**Enhancing the Accuracy of International Corn Trade Forecasts using Machine Learning Methods**

**SATMEET SINGH SALUJA**

**Plagiarism: 6%**

**6Introduction**

Agriculture and food sectors that are efficient and agile are critical to human life. Natural catastrophes, trade disputes, and pandemics have all wreaked havoc on global agriculture in the last three years. Unprecedented uncertainty has influenced a wide range of actions, beginning at the farm and ending at the consumer's home or port (trade). Such unusual tendencies necessitate a large volume of data.

Increased availability of large data and advancements in computer hardware have posed a challenge to traditional statistical and econometric methodologies in predicting complicated patterns or economic correlations over the last decade. Machine learning (ML) has now been proposed as a replacement to solve many of these issues. With recent trade disruptions such as Brexit and tariffs between the United States, China,Japan and South Korea, there is a greater need than ever for alternative techniques to analysing and projecting (modelling) trade flows. Serious supply-side disruptions, significant uncertainty in demand, and sudden collapse of both supply of inputs and demand for output have all influenced decisions to plant, maintain crop progress, harvest, and market in the near term, and to invest in farm assets in the medium term during such disruptions (e.g., Covid-19). These events have produced a level of uncertainty and complexity not seen in decades, particularly in the context of agricultural trade.The static character of most trade models, which frequently conduct comparative static analysis of trade outcomes from deterministic trade policy changes, exacerbates the dilemma. There is a scarcity of theoretical guidance on how to simulate trade policy uncertainty and its implications for producer and consumer preferences and behaviour. The gravity model of bilateral (aggregate or industry) trade flows is subjected to machine learning techniques in this work. Because of its popularity and success in quantifying the effects of numerous factors of international commerce, the gravity model is sometimes referred to as the workhorse in international trade. The gravity model (Santos Silva and Tenreyro, 2006; Yotov et al., 2016; Athey and Imbens, 2019) was developed by Anderson (1979) and applied to data by Anderson and van Wincoop (2003). By relying on boosting methodologies, this study provides an alternative to time-series projections and expert judgement studies, allowing for alternative and robust descriptions of complicated economic linkages (Baxter and Hersh, 2017; Storm et al., 2019). ML models can also make accurate predictions, which has become a goal for many economists in recent months as a result of trade disruptions among the world's main economies.

International trade regulations have recently gotten a lot of attention for restricting cross-border commerce of vital items (e.g. steel, aluminum, soybeans, and corn).Because trade has such a significant impact on employment and income, forecasting future trade patterns is a top priority for policymakers all around the world. While traditional economic models strive to be accurate forecasts, we investigate whether Machine Learning (ML) techniques might help policymakers make better predictions. Open government data is used to fuel algorithms that can explain and forecast trade patterns and help policymakers make better decisions. The data in this article describes international trade transactions and the economic aspects that are usually connected with them.GBoosting, XGBoosting, and LightGBM are some of the machine learning (ML) models used to forecast future trade patterns. ML approaches enable a spectrum of data-driven and interpretable projections for particular commodities, unlike short-term and subjective (straight-line) estimates and medium-term (aggregated) projections. Models, their results, and policies are introduced, and their accuracy is assessed.

**Literature Review**

ML is only lately being applied to econometrics, according to a recent study by the National Bureau of Economic Research (NBER). ML has been used to solve problems in a variety of fields, including healthcare (Reddy and Aggarwal 2015), education (Niemi et al. ), and sports (Niemi et al). (Alamar 2013). So far, there are just a few applications for studying international trade trends (except for a few narrow studies that are referenced in this paper). Storm et al. offered a thorough examination of machine learning approaches used in applied economics, focusing on their potential for guiding policy decisions (Storm et al. 2019).

