**NutTix: An AI and Data analytics-powered solution for demand prediction and price forecasting of Groundnut**

**Abstract**

**Introduction**

**Methods**

**Results and discussion**

**Conclusions**

**References**

**Abstract**

The Indian agriculture sector is undergoing a seminal shift in the post-COVID-19 pandemic era from being a sector driven by traditional techniques to a technologically advanced, well-equipped industry, driven by Smart and Precision Farming. India is the 2nd largest producer (9.95 million tonnes) and 3rd biggest exporter of groundnut in the world. Groundnut is the highest exported oilseed crop in India, exporting it to China, Vietnam, Indonesia, Philippines, Malaysia, Russia, Ukraine and UK. The exports have increased by 142% from last year. The groundnut production estimate was 8.25 million tonnes for 2021-22. Majority of its export goes out as bird feed in different countries which limits its usage as an edible vegetable oil. Besides that, the south Asian market doesn't fetch good prices thus affecting the income of farmers. Groundnut needs a proper layout to drive it to a market oriented high-income generating solution. Despite its increased demand in the international market, the groundnut production in the country is facing a backlash from the farmers as production area is decreasing in many parts of the country. Fluctuating weather, subsequent increase in the cost of cultivation, erosion of river beds, poor irrigation system and lack of government support in decreasing the groundnut production acreage. This paper by understanding the market dynamics unfolds predictability of demand and forecasting of groundnut prices in the domestic as well as international market based on advanced technology of machine learning and deep learning. Nuttix can enhance the farmers income by two-three folds with its disruptive solution. Forecasting of prices of commodities, especially those of agricultural commodities is very difficult because they are not only governed by demand and supply but by so many other factors which are beyond control like weather vagaries, storage capacity, transportation etc. Forecasting of food prices are intended to be useful for farmers, policy makers and agribusiness industries. In the present era of globalization, management of food security in the agriculture dominated developing countries like India needs efficient and reliable food price forecasting models more than ever. Sparse and time lag in the data availability in developing economies, however, generally necessitate reliance on time series forecasting models. The recent innovation in Artificial Neural Network (ANN) modeling methodology provides a potential price forecasting technique that is feasible given the availability of data in developing economies. The fluctuations of agricultural commodity prices can significantly affect the global economies and living standards in many countries. Under COVID-19 crisis pressure, an accurate AI/ML forecast solution has never been more important for business leaders and agricultural authorities to make better decisions. Nuttix will study the domestic as well as international groundnut market and forecast the pricing of it.

*Keywords: Nuttix, Groundnut, prices, price forecasting, LSTM, exports*

**Introduction**

*Background*

The current scenario of Indian agriculture stands at promoting organic farming, integrated farming system, smart and precision farming which includes satellite farming, site specific crop management, robotics, drones and IoT, micro irrigation, conservation agriculture and nanotechnology. The usage of artificial intelligence, machine learning, blockchain technology in agriculture and allied sectors are increasing at an exponential rate. Smart and precision farming is the need of the hour. From 2017 to 2020, India received ~US$ 1 billion in agritech funding.

In India, annual demand for oil is 25 million tonnes, 60% of which is being imported. Palm oil, soybean oil and sunflower oil accounts for 62%, 21% and 16% share in imports. The current rise in price of oil in the Indian market is due to oil price hike in the international market. The country exports edible oils in small quantities to meet expatriate demand. Groundnut oil leads the export chart followed by Soybean, Coconut, Sesame and Mustard oil. India is the 2nd largest producer and 5th largest exporter of groundnut oil in the world. The diverse agro-ecological conditions in the country are favorable for growing 9 annual oilseed crops, which include 7 edible oilseeds (groundnut, rapeseed & mustard, soybean, sunflower, sesame, safflower and niger) and two non-edible oilseeds (castor and linseed). Oilseeds cultivation is undertaken across the country in about 27 million hectares mainly on marginal lands, of which 72% is confined to rainfed farming.

