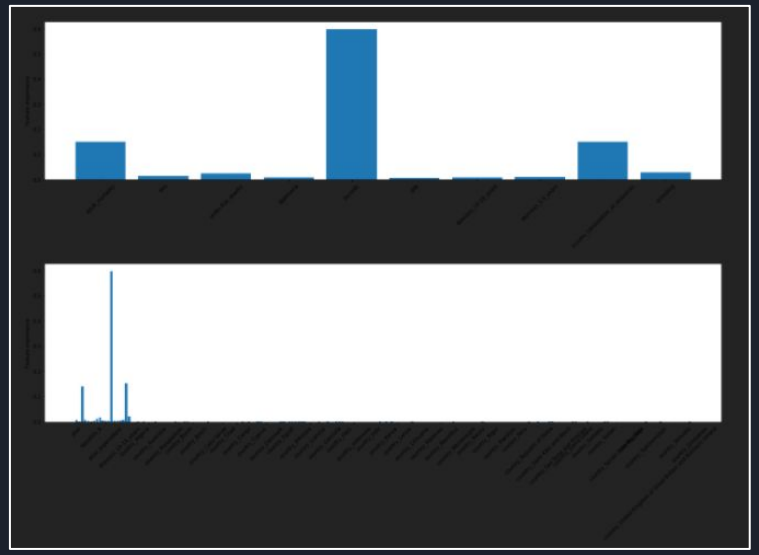
Artificial Intelligence

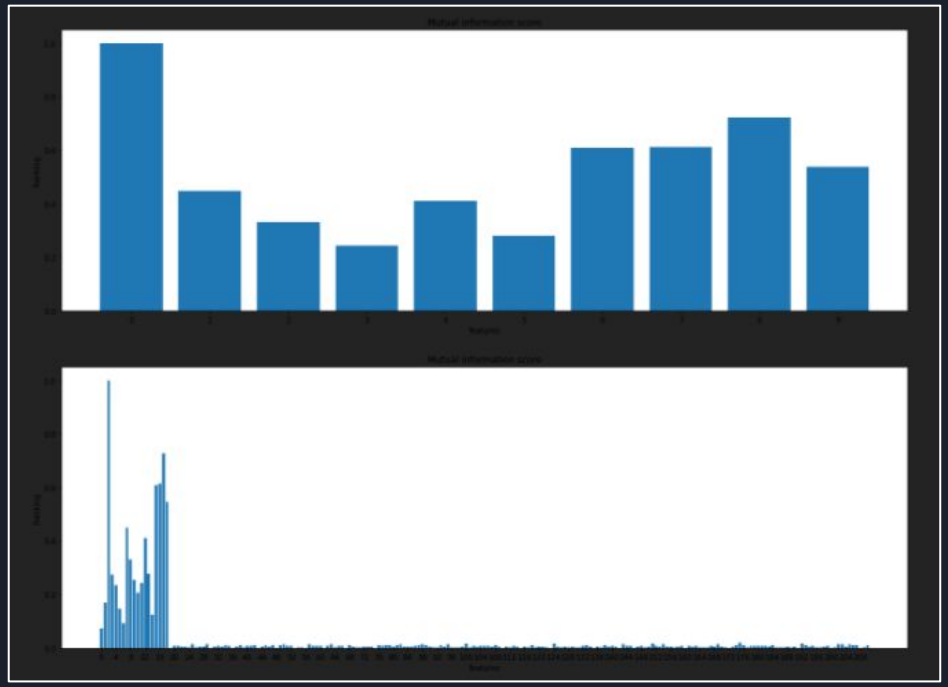
Laboratory 5

Isabella Quicaño
Raúl López

5.1 # Filter methods



Mutual information



Random forest regressor

```
Index(['adult_mortality', 'hiv/aids', 'income_composition_of_resources'], dtype='object')
Index(['adult_mortality', 'hiv/aids', 'income_composition_of_resources'], dtype='object')
```

5.2 # Features would you use for life_expectancy_data?

