Machine Learning report

Candidate Number:277244

May 2024

1 Introduction

Predicting crop export values is an vital task for agricultural stakeholders.Planning,resource allocation,and decision-making to maximise crop productivity and commerce can all benefit from this accurate forecasts.In order to predict the export value of crop for a given geographical region three years from now, machine learning used in this research. The main objective of this study is to build a robust multi-layer perceptron(MLP) model that can accurately forecast the export values three years from now on the basis of historical data.This model is selected because of its capacity to identify intricate linkages and patterns in the data.The procedures involves preparing the data,creating and training the model,and assess the model's performance are all covered in this paper. Crop-production Indicator,Food-trade indicators,Land-temperature,Consumer price indicators,emissions, employment in agriculture, fertiliser use, food balance indicators, foreign direct investment, land use, pesticide use, exchange rates, and food security are included in this dataset used for this project.These characteristics are necessary to comprehend the complex relationship between crop export values.There are 13 files consists of these different characteristics.It consists of the data which are extracted from FAOSTAT database.

2 Feature and Labels

From all the files,relevant features has been taken.Total crop export values has also being considered which is the target variable here.It is required to predict the future crop export value for each year and each geographical areas.As 'Area','Area Code (M49)' and 'Year' is common in all the files,it will used in all the processing.It will be helpful during merging all the dataframe.Following data has been considered for prediction.

1. **crop production indicator**: "Official figures data" has been filtered out which ensures only reliable data points are considered.The filtered data us then grouped by 'Area Code (M49)', 'Area', 'Year', and 'Element', with the mean 'Value' calculated for each group.This 'Value' column has been renamed as 'yield' and the element column is subsequently dropped.

   Higher yields can boost the supply of agricultural products available for export,increasing the overall export value and pocitively contributing to a country's gross domestic produce(GDP).Higher yield reduce the per-unit cost of production through economies of scale,making crops more competitive in the international markets.Including these feature provides valuable insights into the agricultural productivity levels of different region over time.

2. **Emission** "Emission(CO2)" has been filtered out from element columns of emission dataframe.It then grouped the filtered data by 'Area Code (M49)', 'Area', 'Year', and 'Element',calculates the mean 'Value' for each area and year.Element coloums has been dropped and Value is renamed as 'EmissionCO2".

   Increased CO2 levels have a direct impact on crop yields and productivity because they may indicate intensive agricultural practices or inefficiencies in resource management.The dataset contains highest distribution of 'emission CO2' when compared to other types of emission.Higher CO2 levels can increase crop growth rates for some crops due to enhanced photosynthesis.Extreme weather events can negatively impact crop yields.Crops produced in regions with lower CO2 emission might be more desirable in markets prioritizing environmentally friendly practices.Countries with high emissions might face trade barriers affecting export values.That's why this feature has been considered.

3. **Exchange Rate**The group-by operation has been done on data based on Area Code (M49)', 'Area', and 'Year', and then mean 'Value' calculated for each group.Subsequently,the column 'Value' is renamed to 'Average_exchange_rate' to reflect the average values.

Because the currency rate directly affects how competitive agricultural exports are on global markets, it is imperative to take this into account. Changes in exchange rates have an effect on foreign buyers' purchasing power, which in turn affects agricultural demand and, eventually, export values. Higher exchange rates usually translate into more expensive exports in foreign currencies, which may lower export earnings and demand. In order to predict changes in export values and modify tactics appropriately in order to stay competitive in international markets, it is helpful to keep an eye on and factor the average exchange rate into analysis.

4. **Fertilizers** Average of all types of fertilizers has been considered here for each year and area.Only the 'official figure' from flag-description has been filtered out and it then groups the filtered data by 'Area Code (M49)', 'Area', and 'Year', and mean value for each country and each year is calculated.Value is renamed as 'Avg_fertilizer'.

