MODELLING AND PREDICTING THE SPREAD OF TUBERCULOSIS

Research objective: the objective of this research work is to develop a model that can predict the spread of tuberculosis over time.

Data collation: various data was collected as regards the spread of tuberculosis in previous years and were subjected to a-priori and priori tests for significance so as to avoid over-fittings and under-fittings in our model for the best accuracy in predictions of futuristic outcomes

The data used in our model for prediction include;

i. **Facebook tuberculosis data**
ii. **Poverty index**
iii. **Social connectedness**
iv. **School enrolment Facebook data (created)**

Target data: Target data was set as rate of prevalence of tuberculosis (this was taken from **Facebook tuberculosis data)** which was taken as the values we wish to predict in our model.

Training data: 70% of our data was used in training our model to help our model familiarize with our data and create a good fit for accurate predictions.

Testing data: **30%** of our data was used in testing the accuracy in prediction of our model or rather the predictive power of our model.

MODEL DEVELOPMENT

After splitting our data and using various machine learning models to fit our training data and also carrying out a descriptive analysis on our data to check for correlation among the explanatory variables (i.e. variables other than our target variable) which can distort our model development.

We were able to come up with a model that scored 78% accuracy in predicting our test data outcomes (30% of our target data) i.e. the developed model can predict future rate of prevalence of tuberculosis with an accuracy of 78%.