In the paper “International agricultural trade forecasting using machine learning” by Feras A. Batarseh , he focuses on seven major agricultural commodities with a long history of trade. To interpret trade trends, this study uses data-driven analytics, namely supervised machine learning (ML) and neural networks. Data is used to train supervised machine learning and neural network approaches till 2010 and 2014, respectively. In comparison to traditional methodologies, which are generally subjective assessments or time-series projections, the results reveal that ML models are highly relevant to forecasting trade trends in the short and long term. While supervised machine learning techniques assessed fundamental economic elements that underpin agricultural trade flows, neural network approaches offer greater long-term fit.

Many international institutions and government agencies forecast economic variables, including trade flows, to inform decisions in national and multilateral contexts (World Economic Outlook—International Monetary Fund, 2019; Organization for Economic Cooperation and Development, Trade in Goods and Services Forecast, 2019; United States Department of Agriculture, 2019; World Trade Organization, 2019). Because these forecasts are based on a mix of model-based analyses and expert judgement, several sources have pointed out their flaws, such as forecast accuracy of less than 35 percent (US Department of Agriculture, 2019) and quantifying the contribution of underlying economic factors (Chapter 4, World Economic Outlook—International Monetary Fund, 2019). With recent trade disruptions such as Brexit and tariffs between the United States and China and Japan and South Korea, there is a greater need than ever for alternative techniques to analysing and projecting (modelling) trade flows.

During such disruptions, decisions to plant, maintain crop progress, harvest and market in the near term and to invest in farm assets in the medium term have all been impacted by serious supply-side disruptions (e.g., flooding or drought), significant uncertainty in demand (e.g., soybean purchases by China) and sudden collapse of both supply of inputs and demand for output (e.g., Covid-19). These events, especially in the context of agricultural trade, have created a level of uncertainty and complexity unknown over the past several decades. The static character of most trade models, which frequently conduct comparative static analysis of trade outcomes from deterministic trade policy changes, exacerbates the dilemma. There is a scarcity of theoretical guidance on how to simulate trade policy uncertainty and its implications for producer and consumer preferences and behaviour.

Only a few researches have used machine learning approaches in economics. “The Nexus Between Artificial Intelligence and Economics,” by Gevel et al., was published in 2013. It was one of the first books to present agent-based computational economics (Gevel et al. 2013). Feng et al. used a neural networks model to study economic growth in the Chinese province of Zhejiang a year later. Their method, on the other hand, is limited in scope and impossible to apply across other Chinese provinces or geographical entities in other countries (Feng et al. 2014). Abadie et al. devised a similar model, but they applied it to California's booming tobacco industry (Abadie et al. 2010). Milacic et al. widened the scope in 2016, creating a model for GDP growth that included all four components: agriculture, manufacturing, industry, and services (Milacic et al. 2016). Falat et al. created a series of machine learning models for deciphering economic trends, although they did not make any forecasts (Falat et al. 2015). The experimental work given in this study optimises machine learning approaches to provide trade predictions for specific commodities and countries. Two models are developed and compared to explain international trade trends, given the dimensions of the data and the large number of variables involved.

**Methodology and model**

The term "machine learning" refers to a set of computational algorithms for extracting hidden insights from enormous volumes of data. Multiple machine learning algorithms are applied to a large data set of international trade (imports/exports) in this study. The distinction between ML and econometrics is that the former is concerned with regressions, classifiers, clusters, associations, and a variety of other actionable results, whilst the latter is focused with identifying casualties.

In recent years, decision trees and their extensions have grown highly popular. Because data is stratified or segmented into branches (splits) and leaves, they are referred to as trees (nodes). The number of predictors and cut-off values for predictors are used to stratify the data. If X comprises two column vectors, for example, stratification will be based on both constituents, either sequentially or randomly, for all conceivable cut-off values for each of these two predictors, such as X1<c1. Trees often utilize the mode or median of the outcome Y in the region to which the new X belongs to make predictions for new values of X.

