Technical Document

Project Title:	
Student ID:	21262085
Student name:	Sheetal Kumar
Student email	sheetal.kumar8@mail.dcu.ie
Chosen major	Data Analytics
Student ID:	21267100
Student name:	Patrick Travers
Student email	patrick.travers3@mail.dcu.ie
Chosen major:	Artificial Intelligence
Supervisor	Prof. Martin Crane
Date of Submission	31/07/2022

Contents

Section 1 - Introduction

This Technical document is a manual which consists of the detailed description about the project and its components which are created to achieve the result.

The execution of the project follows the same sequence as the section mentioned below.

Section 2 – Data information

The data which we used consist of the agriculture commodity Corn and Soybean and weather data from US region. The crop data is from CBOT which consist of futures contracts price data for corn and soybean from 1 January 1974 to 31 December 2017.
Weather data is from three weather stations across the state of Iowa (Cherokee, Mason, and Le Claire) is freely available from the National Climatic Data Center's website (ncdc.noaa.gov).

Below table shows the location of data files:

Data	Location
Corn	src/inputdata/corn_data
Soybean	src/inputdata/soybean_data
Weather	src/inputdata/weather_data

Section 3 – Data transformation

Go to '*src/datapreprocessing*' location in git repository and execute the below described files which intern generate the desired output data mentioned in '*Final data output file'* column.

We have transformed the data as per below:

1. Down sampling of the data is done from daily to weekly to check for the one more variation of model of prediction of prices. The weekly down sampling is done by taking the weekly mean of the sample.

   Weekly data is generated by executing the below files:

Type of Data	File Name for data generation	Final data output file
Corn weekly data	*process corn_weekly.csv.ipynb*	corn_weekly.csv
Soybean weekly data	process soybean_weekly.csv.ipynb	soybean_weekly.csv
Weather weekly data	weather_engineer_2.ipynb	iowa_weather_train.csv, iowa_weather_valid.csv, iowa_weather_test.csv

2. Empirical mode decomposition (EMD) is done to reduce noise present in the data. The noise is removed from both daily and weekly data using this technique.
   Below are the details of notebooks and data output files.

Type of Data	File Name for data generation	Final data file name
Corn data	**Weekly** – process corn_emd_weekly.csv.ipynb **Daily** - process corn_emd_daily.csv.ipynb	**Weekly** -corn_emd_weekly.csv **Daily** - corn_emd_daily.csv
Soybean data	**Weekly** - process soybean_emd_weekly.csv.ipynb **Daily** – process soybean_emd_daily.csv.ipynb	**Weekly**- soybean_emd_weekly.csv **Daily** - soybean_emd_daily.csv

3. Similarly Variational mode decomposition (VMD) technique is applied on the data to reduce noise from daily and weekly data. It is more flexible technique than EMD in reducing noise.

   Below are the details of notebooks and data output files.

Type of Data	File Name for data generation	Final data file name
Corn data	**Weekly** – process corn_vmd_weekly.csv.ipynb **Daily** - process corn_vmd_daily.csv.ipynb	**Weekly** -corn_vmd_weekly.csv **Daily** - corn_vmd_daily.csv
Soybean data	**Weekly** - process soybean_vmd_weekly.csv.ipynb **Daily** – process soybean_vmd_daily.csv.ipynb	**Weekly**- soybean_vmd_weekly.csv **Daily** - soybean_vmd_daily.csv

Section 3 – Model Creation

This section contains the description of how models are generated and trained for the final data which is generated after data decomposition step.

Go to '*src/trainedmodel'* location in git repository and execute the below described files which intern generate the desired output data mentioned in '*Description and output file***'** column.

We have used ARIMA and RNN models for the price prediction and created three different combinations for both daily, weekly and data from Wang's original paper for extracting best model on the basis of MAPE.

Below are the notebooks which can be used to for the same.

