

CDS 506:

**Research, Consultancy and Professional Skills**

Consultancy Project Proposal:

**Time Series Analysis on**

**Agriculture Food Production in Malaysia**

Prepared by:

**Muhammad Azzubair bin Azeman P-COM0019/19**

Submitted for:

**Dr Nasuha Lee Abdullah**

**Ts. Dr. Chew XinYing**

Date of Submission:

**26 June 2020**

School of Computer Science

Academic Session 2019 / 2020

**Abstract**

Agricultural Food Production (AFP) serves as the backbone of food resources in Malaysia. It needs to be produced at large scale and continuously monitored to ensure that growing demands of Malaysian population for food is fulfilled. AFP prediction is also a requirement for it to be used by the client (Ministry of Agriculture and Food Industries) as reference values in their annual budget planning. In this project proposal, Related works in agricultural technology, data science techniques, and analytical tools were briefly reviewed. In order to monitor and predict AFP, Time Series Analysis and Forecasting using ARIMA and Neural Network models are proposed and justified. The project’s data source and risks are defined and explain respectively. Finally, the project’s deliverables were scheduled in Gantt Chart proposed.

1. **Background of Selected Company**

DataMicron Systems Sdn Bhd is a technology company which offers consultant services for business intelligence and big data-related solutions. As many companies nowadays have their own databases, they face challenges and difficulties in gaining insights from their big and complex data using traditional techniques. Therefore, this is where DataMicron comes in whereby their managing director said in his interview with *The Star* newspaper publication:

“*We extract data from various databases provided to us by our clients and merge them in the data warehouse, where from there, we do analysis of the data to provide our clients with business intelligence and predictive analysis, which in turn, would help them in their decision-making process*” (Hooi, 2014).

On 2004, DataMicron company was granted by Government of Malaysia through Malaysia Digital Economy Corporation (MDEC) with Malaysia Status Services (MSC) status which enables their company to enhance their product and service developments on multimedia technologies. As a result, the company has extended their scope of services to more than five countries as in 2014. The success of this company was reflected by their Microsoft Asia Pacific Keystone Award on 2005, and SME Corp Innovation Award (ICT) on 2013 (Hooi, 2014).

The key to their success is their user-friendly dashboards which is easy for their clients to understand. Their manager said in the interview “*Once it is finalised, we move on to* develop *the front-end tools, which come in a user- and visual-friendly dashboard with predictive capabilities to display the analysis*” (Hooi, 2014).

* 1. **Type of ownership**

As the name implies at suffix of DataMicron Systems Sdn Bhd, this company is a private company which is owned by Jimmy Ting (Managing Director of DataMicron) since 2002. This organisation of this company is compiled correctly with guidelines published by Companies Commission of Malaysia (SSM) whereby “*At least one (1) director who ordinarily resides in Malaysia by having a principal place of residence in Malaysia and minimum of one (1) promoter*” (Companies Commission of Malaysia, 2018). Hence, DataMicron Systems Sdn Bhd has been standing as a private company for almost two decades.

* 1. **Products or services**

DataMicron provides innovative solutions for Big Data, and Business Intelligence for many local and international organisations. As data value is significantly increasing, DataMicron offers three types of data-related services which are Training, Consultancy, and Support. In terms of training, DataMicron together with other industry partners agreed to develop future talents by conducting one-year placement under their company for Bachelor students of Universiti Teknologi Malaysia under 2u2i mode programme (Malaysia Digital Economy Corporation, 2019).

Meanwhile for consultancy, this company is strong in providing business intelligence for banks, governments, and retailers using their cost-effective product known as Insta BI (Guiam, 2015). This product is introduced in packaged features that can be upgraded vertically for their business. Furthermore, they also provide technical support for companies investing Big Data Analytics, Business Intelligence, and Internet of Things on them (DataMicron, 2016). Overall, these services aimed to assist their clients in gaining insights for making better decisions.

* 1. **Business model of the company**

DataMicron uses Bimodal Delivery model for executing their business processes. This model is reflected by their Insta BI product where it integrates between Self-Service with Enterprise Business Intelligence (refer Figure 1). In return, clients invest on fast providing solutions, together with flexible cloud or on-premise deployments. This surely fulfils clients’ needs, and guarantees their long-term investment on Business Intelligence, and Data Analytics (DataMicron, 2016).