The objective function for decision tree is:

Q = =

When splits are made on the basis of kth predictor,

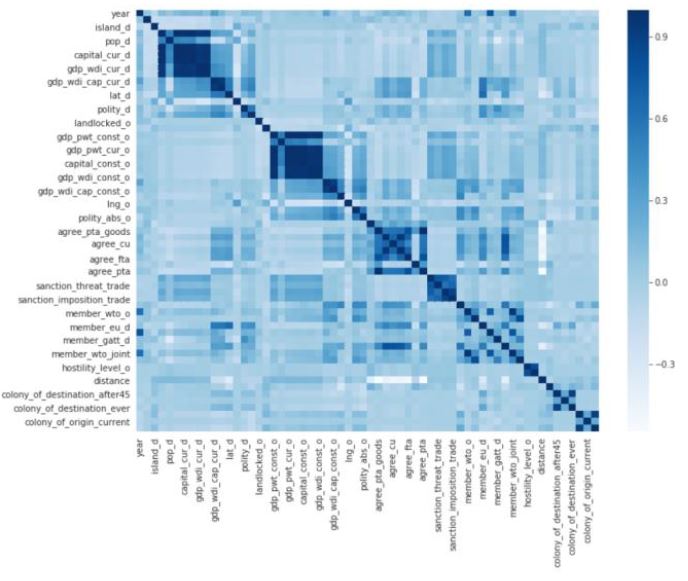
Q(k,c) =

The predicted Y is given by:

= , =

Hence two types of boosting are considered in this paper namely XGBoost and LightGBM. The data has been taken from Dynamic Gravity Dataset (DGD) provided by theU.S. International Trade Commission. It provides 70 gravity variables, but we will employ the 35 variables based on correlation analysis. The commodity that we will take into account will be the Corn trade across the world. The training data will consist of variable values from 1962-2010 and the predictions will be tested on the year 2011-2016.