1. **ARIMA Models -**

File Name	**Description and output file**
ARIMA_corn_weekly.ipynb	This notebook will generate ARIMA model for weekly ***corn*** data. This book is basically covering all the aspect of creating an ARIMA model by making the data stationary and finding the correct differencing order. Final output of this notebook is a validation dataset which will be used in final model to check accuracy of test data prediction. Output files – vmd_ARIMA_weekly_valid.json, emd_ARIMA_weekly_valid.json
ARIMA_soybean_weekly.ipynb	This notebook will generate ARIMA model for weekly ***soybean*** data. This book is basically covering all the aspect of creating an ARIMA model by making the data stationary and finding the correct differencing order. Final output of this notebook is a validation dataset which will be used in final model to check accuracy of test data prediction. Output files – vmd_ARIMA_weekly_valid.json, emd_ARIMA_weekly_valid.json
ARIMA_corn_daily_Wang.ipynb	This notebook will generate ARIMA model for daily data used by Wang's[1] paper corn data. This book is basically covering all the aspect of creating an ARIMA model by

	making the data stationary and finding the correct differencing order. Final output of this notebook is a validation files which will be used in final model to compare with new model created in this project. Output files – vmd_ARIMA_daily_valid.json emd_ARIMA_daily_valid.json
ARIMA_soybean_daily_Wang.ipynb	This notebook will generate ARIMA model for daily data used by Wang's[1] paper ***Soybean*** data. This book is basically covering all the aspect of creating an ARIMA model by making the data stationary and finding the correct differencing order. Final output of this notebook is a validation files which will be used in final model to compare with new model created in this project. Output files – vmd_ARIMA_daily_valid.json emd_ARIMA_daily_valid.json

2. RNN model –

File Name	Description
RNN_corn_weekly.ipynb	This notebook will generate RNN model for weekly ***corn*** data. This book is basically covering all the aspect of creating an RNN model by dividing the data in train and validation set. Final output of this notebook is a validation dataset which will be used in final model to check accuracy of test data prediction. Output files – emd_RNN_weekly_valid.npy vmd_RNN_weekly_valid.npy true_RNN_weekly_valid.npy
RNN_soybean_daily_Wang.ipynb	This notebook will generate RNN model for daily ***soybean*** data and other parameter for the model same as used in Wang's paper. This book is basically covering all the aspect of creating an RNN model by dividing the data in train and validation set. Final output of this notebook is a validation dataset which will be used in final model to check accuracy of test data prediction. Output files – emd_RNN_daily_valid.npy vmd_RNN_daily_valid.npy true_RNN_daily_valid.npy

RNN_corn_daily_Wang.ipynb	This notebook will generate RNN model for daily ***corn*** data and other parameter for the model same as used in Wang's paper. This book is basically covering all the aspect of creating an RNN model by dividing the data in train and validation set. Final output of this notebook is a validation dataset which will be used in final model to check accuracy of test data prediction. Output files – emd_RNN_daily_valid.npy vmd_RNN_daily_valid.npy true_RNN_daily_valid.npy
RNN_soybean_weekly_weather.ipynb	This notebook will generate RNN model for weekly ***soybean*** data mapped with the weather data drawn over the years. This book is basically covering all the aspect of creating an RNN model by dividing the data in train and validation set. Final output of this notebook is a validation dataset which will be used in final model to check accuracy of test data prediction. Output files – emd_RNN_weekly_valid.npy vmd_RNN_weekly_valid.npy true_RNN_weekly_valid.npy
RNN_soybean_weekly.ipynb	This notebook will generate RNN model for weekly **soybean** data. This book is basically covering all the aspect of creating an RNN model by dividing the data in train and validation set. Final output of this notebook is a validation dataset which will be used in final model to check accuracy of test data prediction. Output files – emd_RNN_weekly_valid.npy vmd_RNN_weekly_valid.npy true_RNN_weekly_valid.npy
RNN_corn_weekly_weather.ipynb	This notebook will generate RNN model for weekly ***corn*** data mapped with the weather data drawn over the years. This book is basically covering all the aspect of creating an RNN model by dividing the data in train and validation set. Final output of this notebook is a validation dataset which will be used in final model to check accuracy of test data prediction. Output files – emd_RNN_weekly_valid.npy vmd_RNN_weekly_valid.npy true_RNN_weekly_valid.npy

After creating all the required validation test data by executing the above models we created forecast combination of various model with Artificial Bee Colony algorithm for calculating the weights for our test model for final prediction. Next section will follow the same.