   Calculating the average usage of fertilizers across various types is a good strategy here due to its comprehensive representation of overall fertilizer application trends.This method simplifies the combined effects of various fertiliser kinds on crop yield and environmental dynamics, making it easier to analyse agricultural practices in a clear and concise manner.A fair and comprehensive assessment of agricultural practices and their effects on crop yields is ensured by averaging fertiliser usage, which reduces any biases associated with different fertiliser categories.

5. **Food Balance**The rows where element is 'Export Quantity' and 'Item' is not Meat','Eggs','Milk - Excluding Butter','Fish, Seafood','Alcoholic Beverages' has been filtered out, that is only the export quantity of crop-products has been considered and then it groups the data by 'Area Code (M49)', 'Area', and 'Year'.Mean value for each group is calculated and columns has been renamed as 'food_balance_export'.

   When export quantities increases,assuming constant or increasing prices export value tends to increase as well.Additionally,external factors such as trade policies,tariffs,and geopolitical events can also influence the relationship between export quantity and value. As Export value of crop products need to be predicted,export quantities has been considered as an important feature here.

6. **Foreign Direct Investment** Rows where the Item column contains 'Total FDI inflows' and 'Total FDI outflows' has been filtered out.It then groups the filtered data by 'Area Code (M49)', 'Area', and 'Year',which calculated the mean 'Value' of each group,the results are presented in a pivot table with 'FDI_inflows' and 'FDI_outflows' as the columns.

   Creating a pivot for Foreign Direct Investment inflows and outflows values facilitates the analysis of capital movements,which are crucial indicators of economic and investment climate.**FDI inflows** represents capital investment made by foreign entities into the agriculture field of a area.These will lead to technological advancement,infustructure development and increased production which will contribute to higher yield and export values.**FDI outflow** indicates investments made by domestic agricultural enterprises in foreign markets.Outbound investments can provide access to new markets,resources,technologies,increasing competitiveness.So monitoring both feature can provides potential insights and can directly inflluence export performance.So both feature is important for export crop value prediction and that's why it has been considered here.

7. **Land Temperature** The rows containing element as "Temperature change" has been filtered out and grouped by 'Area Code (M49)', 'Area', and 'Year',which calculated the mean 'Value' of each group. This value is renamed and represented as a average temperature change for each area and year.//[3pt] Calculating Average Land temperature is an important factors for agricultural productivity and crop yields of a country and year.By assessing temperature changes over time,insights can be gained about climatic conditions experienced by different regions,which directly impact crop growth,development and overall agricultural output.Fluctuation in temperature can affects various aspects of cultivation,growth rates,susceptibility to pests which can direct impact yield and export value of crops.So its a important feature for prediction export value of crop production.

8. **Pesticides Use** The row containing items as "Pesticides(Total)" has been filtered out and grouped by 'Area Code (M49)', 'Area', and 'Year',which calculated the mean 'Value' of each group.This value is renamed and represented as 'Average pesticides value'.

   Using average value pesticides usage played a important role in agricultural practices,helping to manage pests,disease and weeds that can negatively impact crop yields.Excessive or improper use of pesticides can have negative effects on environment,human health and trade regulation,which may ultimately impact export values. So for export value of production this feature is very much important.The overall level of pest management practices directly affects the crop healths and productivity.

9. **Land Use** Rows containing official figures and cropland item has been filtered out and average cropland use has been considered using group by method on 'Area Code (M49)', 'Area', and 'Year'.To make the dataset concise,only average cropland area has been filtered out.

   Cropland is considered as it represents the extent of land dedicated to agricultural cultivation,directly influencing crop production capacity and consequently export values.Increased crop land use typically leads to higher agricultural productivity,resulting in larger quantities of crops available for export.This expansion of crop land can result from various factors like technological advancements,improved farming practices,and government policies promoting agricultural growth.Diversity of crops grown on the expanded land can also influence export values.Region with a wide variety of crops cultivated on their land can enhance export revenue.Quality and yield of crops produced on the land can directly impact their marketability and consequently, export value.To put it briefly, monitoring cropland area offers important contextual data that is necessary to comprehend the dynamics of agricultural trade and production.