**Tools and software to be used:** Python, Pandas, Scikit-learn, Matplotlib, Seaborn, Keras.

****

**Python Code:**







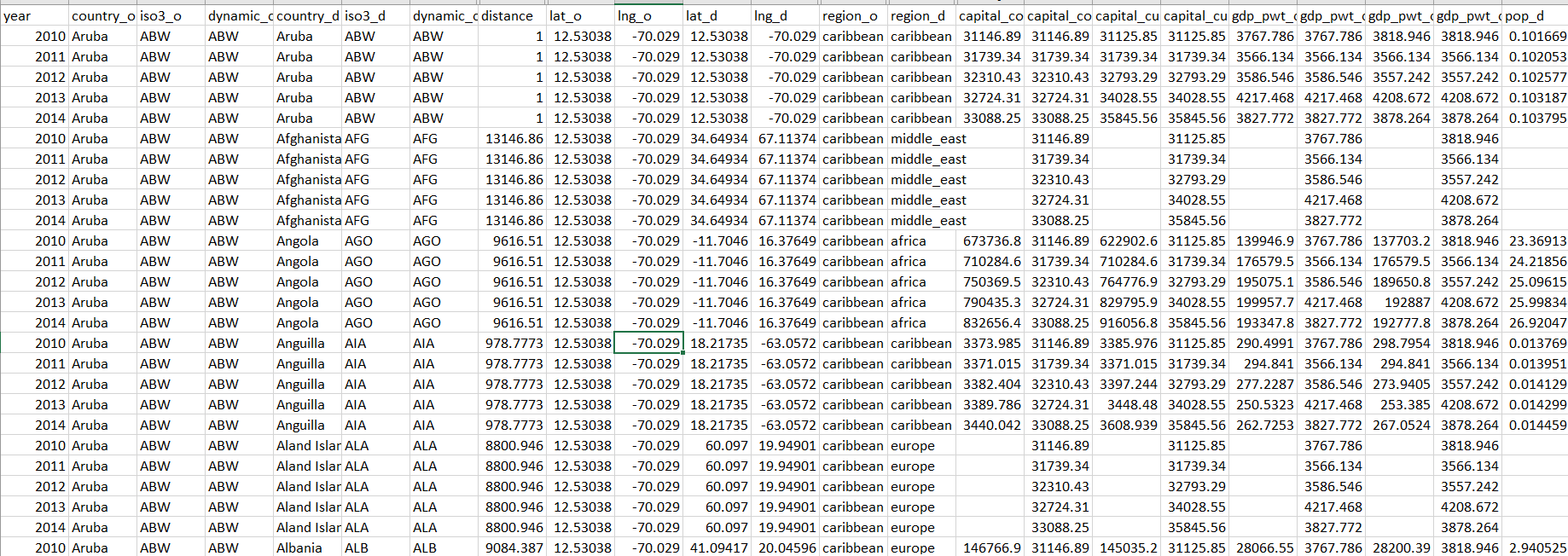


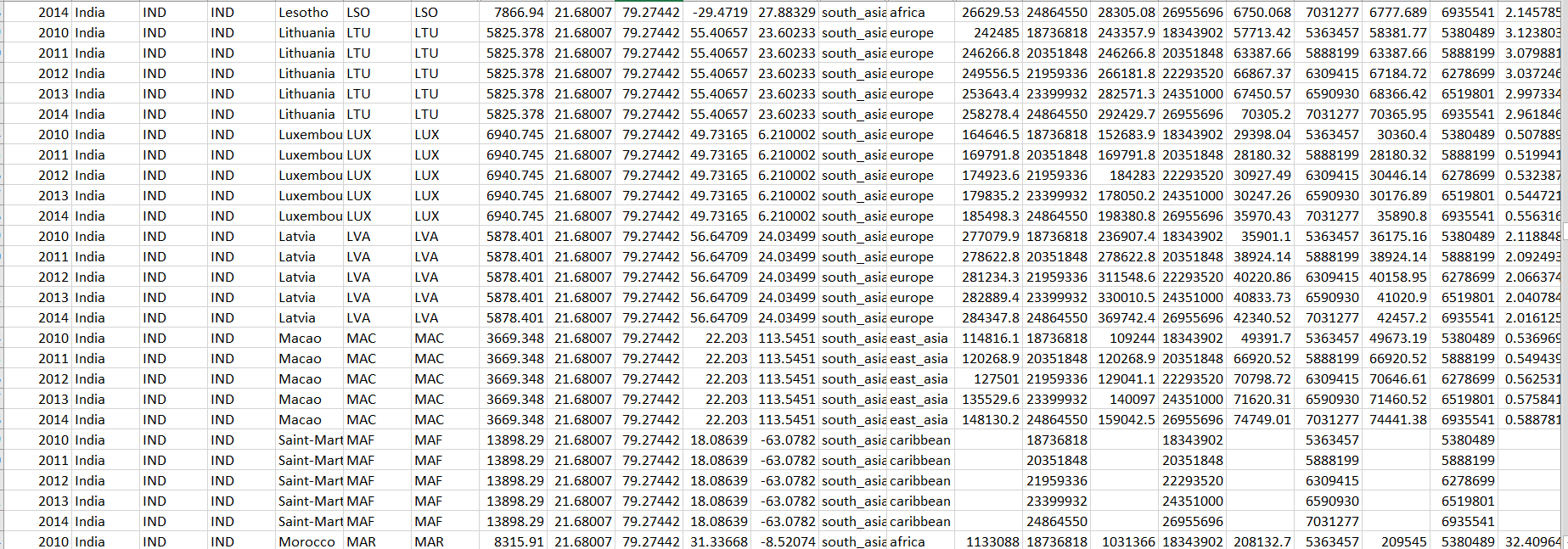






**Sample Dataset:**

****

****

**XGBoost Model:**

Gradient Boosted decision trees are implemented in XGBoost. Decision trees are constructed sequentially in this approach. In XGBoost, weights are very significant. All of the independent variables are given weights, which are subsequently fed into the decision tree, which predicts outcomes. The weight of variables that the tree predicted incorrectly is increased, and the variables are then fed into the second decision tree. These various classifiers/predictors are then combined to create a more powerful and precise model. It can be used to solve problems including regression, classification, ranking, and user-defined prediction.