A screenshot of a cell phone

Description automatically generated

Figure 1 Bimodal Delivery model implemented by DataMicron

1. **Introduction**

Agriculture Food Production (AFP) is defined as the large-scale land cultivation of domestic plants or animals to be contributed majorly for human diet, crops production and livestock (Harris & Fuller, 2014). This sector is vital to be fully utilised especially for developing countries such as Malaysia as the number of populations is consistently growing throughout the years (Masron *et al.*, 2012). However, local AFP in Malaysia is still considered as a minority whereby it is not enough to satisfy consumers’ demand. Therefore, imported organic foods were brought into market and in fact, occupies for more than 60% of overall organic foods (Dardak *et al.*, 2009; OTA, 2017). The main factors contributing to this is due to lack of government support, land issue, and shortages of labour (Etingoff, 2017). Thus, as an effort to reduce food imports, local AFP must be boosted through effective decisions made by local government. This is reflected by small changes in trends in time series plots.

Time series analysis is an effective approach to monitor AFP trends with respect to time. Essentially, time series analysis requires a time series dataset in which here, the important variables that are monitored regularly are Agricultural Production Index (AGRICPI), and Food Production Index (FPI) (Murad, 2017). In particular, the time series is a collection of AGRICPI, and FPI which are recorded at fixed time intervals either in hours or days or even years. Moreover, AFP trends can also be forecasted either via statistical techniques or machine learning techniques. A popular statistical technique for predicting time series is Autoregressive Integrated Moving-Average (ARIMA) model (Brockwell & Davis, 2002). Meanwhile for machine learning, Multilayer Perceptron (MLP), and Bayesian Neural Network (BNN) are known as the best models for forecasting time series (Makridakis *et al.*, 2018).

* 1. **Problem statement**

Much studies and technologies have been proposed for AFP such as nanotechnology implementations (Mousavi & Rezaei, 2011), urban AFP (Mensah, 2018), and recently the Smart AFP (Prathibha *et al.*, 2017). However, studies done on time series analysis on AFP particularly for Malaysia are very limited. Without sufficient insights of this analysis, it would be a challenge for decision makers to come out with effective decisions in boosting local AFP as well as reducing import organic foods.

* 1. **Research question**

This proposal makes an attempt to analyse time series of Agriculture Food Production, which consequently results in making the best decision for improving local AFP along with the effort of lessening import organic foods. It proposes the techniques used for time series analysis and metrics used for model evaluation. In particular, this proposal suggests for model comparison between ARIMA model with Neural Network (NN) model. This brings to the following research questions:

* + - **What is the recent trend of the Agriculture Food Production time series?**
    - **Between ARIMA with NN models, which of them performed better?**
    - **What are the forecasted AGRCPI and FPI for the next 2 years?**

Throughout this proposal, the techniques for answering the research questions will be reviewed from literatures, to enlighten on the time series analysis effectiveness, to offer valid justifications, and to equally propose the best model as a solution for time series analysis and forecasting that suits best for agricultural domain in Malaysia.

* 1. **Objectives of project**

The main goal of this project is basically to provide time series analysis and forecasting of Agricultural Food Production in Malaysia. In order to achieve this, the objectives of this project are highlighted which are:

* + - To determine the recent trend of Agricultural Food Production time series
    - To evaluate model performance between ARIMA and NN models
    - To forecast Malaysia’s AGRCPI and FPI for the next 2 years
  1. **Benefits of the project**

This problem is actually a consultation project between DataMicron with Malaysia’s Ministry of Agriculture and Food Industries (MOA). Therefore, this project will benefit DataMicron in providing proposed solution for their client’s problem. In particular, this project will deliver the insights of previous AGRCPI and FPI trends and will develop a reliable model for suggesting effective decisions to be used by MOA’s top management. In return, an optimised budget planning for MOA will be constructed.

1. **Related Works**

In this section, agriculture literatures will be referred to explore the details of established and proposed methods in agricultural practice. Published researches demonstrate on the theories and analysis behind every agricultural practice and also some others include on proposed innovations potentially applicable in the future.

