.84653634, -0.22617998,
        -0.68341151, -0.59661724],
       [ 1.19276569,  0.47178651,  0.        ,  0.84653634, -0.22617998,
        -0.52129994,  0.25862449],
       [ 0.87115137,  0.08370406,  0.        ,  0.84653634, -0.22617998,
```

```
 -0.35918837,  1.08237552],
[ 0.87115137,  0.08370406,  0.        , -0.98284304,  1.08463581,
 -0.35918837,  1.10153914],
[ 0.87115137,  0.08370406,  0.        ,  0.11478459,  1.34679897,
 -0.35918837,  0.15789492],
[ 0.45764726,  0.85986897,  0.        ,  0.84653634, -0.22617998,
 -0.19707681,  0.21066839],
[ 0.45764726,  0.85986897,  0.        , -0.98284304,  1.08463581,
 -0.19707681,  0.6678502 ],
[ 0.45764726,  0.85986897,  0.        , -0.25109129,  1.34679897,
 -0.19707681,  1.17931899],
[ 0.41170236,  1.24795142,  0.        ,  0.84653634, -0.48834314,
 -0.03496524, -0.74218844],
[ 0.41170236,  1.24795142,  0.        , -1.34871892, -0.75050629,
 -0.03496524, -0.57050462],
[ 1.10087588,  0.08370406,  0.        ,  0.84653634, -1.01266945,
  0.12714633, -0.77544256],
[ 1.10087588,  0.08370406,  0.        ,  0.11478459, -0.75050629,
  0.12714633, -1.31346715],
[ 1.10087588,  0.08370406,  0.        , -0.61696716,  0.82247265,
  0.12714633,  0.11654809],
[-1.19636921, -0.30437839,  0.        , -0.25109129, -1.01266945,
  0.28925789, -2.6686866 ],
[-1.19636921, -0.30437839,  0.        ,  0.48066047,  1.34679897,
  0.45136946,  0.89516893],
[-1.19636921, -0.30437839,  0.        ,  0.11478459, -0.75050629,
  0.45136946,  0.51002923],
[-1.19636921, -0.30437839,  0.        ,  0.48066047, -0.48834314,
  0.45136946, -1.57727317],
[-1.19636921, -0.30437839,  0.        ,  0.48066047,  1.60896212,
  0.45136946,  0.41640915],
[-1.19636921, -0.30437839,  0.        , -0.98284304,  0.82247265,
  0.45136946, -1.10097381],
[ 0.22792275,  0.47178651,  0.        , -0.25109129,  1.34679897,
  0.61348103, -0.16001844],
[ 0.22792275,  0.47178651,  0.        ,  0.84653634, -1.01266945,
  0.61348103,  0.79928405],
[ 0.22792275,  0.47178651,  0.        ,  0.84653634, -1.01266945,
  0.61348103,  0.14857688],
[ 1.10087588,  0.08370406,  0.        , -2.44634655,  0.56030949,
  0.77559259,  0.98519776],
[ 1.10087588,  0.08370406,  0.        ,  0.84653634, -0.75050629,
  0.77559259, -2.45688246],
[ 0.31981255,  0.08370406,  0.        , -0.25109129,  1.34679897,
  0.93770416, -1.61222666],
[ 0.31981255,  0.08370406,  0.        ,  0.84653634, -0.75050629,
  0.93770416,  0.81440854],
```

```
       [ 0.31981255,  0.08370406,  0.        ,  0.84653634, -1.01266945,
         0.93770416,  0.98519776],
       [ 0.41170236,  1.24795142,  0.        , -0.98284304,  0.82247265,
         1.09981573,  1.04535678],
       [ 0.45764726,  0.47178651,  0.        ,  0.11478459,  0.82247265,
         1.26192729,  0.35655197],
       [ 0.45764726,  1.63603387,  0.        ,  0.84653634, -1.01266945,
         1.42403886,  0.24450666],
       [ 0.45764726,  1.63603387,  0.        , -2.44634655, -0.75050629,
         1.42403886, -1.68007619],
       [-0.64503039, -0.30437839,  0.        ,  0.84653634, -1.01266945,
         1.58615043, -1.48790854],
       [-1.51798352,  2.02411633,  0.        , -2.08047067,  1.60896212,
         1.74826199,  1.4155516 ],
       [-1.51798352,  2.02411633,  0.        , -2.08047067,  1.60896212,
         1.91037356,  1.4155516 ],
       [-1.51798352,  2.02411633,  0.        ,  0.84653634, -1.01266945,
         2.07248513,  0.14101718]])
```

[17]:
```
from sklearn.decomposition import PCA

pca = PCA(n_components=5)

pca.fit(featuresTransformed)

result=pd.DataFrame(pca.transform(featuresTransformed), columns=['PCA%i' % i␣
↪for i in range(5)])

result.head(10)
```

[17]:

	PCA0	PCA1	PCA2	PCA3	PCA4
0	-3.149930	-0.354331	-0.429642	0.398361	0.715131
1	-1.849930	-2.291115	-0.875543	0.750490	-1.002239
2	-1.339644	-1.848462	1.486153	0.957123	1.752901
3	-1.168393	-1.304228	0.969168	1.481622	1.631511
4	-2.193722	-0.367401	0.201702	-0.490824	0.143486
5	-1.949279	0.251310	1.872268	-0.348433	-0.477504
6	-1.459827	-0.731184	1.566880	-1.004888	0.762812
7	-1.459827	-0.731184	1.566880	-1.004888	0.762812
8	-0.750723	-1.795165	1.399297	-0.314486	-1.145681
9	-1.450976	-0.361905	-1.253764	-0.832103	0.674737

[18]:
```
print('explained variance ratio (first three components): %s' %
str(pca.explained_variance_ratio_))
print('sum of explained variance (first three components): %s' %
str(sum(pca.explained_variance_ratio_)))
```

```
explained variance ratio (first three components): [0.32833981 0.23094825
0.17265137 0.14192254 0.09557452]
sum of explained variance (first three components): 0.9694364891147347
```

[19]:
```
pd.DataFrame(pca.components_,columns=features,index =␣
↪['PC1','PC2','PC3','PC4','PC5'])
```

[19]:

	pedido.data.attributes.age	pedido.data.attributes.diagnostic_main	\
PC1	-0.094054	0.603912	
PC2	0.135318	0.327845	
PC3	-0.891259	-0.174252	
PC4	0.382620	-0.115174	
PC5	0.098479	0.051453	

	pedido.data.attributes.gender	respuesta.articlesRevisedYear	\
PC1	-1.110223e-16	-0.365651	
PC2	-5.273559e-16	0.362710	
PC3	-5.828671e-16	-0.236812	
PC4	-1.929013e-15	-0.621075	
PC5	3.774758e-15	-0.540677	

	respuesta.articlesRevisedMonth	respuesta.pubmed_keys	articulo
PC1	0.338374	0.585481	0.188307
PC2	-0.556061	0.399721	-0.523132
PC3	-0.133609	0.066053	-0.311447
PC4	0.162456	-0.093185	-0.647709
PC5	-0.729252	-0.080884	0.396197

PC1 => diagnostic_main PC2 => pubmed_keys/Year PC3 => pubmed_keys PC4 => age PC5 => articulo