

## Data Collection and Preprocessing Phase

Date	01 December 2024
Team ID	739791
Project Title	Rice Crop Monitoring-Time Series Analysis
Maximum Marks	6 Marks

### Preprocessing Template

The images will be preprocessed by resizing, normalizing, augmenting, denoising, adjusting contrast, detecting edges, converting color space, cropping, batch normalizing, and whitening data. These steps will enhance data quality, promote model generalization, and improve convergence during neural network training, ensuring robust and efficient performance across various computer vision tasks.

Section	Description
Data Overview	The dataset consists of data of rice crop production by various aspects, such as area, year, unit, value, flag , flag description. The data is collected from publicly available Kaggle datasets (e.g., Rice Crop Monitoring-Time Series Analysis )
Label Encoder	A tool in machine learning to convert categorical labels into numerical form.
Time Stamp	A sequence of characters representing a specific date and time.Useful for time series analysis, logging events, and tracking.
Bar plot	A graphical representation of data using rectangular bars. Used to compare different categories of data
<b>Data Preprocessing Code Screenshots</b>	

## Loading Data

```
df = pd.read_csv('rice production across different countries from 1961 to 2021.csv')
df.head()
```

	Area	Year	Unit	Value	Flag	Flag Description
0	Afghanistan	1961	tonnes	319000.0	A	Official figure
1	Afghanistan	1962	tonnes	319000.0	A	Official figure
2	Afghanistan	1963	tonnes	319000.0	A	Official figure
3	Afghanistan	1964	tonnes	380000.0	A	Official figure
4	Afghanistan	1965	tonnes	380000.0	A	Official figure

## Label Encoder

```
le=LabelEncoder()
df['Area']=le.fit_transform(df['Area'])
df['Flag Description']=le.fit_transform(df['Flag Description'])
df['Unit']=le.fit_transform(df['Unit'])
df['Flag']=le.fit_transform(df['Flag'])
df.info()
```

[15]

```
... <class 'pandas.core.frame.DataFrame'>
RangeIndex: 7324 entries, 0 to 7323
Data columns (total 6 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   Area                7324 non-null   int64
 1   Year                7324 non-null   int64
 2   Unit                7324 non-null   int64
 3   Value              7324 non-null   float64
 4   Flag                7324 non-null   int64
 5   Flag Description    7324 non-null   int64
dtypes: float64(1), int64(5)
memory usage: 343.4 KB
```

## Time Stamp

```
type(dates[0])
```

[9]

```
pandas._libs.tslibs.timestamps.Timestamp
```

## Bar Plot

