FORECASTING PRICES OF MAJOR UKRAINIAN EXPORT PRODUCTS WITH MACHINE LEARNING TOOLS (LSTM-MODEL)

by

Danyil Dehtiarenko

A thesis submitted in partial fulfillment of the
requirements for the degree of

MA in Business and Financial Economics

Kyiv School of Economics

2024

Thesis Supervisor: ______________________________ Professor Oleh Nivievskyi

Approved by ____
Head of the KSE Defense Committee, Professor [Type surname, name]

____

____

____

Date ______________________________________

ACKNOWLEDGMENTS

The author wishes to thank all the departments of KSE, all the professors and their assistants, as well as the students who contributed to my knowledge. Special gratitude to Oleh Nivievskyi for his contribution to my scientific achievements and the experience gained.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

LSTM Long Short-Term Memory

RNN Recurrent Neural Network

MAE Mean Absolute Error

RMSE Root Mean Squared Error

DXY USA Dollar Index

S&P 500 Standard & Poor's 500 (Index of 500 largest USA companies)

NN Neural Network

ML Machine Learning

VAR Vector Autoregression

ARIMA Autoregressive Integrated Moving Average

CHAPTER 1. INTRODUCTION

This research paper studies the modeling and forecasting of important Ukrainian major export commodities such as wheat, corn, soybeans, and steel. Commodity pricing for these commercially important products directly relates to global food security, inflation, and energy stability. Developing and improving the accuracy of price forecasts for key Ukrainian export commodities can also have a significant impact on Ukrainian exporting enterprises, policymakers, and policy institutions, since the effectiveness of applying high-quality analysis of future price behavior determines the foundation for decisions in marketing, risk management, and investment strategies. Indeed, the following development is called for as a means of reducing economic risk and ensuring world stability, especially in developing economies such as Ukraine.

This research uses data visualization and econometric modeling of time series via application of LSTM (Long Short-Term Memory) model as a base for analysis and price forecasting. Such a model allows the identification of temporal dependencies in the data for the extraction of patterns, which can be further used to predict prices in the future.

Statistics covering the periods 2000 to 2024 will be used to analyze daily futures prices of its constituents. The results of the study will reveal key trends and patterns existing in the price dynamics of wheat, corn, soybeans and steel, and will also provide methodologies for effective price forecasting that can be applied in other economic analyses. Such a conclusion will undoubtedly be effectively used by economists, analysts, agricultural producers and policymakers trying to improve the decision-making process based on price forecasts for key Ukrainian export commodities.

The purpose of this study is to contribute to the dissemination of machine learning methods in the field of standard methods of economic analysis. I hope that my successful application of the LSTM model will make a relevant contribution to creating a foundation for the application of machine learning in Ukrainian and international research in the field of economics and time series analysis.

This study also aims to increase the relevance of the topic of data collection for key economic indicators and indices in the book in the emerging economies of the world. After all, one of the key guarantees of the success of the economy now is the presence of institutions and open data sources that collect historical data on key macro and microeconomic indicators. After all, it is these indicators and datasets that allow analysts and economists to conduct research and expand the scope of their research.

Most post-Soviet countries differ from developed economies in the presence of high-quality historical data on their key export products and industries that are the drivers of the economy.

Every subsequent day during which prices and indices for domestic products are not saved is an irretrievable loss for all analysts and economists who will further contribute to the development of the country and industry.