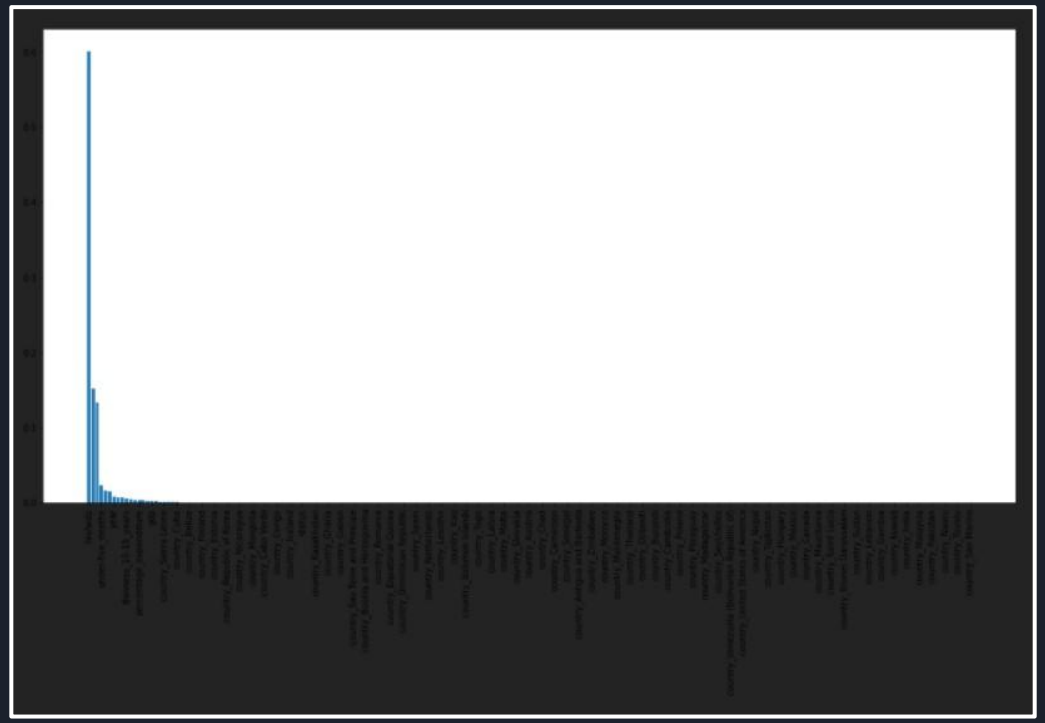
STEP 1:

```
Predictors:  (2938, 211)
Target:  (2938,)
Index(['year', 'status', 'adult_mortality', 'alcohol',
       'percentage_expenditure', 'hepatitis_b', 'measles', 'bmi',
       'under-five_deaths', 'polio',
       ...
       'country_United Republic of Tanzania',
       'country_United States of America', 'country_Uruguay',
       'country_Uzbekistan', 'country_Vanuatu',
       'country_Venezuela (Bolivarian Republic of)', 'country_Viet Nam',
       'country_Yemen', 'country_Zambia', 'country_Zimbabwe'],
      dtype='object', length=211)
```

```
Train data shape: (2056, 211)
Test  data shape: (882, 211)
```

Importance plot

STEP 2:



5.2 # Features would you use for life_expectancy_data?

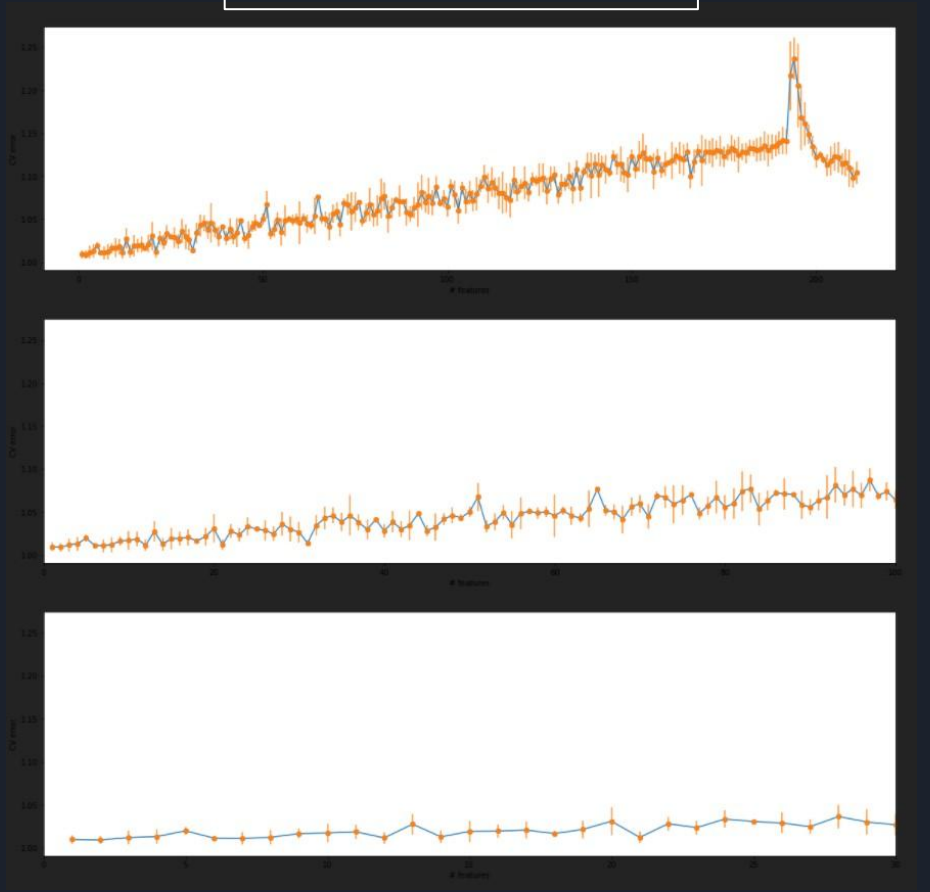
STEP 3:

