Work Flowchart-DRAFT:
Machine Learning Based Prediction To Detect Anti-Cancer Polypeptides Using Deep Learning, Support Vector Machine Sigmoidal Function & Naive Bayes.

Start

Obtain Tyagi-B dataset & Controls, {2 x *.fasta Format}

Convert
Polypeptides to % AA & Dipeptide Composition
2D-matrix [~500 polypeptides, ~423 Features]

Exploratory Data Analysis

Testing
1. Investigate: 1st, 2nd 3rd moments of AA & DP,
2. Check for Near Zero Variance of features,
3. Use Boruta RF for Feature Selection,
3. Check Pearson's Correlation Coefficients for multi-collinearity,
4. Conduct Principle Component Analysis for feature selection,
5. K-Means Cluster analysis

Remove
N.Z.V. Features

Remove
Boruta RF Features

|R| >= 0.75

YES

Test & Remove
Boruta RF
Features

NO

Attribute Columns:
NO Change

Combine Columns:
To Obtain Working Dataset.

Skewness >= 2.0

YES

Test:
Box-Cox
Transformation,
λ = xi
&
Transform
Feature i

NO

Attribute Columns:
NO Change

Combine Columns
To Obtain Working Dataset.

Generate
Skree Plot from PCA

Test
PC(i) VS Model Accuracy

Model Accuracy >= 0.9

NO

ADD
PC(i+1)

YES

Combine Columns
PC(i) + Indicator Variable
To Obtain Working Dataset.