The oilseed production in the country has increased seven folds since independence. The production of groundnut oil, rapeseed & mustard, and soybean oil production has also increased significantly. Oilseed crops are the second most important determinant of the agricultural economy, next only to cereals within the segment of field crops. Though the self-sufficiency in oilseeds was attained through “Yellow Revolution” during early 1990’s, it could not be sustained beyond a short period. Despite being the fifth largest oilseed crop producing country in the world, India is also one of the largest importers of vegetable oils today. There is a spurt in the vegetable oil consumption in recent years in respect of both edible as well as industrial usages. Despite commendable performance of domestic oilseeds production of the nine annual crops (Compound Annual Growth Rate of 3.89%), the demand-supply gap could not match with the galloping rate of per capita demand (~6%) due to enhanced per capita consumption (19 kg oil per annum).

*Market Dynamics*

Import growth in respect of edible oils during the last decades is about 174%. The import figure of edible oils during 2019-20 reveals that India imported a total of 13.35 million tonnes of vegetable oils costing Rs. 61,559 crores. Among nine major oilseeds, soybean (33.5%), groundnut (30%) and rapeseed & mustard (27%), contribute to more than 90% of total oilseeds production in the country. However, in terms of vegetable oil production mustard, soybean and groundnut contribute 27%, 34% and 30% respectively. India’s export of oilseeds and products dates back to many decades. Groundnuts, Sesame seeds, Safflower seeds, Sunflower seeds, Niger seeds, Mustard seeds, found their way from India to the international market for many years. Groundnut and Sesame together account for over 80% share in India’s oilseed export basket. The Indian climate is highly congenial for cultivation of the oilseed crops.

India is the largest exporter of groundnuts seed in the world with a share of about 29% in world trade. The country exported 542730 tonnes of groundnut in 2015-16 valuing 4075.63 crores. The food and nutritional security to all the citizens of the country is the prime objective of the Government. ICAR, thus released 89 HYVs and 4 Biofortified varieties of oilseeds. Southeast-Asian countries are the major importers of groundnut from India.

*Previous works*

A research paper titled ‘Modeling and forecasting of oilseed production of India through artificial intelligence techniques’ by SANTOSHA RATHOD et al, tried to forecast the oilseed production of India using ARIMA, TDNN and NLSVR models. Based on the results obtained in this work one can infer that application of artificial intelligence techniques like time delay neural networks and nonlinear support vector regression techniques in modeling and forecasting of time series can increase the forecasting accuracy, in particular, the nonlinear support vector regression model performed better for forecasting oilseed production of India as compared to other models. They also suggested that clubbing artificial intelligence techniques and machine learning techniques together will help in extending the approach for varying autoregressive and moving average orders in other agricultural crops.

A review paper titled ‘Oil Palm and Machine Learning: Reviewing One Decade of Ideas, Innovations, Applications, and Gaps’ by Nuzhat Khan et al, provides an opportunity to understand the significance of Machine Learning techniques in the oil palm agricultural industry and provides a roadmap for future research in this domain. Machine learning (ML) offers new technologies in the precision agriculture domain with its intelligent algorithms and strong computation. The paper also infers that the existing research techniques are too inadequate to design practically supportive tools that are capable of increasing yields and improving quality and plantation sustainability in an environmentally friendly way. On the other hand, the integrated ML practices such as big data, remote sensing, data analytics, image-processing, and automated information extraction are progressing to achieve knowledge-based oil palm agriculture.

A research paper titled as ‘Application of Artificial Intelligence in Indian Agriculture’ by Anurag Saxena et al, was written to show how AI techniques can be used in different fields of agriculture, like smart irrigation (using IoT based devices), monitoring crop and soil health, field management, disease detection etc. It also suggests that a direct application of AI (Artificial Intelligence) or machine intelligence across the farming sector could act to be an apotheosis of shifting of traditional farming practice today. Artificial intelligence technology is supporting different sectors in agriculture to boost productivity and efficiency. AI solutions are assisting to overcome the traditional challenges in every field. Intervening of AI in agriculture is helping farmers to improve their farming efficiency and reduce environmental hostile impacts. The agriculture industry strongly and openly grasped AI into their practice to change the overall outcome. AI is shifting the way of food production where the agricultural sector's emissions have decreased by 20%. Inculcating AI technology in agriculture is helping to control and manage any uninvited natural condition.

A research paper titled ‘Interactive machine learning for soybean seed and seedling quality classification’ concluded that the interactive machine learning method for classification of soybean seeds from their appearance is highly accurate. This approach effectively identifies seeds with damage and classifies seedlings in vigor levels. The use of LDA, RF, and SVM algorithms is recommended for classifying soybean seeds and seedlings based on data generated with the Ilastik software. The aim of this study was to propose an approach based on interactive and traditional machine learning methods to classify soybean seeds and seedlings according to their appearance and physiological potential.

