**CCT College Dublin**

MSc in Data Analytics

Integrated CA2 – Individual

Big Data and Agribusiness

**Lecturer Names:**  Marina Iantorno/Vladimir Milosavljevic

Muhammad Iqbal & David McQuaid

**Student:** Claudia Oliveira da Silva

**Student Number:** sbs22026

**Assessment Due Date:** 22nd May2022

**Date of Submission:** 21st May2022

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

**Table**

|  |  |  |
| --- | --- | --- |
| Table | …………………………………………………………… | 2 |
| 1. Abstract | …………………………………………………………… | 3 |
| 2. Introduction | …………………………………………………………… | 3 |
|  | …………………………………………………………… | 4 |
|  | …………………………………………………………… | 4 |
|  | …………………………………………………………… | 4 |
|  | …………………………………………………………… | 5 |
|  | …………………………………………………………… | 6 |
|  | …………………………………………………………… | 8 |
|  | …………………………………………………………… | 8 |
|  | …………………………………………………………… | 9 |
|  | …………………………………………………………… | 10 |
|  | …………………………………………………………… | 10 |
|  | …………………………………………………………… | 11 |
|  | …………………………………………………………… | 13 |
|  | …………………………………………………………… | 14 |
|  | …………………………………………………………… | 14 |
| 7. References | …………………………………………………………… | 15 |

1. **Abstract**

In general technological advancement alludes to the disclosure of better methods of producing goods. Changes in innovation lead to an expansion in the efficiency of work, capital, and different variables of production. The use of software algorithms created an avalanche of modernization in the Agriculture business. As the global population increase is decisive to leverage productivity, the profitability of businesses, food production, and the well-being of rural and urban society.

Buying rural areas is a very profitable investment than investments in foreign currency, fixed income, and gold especially in the period since 2008. This research paper is a comparative study of the price and size of the land designated for farming and the price the relation to the increase in milk productivity versus the quantity of fertilization used in the last years.

All data were collected on government and open websites, which we understand as a reliable sources. as Agricultural Production is a very specific topic and not linked to being part of our daily routine conversation, this research was not used web scrapping, instead was made a programmatic request for sentiment analysis.

Performance was compared using evaluation machine learning linear regression prediction. Upon successful completion of experiments, results show that regression-based machine learning models generally showed better results for modeling with lengthier historical data (more than three years). It was used different Datasets for training and test, we found good performance in almost all models (RandomForest Regressor, KNeighbors Regressor, AdaBoost Regressor, Linear Regression, MinMaxScale).

1. **Introduction**

Since the Farmers began to cultivate the land with the support of technology in agriculture, farming has never been the same. Instead, the open data available shows that, from time to time, technological innovations in agriculture provide a paradigm shift in the agricultural production model.

This means that the innovations in agriculture that characterize technologies such as Big Data on internet mobility, soil analysis, biofortification, GPS become part of life in the countryside, revolutionizing modern agriculture.

For us the majority for people that living in the big cities, it may seem obvious but the wireless and high-speed connection has allowed the introduction of numerous new technologies in agriculture. Now farmers can monitor the farm in real time, follow machines during work, manage staff and access data, KPIs and plantation information on their smartphone screen. All business details can be seen through the cloud, making it much easier to manage the production. The use of technology to management agriculture industry for change the production vision, reduce costs, increase productivity and reduce environmental damage.

Every year are new technologies, software(weather forecast, grain price, price of livestock markets), and of machines(GPS, tractors), soil analysis( fertilizers ) to accelerate the production process, new management procedures have been developed to face the sector's competition.

In this research we choose try to find a correlation between, price of the land for farming, use of fertilizers and increase the amount in agricultural production in the same period.

In view of the huge variety of agricultural products, was defined in our research analysis of a single product cow milk and than choose others to products work on. The reason for that was to create a challenge learning process. Instead, look for one single product, where we had much more guaranteed correlation if for example, cattle and pasture land price only, or looking for a date on cereal production. We prefer to work with products are a from a view of part without immediate correlation for a person with no knowledge in the agricultural sector.

So, in this paper, was analysed Price of Land, Type and Ages of Cows and Price of Fertiliser. The main challenge was to show the possibilities within Data Analysis in Marchine Learning and that the most important thing is not the amount of data, but what is done with it.

