National College of Ireland

Configuration Manual

MSc Research Project
MSCDAD

Lynnet Grace Nakagiri
Student ID: X 23217677

School of Computing
National College of Ireland

Supervisor: Harshani Nagahamulla

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Lynnet Grace Nakagiri		
Student ID:	x23217677		
Programme:	MSc. Data Analytics	**Year:**	2024
Module:	MSc. Research Project		
Lecturer:	Harshani Nagahamulla		
Submission Due Date:	12/12/2024		
Project Title:	A Hybrid Machine Learning Approach for Crop Classification, Yield and Fertilizer Prediction for Sustainable Agriculture		
Word Count:	557	**Page Count:**	4

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Lynnet Grace Nakagiri

Date: 11th December 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Lynnet Grace Nakagiri
Student ID: x23217677

1. Introduction

This configuration manual provides comprehensive details on the hardware and software requirements, datasets, and implementation steps for the research project titled " A Hybrid Machine Learning Approach for Crop Classification, Yield and Fertilizer Prediction for Sustainable Agriculture." The tools and methods used to create the hybrid machine learning is outlined. A step-by-step guide is provided in the following manual with more details to ensure reproducibility, including configuration details, explanations and code snippets.

2. System Configuration

The Hardware requirements used in this project are as follows:

- Processor: 13th Gen Intel(R) Core(TM) i5-1335U 1.30 GHz
- RAM: 8 GB
- System Type: Windows 11 OS, 64-bit
- Storage: 475GB SSD

In order to implement the project, the following Software libraries and tools are used:

- Microsoft Excel 2016 was used to display datasets downloaded in flat files as CSV.

- IDE/Notebook: Google Colab (used for cloud-based computation), Visual Studio Code and Jupyter Notebook. Google Colab has an option where we are able to choose GPU or TPU settings dependent on how big the dataset we are to work on is. In order to select the GPU settings which were used for this research, select the change runtime type option in the Runtime tab as seen in Figure 1

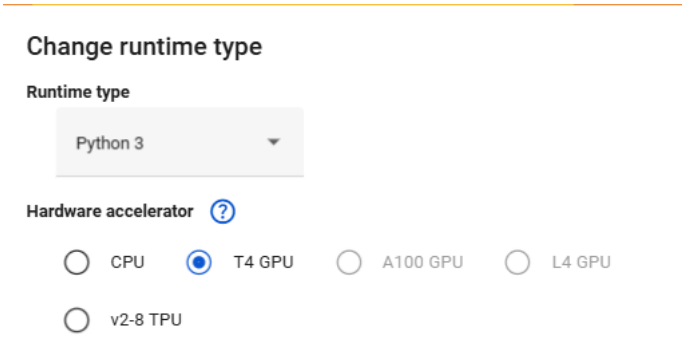

Figure 1: Setting of the GPU in Colab

The Programming Language used in this project is Python 3.8+

3. Libraries and Tools

The necessary libraries are loaded in the jupyter notebooks are seen in Figure 2 and these include:

```
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.ensemble import RandomForestClassifier, RandomForestRegressor, GradientBoostingRegressor
from xgboost import XGBRegressor, XGBClassifier
from sklearn.svm import SVR
from sklearn.neural_network import MLPRegressor
from sklearn.model_selection import train_test_split, RandomizedSearchCV
from sklearn.feature_selection import RFE
from sklearn.metrics import ConfusionMatrixDisplay
from sklearn.metrics import mean_squared_error, accuracy_score, confusion_matrix, classification_report, mean_absolute_error, r2_score, mean_absolute_percentage_error
from tensorflow.keras.models import Model
from tensorflow.keras.layers import LSTM, Dense, Input, Dropout
import tensorflow as tf
import warnings

warnings.simplefilter("ignore")
```

Figure 2: Python Libraries used.

Below is a brief explanation of the libraries used:

Data Handling: pandas, NumPy for data manipulation and processing.
Visualization: matplotlib and seaborn for data visualization and evaluation metrics.
Machine Learning: scikit-learn for model training, feature selection, and hyperparameter tuning.
Deep Learning: TensorFlow/Keras for training neural network models (LSTM and hybrid neural network).
Utility: warnings, os

4. Project Development

The First stages carried out in the development of the research project include column standardization, merging the datasets and data preprocessing. The predictive model is implemented using the keras and sk-learn (scikit-learn) python libraries.

3.1 Data Preparation

The data preparation step is carried out using the numPy and pandas libraries. For the research, the two datasets that were used are in the CSV format.
These datasets used in this are loaded into the Jupyter Notebook as illustrated in Figure 3 below:

```
# Load datasets
fertilizer_data = pd.read_csv('Fertilizer Prediction.csv')
crop_yield_data = pd.read_csv('crop_yield.csv')
```

Figure 3: Loading of the datasets

Crop Yield Dataset.[1] contains crop production data by state and year, including features like area, production, and yield as seen in figure 4.

	Crop	Crop_Year	Season	State	Area	Production	Annual_Rainfall	Fertilizer	Pesticide	Yield
0	Arecanut	1997	Whole Year	Assam	73814.0	56708	2051.4	7024878.38	22882.34	0.796087
1	Arhar/Tur	1997	Kharif	Assam	6637.0	4685	2051.4	631643.29	2057.47	0.710435
2	Castor seed	1997	Kharif	Assam	796.0	22	2051.4	75755.32	246.76	0.238333
3	Coconut	1997	Whole Year	Assam	19656.0	126905000	2051.4	1870661.52	6093.36	5238.051739
4	Cotton(lint)	1997	Kharif	Assam	1739.0	794	2051.4	165500.63	539.09	0.420909

Figure 4: Crop and Yield dataset

Fertilizer Dataset.[2] consists of data on fertilizer usage, environmental factors like rainfall, temperature, and soil conditions as seen in figure 5.

	Temparature	Humidity	Moisture	Soil Type	Crop Type	Nitrogen	Potassium	Phosphorous	Fertilizer Name
0	26	52	38	Sandy	Maize	37	0	0	Urea
1	29	52	45	Loamy	Sugarcane	12	0	36	DAP
2	34	65	62	Black	Cotton	7	9	30	14-35-14
3	32	62	34	Red	Tobacco	22	0	20	28-28
4	28	54	46	Clayey	Paddy	35	0	0	Urea

Figure 5: Fertilizer Dataset

[1] https://www.kaggle.com/datasets/akshatgupta7/crop-yield-in-indian-states-dataset
[2] https://www.kaggle.com/datasets/gdabhishek/fertilizer-prediction

Before merging the fertilizer data with the crop yield dataset, column names in the fertilizer dataset are standardized to match the naming conventions in the crop yield dataset as seen in Figure 6 below.

```
# Standardize column names in fertilizer data for merging compatibility
fertilizer_data = fertilizer_data.rename(columns={
    'Temparature': 'Temperature',
    'Phosphorous': 'Phosphorus',
    'Crop Type': 'Crop'
})
fertilizer_data.columns = fertilizer_data.columns.str.strip()
```

Figure 6: Standardizing of Column names

After standardizing the column names, the fertilizer and crop yield datasets were merged on the common 'Crop' column. This allows for combining fertilizer and crop yield features into a single dataset.

```
# Merge fertilizer and crop yield data on 'Crop' and environmental factors
merged_data = pd.merge(crop_yield_data, fertilizer_data, on='Crop', how='inner')
```

Figure 7: Merging of datasets

This step ensures that data related to crop types and environmental factors (like temperature, humidity, etc.) are aligned in a unified dataset as seen in Figure 7.