Loss function in XG Boost:

L(Փ)=

Where γT + (1/2)λw2

L(t) =

L(t) =

After removing the constant term,

L(t) =

L(t) =

=

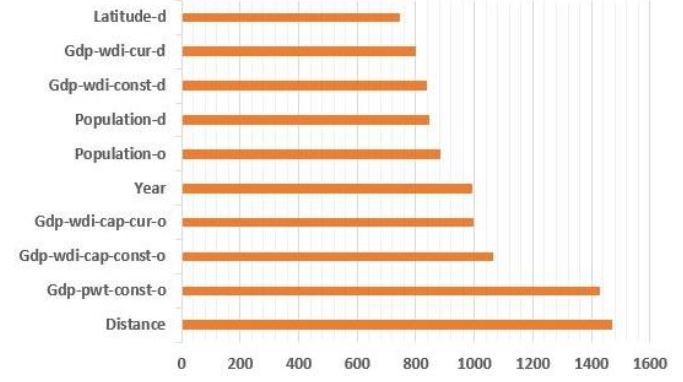
On applying FOC,

w\*j = -

L(t)(q) = -

The loss reduction after split is given by

Lsplit = -



**Relative importance of Variables for predicting trade**

**LightGBM Model:**

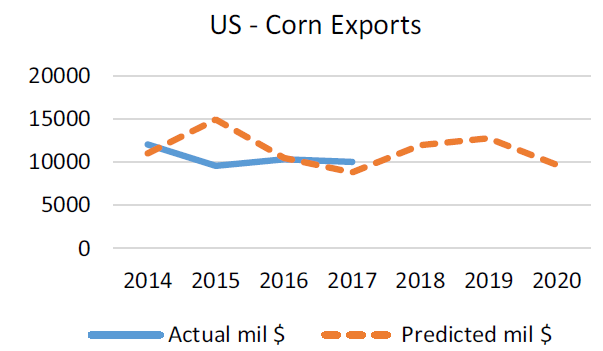
LightGBM scans all data instances to calculate the Gain from all possible Split points, which is evaluated in terms of the reduction in the total of squared errors. Rather than adjusting the weights for each erroneously predicted observation at each iteration, as other approaches such as GBoost do, LightGBM aims to match the new predictor to the prior predictor's errors.

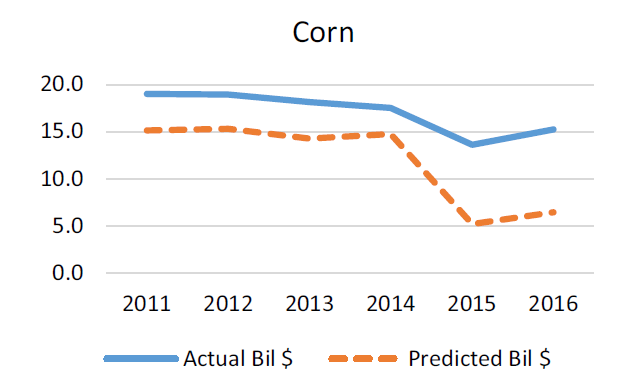
LightGBM splits the tree leaf-wise, whereas XGBoost methods split the tree level-wise with the best fit. The leaf-wise approach can decrease more loss than the level-wise algorithm, resulting in more accurate predictions.

**Result and Conclusion**

When the models have a high-adjusted R-square and trade data spans a wide number of nations and years, supervised ML results demonstrate that the models fit well in the near to medium term (2011–2014), that is, forecasts closely match actuals. The addition of extra trees or boosting to change weak supervised learners into strong ones, on the other hand, causes predictive quality to plummet 3–4 years after the training data cut-off due to the high presence of zero trade values. The top 10 variables in terms of information gain (reduction in the total-sample sum of squared errors of the loss function attributable to a predictor) and relative importance show that the size of economies, distance between them, location of countries, and time are more important in explaining trade flows than other gravity variables. While these findings are in line with the literature on trade and gravity models, ML excels in variable selection, prediction, and economic importance. Furthermore, supervised machine learning models have made it possible to address time- and space-varying effects, such as the distance puzzle. While changing the training sample size is likely to generate various contributions by features, additional model criteria must be carefully examined to fully solve such puzzles.