*Gaps in the State-of-the-art*

While studying the current technology stacks used to predict oilseed prices and manage oilseed production, we noticed a big gap in the problems that are being solved and the technology used to solve the said problems. Models such as ARIMA and SVM are simple and scalable solutions that provide a low variance model to predict oilseed market prices. However, these models have extremely high biases, which suggests that these models are not sophisticated enough to capture the quick movements in market prices. On the other end of the spectrum, we observe existing Recurrent Neural Network (RNN) architectures such as Long Short-Term Memory (LSTM) RNN that are versatile enough to predict any and all time series dataset, making them incredibly popular among existing commercial solutions that provide farmers with price predictions and updates. However, these models have relatively high variance and low bias, in the sense that these architectures cannot properly predict market crashes that are going to happen in the future, since they look at just the time series data. This can result in disastrous consequences for the average farmer, who might depend on highly accurate market forecasts to determine his next cropping cycle. Small inconsistencies in prediction can result in catastrophic losses.

*Problem Statement*

Groundnut is one of the most important crops around the world. India is the largest grower and second largest producer of groundnut in the world. The demand of groundnut is not steady and hence farmers face a lot of problems. Though after the pandemic the demand is expected to rise by 4.5% around 2021-26. Due to the pandemic, the peanut exporting countries have been hard hit with the decline in demand in importing countries amid the lockdown which resulted in a decline in demand for peanuts globally. However, the peanut demand has returned to pre-COVID-19 levels currently and is expected to increase as exporters are experiencing demand from Southeast Asian countries and the European region.

As of 2019, China and India are the largest consumers and exporters of peanuts in the world, accounting for more than 36.0% of the global consumption. However, peanut consumption increased dramatically in the world due to its dual usage as pulses and oilseeds. Groundnuts are widely used in the food and beverage industry in the form of oil, flour, snacks, and peanut butter. The peanuts market includes the production analysis (volume), consumption analysis (value and volume), import analysis (value and volume), export analysis (value and volume), and price trend analysis. India, the United States, and Argentina are the major exporters of peanuts globally. The increasing demand for nut-based snacks, nut butter, and protein-rich foods is expected to drive the high demand for peanuts globally thus encouraging higher exports. The Netherlands, Germany, and the United Kingdom are the three countries that offer ample opportunities to exporters of groundnuts in developing countries. In the future, the European market for peanuts is expected to grow due to the changes in the consumption patterns of customers, as plant-based protein is gaining popularity instead of meat-based protein. With people becoming more interested in healthy eating, groundnuts are expected to become an important source of unsaturated fats, fiber, proteins, vitamins, and minerals. Though in 2018, production of groundnuts decreased in India, the United States, and Senegal due to adverse weather conditions, especially delayed and irregular rainfall, it came back to the pre-covid levels in 2021. Also due to bumper production in 2020, the Indian groundnut farmers could offer a lower export price than US and Argentina. It has been reported that Chinese demand for Indian nuts has increased by 2,500-3,000 tons. Since groundnut is among the highest exported crops from the country, the demand and price of it depends heavily on its export potential. The production of groundnut depends upon timely rainfall and pleasant weather conditions. This would increase the quality of nuts inside. The present MSP of groundnut is ₹5550 which increased from ₹5090 in just 2 years which gives a good backup to the farmers. But the government need to magnify the production levels as it’s a very high value crop. Also, it should encourage the farmers for further value addition to ensure a good return.

### Proposed Solution

Based on our in-depth domain research, landscape exploration, and competitive analysis, we clearly understood that oilseed farmers desperately need an intelligent, farsighted, and more accurate price prediction for oilseed crops to better plan for future costs, crop insurance premiums, strain availability, etc. Considering all these requirements and nuances with the utmost care, we present Nuttix, a disruptive intelligent market price prediction system custom-designed to predict the market price for oilseeds for up to two cropping cycles. Unlike other predictive systems like ARIMA or vanilla LSTMs, Nuttix takes into account the socio-economic factors governing the commodity price index (CPI) and thus can provide a more accurate insight into the future prices of oilseeds. It combines the predictive power of Recurrent Neural Networks and expert systems to provide the most accurate results far beyond the current cropping cycle, providing farmers, wholesalers, retailers, and policymakers ample time to respond to future disasters.