* 1. **Background of the domain**

Ever since science and technology applied in farming practices, Agriculture Food Production (AFP) increased significantly worldwide. Conventional technologies such as inorganic fertilisers, pesticides, feedstuff, and tractor machines lead to enhancement on AFP in which those technologies provided a systematic and controlled environment against disease and pest challenges (Royal Society, 2009). As technology develops, new and modern techniques are now commercialised to further enhance on AFP with higher efficiency even though with limited farming space such as the use of hydroponic (Bridgewood, 2003) and urban farming (Platt, 2007).

Nowadays, AFP has integrated together with information technology resulting in smart farming whereby lesser workforce labour is required such as the use autopilot tractors, crop sensors, and digital farming (Baseca *et al.*, 2019). This reflects for the importance of Internet of Things (IoT) devices in optimising AFP to fulfil growing population demands. Recently, data-driven agriculture has been proposed to enhance crop productivity in AFP. This approach suggests for utilisation of real-time agricultural data captured by agricultural sensors. Insights obtained by big data approach would be easily accessed and analysed for further AFP optimisation (Sarker *et al.*, 2019).

* 1. **Related works on a few Data Science & Analytics Techniques**

Statistical techniques were reported in earlier literatures for AFP Time Series Analysis and Forecasting (TSAF). Such methods are the Least Square Method, Moving Average Method, and Winter’s Method. These techniques were compared in a study conducted for crop management whereby ­­ (Sodha & Saha, 2016). This was reflected by Winter Method’s lowest Mean Square Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) than others.

­ (Jadhav *et al.*, 2017). The main requirement for this model is the time series must be stationary in which the mean and variance are constant throughout the time series. Stationary time series are viewed by Autocorrelation Function (ACF) and Partial Auto Correlation Function (PACF) plots. For non-stationary time series, the dataset must undergo differencing. Then, several combinations of ARIMA were tested and each model was compared based on their Q-statistics, Akaike’s Information Criteria (AIC), and Schwartz Basic Criteria (SBC). Q-statistics was used for determining residual significance from zero; while AIC and SBC are standard metrics that enumerate the models’ quality (Jadhav *et al.*, 2017).

Since the emergence of Industrial Revolution 4.0, global agriculture researches began to propose machine learning techniques for forecasting AFP time series. On 2018, Neural Network (NN), Linear Regression (LR), and Polynomial Regression (PR) were applied for modeling apple and pear production amount in AFP. The study found out that NN outperformed LR and PR with having the least prediction error when sufficient time series is supplied. However, in case of small data for model training, LR is better than NN. In addition, NN model performance were outstanding whereby when sufficient monthly time series is trained, the performance error gap between NN with LR and PR was more than 10.00%. Overall, this study reveals the advantage and disadvantage of the proposed algorithms for TSAF (Balducci *et al*., 2018).

* 1. **Comparison and discussion on a few Data Science & Analytics Techniques**

Comparison between statistical technique with machine learning technique was extensively discussed by Makridakis *et al*. (2018). The dataset used in the study was 1045 monthly time series from the M3 competition open source data. 8 statistical and 8 machine learning techniques were tested in the study. In short, the outcome of the study found out that statistical methods have better forecasting performance. This is shown by the lesser symmetric Mean Absolute Percentage Error (sMAPE) of statistical methods compared to machine learning. This is a significant finding which explains Machine Learning algorithms need improvement for it to surpass statistical techniques. Figure 2 below shows the model performance comparison between statistical methods with machine learning methods in forecasting M3 competition time series.

