Project Procedure

- Firstly Data preprocessing has been done by removing all the outliers using **IQR** method with capping so that the number of rows remain intact and robust standardization has been used as the data was slightly skewed in nature
- PCA has been applied rather than t-SNE (**t-Distributed Stochastic Neighbor Embedding**) as it is not perfect fit for regression data
- MLP (Multilayer perceptron) has been chosen over other complex architectures as it was giving the better results form the CNNs and LSTM which are very specific like for image classification and time-series analysis respectively

Limitations

- The model has limitations as by reducing features using PCA can miss out important features
- The vomitoxin_ppb is highly variable which makes the prediction difficult and less r2 score

Improvement areas

- Instead of a neural network, machine learning models like RandomForest regreesor can be used for better results
- Hyperparameter tunning can be optimized
- Data Augmentation to improve generalization