|  |  |  |  |
| --- | --- | --- | --- |
| Commodity | Observations | LightGBM | XGBoost |
| Corn | Training—29,500  Test—10,583 | 0.880 | 0.924 |





Predicting agricultural trade patterns is crucial for public and private sector decision-making, particularly in the current environment of trade conflicts involving tit-for-tat tariffs. Farmers, for example, are likely to assess prospective demand from other overseas sources before planting crops, particularly in significant exporters. Similarly, countries that set budgets for farm programmes require improved price and trade flow forecasts in order to analyse domestic production and consumption demands, as well as the tools used to attain those goals. The ML models allow for alternative and robust descriptions of complicated economic linkages by relying on data and deep learning. Furthermore, the ML models are cross-validated and may be used to simulate trade outcomes under various policy scenarios, including recent uncertainties. Modeling alternative policy scenarios can be aided by ML, much as it can be aided by computable general equilibrium models, which are useful in assessing the effects of trade policy changes. Changing training or testing data with prohibitive tariffs or trade prohibitions, for example, can produce predictions that can be compared to standard models.

Future work on additional industries, such as manufacturing, data/matrix completeness (a big issue when dealing with zeros in trade and tariffs), multi-variate response variables, and prescriptive ML techniques to compare with current causal models would considerably improve public and private decision making.

Our findings demonstrated that there is high relevance of advanced machine learning model for predicting bilateral trade flow. Our Models- XGBoost and lightGBM improved the prediction quality and will lead to better and accurate results. This is noteworthy since our findings showed that utilizing basic time agnostic information, machine learning approaches are effective in predicting economic variables. Methods used in this paper allowed us the extractions of the best economic variables that would affect trade of specific commodities. We provide a superior alternative to current public-sector agriculture trade flow forecasting methodologies. We use data and deep learning from data to enable for alternative and robust definitions of complicated economic interactions, rather than complex behavioural models with assumptions answered by accessing information from a variety of research. In alternative policy situations, ML models can be used to simulate trade consequences. This research will be broadened to include other commodities and models, with a focus on those that provide policymakers with interpretable outcomes.

**References**

1. Batarseh, F.; and Yang, R. 2018. Federal Data Science: Transforming Government and Agricultural Policy us-ing Artificial Intelligence. Elsevier’s Academic Press. ISBN: 0128124431.

2. The White House. 2008. Open Data Initiative.

3. Gevel, J.; and Noussair, C. 2013. The Nexus between Artificial Intelligence and Economics. Springer Briefs in Economics. DOI: 10.1007/978-3-642-33648-5\_1

4. Gunning, D. 2017. Explainable Artificial Intelligence. DARPA’s update on XAI.

5. National Bureau of Economic Research (NBER). https://www.nber.org/

6. Reddy, C.; Aggarwal, C. 2015. Healthcare Data Analyt-ics. Taylor and Francis Group CRC Press. ISBN: 1482232111.

7. Niemi, D.; Pea, R.; Saxberg, B.; Clark, R. 2018. Learn-ing Analytics in Education. Information Age Publish-ing. ISBN: 978-64113-369-2.

8. Bajari P, Nekipelov D, Ryan SP and Yang M (2015) Machine learning methods for demand estimation. American Economic Review 105(5), 481–485.

9. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q and Liu TY (2017) LightGBM: A Highly Efficient Gradient Boosting. In the 31st Conference on Neural Information Processing Systems, NIPS. Long Beach, CA, pp. 1–9.

10. Dynamic Gravity Dataset (DGD): <https://www.usitc.gov/data/gravity/dgd.htm>

11. International agricultural trade forecasting using machine learning , Munisamy Gopinath, Feras A. Batarseh and Sei Jeong.