Potato Research
https://doi.org/10.1007/s11540-024-09744-x

An Enhanced Long Short-Term Memory Recurrent Neural Network Deep Learning Model for Potato Price Prediction

Sarah A. Alzakari[1] · Amel Ali Alhussan[1] · Al-Seyday T. Qenawy[2] · Ahmed M. Elshewey[3] · Marwa Eed[4,5]

Received: 25 April 2024 / Accepted: 30 May 2024

Abstract
Regarding the potato market, pricing fluctuations are a significant factor, and unfortunately, they cause many issues for producers and consumers. It happens to result in food insecurity and economic instability. This study brings in an advanced LSTM-RNN model built to predict potato prices, which might alleviate the mentioned challenges. We gathered a historical potato price database and other economic variables, normalized by the Z-score normalization method to ensure all the data was consistent and credible. The model's effectiveness was benchmarked against five traditional machine learning models: we used K-nearest neighbor, random forest, support vector regressor, linear regression, and gradient boosting regressor to classify isolated households and determine their socioeconomic status. The empirical data implied that our proposed LSTM-RNN model was more efficient than all comparison models, leading to an R^2 value of 0.98. The paper not only substantiates the plausibility of applying deep learning to address the agricultural market prediction issue but also serves as a guideline noting the capabilities of the LSTM-RNN routine in improving the decision-making processes for the farmers participating in the sector. This model supports a sustainable food system and a balanced economy by bringing price stability integral to designing and implementing strategies to address food security.

Keywords Agricultural economics · Deep learning · Long short-term memory · Predictive analytics · Recurrent neural network

Background

The topic of potato pricing may seem straightforward at first glance, but it encompasses a complex interplay of economic, agricultural, and social factors. As a staple food for millions worldwide, potato price fluctuations broadly affect consumer

Extended author information available on the last page of the article

Published online: 29 June 2024

demand and international trade (Pavlista and Feuz 2005). The fundamental economic principle of supply and demand largely governs these price movements. Various elements critically influence potato production, including weather conditions, pests, diseases, and farming techniques. For instance, a surplus harvest leads to a price drop, whereas poor yields due to adverse conditions can cause prices to spike sharply.

Further complexity arises from changes in consumer preferences and dietary habits and the overarching impact of population growth, which drive demand for potatoes together. Geographical diversity also plays a crucial role; different regions have varying capacities for potato production, influenced by local climate, soil fertility, and water availability for irrigation. These regional disparities can lead to significant price fluctuations, often exacerbated by the cost and logistics of transportation (Salmensuu 2021a).

Government policies, too, have a hand in shaping potato prices. Subsidies, tariffs, and import/export regulations can artificially stabilize or distort market prices. For example, a potato tariff might force local consumers to pay more than they would in a free market, affecting affordability and consumption patterns. Moreover, major potato-producing countries like the USA, China, India, and Russia can influence global prices through their national policies on trade and production. Exchange rates and international trade agreements further affect the competitiveness of potato imports and exports' competitiveness, making potatoes' pricing a matter of international economic strategy (Bolotova 2017; Şahinli 2020).

Cultural and social dynamics also contribute to the pricing puzzle. In some cultures, potatoes are an essential dietary staple; in others, they are considered a luxury item. This cultural valuation can significantly influence how potatoes are priced and consumed in different parts of the world. Thus, understanding potato pricing requires a nuanced appreciation of a wide array of factors, from the local to the global, all of which contribute to the final price tag consumers see in the market (Dhakre and Bhattacharya 2016).

Through the innovation of technology in agriculture, such as using better farming methods and genetic modification, productivity and payment for food either increase or decrease, affecting food prices. In producing via precision technologies in farming, high yield and efficiency may be the attributive factors of stable prices (Thorne 2012). The two major environmental problems regarding the potato industry are climate change and water shortage, which pose the challenge of uncertain and fluctuating prices owing to the deranged weather pattern and the decline in yield. Usually, the cost of potatoes is influenced by a broad range of determinants such as supply, demand, and consumption, types of farming, government decision-making, international trade, new technology, and environmental issues (Kumar and Baishya 2020). How these factors influence the potato industry's decisions has become very important because, even today, decision-makers are experiencing high market volatility. Prices of potatoes are determined through different factors such as the supply–demand paradigm, agricultural practice, weather patterns, government policy, trade patterns, and technology in addition to people's tastes (Anwar et al. 2015).

The supply side response to potato production comes from various agricultural factors, such as the weather. Successful growing seasons and good farming practices enable high yields, which drive prices down because of the increased quantity. On the other hand, adverse conditions, including drought and pest infestation, usually result in reduced yields, leading to supply declines and higher prices. Consumer tastes, dietary preferences and patterns, population growth, and economic situation are the main factors behind the consumption of potatoes. Many people hold potatoes as a staple food in several areas because of their versatility and affordability. Nevertheless, the likes of consumers of other offbeat food items or dietary habits would cause them to shift their demands, and as a result, the product prices could also be affected. Moreover, location influences are also noteworthy issues (Koblianska et al. 2022). Some areas have better productivity of potatoes, whereas others may have different climatic, soil, and water conditions, e.g., require irrigation. That is, widespread oversupply or shortage in the local markets and transportation problems are further aggravated by the high costs of moving potatoes from a surplus point to a deficit point (Lu et al. 2023). Government actions through subsidies, tariffs, and import–export regulations have a profound role in determining potato prices.

As an illustration, domestic suppliers are behind high tariffs that hike consumer prices, whereas subsidies can smoothen producers' prices by compensating production costs. Exchange rates, trade treaties, and geopolitical factors distort trade, affecting global potato export and import prices. Social and cultural elements, in addition to this, play a role in determining the cost of potatoes. Our cultural preferences drive consumption behavior, traditions pass on, and food habits emerge in different parts of the globe, influencing the demand and prices of goods in other regions. Technology in agriculture, which can be seen through the development of farming techniques or biotechnology, affects productivity and production costs that also affect prices (Arısoy and Bayramoğlu 2017).

