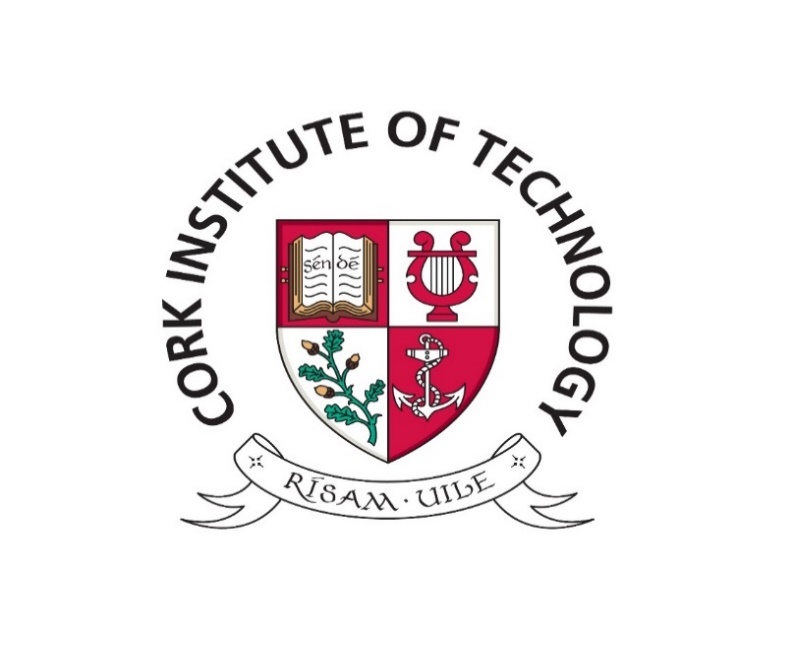
**CORK INSTITUTE OF TECHNOLOGY**



Department of Mathematics

**MSC IN DATA SCIENCE & ANALYTICS 2019-2020**

**CROP SELECTION METHOD TO MAXIMISE CROP YIELD RATE USING LONG SHORT-TERM MEMORY NEURAL NETWORKS(LSTM-NN) TIME SERIES FORECASTING**

|  |  |
| --- | --- |
| STUDENT NAME | SHIVAANI KATRAGADDA |
| STUDENT NUMBER | R00183214 |
| MODULE NAME | RESEARCH METHODS |
| MODULE CODE | MATH9001 |
| ASSIGNMENT NUMBER | 1 |
| PROFESSORS | ANGEUS DALY AND VINCENT CREGAN |
| DATE | 10TH MARCH 2020 |
| DAY | TUESDAY |

**INTRODUCTION:**

Agriculture is the backbone of every country. For economic growth agriculture depends on various crop types which rely on factors including weather conditions, soil classification and marketing price. Due to improper crop selection there is a deep reduction in crop yield which leads to insecurities and food shortages that can be experienced across the world. Analyzing these issues will potentially have the impact of increased and improved crop yield.

**INTRODUCTION OF DOMAIN:**

Data mining[1] is useful for predicting the crop yield production. It is defined as a process of finding patterns among multiple fields in large relational databases which can be converted into historical patterns and future trends.

**RESEARCH AREA:**

Throughout the growing and harvesting cycle of agriculture, machine learning has

become popular in this multi-disciplinary agriculture domain. It begins with the

plantation of a seed in the soil, followed by seed breeding and water fed measurement.

Which concludes at the harvest with the help of computer vision. The literature

review [2] shows that the most popular models in agriculture are Artificial and Deep

Neural Networks (ANNs and DL), Support Vector Machines (SVMs), Ensemble learning

(EL). The primary focus of this research is to implement crop selection method to

maximize crop yield rate using deep learning architecture of ANN, Long Short Term

Memory Neural Networks (LSTM-NN) time series forecasting [3].

**RESEARCH QUESTION:**

The primary aim of this research is to maximize crop yield rate of different crops and to investigate the crop trends of different countries in the world. The objective of the study will be to answer the following questions concerning the study area:

* Does crop production correlate with increasing population?
* Is crop yield is sufficient for Population?