10. **Food Trade**This is one of the most important data as the target variable is present here.Row containing non-crops item like 'Meat and Meat Preparations', 'Dairy Products and Eggs', 'Non-food', 'Other food', 'Alcoholic Beverages' has been removed and only the element containing export value has been considered.Total export values has been calculated after grouping it based on 'Area Code (M49)', 'Area', and 'Year'.

    Total Export value for all area and year has been filtered which is utilized here to predict the export value of crops.**So this variable will be act as target here**.It is to be noted that this "target" variable is considered dependent as its value depend on all values of the features.The variables which are discussed just before this acts as a independent variables as their values are used to predict the value of target variable.

11. **Some of the features has not been used here-**

    - **consumer production Indicators**Food price inflation is considered at first considering the market dynamics and economical impact.But in the final dataframe food price inflation has not been considered.It is often correlated to yield,weather conditions and costs which can influence both food prices and export values of crops.If other features already captures similar information or have stronger correlation with export values,adding food price inflation may not provide additional predictive capability to the model.
    - **Employment** Employment factors may not directly impact the export values of crops as other factors such as yield,weather conditions.Therefore prioritizing features that have a more direct influence on crop export values would be more appropriate in this scenario.
    - **Food Security** while food security may be good for economic stability,its direct impact on crop export values might be less pronounced compared to other factors like yield,pests and climate condition.Additionally,this may not capture the nuances of international trade dynamics.

After preparing the files,merging has been done by taking all the relevant dataframes utilizing an outer join.This approach is chosen to construct a comprehensive dataset for the prediction of crop export values.Common identifiers like 'Area','Area Code (M49)','Year' has been used for merging the data.Utilizing the outer joins ensures that no information has been lost during the merging process,enabling a thorough analysis of the interplay between various factors and their impact on crop export values.
After merging the dataset the shape of the merged dataframe is **(9731, 14)**.

3 Data Pre-Processing

After preparing the feature and target variables,data pre-processing will be done which includes checking null values,outliers,skewness.

3.1 Checking and removal of Null Values

Before merging the file,each dataframe has been checked to ensure if there is any null value.After merging,many null values can be observed.Huge number of null values has been observed after merging the dataframe as for some area and year, feature values are missing.The distribution of null values shown below1

