# Analysis for Sub-problem 5

## 1. Correlation between the general climate of a county with the health of its population

To find correlation between climate of a county and the health of the population, we tried to predict the climate of the county by using health indicators i.e. prevalence of diseases in the county as features.

Features: 'Prevalence of obesity', 'Hypertension', 'Diabetes', 'CVD', 'HIV/AIDS', 'cvd\_100k', 'hypertension\_100k'

Label: 'IECC\_Zone'

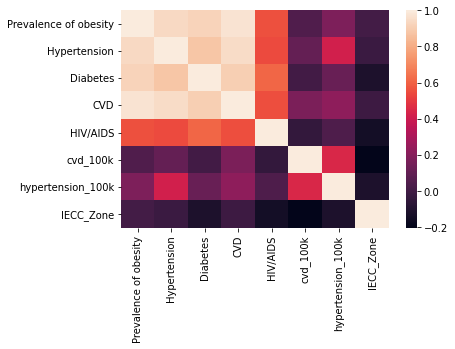
PCA was used to determine how many features best explained the classification.

Results:

|  |  |
| --- | --- |
| CLASSIFIER | ACCURACY |
| Support Vector Machines (Number of features: 4) | 37.3% |
| Decision Trees (Depth: 7) | 35.8% |
| Random Forest (Depth: 7) | 38.7% |
| Multi Layer Perceptron (Number of features: 5) | 41.2% |

Using various classifiers, a maximum test score of 41.2% could be achieved which indicates that there is weak or no relation between the health of the county and the climate.

We can also see in the correlation heatmap that the climate zone has a weak correlation with the diseases:



## 2. Correlation between Climate of the county and number of pharmacies

To find correlation between climate of the county and number of pharmacies in the county, we tried to predict the climate of the county based on the number of pharmacies.

Results:

|  |  |
| --- | --- |
| CLASSIFIER | ACCURACY |
| Decision Trees (Depth: 2) | 23.5% |
| Random Forest (Depth: 1) | 22.6% |
| Logistic Regression | 22.2% |

Using various classifiers, a maximum test score of 23.5% could be achieved which indicates that there is weak or no relation between the number of pharmacies in a county and its climate.

We can see in the correlation heatmap that there is low (or no) correlation between Climate of the county and number of pharmacies:

