MSc. in Computing
Practicum Approval Form

Section 1: Student Details

Project Title:	Crop Yield & Price Prediction using Machine learning model and crop models with adoption of remote sensing.
Student ID:	22265263, 22262962
Student name:	Hemant Kumar Singh, Sailesh Krishnan
Student email	Hemant.singh5@mail.dcu.ie, sailesh.krishnan2@mail.dcu.ie
Chosen major:	Data Analytics, Artificial Intelligence
Supervisor	Dr. Andrew McCarren
Date of Submission	30/11/2022

Section 2: About your Practicum

Please answer all questions below. Please pay special attention to the word counts in all cases.

What is the topic of your proposed practicum? (100 words)

Crop yield/price prediction using machine learning and crop modelling with remote sensing.

The use of machine learning models for crop yield prediction has yielded progressive results; however, there are aspects that could be taken into account for further improvement. For example, combining crop models with ML models is likely to generate even better results when satellite imaging is included. Additionally, the results of this work are not yet being used to predict the price of the crops. Our study will seek to use crop modelling, machine learning, and remote sensing to more accurately predict prices and ultimately the vendor's price

Please provide details of the papers you have read on this topic (details of 5 papers expected).

1. Rashid, Mamunur, et al. "A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction." IEEE Access 9 (2021): 63406-63439.

This study provides an overview of the use of machine learning algorithms to predict crop yield, with special emphasis on palm oil yield prediction. It discusses the current status of

palm oil yield around the world, widely used features and prediction algorithms, and the advantages and difficulties related to machine learning-based crop yield prediction. It also identifies current and future challenges to the agricultural industry and proposes potential solutions. Finally, it proposes a prospective architecture of machine learning-based palm oil yield prediction and suggests new research.

2. Nevavuori, Petteri, Nathaniel Narra, and Tarmo Lipping. ”Crop yield prediction with deep convolutional neural net-works.” Computers and electronics in agriculture 163 (2019): 104859.

In this study, Convolutional Neural Networks (CNNs) were applied to build a model for crop yield prediction based on NDVI and RGB data acquired from UAVs. Different aspects of the CNN such as selection of the training algorithm, depth of the network, regularization strategy, and tuning of the hyperparameters were tested.

3. Shahhosseini, Mohsen, et al.”Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt.” Scientific reports 11.1 (2021): 1-15.

This study explores the potential of combining crop modeling and machine learning to improve corn yield predictions in the US Corn Belt. Five ML models and six ensemble models were designed to test the hypothesis. Results suggest that adding simulation crop model variables as input features to ML models can reduce yield prediction root mean squared error (RMSE) by up to 20%. Further analysis revealed that soil moisture related APSIM variables are most influential on the ML predictions, followed by crop-related and phenology-related variables.

4. Gastli, Mohamed Sadok, Lobna Nassar, and Fakhri Karray. ”Satellite images and deep learning tools for crop yield prediction and price forecasting.” 2021 International Joint Conference on Neural Networks (IJCNN). IEEE, 2021.

This paper presents an investigation into the use of satellite images and deep learning models to predict crop yields and forecast farmers' prices. The images were converted to histograms to reduce the dimensionality. The models tested included LSTM, CNN, CNN-LSTM, and CNN-LSTM ensemble, each with a Gaussian Process for enhanced performance. The proposed ensemble of CNN-LSTMs was found to be the best model for predicting yearly soybean yields, forecasting daily strawberry yields and prices, and outperformed models from the literature with an improvement of 31% in terms of average Root Mean Square Error (RMSE).

5. Gavahi, Keyhan, Peyman Abbaszadeh, and Hamid Moradkhani. "DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting." *Expert Systems with Applications* 184 (2021): 115511.

This study proposes DeepYield, a combined structure of ConvLSTM and 3D CNN, to more accurately and reliably extract spatiotemporal features. The models are trained with historical yield data and MODIS LST, SR, and LC data from 1836 primary soybean growing counties in the US. Results show that DeepYield outperforms Decision Trees, CNN + GP, and CNN-LSTM, as well as ConvLSTM and 3DCNN.

How does your proposal relate to existing work on this topic described in these papers? (200 words)

All of the above mentioned work provides us multifarious views on the topic of crop yield prediction. [1] This research paper explores the adoption of machine learning algorithms like ANN, RF, CNN, DNN, LSTM, MARS, SVM, DT, etc., to forecast crop yield with particular prominence on palm oil yield prediction.[2] This study delves into specificity by suggesting the usage of Convolutional neural networks(CNN) with remote sensing and UAVs for forecasting crop yield.[3] This study brings up a very interesting idea of the creation of a hybrid model for crop yield prediction by merging machine learning with crop modelling. This study closely relates to our idea for the future work. [4] This study employs remote-sensing technology with data learning models like LSTM, CNN, CNN-LSTM, and CNN-LSTM ensemble for crop yield forecasting together with Farmers' price prediction. [5] This study adopts combined deep convolutional network based on ConvLSTM and 3DCNN, called DeepYield, for county-based crop yield forecasting across the CONUS. The findings of this study indicated that the most efficient deep features were determined by the proposed approach, outperforming other individual methods Decision Tree, CNN + GP, and CNN-LSTM**.** Conclusively, each study provided us better clarity about the approach we could follow in our future work.

What are the research questions that you will attempt to answer? (200 words)

1. How implementation of remote sensing along with Hybrid Model (crop modelling + machine learning) affects the crop yield prediction?
2. How can we utilise the results of crop yield prediction in crop price prediction?
3. Can a standard architecture be created for crop yield prediction?

How will you explore these questions?

- What software and programming environment will you use?

Python, APSIM and cloud platforms like google colab for virtual GPU and Tensorflow packages for deep learning techniques.

- What coding/development will you do?

We will be creating Machine learning models using the available data set. Prebuilt models will be taken into the workflow to create new feature to increase the efficiency of our model.

Moreover Deep Learning and Neural Network Programming will be implemented from our side.

- What data will be used for your investigations?

Our data pool will comprise of Yield Data, Weather Data, Soil data, Management Data and Satellite data inclusive of surface reflectance, temperature, moisture.

- Is this data currently available, it not, where will it come from?

Yes, Data is available. These are the few sources:

- temperature data : http://www.weather.gov/coop
- precipitation data: http://www.nssl.noaa.gov/projects/mrms
- radiation data : https://power.larc.nasa.gov
- Soil data: https://websoilsurvey.sc.egov.usda.gov/App/HomePage.html
- Satellite data : https://modis.gsfc.nasa.gov/data/dataprod/mod09.php , https://developers.google.com/earth-engine/datasets/ , https://www.climatologylab.org/terraclimate.html , https://earth.esa.int/eogateway/catalog
- Corn Yield data: https://www.nass.usda.gov/Data_and_Statistics/index.php

- What experiments do you expect to run?

Creating a hybrid model containing crop modelling and ML and integrating and determining the features from the crop modelling that are the most effective to be integrated. Utilisation of remote sensing data for forecasting yield price instead of price time series.

- What output do you expect to gather?

A more efficient model comparative to the existing models with more precise predictions, reflecting a decrease in root mean squared error for yield prediction.

- How will the results be evaluated?

The results will be evaluated based on the improvement of results in values of Performance metrics like Root Mean Squared Error (RMSE), Relative Root Mean Squared Error (RRMSE), Mean Bias Error (MBE) and Coefficient of determination (R2)