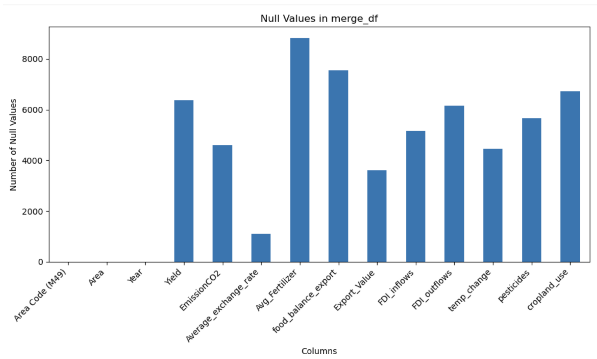


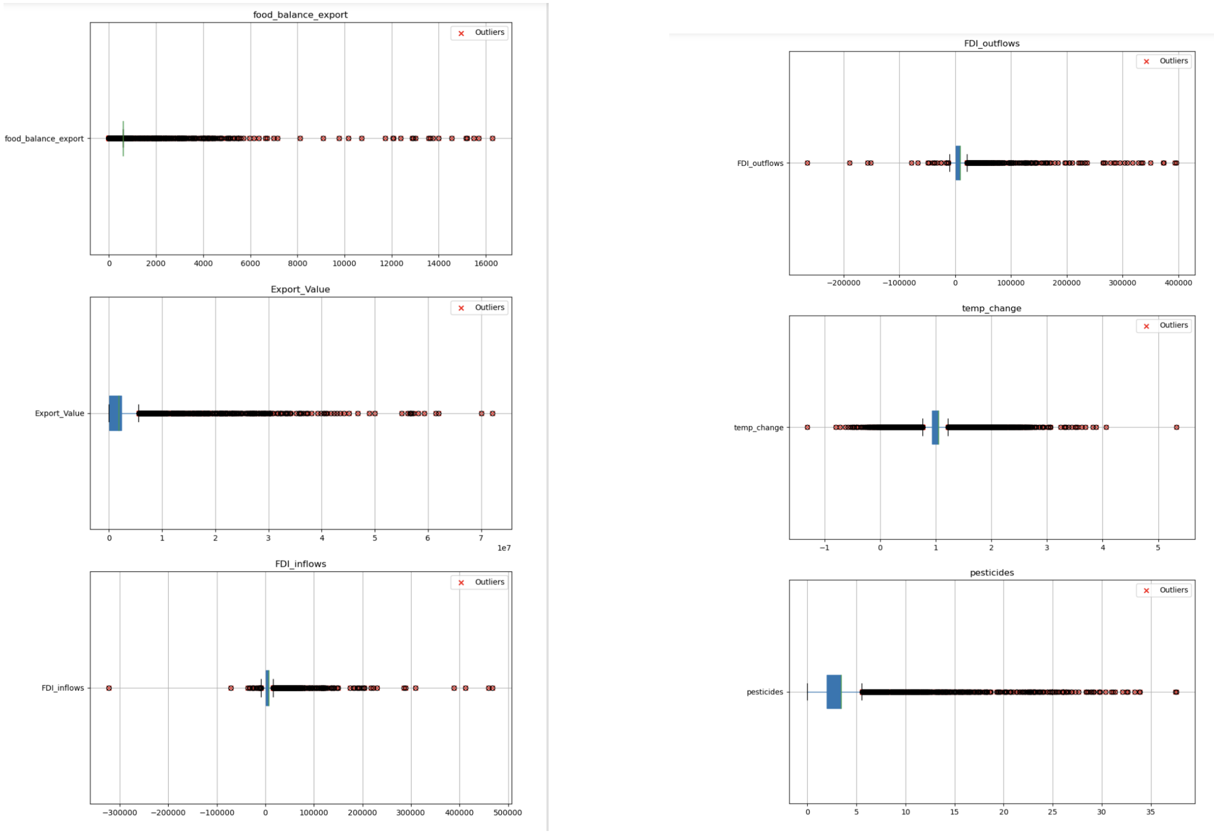
Figure 1: Null Values distribution

3.1.1 Zero-value imputation

Zero-value imputation has been used which is a method to handle these missing values in a dataset by replacing them with zeros.This approach is often used when missing values are assumed to represent a true absence of the attribute being measured,rather than data entry errors or other anomalies.For example:if a particular country does not report any Co2 emission for a specific year,it could imply that there were indeed no emission recorded rather than an missing data.In such cases,replacing missing values with zeros allows us to preserve integrity of the dataset and proceed with the analysis without introducing bias or distortion.This is to mention that all the features and target except Year and Area Code is float object. 'Area' column has been removed from the dataset as it is not a float or int datatype.

3.2 Outliers Removal

Outliers are data points which needs to be noted as they significantly differ from other observation in a dataset.They can manifest as unusually high or low values and may indicate variability,errors or anomalies in the data.It can skew the statistical analysis which can lead to inaccurate results or biased interpretations, so needs to be addressed.Many number of outliers has been observed here which needs to be handled carefully.Some outliers has been given below using a boxplot.



Outliers outliers

Figure 2: Outliers in the dataset

3.2.1 z-score Method

To identify and removing outliers,z-score methods has been used.This function calculates the z-score for each data point and identifies outliers based on a specified threshold.Here,threshold value of 3 has been taken.Data points with a z-score greater than 3 or less than -3 are considered outliers.Mean μ. and Standard Deviation σ computes means and standard deviation of the data,denoted by:

$$z = \frac{x_i - \mu}{\sigma}$$

The condition for identifying outliers based on the z-score can be written as:

$$|z| > 3 \tag{1}$$

After identifying the rows containing outliers,these rows are filtered out from the specified column.