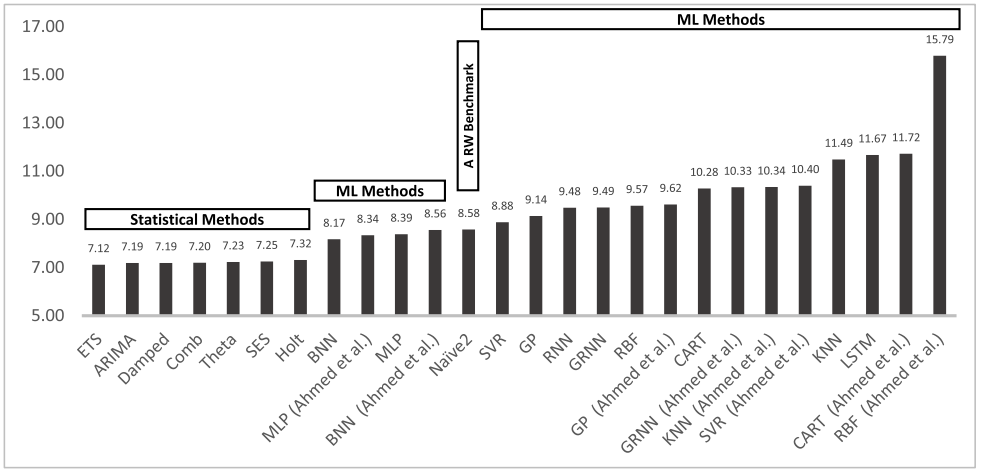


Figure 2 Comparison of model forecasting performance

between statistical and machine learning techniques for AFP TSAF

In particular, Makridakis *et al*. (2018) further study on the overfitting issue between statistical and machine learning techniques. They discovered that models with high sMAPE such as Classification and Regression Tree (CART) and Radial Basis Function (RBF) are due to their overfitting to the time series. This is because there is no threshold was made in their parameter, thus results for high model complexity. In contrast, they also unveiled that models with lower sMAPE such as ARIMA has better accuracy due to model parameterisation which is to minimise their AIC and SBC. Indirectly, with the minimisation of AIC and SBC, ARIMA model has optimised between the model’s goodness of fit with its complexity. At the end of the study, they advised for data deseasonalization and simpler models for better predictions.

* 1. **Relevant/related works on a few analytical tools**

Analytical tools for TSAF are abundant nowadays. They are available online as basic free software or advanced premium software. A well-integrated, free, and popular data science tool among data scientists is the Python programming language. Multiple studies have been using Python for implementing their application and others have been using it for proposing their solution. Another analytical tool that is also well known, and user interactive for data science beginners is the RapidMiner. This tool eases for users who have no basic programming in applying data science techniques.

In 2016, Agricultural data was beginning to be used for analysis in Python (Kim *et al.*, 2016). The study analysed farming households and their demographics in yearly basis which was then visualised in graphical bar charts. Also, simple calculations for obtaining farmer populations and growth rate was done. As years advanced, agricultural analysis using data mining and machine learning techniques was introduced to increase AFP (Vanitha *et al.*, 2019). They aimed to find hidden patterns in their collected data to come out with relevant insights which would help farmers optimise their costs and increase AFP. Support Vector Machine (SVM), Bayesian Network, and K-Nearest Neighbour (KNN) algorithms were used for classification task while k-Means Clustering was applied for clustering task. In short, this study presented that Python software is very well suited for data science application in agriculture.

Besides that, RapidMiner is also common to be used for agriculture data mining. Sudirman *et al.* (2018) used RapidMiner to identify the most optimised groups of rice crop as a start-up to assist Indonesian government overcoming unbalanced food land availability. The study used k-Means Clustering to determine the number of the most optimised group which is reflected by k value with the lowest Davies-Bouldin index. Performance of the clustering model was evaluated by performance accuracy percentage. In other study, Balducci *et al.* (2018) proposed to use RapidMiner for modeling data from IoT devices which hopefully increase farmers’ profit and reduce their costs. The proposed model includes multi-layer perceptron for forecasting time series, and KNN, Decision Tree, and PR for predicting missing data of IoT sensor devices. In brief, RapidMiner is a powerful tool for data science tasks in agriculture.

* 1. **Comparison and discussion on a few analytical tools**

Analytical tools comparison was briefly discussed by Dušanka *et al.* (2017) whereby five data mining tools were compared and discussed. The five data mining tools are Weka, Azure ML Studio, RapidMiner, H2O, and Apache Spark. Python itself was not included in the study because it was already integrated and embedded with the five tools except for Weka. The study found out that all of the five tools can perform full functionality required by data science technique from data pre-processing until data modelling and visualisation. However, the study figured out that RapidMiner, Azure ML Studio, and H2O could not produce AUC output for multiclass classification.