Section 4 – Forecast Model combination

In this section we have created models which used the forecast data generated as part of section 2 and 3 steps as an input which then combined with Artificial Bee Colony (ABC) algorithm to calculate the weights for ABC and EIDS which will be used in the final test combination to calculate the accuracy of prediction on the basis of MAPE percentage.

Go to '*src/forecastcombination'* location in git repository and execute the below described files which intern generate the desired weight for proposed models to be used in test.

Models	Description
1. ABC_corn_weekly.ipynb 2. ABC_corn_weekly_weather.ipynb 3. ABC_corn_weekly_pruned.ipynb 4. ABC_corn_weekly_plusDUD.ipynb 5. ABC_corn_daily_Wang.ipynb 6. ABC_corn_daily_Wang_plusDUD.ipynb 7. ABC_corn_daily_Wang_pruned.ipynb	These model combinations first take the individual forecast data for corn (daily and weekly including weather) generated as part of section 3 model execution step and check for the individual MAPE % to check accuracy individually first. Then these are combined with ABC algorithm to create new forecast combination to calculate new weights for the test model. The weights are calculated by comparing ABC with Naïve and error-informed discritised search which will act as input for the test model
1. ABC_soybean_weekly.ipynb 2. ABC_soybean_weekly_weather.ipynb 3. ABC_soybean_weekly_pruned.ipynb 4. ABC_soybean_weekly_plusDUD.ipynb 5. ABC_soybean_daily_Wang.ipynb 6. ABC_soybean_daily_Wang_pruned.ipynb 7. ABC_soybean_daily_Wang_plusDUD.ipynb	These model combinations first take the individual forecast data for soybean (daily and weekly including weather) generated as part of section 3 model execution step and check for the individual MAPE % to check accuracy individually first. Then these are combined with ABC algorithm to create new forecast combination to calculate new weights for the test model. The weights are calculated by comparing ABC with Naïve and error-informed discritised search which will act as input for the test model

Section 5 – Final Test Models creation

In this section the final test model is created which take the input from section 3 and 4 generated files.

Go to '*src/corntestmodel and src/soybeantestmodel'* location in git repository and execute the below described files to test proposed models in the project.

	Files	Description
1.	test_ARIMA_corn_daily_Wang.ipynb test_ARIMA_corn_weekly.ipynb test_RNN_corn_daily_Wang.ipynb test_RNN_corn_weekly.ipynb test_RNN_corn_weekly_weather.ipynb	For test files name with ARIMA contains the logic used to prepare data as non-stationary for using in the ARIMA model by considering seasonal differencing with most prominent value. For which it checks for ACF and PACF order calculation, visualization of ACF and PACF graphs. At

		the end it generates validation and test set prediction for the final test file. Similarly, test files having RNN in it contains the logic of RNN model with required parameters to generates validation and test set prediction for the final test file.
2.	test_RNN_soybean_weekly.ipynb test_RNN_soybean_weekly_weather.ipynb test_RNN_soybean_daily_Wang.ipynb test_ARIMA_soybean_weekly.ipynb test_ARIMA_soybean_daily_Wang.ipynb	Similarly, these files do the same of soybean data.

At the end we have final test model which is created by using output files generated form above step.

Go to '*src/corntestmodel/test_combinations and src/soybeantestmodel/test_combinations'* location in git repository and execute the below described files to compare the proposed models with ABC combined models.

File	Description
corn_test_MAPEs.ipynb	This notebook basically combines the forecast of corn on individual, weekly and combined with weather and evaluation of those results with ABC and EIDS using MAPE % to identify which model provides better results.
Soybean_test_MAPEs.ipynb	Similarly, this notebook basically combines the forecast of soybean on individual, weekly and combined with weather and evaluation of those results with ABC and EIDS using MAPE % to identify which model provides better results.