3.3 Skewness

Skewness indicates the degree to which data deviates from symmetry.positive skewness indicates that right tail of the distribution is longer than left one,while negative skew is the opposite.A skewness of 0 indicates perfectly symmetric distribution.From a initial distribution plots,it is observed that many columns are right-skewed which indicates most of the data points are concentrated on left side of the distribution with a few larger values extending the tail to the right.From the below3,column like-EmissionCO2,Average_exchange_rate,Avg_Fertilizer,food_balance_export showing right skew distribution.

3.3.1 Log transformation

To handle the skewness,**log transformation** has been applied only for the float columns.This transformation can make the data more normally distributed and stabilized the variance of data,which is often an assumption for many machine-learning algorithm.

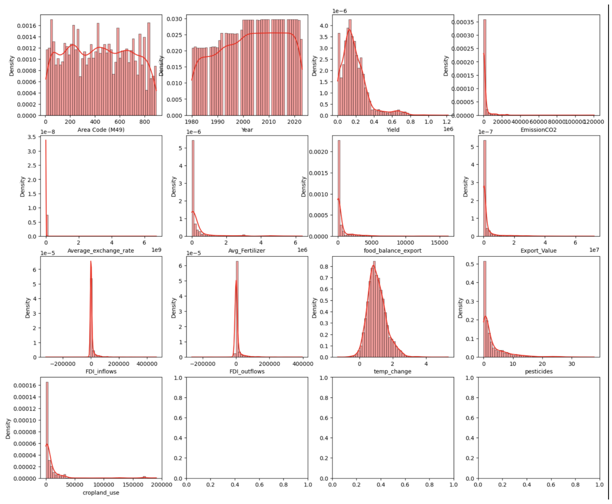


Figure 3: Skewness check before applying log transformation

The transformation can be expressed mathematically as follows:

$$y_i' = \log(y_i) \tag{2}$$

y_i is the original data point and y_i' is the transformed value of datapoints.In this scenario,

$$\text{If } i > 0, \text{ it is transformed to } \log(i).$$

$$\text{If } i \leq 0, \text{ it is set to 0 to avoid taking the logarithm of non-positive numbers, which is undefined.}$$

After this operation, the variables get normally distributed pattern.

3.4 Correlation Matrix

A correlation matrix is a table that displays the correlation coefficients between multiple variables.Each cell in the table shows the correlation between two variab;es.The va;ues in the correlation matrix can range from -1 to 1.Below is the correlation matrix4 Some of the insights taken from correlation matrix:

1. Higher crop yields are strongly associated with higher export values.
2. high Foriegn direct investment inflows strongly associate with high crop yield and high export values,high outflow.
3. show a positive relationship between food balance export quantities ans crop yield.
4. Higher CO2 emissions are moderately associated with higher yields,possibly due to more intensive agricultural practice.
5. Temperature change shows generally weak correlation, indicating it may not be a primary driver but still relevant for specific variables like food balance.
6. exchange rates is weakly correlated with most variables,indicating other factors plays a more important role.

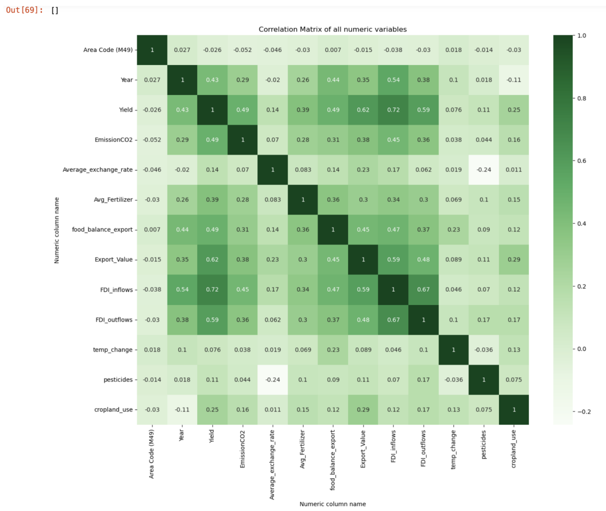


Figure 4: Correlation Matrix

The correlation matrix reveals several important relationships between agricultural variables,investment flows and other environmental factors.