First 20 features selected
1 = 1.0095859000945264
2 = 1.0090801638488396
3 = 1.0118315180602713
4 = 1.0130923123958728
5 = 1.0196972939725235
6 = 1.0112263577459448
7 = 1.010979800981098
8 = 1.012205460303692
9 = 1.016447849519072
10 = 1.0171658184926262
11 = 1.0185699436378235
12 = 1.0115068901231326
13 = 1.0274179605692704
14 = 1.0126929257041157
15 = 1.0190446725877085
16 = 1.01935829773516
17 = 1.020567300113745
18 = 1.0164886294847844
19 = 1.0213664805243388
20 = 1.0308496872303003
21 = 1.012038591037621
22 = 1.027978361234857
23 = 1.0233609791868359
24 = 1.0333897117487254
25 = 1.0303942746799828

Minimun error = 1.0090801638488396

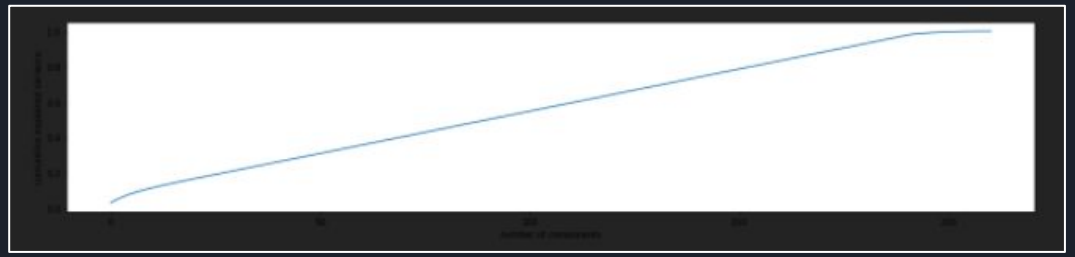
From 200 to 211 features selected
211 = 1.1253027839080747
210 = 1.1196832892918303
209 = 1.1130376468413101
208 = 1.1176255928608934
207 = 1.1228597990250895
206 = 1.1220764500837577
205 = 1.114298329240061
204 = 1.1162644452440205
203 = 1.1097309029120634
202 = 1.0984019193245491
201 = 1.1041673865562462