In addition, Dušanka *et al.* (2017) also pointed out that Weka is unable to process large datasets (typically 20 megabytes) unless command line interface (CLI) was used. Moreover, RapidMiner’s open source version has limited data processing capabilities whereby only 10,000 instances can be processed. Nevertheless, Apache Spark, H2O, and Microsoft Azure ML Studio have limitless processing capabilities which makes them able to process large amount of data instantly. However, not all of the five data mining tools have the ideal features as a data science software. tabulated usability of the five data mining tools in terms of Graphical User Interface (GUI), Command Line Interface, Business Application, Applied Research, and Educational uses. Azure ML Studio did not have CLI, and Apache Spark is lacking GUI, and also inapplicable for educational purposes (Dušanka *et al.*, 2017).

1. **Proposed Methodology**

In this section, proposed solution and analytical tool to be used will be stated, justified, and data source will be described in detail. Finally, the potential risks and issues in this project will be addressed and highlighted as a precaution

* 1. **Proposed solution**

For solving the client’s problem as stated in problem statement, statistical and machine learning techniques are proposed. In this project, only the best algorithms will be used for TSAF as determined by Makridakis *et al*. (2018) which are ARIMA representing for statistical technique and Neural Network for machine learning technique. For ARIMA, the order of auto-regressive (p), order of differencing (d), and order of moving average (q) will be determined by ACF and PACF plots. Also, several p, d, and q combinations will be tested, and the best model will be identified with the least AIC and SBC metric values. Using the best ARIMA model, AFP will be forecasted for the next 2 years. Meanwhile for Neural Network, the input layers will be set based on available data attributes provided by client, the number of hidden layers will be optimised through iterative process of model training, and the number of outputs will be the forecasted values of AFP for the next 2 years. The time series will be divided into training, validating, and testing sets by a ratio of 60:20:20. The outcome of the proposed solution will be presented to mentor for forecasting evaluations.

* 1. **Justification on the selected Data Science & Analytics Technique**

It is important to test and model AFP time series using both statistical and machine learning techniques. This is because previously, statistical technique was proven to be much reliable for TSAF compared to machine learning technique (Makridakis *et al*., 2018). However, since there was no study had been done on Malaysian AFP time series, it is still unknown whether statistics or machine learning would perform better particularly in agriculture domain. For statistical technique, ARIMA was selected to be used in this project as it is having the least error statistical models and robust against overfitting issues faced by other statistical models (Makridakis *et al*., 2018). Meanwhile for machine learning technique, Neural Network algorithm was chosen to be modelled also due to its resistance against overfitting issue suffered by other algorithms (Makridakis *et al*., 2018). Hence, it is justified that ARIMA and Neural Network algorithms will be selected in this project due to their robustness against overfitting.

* 1. **Justification on the selected analytical tool**

Despite of having of abundant analytical tools online, RapidMiner software integrated with Python language will be selected in this project. User-interface of RapidMiner plus with universality of Python will be combined together to provide rapid effective insights and AFP predictions. Even though open source RapidMiner has limitations of only 10,000 data instances can be used for processing, however, in this project, RapidMiner Educational Edition will be utilised enabling for full functionality of features for data science tasks particularly for TSAF (RapidMiner, 2020). Moreover, extensions required for TSAF in RapidMiner such as “Time Series version 0.2.2” will be installed. Meanwhile in Python, “Numpy”, “Pandas”, “Scikit Learn”, “Neuralnet”, “Matplotlib”, and “Datetime” libraries will be installed for data management, and difficult time series modelling where RapidMiner could not perform (Vanitha *et al.*, 2019).

* 1. **Steps you performed to address the problem**

As a consultant, the problem will be understood from an online interview with the client. Then, the problem will be researched in depth by going through literatures to obtain related solutions proposed. In order address the problem, a proper documentation will be prepared afterwards and presented to client in the next meeting.

* 1. **Data source**

AFP time series dataset will be given privately after the online meeting with the client. The client stated that the dataset will be related to agricultural data provided by Ministry of Agriculture and Food Industries (MOA). Non-Disclosure Agreement (NDA) will be signed to prevent data disclosure to the public.