## The Nuttix Model

Nuttix uses a combination of socio-economic and climatic factors, and presents a hidden non-linear relationship between these factors. These factors are then combined with the historical data points and fed into the LSTM RNN to produce highly accurate predictions.

We present the incidence score (I) as a measure of how other market factors are affecting the prices of oilseeds at a particular point of time t. The incidence score can be measured as:

I = 𝛽Fa + 𝛾Gb + 𝛿Tc + 𝜀Sd + 𝛼𝜂e

Where 𝛽, 𝛾, 𝛿, 𝜀, 𝛼, a, b, c, d, e are real numbers automatically determined by the LSTM model. F is the fuel indicator, which is typically a value indicating the price of fuel. It can be petrochemicals, coal, natural or petroleum gas, renewables, electricity, human labour, etc., or anything that is used to directly or indirectly transport, harvest or maintain oilseed crops. Without loss of generality, it can be set to the price of petroleum or crude oil, since its price is relatively unchanged from region to region and since it is used almost universally to power mechanised agricultural equipment and to transport produce. The price of electricity can also be used as a substitute. However, the inconsistencies in the cost structure of electricity for consumer and industrial use, as well as its low use in goods transport worldwide as of 2022 makes it a poor choice for the calculation of I.

G is the Gross Domestic Product (GDP) indicator, which can be safely assumed to be the ground truth GDP for the current fiscal year, GDP forecast for the next or current fiscal year (announced previous fiscal year) or inflation expressed as a percentage for the current fiscal year. T is the import duty on the product imposed by the top 3 importers of the oilseed, expressed as an average or a weighted average.

S is the sentiment score expressed as a real number between 0 and 1, where 1 meaning that the general population is “happy” or “willing to spend” as a whole and 0 implying that the general population is “not happy” as a whole. Since this is a largely subjective measure, we have used Natural Language Processing to provide a sentiment score using the news articles of time t. This is in no way a major factor while determining price. However, it can serve as a useful measure in determining how and when market crashes can occur, thereby providing some early warning to the user.

𝜂 is the number and severity of environmental disasters in the country chosen. This is typically a normalized weighted average of the disasters, where

𝜂 = 1 / (1 + exp(ΣiDiLi))

Where for every disaster i, Di is the number of deaths and Li is the amount of loss in US dollars caused by the disaster i.

Since the whole model is based on the LSTM architecture, there is a large degree of freedom in the units and indicators that can be chosen to influence the incidence score (I). The LSTM treats all the units equally and effectively moulds around the units, multipliers and exponents, thus making them a part of the RNN architecture. However, we have observed these indicators to provide the most comprehensive indication of the socio-economic conditions of the country in which the price is predicted, thereby giving the most accurate result.

### Methods

### Dataset Used

We have tested Nuttix on available historical datasets that describe the US oilseed market without loss of generality. The same tech stack can be used for any other country or economic zone, subject to the condition that adequate data as described below, can be obtained. The data used for testing and evaluating Nuttix can be divided into three segments:

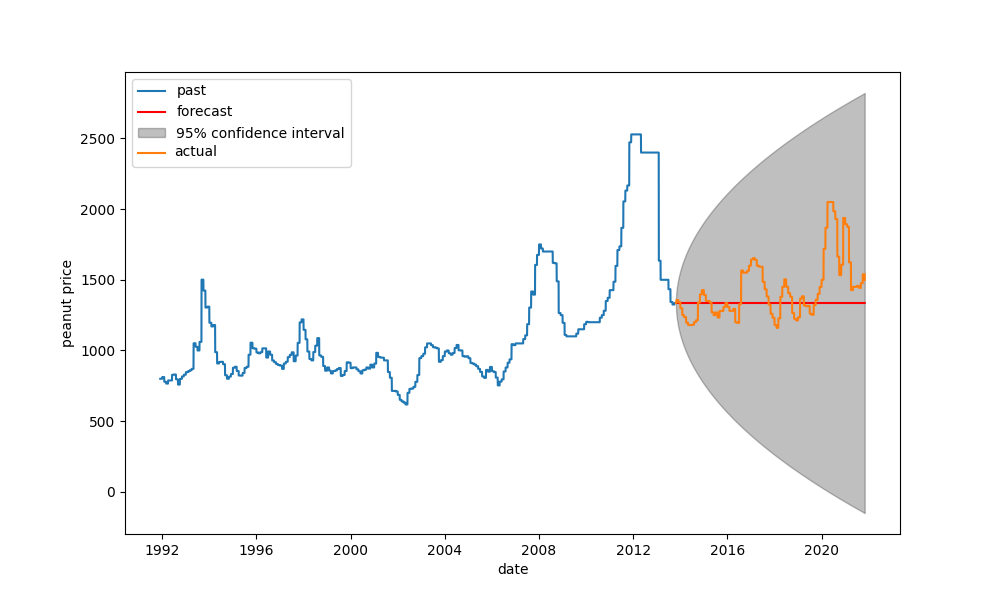
1. Historical oilseed prices: In this case, the historical monthly price of shelled Groundnuts (in US Dollars per Metric Ton) in the United States, between December 1991 and November 2021, sourced from the Agricultural data catalog of the World Bank, has been used. This data has been expanded into daily price data by data filling between the dates of December 1, 1991, and November 31, 2021.
2. Socio-economic data: We hypothesize that the socio-economic and environmental factors affecting the price of oilseeds can be detected and measured by looking at three indicators. These are as follows:
   1. The global price of crude oil: We have considered the daily price of Brent Crude oil in US Dollars per barrel between the dates of December 1, 1991, and November 31, 2021. This data has several gaps, including federal holidays, weekends, etc. which have been filled by data filling. We have followed the convention of filling the price of the previous date if no price data for the said date is available.
   2. The GDP of G20 countries: We have considered the yearly GDP of G20 countries, obtained from open-source International Monetary Fund Datasets. We considered the GDP value of the current fiscal year between the dates December 1, 1991, and November 31, 2021, with some added random noise to simulate yearly GDP forecasts. GDP forecasts serve an important role in determining the confidence of the average consumer, besides reflecting the political and economic stance of the biggest economies, which in turn are significant factors that determine prices of foodgrains across the globe.
   3. Consumer tax: We have considered import duties on oilseeds for the biggest net purchasers of US groundnuts: Canada, Mexico, and the People’s Republic of China. Import duties of groundnuts imposed by importing countries serve an important role in determining the flow of groundnuts as a traded commodity and will serve as an important factor in the determination of market selling price. The import duties of all these countries for the current fiscal year between the dates December 1, 1991, and November 31, 2021, are taken and are converted into daily data by data filling.
   4. Sentiment analysis data: We have scrapped the Twitter feeds of the 12 largest English news outlets in the United States and have assigned a happiness index to the news feed of a particular day by using sentiment analysis to gauge the “Market Happiness” of the average adult in the United States. Since this dataset was not available before 2008, we supplemented this dataset by using open-source news datasets of major headlines of the Reuters and ABC news feed between 1991 and 2008 to supplement this dataset. The sentiment analysis assigns a real number score between 0 and 1 to the public sentiment, which is fed into the Nuttix system.
3. Droughts, floods, and environmental crises: We have constructed a dataset that indicates the number and severity of the environmental crises that took place between December 1, 1991, and November 31, 2021, in the United States. Environmental crises drive up the prices of foodgrains and thus serve as an important factor in determining the market price of important cash crops, including groundnuts.

These datasets are fed into the Nuttix system, which when implemented as a free-standing API, continuously collects any future changes to the data using custom web scraping modules. The predictive LSTM model retrains and reconfigures itself once every 24 hours to provide the most up-to-date and accurate prediction to any user.