Technologies like precision farming can bring over efficiency and economy, although prices will be stabilized. Among the environmental challenges the potato industry faces are climate change and water scarcity, which fuel price fluctuations (Rosen 1999). Sudden weather transformations with different cultural practices lead to a shorter growing season and a rise in production costs. Potato markets are influenced by many factors, such as availability and demand, cultural practices, government policies, international trade relations, people's food choices, new technologies, and external environmental conditions. Market uncertainties can be effectively dealt with if the potato industry operators understand all these factors that may interfere with the smooth running of market processes.

Informed stakeholders are put in a position to make sound decisions when making such decisions. Crop disease and severe climate conditions are the primary triggers of a decline in potato growth, which can lead to lower yields and quality. However, the climatic changes are more so for potato growers and can, as a result, affect economic growth, food shortage, and supply chain disruption (McGregor 2007). Potato crops are subject to numerous diseases of different kinds of pathogens, including bacteria, fungi, viruses, and others. Among the most common diseases are late blight, early blight, bacterial wilt, and potato virus Y (PVY). Many diseases may occur in potato crops and these diseases can affect foliage, tubers, and roots, and are transmitted in one way or another thus reducing crop yields and quality of the potatoes (Bolotova et al. 2010).

Late blight, caused by the oomycete *Phytophthora infestans*, is known to be one of the most destructive diseases in potatoes globally. The oomycete is aggressive, especially under humid conditions, and causes foliage decay and tuber rot. Other historical periods show the destruction caused by late blight on farms and nations. *Alternaria solani* is the main causal agent of early blight and also affects the foliage. Although both types of blight are not equally severe, one can expect severe yield and quality loss. Rather than transferring nutrients and water, which healthy plants require, the vascular system of the potato is breached by the bacterium *Ralstonia solanacearum*. Thus, the plants rapidly wilt and eventually die. This causal agent of bacterial wilt is difficult to control as it is both soil-borne and seed-borne. The potato virus Y (PVY) may appear in the form of leaf discoloration, stunting the growth of the potato and decreasing its tuber quality (Mitra et al. 2018). PVY quickly transmits through infected seed potatoes and aphid vectors affecting the whole crop, causing high yield losses. Full-scale natural disasters such as droughts, floods, heat waves, and severe storms make potato plants fragile. Drought is mainly responsible for water shortage, dwindling potato production by depleting yield and fostering small tubers due to insufficient water. Furthermore, drought-stressed plants increase their vulnerability to diseases and pests, accelerating mortality rates (Fatima et al. 2015).

Floods also cause the soil to be saturated with excessive water, meaning that the potato plants are deprived of oxygen and vital nutrients for healthy root growth. This usually leads to damaged roots and short or stunted plants. Disease infection is more prominent due to this stress. Heat waves enhance haulm growth but impede tuber growth while reducing tuber quality (Goodwin Jr 1988). Irrigation during drought spells increases water consumption. Heavy storms with strong winds, heavy hail, and lots of rain cause physical damage to potatoes, such as snapped stems (Salmensuu 2021b). This reduces photosynthesis ability, reducing yield and increasing susceptibility to disease and insects. Potato growers across the globe are confronted with devastating consequences of the spread of diseases and adverse weather events, urging the implementation of preventive strategies such as crop rotation, disease-resistant cultivars, integrated pest management, and irrigation systems to limit the potential losses and secure a stable harvest (Hee 1958). Progress in this direction relies on ongoing research and development initiatives in reliable disease protection and sustainable farming, resulting in the withstanding of threats in the agricultural sector even as they continue to change. The potato price can be affected by consumer demand, production costs, market tendencies, and potato-to-food use relation (Koblianska et al. 2021). The main reasons for price differences between potato varieties are:

1. Consumer demand: Consumer preferences regarding taste and culinary versatility highly influence potato prices. Varieties are appreciated for flavor, texture, or culinary flexibility fetch reasonable prices. Naturally, this puts a relatively higher

demand on specialty varieties like Yukon Gold, which have a velvety texture and a nutty flavor and often command premium pricing over standard varietals.

2. Production costs: The costs of growing any variety of potatoes differ depending on seed availability, labor intensity, growing conditions, and yield potential. Some may require different growing methods or give low yields; their cost of production may, in turn, be relatively higher than that of other varieties.
3. Market trends: The price of potatoes is bound to fluctuate with demand and supply, which can have setbacks from time to time due to several factors, such as weather or changes in consumer taste. Where shortages exist, or the variety is in high demand, price increases could be forthcoming; the opposite is usually true where there is an oversupply or decline in demand to encourage sales.
4. Culinary characteristics: Varieties differ in flavor, texture, and suitability of tubers for different cooking methods and these differences will determine the prices. Potatoes are valued for culinary traits and adaptability to many recipes; some potatoes may fetch premium prices in specialty or gourmet markets.
5. Seasonality: Potato prices change every season, and some varieties are abundant during certain times of the year, while others are scarce. Changes in supply and demand, including harvest timing and storage availability, will affect the price movements of a specific variety of potatoes.
6. Geography: Regional variations, including local growing conditions, affect the marketing chain and may influence market competition regarding potato prices. Excellent examples of this situation are the varieties that fit the regions or climates under consideration, and their availability may be abundant and cheaper than in other areas.