* 1. **Risk and issues of the project**

Since the project involves with Government of Malaysia’s data, there are quite a number of risks and issues to be considered which are discussed below.

* + - **Confidentiality**

All data for governmental use is highly confidential which means that the data is strictly protected under data security. Information such as farmer’s demographics, total raw food sales, and total AFP produced are very high risk if it falls under irresponsible people. They can manipulate current existing data and produce false data which would result in casualties in agricultural markets. From other point of view, data confidentiality creates people’s trust and dependency towards data. People will trust the insights of confidential data as it really reflects the actual behaviour and patterns of AFP in Malaysia. Thus, data confidentiality is a matter of concern for data scientists to continuously adhere for.

* + - **Integrity**

Quality of data reflects for its integrity of being accurate and valid data. This also means that the raw AFP data was collected directly from its source. Therefore, the collected data have less error thus, being more authentic for data analysis. There are many factors that contribute to presence of error in data such as viruses, hacking, and hardware problems. In order to keep data at high integrity, two precautionary measures must be implemented. The first precaution is governmental data is collected attentively and stored in a secure database to protect against data loss. The second precaution is collaboration with National Cyber Security Agency of Malaysia (NACSA) has protected governmental data against external threats. Hence, data integrity is vital to be protected at all times to preserve data quality and also to secure individual privacy.

1. **Expected Outcomes**

This project will produce a table of comparison for model performance between ARIMA and Neural Network models. Both models are expected to analyse AFP time series and forecast AFP for the next 2 years. The insights of this analysis will provide a basis for MOA in determining the AFP trend in recent years. Moreover, forecasted AFP values will be used as reference to assist MOA decision makers in making effective decisions for annual budget and project planning in agriculture.

1. **Gantt Chart**

Timeline of this project’s deliverables are tentatively scheduled as displayed below.



Figure 3 Gantt Chart of Consultancy Project and Practicum

1. **Conclusion**

Overall, this project proposal aims to propose a consultancy project for providing solutions to client using data-driven approach. Research questions addressed will be the main topic of discussion circulating throughout the project. Three objectives stated will be answered at the end of the project. Much more literatures need to be reviewed to discover available and proposed solutions in enhancing Agriculture Food Production. The proposed methodology will be performed, and the output analysis will be extensively discussed and supported by available literatures. The outcome of the discussions will be summarised in graphical representations and derived into decision options for the client to evaluate and choose for the benefit of the client’s organisation.

1. **References**

Balducci, F., Impedovo, D., & Pirlo, G. (2018). Machine learning applications on agricultural datasets for smart farm enhancement. *Machines*, 6(3), 38.

Bridgewood, L. (2003). *Hydroponics: Soilless gardening explained*. United Kingdom, UK: Crowood Press.

Brockwell, P. J., & Davis, R. A. (2002*). Introduction to Time Series and Forecasting*. New York, NY: Springer-Verlag.

Cambra B., Sendra, C., Lloret, S., & Tomas, J. (2019). A smart decision system for digital farming. *Agronomy*, 9(5), 216.

Companies Commission of Malaysia (2018). *Guidelines for the Incorporation of Local Companies*. Guidelines. Retrieved June 20, 2020, from https://www.ssm.com.my/Pages/Legal\_Framework/Guidelines.aspx

Dardak, R. A., Abidin, A. Z. Z. & Ali, A. K. (2009). Consumers’ Perception, Consumption, and Preference on Organic Product: Malaysian perspective. *Econ. Technol. Manag. Rev.*, 4, 95–107.

DataMicron (2016). *Insta BI*. Blazing fast data discovery. Retrieved June 21, 2020, from <http://www.datamicron.com>

Dušanka, D., Darko, S., Srdjan, S., Marko, A., & Teodora, L. (2017). A comparison of contemporary data mining tools. In *XVII International Scientific Conference on Industrial Systems (IS'17)* (Vol. 4, No. 6).

Etingoff, K. (2017). Organic Farming and Sustainable Agriculture in Malaysia:Organic Farmers’ Challenges Towards Adoption. In *Sustainable Development of Organic Agriculture: Historical Perspective.* Canada, CA: Apple Academic Press Inc.