Error for#features selected

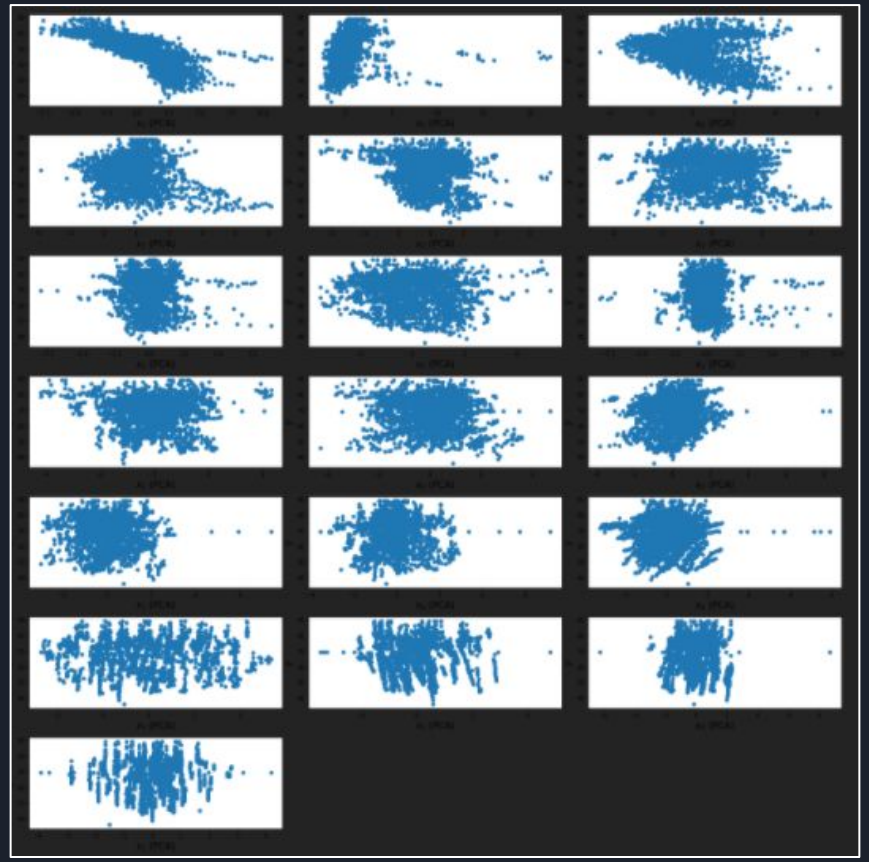


5.3 # Model performance with PCA

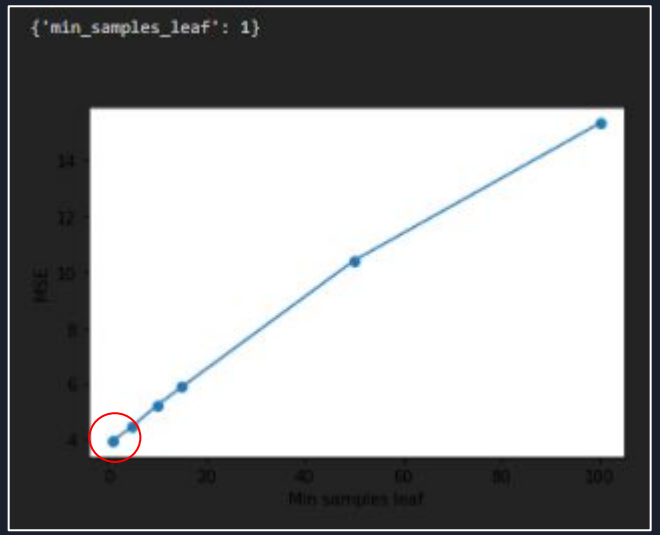
15 components

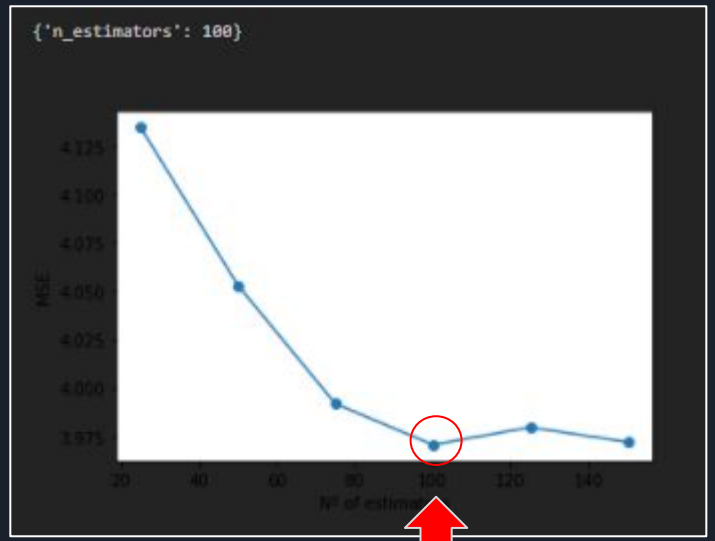


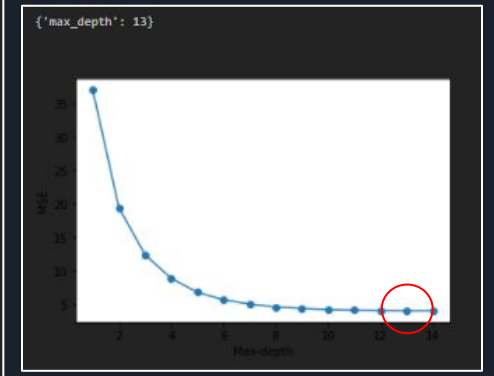
5.3 # Model performance with PCA



5.3 # Model performance with PCA







Optimal values

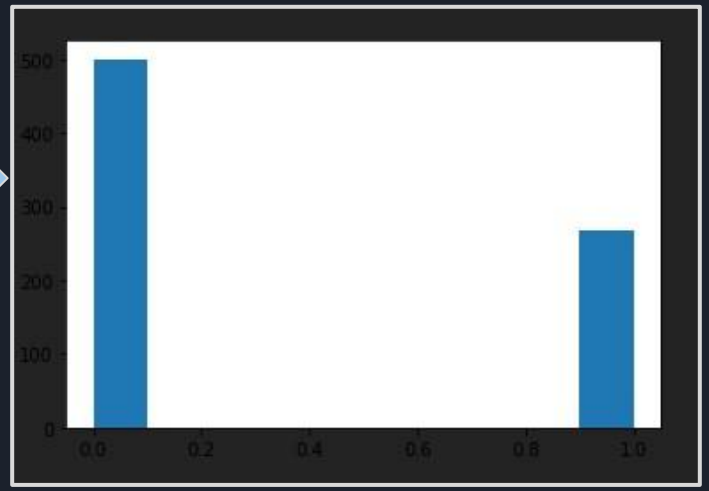
5.3 Model performance with PCA

Model assessed	R^2 score test	R^2 score training
Model with all features (PCA)	0.9819	0.6286
Model with 19 features (PCA)	0.9818	0.7175
Model with all features (original)	0.9892	0.9334