**AIM:**

The main aim of this model is to implement crop selection method to maximize crop yield rate. We can predict the production of each crop by using LSTM-NN time series forecasting which will provide an estimate of the crops yielding rate. Therefore, we can establish which crops have higher demands for farming in the upcoming years.

**METHODOLOGY:**

The dataset is divided into two subsets, the first subset, comprising data from 1961–2008, is used for training the model. The second subset consists of data from 2009–2018 used for validation purpose. Since, the raw data included fluctuations that might affect the forecasting results, a smoothing function is used for training the models. One of the most widely used smoothing functions is Robust-LOWESS [4]. It uses local regression using linear recursive least squares (RLS) which decreases the effects of outliers.

**DATA SET:**

The dataset “Production\_Crops” is taken from Food and Agriculture Organization of the United Nations (FAO) [5]. The dataset consists of 51,154 entries and 123 total columns. Crop statistics are recorded for 173 products, covering the following categories: Crops Primary, Fibre Crops Primary, Cereals, Coarse Grain, Citrus Fruit, Jute Jute-like Fibres, Oil crops Primary, Treenuts, Vegetables and Melons. Data is expressed in terms of area harvested, production quantity and yield. The dataset consists of all primary crops of all countries in the world. The attributes of the dataset are as follows:

|  |  |  |
| --- | --- | --- |
| Attribute No | Attribute Name | Attribute Description |
| 1 | Area code | Area code of all countries of the world and geographical aggregates according to the United Nations M-49 list. |
| 2 | Area | All countries of the world and geographical aggregates according to the United Nations M-49 list. |
| 3 | Item code | Item code of each crop |
| 4 | Item | Name of the Items |
| 5 | Element code | Code of each element |
| 6 | Element | Data is expressed in terms of area harvested, production quantity and yield |
| 7 | Unit | Production Quantity and Seed: tonnes; Area harvested: hectares; Yield: tonnes per hectare. |
| 8-123 | Y1961, Y1961F  to  Y2018, Y2018F | The year indicates that the data refer to production, yield and area harvested during the calendar year |

**SUMMARY:**

The system uses the LSTM-NN based deep learning models it performs significantly better [6] for time-series forecasting. This is due to its capability of handling nonlinear systems due to their specialized LSTM nodes that performs better and it has a natural tendency to extract robust patterns for an input feature space.

WORD COUNT: 680

**REFERENCES:**

**CITATIONS BY USING IEEE STYLE:**

[1] “(1) (PDF) PREDICTING YIELD OF THE CROP USING MACHINE LEARNING ALGORITHM | IJESRT Journal - Academia.edu.” [Online]. Available: https://www.academia.edu/36345441/PREDICTING\_YIELD\_OF\_THE\_CROP\_USING\_MACHINE\_LEARNING\_ALGORITHM. [Accessed: 05-Mar-2020].

[2] “(PDF) Machine Learning in Agriculture: A Review.” [Online]. Available: https://www.researchgate.net/publication/327029380\_Machine\_Learning\_in\_Agriculture\_A\_Review. [Accessed: 05-Mar-2020].

[3] S. A. Haider *et al.*, “LSTM Neural Network Based Forecasting Model for Wheat Production in Pakistan,” *Agronomy*, vol. 9, no. 2, p. 72, Feb. 2019, doi: 10.3390/agronomy9020072.

[4] W. S. Cleveland, “Robust Locally Weighted Regression and Smoothing Scatterplots,” *J. Am. Stat. Assoc.*, vol. 74, no. 368, pp. 829–836, Dec. 1979, doi: 10.1080/01621459.1979.10481038.

[5] “FAOSTAT.” [Online]. Available: http://www.fao.org/faostat/en/#data/Q. [Accessed: 06-Mar-2020].

[6] M. Tan, C. dos Santos, B. Xiang, and B. Zhou, “LSTM-based Deep Learning Models for Non-factoid Answer Selection,” *ArXiv151104108 Cs*, Mar. 2016.