Guiam, A. (2015). *Carving a Competitive Future with Business Intelligence*. Guide to Big Data. Retrieved June 21, 2020 from https://datastorageasean.com/expert-opinions-executive-interviews/briefly-talking-business-intelligence-series-aas

Harris, D. R., & Fuller, D. Q. (2014). Agriculture: Definition and Overview.In *Encyclopedia of Global Archaeology*. New York, NY: Springer

Hooi, L. W. (2014, May 29). *Monetising Intellectual Property*. The Star. Retrieved June 19, 2020, from https://www.thestar.com.my/business/sme/2014/05/29/monetising  
-intellectual-property-plan-to-provide-financing-backed-by-ideas-as-collateral

Hooi, L. W. (2014, June 23). *Big Money in Data Mining*. The Star. Retrieved June 19, 2020, from https://www.thestar.com.my/business/sme/2014/06/23/big-money-in-data-mining-company-provides-business-intelligence-for-more-informed-decision

Jadhav, V., Reddy, B. V. C. & Gaddi, G. M. (2017). Application of ARIMA model for forecasting agricultural prices. In *J. Agr. Sci. Tech.*, 19, pp. 981-992

Kang, M. S., & Jung, Y. G. (2016). Big data Analysis using Python in Agriculture Forestry and Fisheries. *International Journal of Advanced Smart Convergence*, 5(1), 47-50.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*, 13(3), e0194889.

Malaysia Digital Economy Corporation (2019). *Malaysia Takes the Lead as Regional AI and Data Hub*. Retrieved June 20, 2020, from https://www.mdec.my>uploadsPDF/web-results/malaysia-takes-the-lead-as-a-regi  
onal-ai-and-data-hub-MDEC

Masron, T., Yaakob, U., Mohd Ayob, N., & Mokhtar, A. S. (2012). Population and Spatial Distribution of Urbanisation in Peninsular Malaysia 1957 – 2000. In *Malaysia Journal of Society and Space*, 8(2), 20-29.

Murad, Md. W., Siwar, C., Nik Mustapha, N. H., Nik Mohd. Kamil, N. F., Muhamad, S., & Abdul Aziz, A. (2017). Emergence to Develop an Appraisal System for Agricultural Practices in Malaysia. *Journal of Human Ecology*, 28(3), 191-198.

Obosu-Mensah, K. (2018). *Food production in urban areas: a study of urban agriculture in Accra, Ghana*. Routledge.

Organic Trade Association (2013). *U.S. Organic* Trade *Data: 2011 to 2016*. Retrieved June 21, 2020 from https://ota.com/sites/default/files/indexed\_files/OTATradeRepo  
rt\_10-30-2017.pdf

Platt, P. (2007). Vertical farming: an interview with Dickson Despommier. *Gastronomica*, 7(3), 80-87.

Prathibha, S. R., Hongal, A., & Jyothi, M. P. (2017). IoT based monitoring system in smart agriculture. In *2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT)* (pp. 81-84), IEEE.

RapidMiner (2018). *Get Started with RapidMiner.* Retrieved June 23, 2020 from https://rapidminer.com/get-started/

Royal Society (2009). *Reaping the Benefits: Science and the Sustainable Intensification of Global Agriculture.* London, LDN: The Royal Society.

Sarker, M. N. I., Wu, M., Chanthamith, B., Yusufzada, S., Li, D., & Zhang, J. (2019, May). Big Data Driven Smart Agriculture: Pathway for Sustainable Development. In *2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD)* (pp. 60-65). IEEE.

Sodha, D., & Saha, G. (2016). Crop management of agricultural products using time series analysis. In *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)* (pp. 1456-1460).

Sudirman, S., Windarto, A. P., & Wanto, A. (2018, October). Data Mining Tools| RapidMiner: K-Means Method on Clustering of Rice Crops by Province as Efforts to Stabilize Food Crops in Indonesia. In *IOP Conference Series: Materials Science and Engineering* (Vol. 420, No. 12089, pp. 1-8).

Vanitha, C. N., Archana, N., & Sowmiya, R. (2019). Agriculture Analysis Using Data Mining and Machine Learning Techniques. In *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)* (pp. 984-990). IEEE.