The price of potatoes is made up of several factors, including the demand of customers, the production costs, trends, how potatoes are used in cooking, the season of the year, and the location where they are grown (Loy et al. 2011). Awareness of these factors will pay off by helping growers and consumers maneuver the potato market and make well-considered decisions on what varieties to grow or buy (Kubala and Firleja 2019). Pesticide-free, natural potatoes are market players, allowing consumers to opt for products that conform to their environmentally friendly preferences. The trend of organic food consumption triggered concerns about health, environmental factors, and ethical farming practices and became the provoking force of the expanding organic potato market (Sabur et al. 2021). The role of organic potatoes in the market and the price premium associated with them can be explored as follows:

1. Consumer preferences: Many consumers buy organic potatoes due to their health and eco-friendly qualities. Organic cultivation practices do not include the usage of synthetic pesticides and fertilizers but include natural crop rotation, composting, biological pest control, and the application of organic manure. This is very attractive to consumers who want to buy produce with no chemical residues and offers a sustainable ecological farming method.

2. Market growth: Demand for organic food, including potatoes, has grown steadily over the last few years. This, combined with improved consumer confidence in health and environmental issues, increased availability of such products at more mainstream supermarkets, and increased confidence in organic certification standards, drives the market.
3. Supply chain considerations: Considering an organic potato supply requires meeting the demand, which enforces stringent organic certification standards that control every step—from the selection of seeds to production and finally to storage. Usually, organic farming involves high labor costs and produces relatively low yields compared to conventional agriculture. This could, therefore, affect the supply of organic potatoes in the market.
4. Price premium: Organic potatoes usually cost more than conventionally grown ones. This premium includes higher production costs, such as those related to organic certification and labor-intensive weed and pest management. However, it might also imply reduced performance due to the lack of synthetic inputs.
5. Market positioning: In most cases, retailers and marketers position organic potatoes as a premium product aimed at customers who can afford to pay an extra price for organic certification and, by implication, health and environmental benefits. The premium price for organic potatoes is also due to its value-added proposition to health-aware consumers.
6. Economic considerations: Despite the price premium, demand for organic potatoes remains strong, driven by consumer preferences and market trends. Some consumers see the higher cost of organic potatoes as an investment in their health and the environment, which outweighs the price difference compared to conventionally grown potatoes.

Organically grown potatoes are an integral player in the market, as they are in high demand by customers seeking pesticide-free products without chemical additives (Emerson and Tomek 1969). The price of crop-organic potatoes is more about the payment for higher growing costs, certification of organic agriculture, and the superior quality image of organic products among health-concerned consumers (Bielza et al. 2007). Besides this price premium, the number of people who pick organic potatoes increases daily due to consumers' new preferences for healthy and ecologically friendly food. These sustainable potato farming practices can help shape future prices by improving efficiency, lowering environmental impact, and guaranteeing long-term safeness (Singh et al. 2017; Bolotova 2015; Bolotova et al. 2008; Muthoni and Nyamongo 2009; Chandran and Pandey 2007). Here are some practices that could play a significant role:

1. Crop rotation: Crop rotation, where potatoes are rotated with other crops, is very useful in breaking pest and disease cycles, improving soil health, and reducing dependence on chemical inputs. If adopted, this practice can lead to higher yields and better quality potatoes over the years, which may eventually contribute to stabilizing or reducing future prices due to reduced production risks.

2. Soil health management: Soil health management practices include cover cropping, mulching, and organic amendments. All these aim to better the soil's health, fertility, and structure. With the help of healthy soils, potatoes will have a firm stand while soil erosion will have been reduced. The water-retaining of the soil and the absorption of its nutrients will be made possible due to the high resilience of this crop. A more humane soil health could result in lower synthetic fertilizer and pesticide needs, thus reducing the cost of production and future prices.
3. Water conservation: Where water-efficient technologies in irrigation, such as drip or precision irrigation, are adopted, it will lead to reduced losses of the resource, improved efficiency of its use, and reduced susceptibility to drought in potato farming. These sustainable water management practices do more than help protect the environment. These reduce the production cost associated with water use, which may have an impact on prices in the future.
4. Integrated pest management (IPM): It includes several strategies to control pest populations using biological controls, crop monitoring, and cultural practices with minimum pesticides. In addition, IPM practices help promote natural pest suppression and maintain beneficial insect populations, thus promoting environmental sustainability, and they are likely to affect future pricing.
5. Energy efficiency: The potato farm can be placed under an energy-efficient model that utilizes renewable energy sources, complete optimization of machinery used, and improved facilities, such as storage, which are likely to reduce the consumption of energy and gasses from the greenhouse. Other than contributing to environmental sustainability, this kind of incorporation reduces the cost of production and, hence, may have some influence on future potato prices.
6. Conservation of biodiversity: Encouragement toward promoting biodiversity on the farms in every aspect—plantation of windbreaks, planting hedgerows, management of habitats for wildlife, and conservation of native species—all help increase the ecosystem's resilience against natural or any other kind of pest. This way, biodiverse farming systems are more resilient to the stresses of the environment. They can contribute toward elevating yields and crop quality standards, potentially impacting future potato pricing.
7. Certification and traceability: Certifying sustainable farming practices under organic or eco-label certification standards can help gain more comprehensive market access and better prices for sustainably produced potatoes. In addition, traceability systems that track back potatoes' origin and production systems can increase customer trust and eventually support price premiums for sustainably produced products.

In brief, incorporating sustainable potato production techniques in the agricultural sector can enhance efficiency, lower the environmental footprint, and boost the crop's resilience. As an outcome, future prices may change since these techniques may cut production costs, improve quality, and meet upcoming consumer demand for sustainable food.

Methodology

This study will apply the current complex deep learning model of the long short-term memory-recurrent neural network (LSTM-RNN) in making predictions of potato prices. This is because RNNs suffer from problems related to gradients, such as vanishing or exploding gradient problems, which make them ineffective in capturing long-term dependencies in sequence data. This is the problem LSTMs resolve by introducing a gating mechanism that ensures relevant parts are retained and new parts are learned over long sequences. This makes them work immensely for problems where grasping context over time is an issue; language modeling and time-series prediction problems are classic examples.