*Sub RQ: “What Machine Learning approaches are best suited to modelling different types and lengths of data (in terms of model accuracy and processing time), and how to analysis the correlation between apparently different topics.”*

The following sections of the paper are organized as follows. Section 3. presents the review of relevant literature in the field of Agriculture Industry. In section 4. the research methodology is described all process of data manipulation with corresponding outputs in section 5. Machine Learning experiment results and evaluation are presented in section 6, while section 7 presents further discussion of results. Section 8 concludes the research with references.

1. **Literature Review**

Many believe that Big Data is a large volume of information that circulates in the digital environment. Big Data, however, also exists in the data analysis part, in extracting relevant facts and applying the data transforming it in information.

Currently, the Big Data is stored in the cloud, the data can be consulted from anywhere, from the farm's seating room or in the middle of the field , and transfer to wherever the business is. In short, Big Data serves the purposes of traceability, prediction, and management of production software systems. With the ability to handle any type of digital record, some subjects are indispensable, and Big Data can record helping Farmers and Agriculture producers, such as geolocation videos, soil characteristics, and the history of water consumption and crop rotation.

Based on 5 principles (of speed, volume, veracity, variety, and value), Big Data uses diverse information collected in real-time to generate expertise. Applied to Machine Learning, data analysis provides information to farmers and machines ensuring the optimization of fertilizers, seeds, and others. It is also possible to identify threats and vulnerabilities and planning preventive actions that avoid losses.

Some of the benefits that agricultural technology provides are:

1. Productivity: with improved advances, it detects production bottlenecks and increases the number of plants per hectare, or the livestock thus increasing the productivity of the farm.
2. Reducing consumption of water, fertilizers, and pesticides, which, in addition to increasing profit, reduces the value of the product.
3. The action of environmental impacts on their land and the decrease in the production of chemicals that go in rivers and groundwater, make the business more sustainable.
4. Increased efficiency in the application of Fertiliser technology, for all types of Agriculture production
5. Detects soil quantity of nutrients that need to be added to the use as pasture for livestock.

One of the aspects that interfere in a good agricultural productivity is the soil, consequently, its manipulation and capacity of production through the correct application of nutrients and fertilization is of very important. In this context, fertilizers stand out as a technology whose function is to replace and provide the soil with the main nutrients essential for plant growth in order to expand the productive potential of crops and for grazing cattle, sheep, swine, as well as guaranteeing an appreciation of the price of the farm.

In our process of Data Preparation and Visualization was tried different types of charts and graphics in order to locate the most common type (compound) of Fertiliser and from there to test the relation between Price of Land, Productivity and Fertiliser.

|  |  |
| --- | --- |
| In the graphics you can see there are many input attributes, it makes difficult to visualize the data. Basically, it creates a concern that with this dataset with a big numbers of attributes will negatively affects the accuracy and training time of the machine learning model.  After a research the Principal Component Analysis (PCA) shows to be one way to solve this problem and is used for better data visualization.  *Figure Source\_CA2\_sbs2202* |  |

Here we stipulated two main types of Fertiliser to be used in the Machine Learning Models and the Statistics and Data Visualization process together with other datasets regarding Cattle Breed, Cattle age, Milk production, and Land prices.

Also, in the figure above we can see the correlation between Fertilisers (compounds) that shows the majority of the producers in the same period using the same Fertilisers. It is not for a feeling or tradition, it is Big Data has been used effectively in improving soil management and consequently increasing agricultural production.

We can see the correlation between the use of land that uses more Fertilisers has the last chance to be used as Dairy Farming, it means the last chance to be used in the production of milk.

The inclusion of Big Data into Agriculture Industry using cloud data to improve production in agriculture, we can also ask ourselves how this technological integration should be managed, and the advantages it offers. In the Literature and Data raw collection process showed the need to care with quantity and important quality over the data collected through different devices and technologies that make up the software and codes used in modern agriculture. Because, it will make a difference for those who seek better results, which can achieve more effective productions and greater profitability for the sector.

This research process presented us with the challenge of working data in a very specific market, but at the same time with very detailed data without a common structure for data collection. In this scenario we used different datasets within different Jupyter Notebooks, using PCA technique, functions such as ‘drop’, ‘index’, ‘insa’, ‘arima’, ‘merge’,’pivot’, more than we anticipated. But gave us an opportunity to work with different Machine Learning Models and confirmed that one Machine Learning Model is not always the best for everything.

In Resume, in the research literature we learned that when we discuss the Agriculture management system, we are relating data referring to production costs that range from the inputs themselves, labour, machinery, charges, taxes, commercial contracts, stock, transport, land price and much more. All this data can be related to data obtained in the field in your area, in our country and even in your world, to seek better productivity linked to greater operational efficiency.

1. **Methodology**