3.5 Standard Scaling

The StandardScaler is a pre-processing techniques used to standardize features by removing mean and scaling them to unit variance.It is good mainly for algorithms that rely on distance-based measures or gradient descent optimization. Mathematically,its represented as:

$$x_{scaled} = \frac{x - \mu}{\sigma}$$

X_{scale} is the scaled feature values, X is the feature value, μ is mean and σ is the standard deviation.

Scaling has not been done on target variable as regressor will be applied.Scaling the target variable would change its original scale,making it more challenging to interpret the coefficients of the model in the context of the original problem.Most regression algorithm minimizes a loss function such as MSE or MAE to optimize the model parameters.Scaling the target would alter the magnitude of the loss potentially affecting the optimization process and leading to suboptimal results.After the data preparation and pre-processing is done,Model has been build in the next section.

4 Multi-layer perceptron Model

Here,A MLP regressor model has been build to predict the export value of crops.

4.1 Tensors and dataloaders

1. The pre-processed data is converted from NumPy arrays to PyTorch tensors,which are a required step for training neural network in pytorch.

2. A dataloader is created for the training set to facilitate mini-batching which will allows the model to update its parameters using smaller subsets of training data,improving efficiency and convergence.

4.2 Model Description

The MLPRegressor model has been built here.

1. **fc1**:Represent the first fully connected(dense) layer,which perform linear transformation of the input features followed by ReLU activation function.It has input size and 50 output neurons.

2. **dropout1**: Denotes the first dropout layer,which randomly sets a fraction of input units to zero after the first fully-connected layers.Dropout is 0.2

3. **fc2**:Second fully-connected layers which takes the output from fc1,apply another linear transformation followed by a ReLU activation function,and outputs 100 neurons.
4. **dropout2**:represents the second dropout layer,which applies dropout regularization to the output of fc2.
5. **fc3**:This Denotes the final fully connected layer,which takes the output from fc2 which performs a linear transformation and produces a single output neuron for regression.

The forward pass of the neural network where input data 'x' is propagated through each layers.Mathematically,it can be represented as follows:

$$h^{(1)} = \text{ReLU}(W^{(1)}x + b^{(1)})$$
$$\text{where} \quad \text{ReLU}(x) = \max(0, x)$$

This is the output of the fully-connected layer(fc1). $h^{(1)}$ is then passed through the first dropout layer.

$$h^{(2)} = \text{ReLU}(W^{(2)}h^(1) + b^{(2)})$$

This is the output of the second fully connected layer(fc2).h2 is passed through the second dropout layer,similar to first dropout layer.

$$\hat{y} = (W^{(3)}h^(2) + b^{(3)})$$

This is the output of the final fully connected layer(fc3),representing the output for regression.

4.3 Model training

The model is initialized with the input size corresponding to the number of features in the training data.

1. **The Mean Squared Error (MSE)** loss function is used here.It measures the average squared difference between the predicted and the actual target values within a dataset.

   $$\mathcal{L}(\hat{y}, y) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

   where y_i is the true value and $\hat{y}_i$ is the predicted value

2. **An Adam optimizer** is used with a learning rate of 0.01.It is a gradient descent optimizer algorithm.Adaptive Moment Estimation or Adam is an additional technique that calculates adaptive learning rates for every parameter.Adam can be seen of as a hybrid of RMSprop and Adagrad where RMSprop do well in nonstationary and online environment,while Adagrad performs well on sparse gradients.It requires little memory and is computationally efficient.

   $$\theta_{t+1} = \theta_t - \eta\frac{\hat{v}_t}{\sqrt{\hat{m}_t} + \epsilon}$$

   θ_{t+1} is the updated parameter at time-step t+1
   θ_t is parameter at time-step t
   η is the learning rate
   $\hat{v}_t$ is the bias-corrected second moment estimate
   $\hat{m}_t$ is the bias-corrected first moment estimate.
   ϵ is a small constant to prevent division by zero.