There are several critical components in the architecture of the LSTM-RNN model. The model consists of three layers: two LSTM layers and one RNN layer. The first layer of LSTM consists of 128 hidden units, the second layer of LSTM consists of 64 hidden units, and the third layer of RNN consists of 32 hidden units. The batch size is 512, the learning rate is 0.001, the number of epochs is 50, the optimizer is Adam optimizer, the time steps are 64, and the activation function used in the output is linear. Figure 1 depicts the framework of the proposed model in six steps as follows.

- Dataset collection
- Data normalization using Z-score
- Splitting the dataset into 80% training and 20% testing

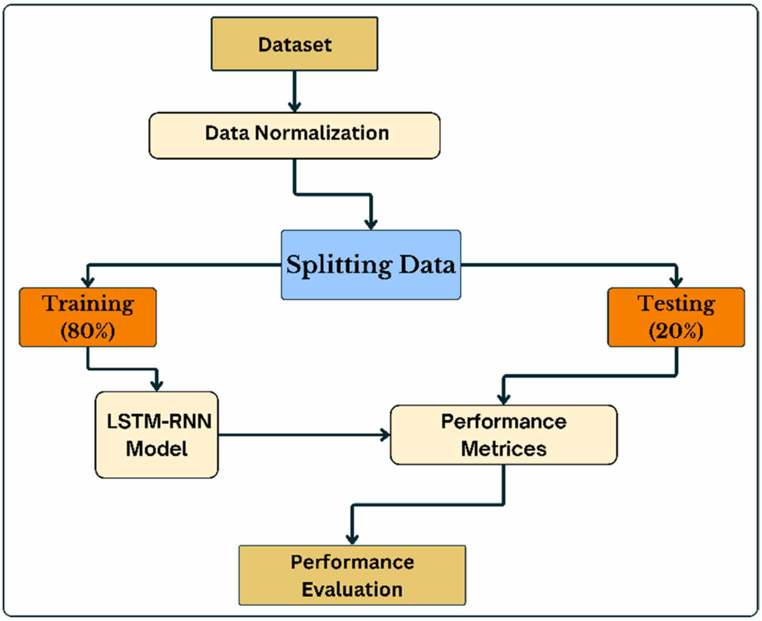

Fig. 1 Framework of the proposed model

- Train the proposed LSTM-RNN model
- Using the evaluation metrics, namely, mean square error (MSE), mean absolute error (MAE), median absolute error (MedAE), and coefficient of determination (R^2).
- Evaluate the performance of the proposed model.

Dataset

The dataset used in this study is available in the literature (Kaggle Potato Pricing n.d.). The dataset likely includes various attributes related to potato pricing, such as the date of pricing, the location (region or market) where the pricing was recorded, and the price of potatoes at that time. The dataset will likely facilitate analysis and research on potato pricing trends, patterns, and factors influencing price fluctuations. Researchers, analysts, and data enthusiasts can utilize this dataset to study the dynamics of potato markets, develop predictive models for potato price forecasting, and explore correlations between potato prices and various factors. Figure 2 demonstrates a plot with potato price over time.

Z-score Normalization

In this study, the critical step that we dealt with was Z-score normalization, which serves as a standard technique known as standardization and, thus, becomes an essential tool for making diverse data ranges compatible and improving the results of different analyses (Fei et al. 2021). This way transforms each possible dataset feature to zero mean (μ) and standard deviation (σ) of one. The normality process is initiated by computing the arithmetic mean and the standard deviation for each parameter, which helps unveil the central tendency and spread of data. After that, every single data point is focused on by subtracting the mean. Thus, these points

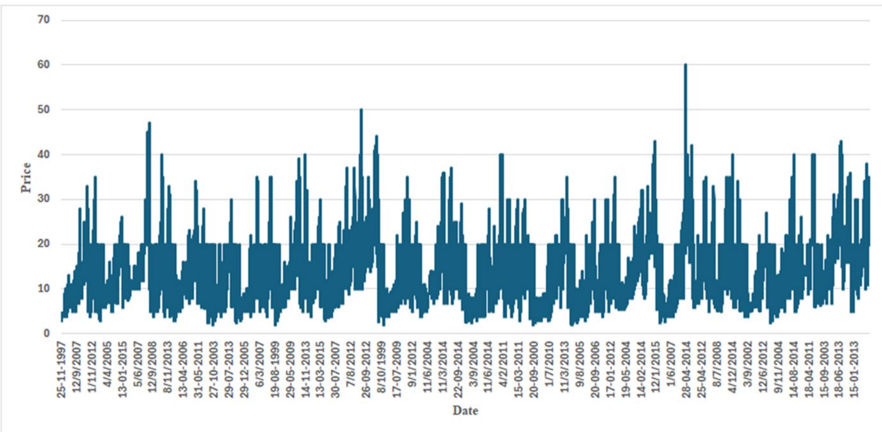

Fig. 2 Potato price over time

build a symmetrical curve that goes around zero (Henderi et al. 2021). This is an essential step as it allows for the removal of the artifact of large-scale feature differences, which, in turn, ensures that no single variable is unduly influencing the decision-making process through an extensive range or scale.

After the center point, each feature is normalized by dividing it by the standard deviation, which results in every feature having a standard deviation of one. This scaling is essential and depends considerably on distance, such as clustering and k-nearest neighbors, and optimization methods like gradient descent. Moreover, Z-score equalization ensures that all units have an equal opportunity to contribute to the analysis process. Thus, the comparison is more straightforward. At the same time, the machine learning algorithms' performance and convergence are improved because of Z-score normalization.

In general, Z-score normalization is unequivocally irreplaceable, especially in improving the performance of our model (the LSTM-RNN), which is envisaged to predict potato prices. It makes the data comparable and compatible across the different parameters, improving both the comprehensibility of the results and the speed of the computations, hence strengthening the dependability of the models.

