**Classification**: Is immunotherapy a good choice for a certain wart patient?

**Big** **Idea**: We will investigate whether the immunotherapy is a good idea for a wart treatment, given patient’s attributes.

**Data Description**: The data can be downloaded in xlsx format from https://archive.ics.uci.edu/ml/datasets/Immunotherapy+Dataset# . The data contains a total of 90 patients on 7 features and 1 response, and there are no missing values.

* **Response**: Categorical with 2 classes: success and failure as the result of immunotherapy
* **Features**: 7 total, two are categorical and the rest are numerical. We anticipate that some interesting features may be **number of warts**, the **type of warts** and the **induration diameter**.
* **Notes**: The type of warts could be treated as either numeric or categorical. Since the there is only three types of warts. Since the type one is the most common one, we consider to change the feature as categorical and set Type one as the baseline.

**Questions of interest**: The primary question of interest will be to determine whether a patient is suitable for immunotherapy to treat wart. It may also be of interest to explore which predictor is more significant according by cross-validation. Since the sample size is small, like most patient sample, we expect high possibility of false positive. However, since the data set is easy to interpret, it is a very good chance to learn how to build a classifier and select features according to R2 and data entropy.

**Interested Parties**: This might be interested to physicians who are looking for more eﬀective and customized treatments for each patient. The ultimate goal is assisting physicians in selecting the best treatment method, saving time for patients, reducing the treatment cost, and improving the quality of treatment.

R code previewing the data:

# download the data from https://archive.ics.uci.edu/ml/datasets/Immunotherapy+Dataset#

install.packages("openxlsx")

library("openxlsx")

my\_data = read.xlsx("Immunotherapy.xlsx", sheet="ImmunoDataset")

summary(my\_data)

attach(my\_data)

pairs(my\_data)