3. **The learning rate** is reduced using a scheduler that reduces the learning rate by a factor of 0.2 if the validation loss does not improve for 5 epochs,with a minimum learning rate of 0.0001

4. **The training loop** run for 500 epochs.

   - **Forward pass**:For each batch of training data,the model computes the outputs.

     $$y = \text{MLP}(x) \tag{3}$$

   - **Loss computation**The MSE loss is computed:

     $$L = \text{MSELoss}(y, \hat{y})$$

- **back-propagation**Gradients are computed using back-propagation.

$$W \leftarrow W - \eta \frac{\partial L}{\partial W}$$

  The optimizers updates the model parameters.This step is implicitly included in the backpropagation and weight update process.

5. Validation loss is evaluated at each epoch to monitor performance and adjust the learning rate.

4.4 Over-fitting Handling

To prevent overfitting, several techniques have been implemented:

1. **Dropout regularization**Dropout layers with a 0.2 dropout probability are added in this model after every hidden layer. This indicates that there is a 20% probability that every neuron in the hidden layers will be "dropped out" or set to zero during training. Dropout regularisation lessens neurons co-adaptation in this way, strengthening the model and lowering its propensity for overfitting. Let x be the input to a dropout layer, p be the dropout probability (i.e., the probability of dropping out a neuron), and x_{dropout} be the output of the dropout layer.

   During training, each neuron in x is set to zero with probability p, and the remaining values are scaled by a factor of $\frac{1}{1-p}$ to maintain the expected value. Mathematically, this can be expressed as:

   For each element x_i in x:

$$r_i \sim \text{Bernoulli}(1-p) \quad \text{(random binary mask)}$$

$$x_{\text{dropout}_i} = \frac{x_i \cdot r_i}{1-p}$$

   Here, r_i is a binary random variable drawn from a Bernoulli distribution with probability $1-p$. x_{dropout_i} represents the value of x_i after dropout.In the context of a neural network, if x is the output of a hidden layer, then x_{dropout} becomes the input to the next layer. This process is repeated for each hidden layer where dropout is applied.

2. **learning rate reduction** A learning rate reduction schedular(ReduceLROnPlateau) is used to adjust learning rate during training.The model can more carefully adjust its parameters by progressively lowering the learning rate, possibly preventing significant oscillations that could cause overfitting. If the validation loss does not improve after a certain number of epochs (patience), the ReduceLROnPlateau scheduler lowers the learning rate by a factor of 0.2 and makes sure the learning rate does not fall below a minimum value (min_lr). By adjusting the learning rate adaptively, the model is able to converge more successfully and avoids becoming trapped in local minima while being trained.

4.5 Hyper-parameter

1. Batch Size: 256 (defined in the DataLoader)
2. Learning Rate: 0.001 (defined in the Adam optimizer)
3. Number of Epochs: 800
4. Dropout Rate: 0.2 (applied after each fully connected layer)
5. Loss Function: Mean Squared Error (MSELoss)
6. Optimizer: Adam optimizer
7. Learning Rate Reduction Scheduler: ReduceLROnPlateau with a factor of 0.2, patience of 5, and a minimum learning rate of 0.001

5 Performance

5.1 Total Number of Instances and Splitting Strategy

The training and test sets were derived randomly from data to ensure that the model learns from a diverse range of instances and can generalize well to unseen data.80:20 random splitting is done ensuring that both sets have a representative distribution of instances.

1. **dataset size before splitting**: The original dataset size after removing outliers was - (8723, 13) that is 8723 instances has been used.