Machine Learning Models

A discussion on various machine learning models employed for regression tasks, including random forest (RF), K-nearest neighbor (KNN), support vector regressor (SVR), linear regression (LR), and gradient boosting regressor (GBR), gives a general idea of their fundamental working principles and observations on their respective gains and weaknesses (Awad and Khanna 2015).

RF regression is a robust model that is comprised of several decision trees, each of which is built from a random set of training data and features. This technique further guarantees model diversity in predictions, so its prediction performances are improved and prevented from being against the model fitting. During this production, the average of several trees' outputs is used to generate the final prediction. Nonetheless, RF yields firm results with the help of which one can analyze the importance of the feature. However, this method is computationally intensive and not a good tool to extrapolate the trends to the unknown lower or higher data ranges than the training dataset.

K-nearest neighbors (KNN) regressor extracts predictions from the closeness of the data points. It saves all training data and has an average new point outcome for neighbors nearest to a point. Here, we see a model that can be easily used and is suitable for datasets with a few dimensions and fewer rows. Nonetheless, as it is a one-dimensional prediction with an entire dataset, each prediction will not be efficient for large datasets, and performance can degrade with the growth in dimensionality due to the curse of dimensionality.

Support vector regressor (SVR) is a generalized method of support vector models that considers regression problems by employing kernel tricks and support vectors to define the decision function. Twin SVR has diverse applications as it is suitable for recording complex and non-linear relationships. On the other hand, a critical part of anything concerns kernels being chosen carefully and hyperparameters like regularization parameters and tolerance level being attuned to the best level, which can be a computationally expensive way of doing that.

Linear regression (LR) represents one of the simplest forms of regressions with its linear modeling of the relationship between dependent and independent variables. The advantage of this approach is its transparency and versatility in implementation, which is particularly suitable for cases where the linearity in the relationship between variables is foreseen. On the other hand, it is simple, which is good, but at the same time, this prevents it from modeling complex relationships, and it can be sensitive to fit outliers that might skew the findings.

Gradient boosting regressor (GBR) has an ensemble of weak learning models, usually decision trees, as a central component that forms a robust predictive model. It is the next generation of the human error of its predecessor's models, and it improves on that in a series of steps, which makes it the best performer in producing highly accurate forecasting (Sipper and Moore 2022). A critical concern with GBR execution is handling overfitting or adjusting the model to the training set data if the model is not tuned appropriately, which could lead to overfitting. This further needs the setting up of parameters such as learning rate and tree constraints to avoid a tradeoff between the growth of the model and training effectiveness.

Different variations of each regression model allow for particular types of tasks and sometimes give rise to crucial limitations. These techniques work as an excellent substitute where the model performance is outstanding in the case of a complex or structured dataset. However, they may not be as efficient as the simplest linear regression model in situations where the patterns are straightforward. Moreover, SVR can be a powerful tool for these problems. However, parameter tuning must be extensively undertaken to achieve the required optimization. Knowing the details of these methods provides the needed equipment for selecting the best model for a particular regression task, considering the computational efficiency, ease of implementation, and superior prediction capacity.

Evaluation Metrics

We applied several key evaluation metrics to assess the performance of the machine learning models developed for potato price prediction in this study (Tatachar 2021). These metrics include mean square error (MSE), mean absolute error (MAE), median absolute error (MedAE), and the coefficient of determination (R-squared, $R2$). Each of these metrics offers a different perspective on the accuracy and reliability of the predictive models, and their mathematical expressions provide the necessary quantitative assessment:

$$\mathrm{MAE}\ \frac{1}{N}\sum_{n=1}^{N}\left|\hat{V}_n - V_n\right|$$

$$\mathrm{MSE}\ \frac{1}{n}\sum_{i=1}^{n}\left(Act_i - pre_i\right)^2$$

$$MedAE \ \text{median}\left(\left|pre_i - Act_1\right|, \ldots \ldots, \left|pre_i - Act_i\right|\right)$$

$$R^2 \ 1 - \frac{\sum_{n=1}^{N}\left(V_n - \hat{V}_n\right)^2}{\sum_{n=1}^{N}\left(\sum_{n=1}^{N} V_n\right) - V_n\right)^2}$$

These metrics collectively evaluate the model's predictive accuracy and robustness, allowing for a thorough understanding of its effectiveness in real-world applications. By examining both the average errors (MAE, MSE) and the distribution of these errors (MedAE), as well as the proportion of variance explained by the model (*R*2), researchers and practitioners can better assess the model's performance and make informed decisions about its deployment in the field of economic forecasting.

Experimental Results and Discussion

This study used a novel deep learning model called long short-term memory-recurrent neural network (LSTM-RNN) to predict potato prices. The LSTM-RNN model's architecture has several key components. The model consists of three layers: LSTM and RNN layers. The first layer of LSTM contains 128 hidden units, the second layer contains 64 hidden units, and the third layer contains 32 hidden units.

The batch size is 512, the learning rate is 0.001, the number of epochs is 50, the optimizer is Adam optimizer, the time steps are 64, and the activation function used in the output is linear. The performance of the proposed LSTM-RNN model is compared using five traditional machine learning models, namely, random forest (RF) regressor, K-nearest neighbor (KNN) regressor, support vector regressor (SVR), linear regression (LR), and gradient boosting regressor (GBR). Evaluation metrics, including mean square error (MSE), mean absolute error (MAE), median absolute error (MedAE), and coefficient of determination (R^2) were utilized to assess the models' performance. Table 1 illustrates the specific hyperparameter values employed for the regression models in this paper. Here is a description of these hyperparameters.