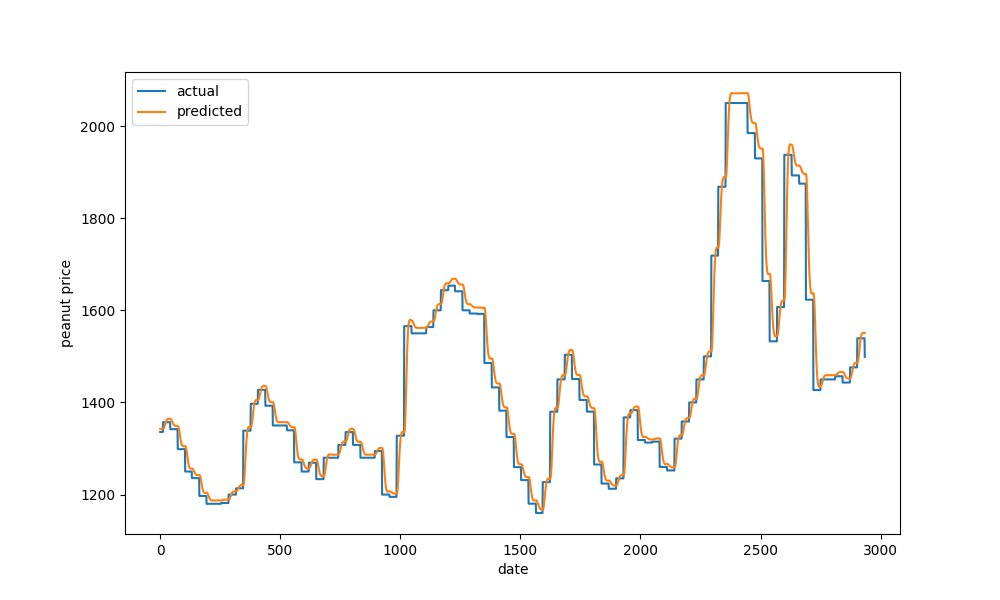
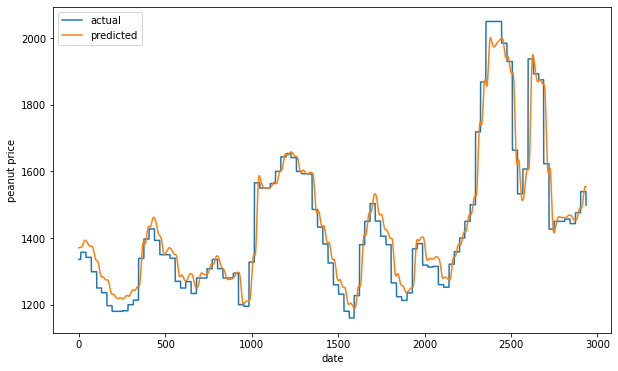
### Hardware Implementation and Mobile App

The Nuttix tech stack is run on a Unix-based server and is implemented using a combination of Flask, Selenium, LAMP stack, and Python. Flask and LAMP stack is used to store data in SQL databases to enable fast data access and data storage, while a combination of Selenium, Python, and NLTK libraries are used to access APIs for dataset access, web scraping, and sentiment analysis. In the end, the Keras library is used to train the LSTM model that is scheduled to run every 24 hours. The predicted data is then displayed as a part of AgroTick™, an end-to-end, AI and cloud-based one-stop portal and app for farmers, FPOs, Agro-vendors and wholesalers that aims to connect farmers to retailers, experts, and dealers, as well as provide them with timely and accurate consultancy on how to manage their products, financials, and prevent diseases. As a part of the AgroTick™ tech stack, Nuttix is beta tested as a part of the FarmersFirst™ software suite.

### Experimental Results

We have observed that Nuttix offers a middle path approach to traditional Machine Learning methods. It is consistently more accurate than ARIMA model, while holding a slight but significant advantage to vanilla LSTM model trained over historic oilseeds prices.

*Figure 1: ARIMA model forecast on oilseed dataset (as described in section IV.A) is woefully inaccurate, providing little insight into the price*

*Figure 2: Vanilla LSTM captures the complexities of the data, but its predictions are delayed by a unit of time, making them slightly more inaccurate*

*Figure 3: Nuttix predicting over the same dataset as used to obtain Figure 2*

When comparing errors for the above three models, we clearly see that Nuttix is the better model.

| Model | Error |
| --- | --- |
| ARIMA | 0.0153739 |
| Vanilla LSTM | 0.0008363 |
| **Nuttix** | **0.0005895** |