2. **Splitting strategy** The original data set was split into train and test data. Training data will be used for training model. Test data will be act as unseen data which is used at a later stage.Training data size is (4151,13) that is only 4151 instances has been used.Testing data has 588 instances.The test-data contains data from 2020-2022 ensures it remains unseen during training.This unseen data is crucial for evaluating and predicting model's performance on new, unseen observation.This will help us to know how well the model can generalizes to unseen data and assess its predictive accuracy.Rest of years will be treated as a train data which will be used to train the model.After the model is trained on historical data, it may be used to forecast new, unknown data by identifying patterns, correlations, and trends in the data.

3. Training is splitted into 12 feature and 1 target variable.Target variable is the export value of the crops.And all the other are the feature variable

4. **Train-val-Test split** The data was further split into training,validation and testing sets using train-test-split function from scikit-learn.In Training data,the feature and target variable is set.Target variable is the export value of the crops and features input are the rest of the variable The training and validation sets are formed by splitting the feature and target variables into 75% training (X train,y train) and 25% for validation(X val,y val).The testing set remain separated for model evaluation(X test,y test).The instances are mentioned below:

   - Training set shape (X_train, y_train): (2490, 12) (2490,) containing 2490 instances
   - Validation set shape (X_val, y_val): (830, 12) (830,) containing 830 instances
   - Testing set shape (X_test, y_test): (831, 12) (831,) containing 831 instances

This splitting strategy ensures that the model is trained on a sufficient amount of data to capture underlying patterns while reserving unseen data for evaluating its generalization performance.

5.2 Performance Report

After training model has been evaluated on the test set.MSE(mean-squared error) has also been calculated for the test-set.

$$\hat{y}_{test} = \mathrm{MLP}(x_{test}) \tag{4}$$

1. **Train Loss** at final epoch was 2.6

2. **Validation Loss** at the final epoch was 2.9.the model performs similarly well on unseen data, suggesting good generalization.

3. **Test set performance**

   - **MSE** The model achieves a MSE score of 2.9 on Test set.
   - **R2 score** The model achieves a r2 score of 0.90 on test data.This statistic measures the proportion of the variance in the dependent variables that is predictable from the independent variables.An r2 value close to 1 indicates a strong fit,meaning that the model explains a high proportion of the variance in target variable.
   - **Mean Absolute Error (MAE)**: The MAE on the test set was 1.1.

Train Loss	Val Loss	MSE	R-2 score	mean absolute error
2.6	2.9	2.9	0.90	1.1

Table 1: Result Analysis of MLP regressor model

The MLP regression model demonstrated strong performance as mentioned by high r2-score and a reasonably low MSE.These results suggest that the model effectively capture the relationship between the features and target variable,providing close good prediction.Overall,the model performs well across all metrics with low losses and a high R2 score indicating effective learning and predictive capability.There is still room for further improvement through techniques such as hyperparameter tuning,additional feature engineering or model architecture refinement.

6 Conclusion

The data has been prepared at first and features,target has been selected based on the problem.Several data pre-processing has been done to ensure data is cleaned and scaled.Multi-layer perceptron Regressor model has been build.The model has been passed through the unseen test data created during splitting and predicted value is somewhere close to the actual one indicating that the model performed well.The output of unseen test data of next three years has been recorded in a csv file(277244_output.csv) mentioning year-area index,actual and predicted output.

7 Referenecs

1. Adam optimizer: https://insideaiml.com/blog/Adam-Optimizer:-In-depth-explanation-1051
2. standard scalar: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html
3. MSE:https://encord.com/glossary/mean-square-error-mse/
4. z-score:https://www.geeksforgeeks.org/z-score-for-outlier-detection-python/
5. Histogram plotting: https://stackoverflow.com/questions/53646710/plot-a-histogram-through-all-the-columns
6. Regressor model in pytorch reference-https://machinelearningmastery.com/building-a-regression-model-in-pytorch/
7. Dropout: https://towardsdatascience.com/simplified-math-behind-dropout-in-deep-learning-6d50f3f47275

8 Appendix

see 277244_ML_assignment.pdf file for codes
see 277244_output.csv file for test output.