Table 2 depicts the performance of the proposed LSTM-RNN model and five individual machine learning regression models, namely, random forest (RF) regressor, K-nearest neighbor (KNN) regressor, support vector regressor (SVR),

Table 1 Hyperparameters for the regression models used in this paper

Models	Hyperparameters
RF	N_estimators = 100
KNN	N_.neighbors = 20, weights = "distance"
SVR	Tol = 0.01, C = 2, kernel = "rbf"
LR	Fit_intercept = true
GBR	Learning_rate = 0.001, n_estimators = 150, max_depth = 4

Table 2 Performance of the proposed LSTM-RNN model and other individual machine learning regression models

Models	Mean square error	Mean absolute error	Median absolute error
LSTM-RNN	0.0145	0.0920	0.0652
RF	0.0182	0.1128	0.1039
KNN	0.0753	0.2418	0.2455
SVR	0.1352	0.3228	0.3437
LR	0.1421	0.3372	0.3464
GBR	0.1457	0.3391	0.3489

linear regression (LR), and gradient boosting regressor (GBR). As demonstrated in Table 2, the best results are obtained by the proposed LSTM-RNN model results with an MSE of 0.0145, MAE of 0.0920, MedAE of 0.0652 and R^2 of 98.45%. The GBR model achieved the worth results with MSE of 0.1457, MAE of 0.3391, MedAE of 0.3489, and R^2 of 83.37%. For the F model, its MSE, MAE, MedAE, and R^2 are 0.0182, 0.1128, 0.1039, and 97.62%, respectively. The KNN model achieved an MSE of 0.0753, MAE of 0.2418, MedAE of 0.2455, and R^2 of 91.81%. The MSE, MAE, MedAE, and R^2 for the SVR model are 0.1352, 0.3228, 0.3437, and 85.74%, respectively. For the LR model, its MSE, MAE, MedAE, and R^2 are 0.1457, 0.3391, 0.3489, and 83.37%.

Figure 3 demonstrates the coefficient of determination (R^2) for the proposed LSTM-RNN model and other regression models, namely, the random forest (RF) regressor, the K-nearest neighbor (KNN) regressor, the support vector regressor (SVR), the linear regression (LR), and the gradient boosting regressor (GBR).

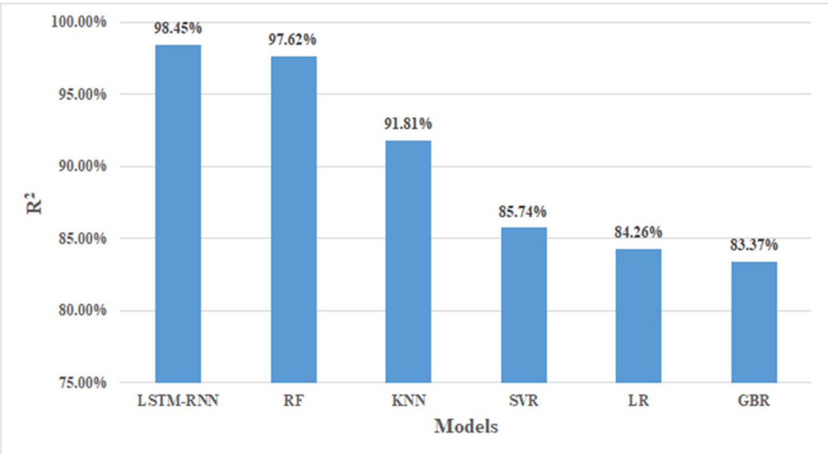

Fig. 3 Coefficient of determination (R^2) for the proposed LSTM-RNN model and several regression models

Figure 4 depicts the training and testing for mean squared error and mean absolute error vs. number of epochs using the LSTM-RNN model for potato price prediction.

An essential point in the comparison study of the LSTM-RNN model with traditional machine learning models for potato price prediction is that several key insights showing the pros and cons of each approach can be generated, giving us a more comprehensive notion of their practical applications.

A common technique among data mining experts is the K-nearest neighbor (KNN) algorithm; it is straightforward to implement and one of the best solutions for small datasets and simple interactions. On the one hand, it is highly efficient when dealing with large datasets since massive computations are required, and the noise can easily be affected. RF, however, is pretty good at the “anti-overfitting” fight and is very capable of working with both categorical and continuous types of data because of its ensemble manner. It is true, however, that its intelligent nature includes some requirements from a computer for high-end computational capabilities, about which it is terrible at extrapolation from unknown data.

Support vector regression algorithm (SVR) works well in high-dimensional spaces and has a significant advantage over various kernel functions, which helps to handle non-linear relationships. Support vector regression algorithms excel in high-dimensional spaces and are equipped with various kernel functions for handling non-linear connections. Notwithstanding this, it is of limited applicability to find the values of parameters and the kernel type correctly, and it can be an issue with computationally intensive requirements. Linear regression (LR) is simply good at linear model interpretation and prediction, but these are, in turn, prone to the bias of linear relationships and noise from outliers. GBR is one of the most developed methods with a flexible boosting procedure that works like a repair/renewal principle. While its ability to fit complex patterns and capture time-order dependencies is its strength, a notable limitation is its propensity toward overfitting and the excessive computational resources required for its sequential training process.