Model with 19 features (original)	0.9898	0.93232

Model assessed	Mean Absolute Error	Mean Squared Error	Root Mean Squared Error
Model with all features (PCA)	4.1938	30.6060	5.5322
Model with 19 features (PCA)	3.5459	22.2832	4.7183
Model with all features (original)	0.6820	0.9583	0.9789
Model with 19 features (original)	0.6513	0.9041	0.9508

5.4 Is PCA transformation useful for classification?

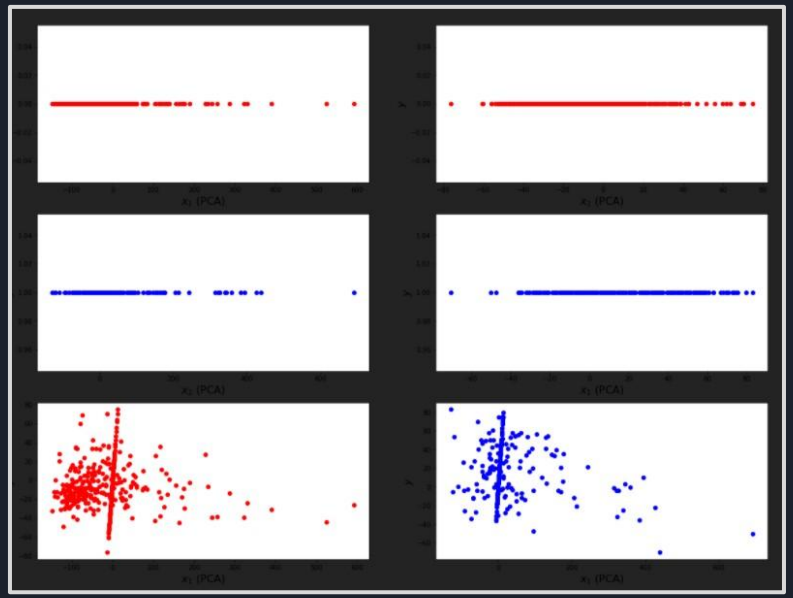
Data histogram




```
Explained variance ratio (first two components): [0.86510267 0.0891348 ]
```

5.4 Is PCA transformation useful for classification?

Individual visualization



Final scatter plot