### Conclusion

Groundnut is an important oilseed crop having immense growth potential in the future as a high valued crop. The demand of groundnut as a vegetable oil is increasing day by day in the health-conscious segment of the world. India being its 2nd largest producer and one of the largest exporters needs to maintain its demand and price balance accordingly. Here, comes the Nuttix which can be used as a price forecasting model, better than most of the existing models like ARIMA. A proper price forecasting model will not only help the farmers to generate a good return domestically but also ensure a good profit margin in the export market. The previous existing models have low variance and high biasedness which earlier has been responsible for huge crop losses. The Nuttix model uses a combination of socio-economic and climatic factors based on non-linear relationship. The data collected are well-fitted in the LSTM RNN to produce highly accurate predictions. Nuttix offers a middle path approach to traditional Machine Learning methods. It is consistently more accurate than ARIMA model, while holding a slight but significant advantage to vanilla LSTM model trained over historic oilseeds prices. The Nuttix tech stack is run on a Unix-based server and is implemented using a combination of Flask, Selenium, LAMP stack, and Python. Flask and LAMP stack is used to store data in SQL databases to enable fast data access and data storage, while a combination of Selenium, Python, and NLTK libraries are used to access APIs for dataset access, web scraping, and sentiment analysis. With all these set of combinations we can easily conclude that Nuttix is a far superior model when compared to the current models used for price prediction and produce management for oilseed crops. This is going to be the game-changer in the oilseed industry for its accurate price forecasting ability. It will not only end the misery of oilseed farmers but also encourage them and the government as well to put more emphasis on the oilseed production of the country.

### References

1. Rathod, Santosha & Singh, Kamalesh & Patil, S & Naik, Ravindrakumar & Ray, Mrinmoy & Meena, Vikram. Modeling and forecasting of oilseed production of India through artificial intelligence techniques. Indian Journal of Agricultural Sciences. 88. 2018.
2. de Medeiros, A.D., Capobiango, N.P., da Silva, J.M. et al. Interactive machine learning for soybean seed and seedling quality classification. Sci Rep 10, 11267 (2020). <https://doi.org/10.1038/s41598-020-68273-y>
3. Khan N, Kamaruddin MA, Sheikh UU, Yusup Y, Bakht MP. Oil Palm and Machine Learning: Reviewing One Decade of Ideas, Innovations, Applications, and Gaps. Agriculture. 2021; 11(9):832. <https://doi.org/10.3390/agriculture11090832>
4. Saxena, Anurag & Suna, Truptimayee & Sa (2020). Application of Artificial Intelligence in Indian Agriculture. 2020
5. Groundnut Outlook Report-January to May2021,Agricultural Market Intelligence Centre, ANGRAU, Lam. 2021
6. Groundnut Price Forecast.October-November 2021
7. R. Sangeetha , M.S. Raman and S. Menaka.An Econometric Analysis on Groundnut Markets in India 16June,2021. https://doi.org/ 10.17485/IJED/v9.2021.5
8. G.Govindaraj and A.P. Mishra Directorate of Groundnut Research (ICAR), Junagadh-362 001, Gujarat.Labour Demand and Labour-saving Options: A Case of Groundnut Crop in India .Agricultural Economics Research Review ,2011 ;Vol. 24, (Conference Number), pp 423-428
9. PEANUTS MARKET - GROWTH, TRENDS, COVID-19 IMPACT, AND FORECASTS (2022 - 2027).
10. SAMEER LOKAPUR, RAVI GURIKAR AND G. N. KULKARNI.Production and export of groundnut from India -An overview.September, 2014 |Volume 5|Issue 2|https://doi.org:10.15740/HAS/IRJAES/5.2/293298
11. Vanilla LSTM captures the complexities of the data, but its predictions are delayed by a unit of time, making them slightly more inaccurate.
12. Author S. Celik,K. Karadas, and E. Eyduran.FORECASTING THE PRODUCTION OF GROUNDNUT IN USING ARIMA MODEL.The Journal of Animal & Plant Sciences, 27(3): 2017
13. S. T. Pavana Kumar and Ferdinand B. Lyngdoh. Prediction of Area and Production of Groundnut Using Box-Jenkins Arima and Neural Network Approach.29 December 2020
14. Jhade Sunil, Abhishek Singh.Structural Change Analysis of GroundnutExport Markets of India: Markov Chain ApproachINDIAN JOURNAL OF ECONOMICS AND DEVELOPMENT ,Published: 16.06.2021 <https://doi.org/10.17485/IJED/v9.2021.5>
15. R. Murugesan, Eva Mishra, Akash Hari Krishnan. Deep Learning Based Models: Basic LSTM, Bi LSTM,Stacked LSTM ,CNN LSTM and Conv LSTM to Forecast Agricultural Commodities Prices. November 1st, 2021.<https://doi.org/10.21203/rs.3.rs-740568/v1> .