The model that uses a recurrent neural network with long short-term memory stands out, with its ability to represent temporal dependences and time-related

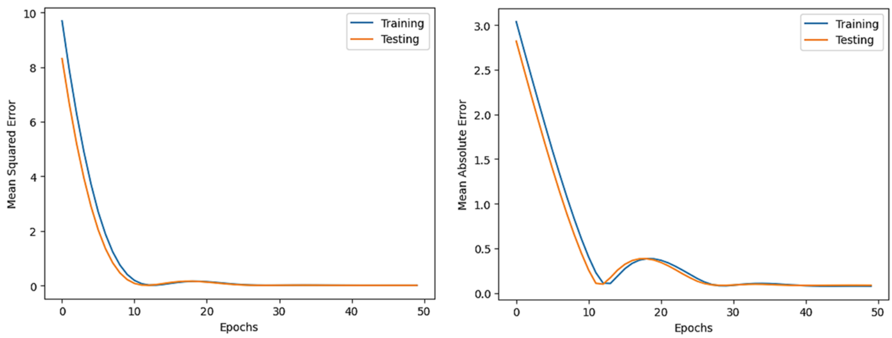

Fig. 4 Training and testing for mean squared error and mean absolute error vs. number of epochs using LSTM-RNN model for potato price prediction

patterns in the sequence data, which are essential in predicting potato prices, which are affected by season and economics. This ability to maintain long-term memory information is one of the most notable features of neural networks, which makes it an exceptional tool for the agricultural economy, where understanding past trends is very important for better forecasting. The practical influence of using LSTM-RNN is significant for investors in agriculture, as implied below. The outcomes song provides a reliable forecasting tool that can support the decision-making process, among other things, by increasing market stability and improving food security approaches. Its outstanding accuracy achieves a very acceptable *R*2 value of 98.45%.

While each type has distinctive traits, focusing them in extreme situations is much better. The preferred model for predicting potato prices is LSTM-RNN because of the diversity of complex, dynamic changes inherent in time-series data, often not covered well by traditional models. This comparison indeed underscores the LSTM-RNN's capability for time-series forecasting. Also, it shows the critical point of settling the model appropriate to the dataset's attributes and forecasting purposes.

Conclusion and Future Work

The globe widely uses potatoes as a basis for agriculture because of the essential nutrition and economic development they facilitate in many developing countries. Because potatoes are so crucial, their price volatility becomes one of the biggest problems, influencing food insecurity and market stabilization. In this study, the LSTM-RNN model, an example of RNN, has been employed to bridge these gaps by predicting potato prices. We combined historical data on potato prices and related economic indicators and normalized them using Z-score normalization to ensure consistency.

In our comparative analysis, the LSTM-RNN model was benchmarked against five conventional machine learning models: random forest (RF), K-nearest neighbor (KNN), support vector regressor (SVR), linear regression (LR), and gradient boosting regressor (GBR). The chosen evaluation metrics—MSE, MAE, MedAE, and R^2—would show that LSTM-RNN outperformed the rest, having an R^2 of 98%.

Practical Implications

The research outcome is significant for stakeholders in the potato industry, such as farmers, policymakers, and market analysts. For farmers, reliable price forecasts will provide the basis for adequate crop care and crop equipment sales, leading to better profits. Policymakers gain the ability to use the knowledge to develop interventions and policies that become effective in the sense of stabilizing food markets and ensuring food supply security. Knowing what affects prices allows market analysts to make more reliable forecasts of future conditions, enhance market efficiency, and make the right investment decisions.

Future Work

Many areas should be explored from a future perspective. Our model plans to be implemented in real-time or near-real-time applications to analyze and predict the market immediately. Due to this feature's real-time nature, decision-makers can access the latest data and act quickly and accurately. Also, we will create intuitive and application-specific interfaces that will present the model predictions, consequently supporting the stakeholders in accepting future market trends.

In conjunction, the model will be continuously engaged with by industry professionals, agricultural experts, and government officials because their validation of the output is essential for verifying the predictions and their accuracy and relevance. Through constant feedback gathering and cooperation with the other key players in the industry, we can keep the model working and make it fit actual, on-farm needs.

Limitations and Discussion

Even though the findings are very inspiring, it is essential to point out that there are certain circumstances. Our predictions' accuracy is greatly influenced by the quality and the quantity of the information applied to them. Data collection irregularities and incompleteness in specific regions may make the model less efficient. Our model has the assumption about the market behavior and external economic factors, which may differ from region to region and condition to condition. Therefore, it may limit the generalizability of this model across different geographic areas or market conditions.

In future revisions to this research, we intend to expand our dataset to accommodate more diverse and detailed worldwide data. Likewise, we will include the use of more diverse variables that can affect potato prices, including weather patterns, political events, and agricultural technologies.

By overcoming these limits and improving the model, we can give it the extra power and precision to make a difference in the worldwide potato market.

Acknowledgements Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R716), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author Contribution All authors have contributed equally.

Data Availability https://www.kaggle.com/datasets/akhilesh84/potato-pricing

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Conflict of Interest The authors declare no competing interests.

References

Anwar DM, Shabbir DG, Shahid MH, Samreen W (2015) Determinants of potato prices and its forecasting: a case study of Punjab, Pakistan

Arısoy H, Bayramoğlu Z (2017) Determination of the effect of price fluctuations on producer income case of potatoes. Turk J Agric-Food Sci Technol 5(11):1342–1349. https://doi.org/10.24925/turjaf.v5i11.1342-1349.1356

Awad M, Khanna R (2015) Efficient learning machines: theories, concepts, and applications for engineers and system designers. Springer Nature. https://doi.org/10.1007/978-1-4302-5990-9

Bielza M, Garrido A, Sumpsi JM (2007) Finding optimal price risk management instruments: the case of the Spanish potato sector. Agric Econ 36(1):67–78. https://doi.org/10.1111/j.1574-0862.2007.00177.x

Bolotova YV (2015) Agricultural supply control: lessons from the US dairy and potato industries. Choices 30(4):1–8

Bolotova YV (2017) Recent price developments in the United States potato industry. Am J Potato Res 94:567–571. https://doi.org/10.1007/s12230-017-9590-4

Bolotova Y, McIntosh CS, Muthusamy K, Patterson PE (2008) The impact of coordination of production and marketing strategies on price behavior: evidence from the Idaho potato industry. Int Food Agribus Manag Rev 11(3):1–30

Bolotova Y, McIntosh CS, Patterson PE, Muthusamy K (2010) Is stabilization of potato price effective? Empirical evidence from the Idaho Russet Burbank potato market. Agribusiness 26(2):177–201. https://doi.org/10.1002/agr.20213

Chandran KP, Pandey NK (2007) Potato price forecasting using seasonal ARIMA approach. Potato J 34(1–2)

Dhakre DS, Bhattacharya D (2016) Price behavior of potato in Agra market-a statistical analysis. Indian Res J Ext Educ 14(2):12–15

Emerson PM, Tomek WG (1969) Did futures trading influence potato prices? Am J Agr Econ 51(3):666–672. https://doi.org/10.2307/1237918

Fatima A, Abid S, Naheed S (2015) Trends in wholesale prices of onion and potato in major markets of Pakistan: a time series analysis. Pak J Agric Res 28(2)

Fei N, Gao Y, Lu Z, Xiang T (2021) Z-score normalization, hubness, and few-shot learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, IEEE, 142–151. https://doi.org/10.1109/ICCV48922.2021.00021

Goodwin HL Jr (1988) Factors affecting fresh potato price in selected terminal markets. West J Agric Econ 233–243

Hee O (1958) The effect of price on acreage and yield of potatoes. Agric Econ Res 10(4):131–141

Henderi H, Wahyuningsih T, Rahwanto E (2021) Comparison of Min-Max normalization and Z-Score normalization in the K-nearest neighbor (kNN) algorithm to test the accuracy of types of breast cancer. Int J Inf Inf Syst 4(1):13–20. https://doi.org/10.47738/ijiis.v4i1.73

Kaggle Potato Pricing. https://www.kaggle.com/datasets/akhilesh84/potato-pricing. Accessed 9 May 2024

Koblianska I, Kalachevska L, Minta S, Strochenko N, Lukash S (2021) Modeling and forecasting of potato sales prices in Ukraine. Agric Resour Econ: Int Sci E-J 7(4):160–179. https://doi.org/10.51599/are.2021.07.04.09

Koblianska I, Seheda S, Khaietska O, Kalachevska L, Klochko T (2022) Determinants of potato producer prices in the peasant-driven market: the Ukrainian case. Agric Resour Econ: Int Sci E-J 8(3):26–41. https://doi.org/10.51599/are.2022.08.03.02

Kubala S, Firleja K (2019) Determinants of variation of potato prices in the European Union. Economia Agro-Alimentare, XXI 3:697–707. https://doi.org/10.3280/ECAG2019-003007

Kumar RR, Baishya M (2020) Forecasting of potato prices in India: an application of ARIMA model. Econ Aff 65(4):473–479. https://doi.org/10.46852/0424-2513.4.2020.1

Loy JP, Riekert S, Steinhagen C (2011) Potato prices as affected by demand and yearly production: a German perspective. Am J Potato Res 88:195–198. https://doi.org/10.1007/s12230-010-9176-x

Lu H, Li T, Lv J, Wang A, Luo Q, Gao M, Li G (2023) The fluctuation characteristics and periodic patterns of potato prices in China. Sustainability 15(10):7755. https://doi.org/10.3390/su15107755

McGregor I (2007) The fresh potato market. Potato Biol Biotechnol 3–26. https://doi.org/10.1016/B978-044451018-1/50043-9

Mitra S, Mookherjee D, Torero M, Visaria S (2018) Asymmetric information and middleman margins: an experiment with Indian potato farmers. Rev Econ Stat 100(1):1–13. https://doi.org/10.1162/REST_a_00699

Muthoni J, Nyamongo DO (2009) A review of constraints to ware Irish potatoes production in Kenya. J Hortic For 1(7):98–102

Pavlista AD, Feuz DM (2005) Potato prices are affected by demand and yearly production. Am J Potato Res 82:339–343. https://doi.org/10.1007/BF02871964

Rosen S (1999) Potato paradoxes. J Polit Econ 107(S6):294–313. https://doi.org/10.1086/250112

Sabur SA, Miah MM, Begum MEA, Rahman MS, Molla MU (2021) Demand, supply and reasons for price hike of potato in Bangladesh. Bangladesh J Agric Econ 42(1):93–108

Şahinli MA (2020) Potato price forecasting with Holt-Winters and ARIMA methods: a case study. Am J Potato Res 97(4):336–346. https://doi.org/10.1007/s12230-020-09788-y

Salmensuu O (2021a) Potato is important for development, focusing on prices. J Risk Financ Publ Manag 14(3):137. https://doi.org/10.3390/jrfm14030137

Salmensuu O (2021b) Macroeconomic trends and factors of production affecting potato producer prices in developing countries. J Dev Areas 55(1). https://doi.org/10.1353/jda.2021.0008

Singh DK, Pynbianglang K, Pandey NK (2017) Market arrival and price behaviour of potato in Agra district of Uttar Pradesh. Econ Aff 62(2):341–345. https://doi.org/10.5958/0976-4666.2017.00018.3

Sipper M, Moore JH (2022) AddGBoost: a gradient boosting-style algorithm based on strong learners. Mach Learn Appl 7:100243. https://doi.org/10.1016/j.mlwa.2021.100243

Tatachar AV (2021) Comparative assessment of regression models based on model evaluation metrics. Int Res J Eng Technol 8(09):2395–56

Thorne FS (2012) February. Potato prices as affected by supply and demand factors: an Irish case study. In: 123rd Seminar

Authors and Affiliations

Sarah A. Alzakari[1] · **Amel Ali Alhussan**[1] · **Al-Seyday T. Qenawy**[2] · **Ahmed M. Elshewey**[3] · **Marwa Eed**[4,5]

✉ Ahmed M. Elshewey
ahmed.elshewey@fci.suezuni.edu.eg

1 Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, 11671 Riyadh, Saudi Arabia

2 Intelligent Systems and Machine Learning Lab, Shenzhen 518000, China

3 Department of Computer Science, Faculty of Computers and Information, Suez University, P.O.BOX:43221, Suez, Egypt

4 Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura 11152, Egypt

5 Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura 35111, Egypt