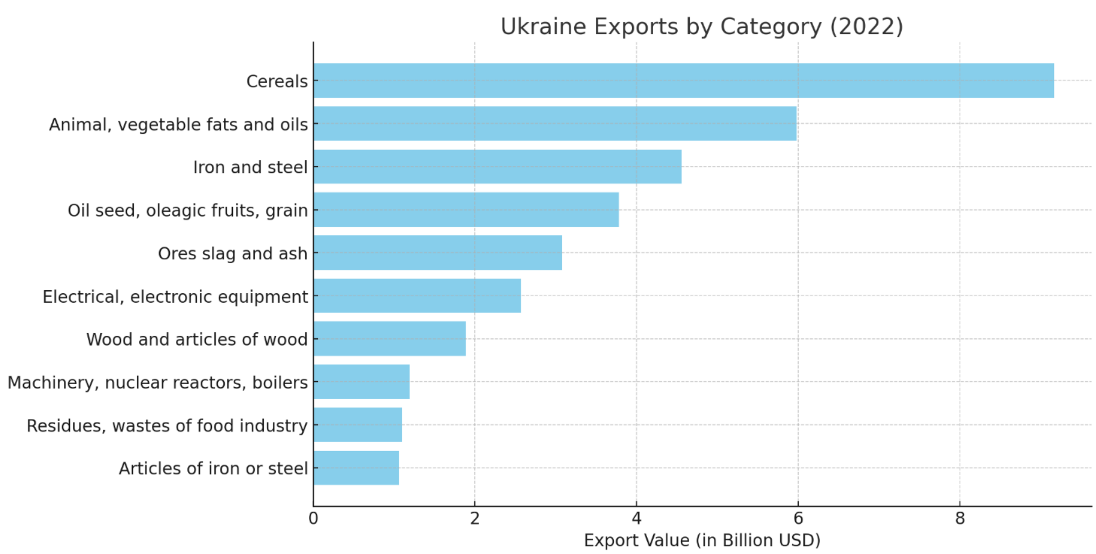
Therefore, in this study, the author proposes that all analysts and economists scale and expand their forecasts using machine learning, which is now available to specialists of all levels and can improve the performance of their analysis like no other tool before.

CHAPTER 2. INDUSTRY OVERVIEW AND RELATED STUDIES

2.1. Industry Overview

The commodity market has a significant impact on the living conditions of all people on the planet, price stability for basic consumer goods is the key to the security of households. To understand the nature of commodity price behavior, it is necessary to analyze the market and key suppliers. First, let's look at which products are the leaders in Ukrainian exports:

Figure 1. Ukraine Exports by Category

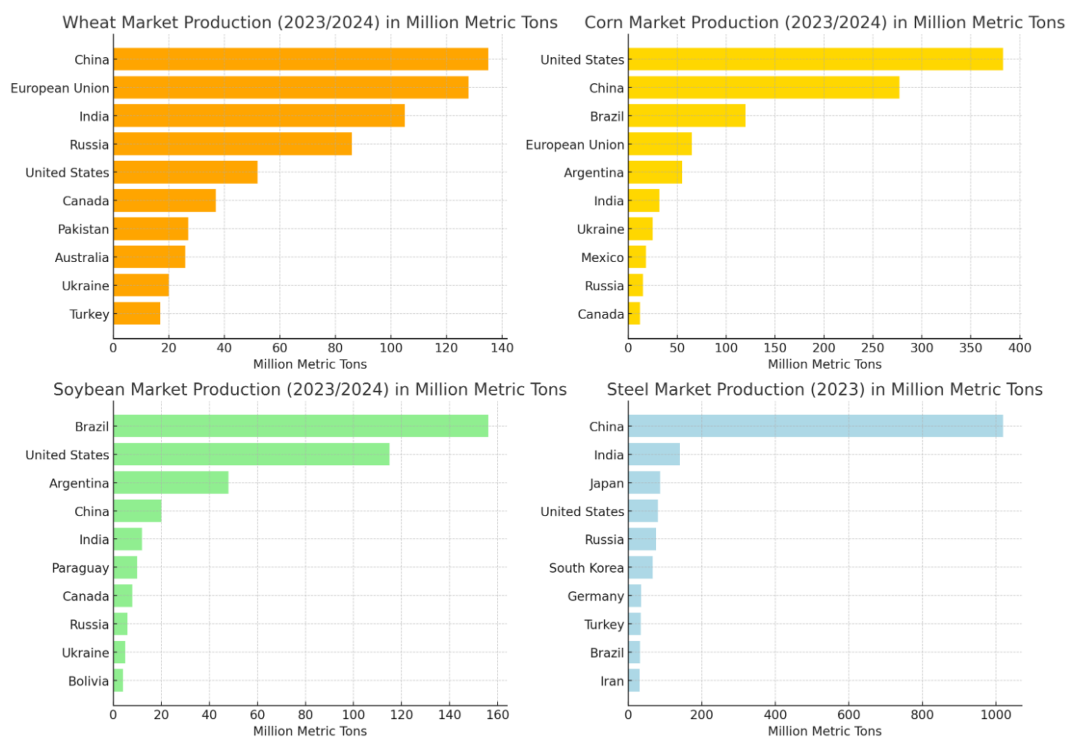


Source: https://tradingeconomics.com/ukraine/exports-by-category

From Figure 1 we can see, grain products and animal products are Ukraine's main exports, with trade volume estimated at $15 billion in 2022. Also worth noting is the impressive volume of iron, steel and ores trade, estimated at $7.6 billion.

To understand the nature of price changes for key Ukrainian export products, it is worth considering each individual global export market for goods in terms of participants and their trade volume:

Figure 2. Commodity market production and producers by country

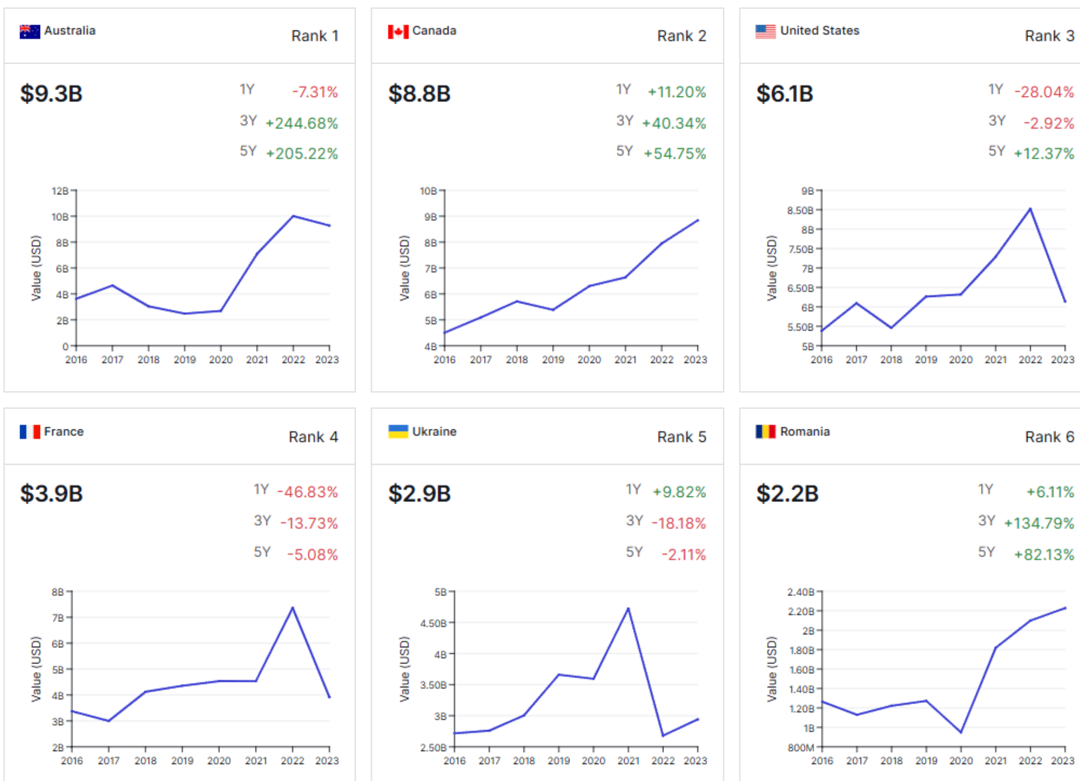


Source: https://fas.usda.gov/data/production, https://worldsteel.org/data/world-steel-in-figures-2024/

- Wheat production:

In 2023, the wheat trade volume is estimated at $50-55 billion. Annual world wheat pressings vary around 1.4 million tons. From Figure 2 it can be observed that The leading wheat producers are China, the EU, India and Russia. On the world stage, these countries are distinguished by their large population and agricultural land. Despite the large volumes of production, most of the wheat is consumed domestically and is not exported, as Figure 3 shows:

Figure 3. Top 6 countries by wheat export values



Source: https://www.tridge.com/intelligences/wheat/export

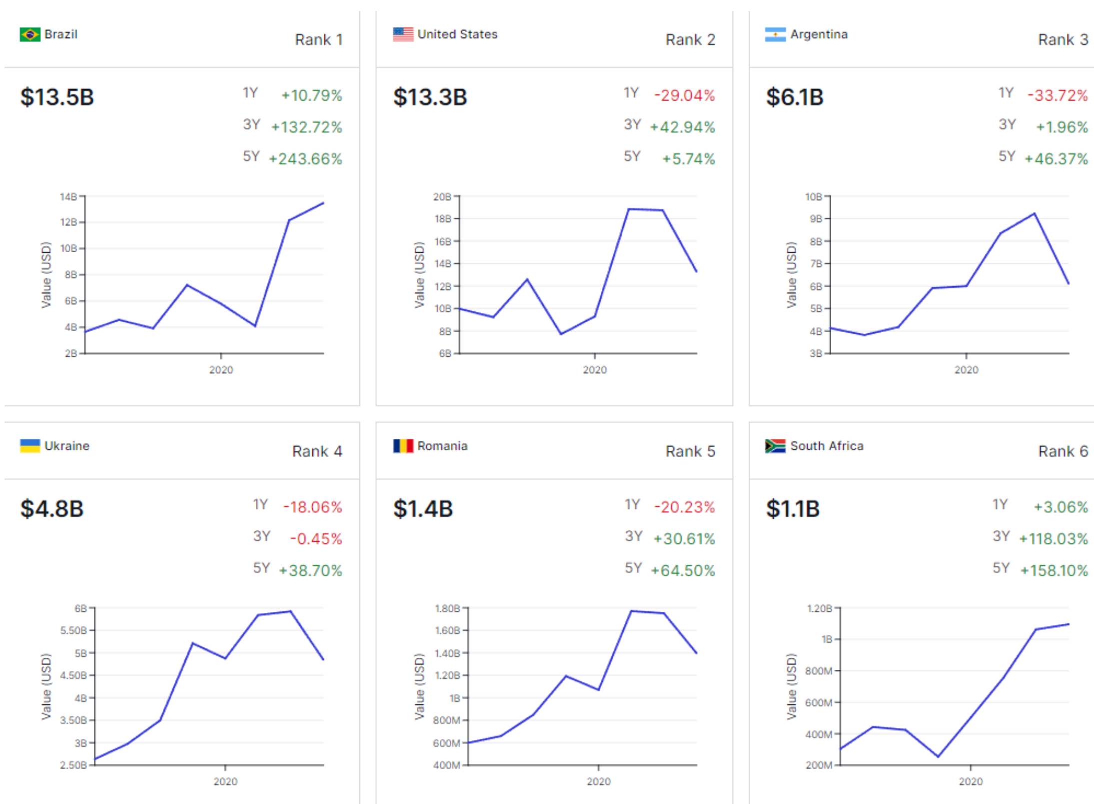
Wheat is a strategically important commodity and the foundation of food security for many countries. Leadership in the trade of such a sought-after raw material, in addition to the obvious advantages in the form of profit, provides opportunities to take a more confident position in the geopolitical arena. It is also worth taking into account that China, which ranks first in wheat production, also ranks first in imports of this raw material. A significant breakthrough in wheat trade has occurred in Australia, Canada and Romania. These countries have shown large and rapid growth in recent years.

Figure 3 shows the significant impact of the outbreak of war in Ukraine on the export of wheat by key global suppliers, which in turn has a significant impact on the stability of wheat prices.

- Corn production:

The corn market for 2023 is estimated at $50-55 billion. The total production volume is approximately 1.4 million tons. The key corn producers in the world are the United States, China and Brazil, which have a significant impact on corn pricing in the world.

Figure 4. Top 6 countries by corn export values



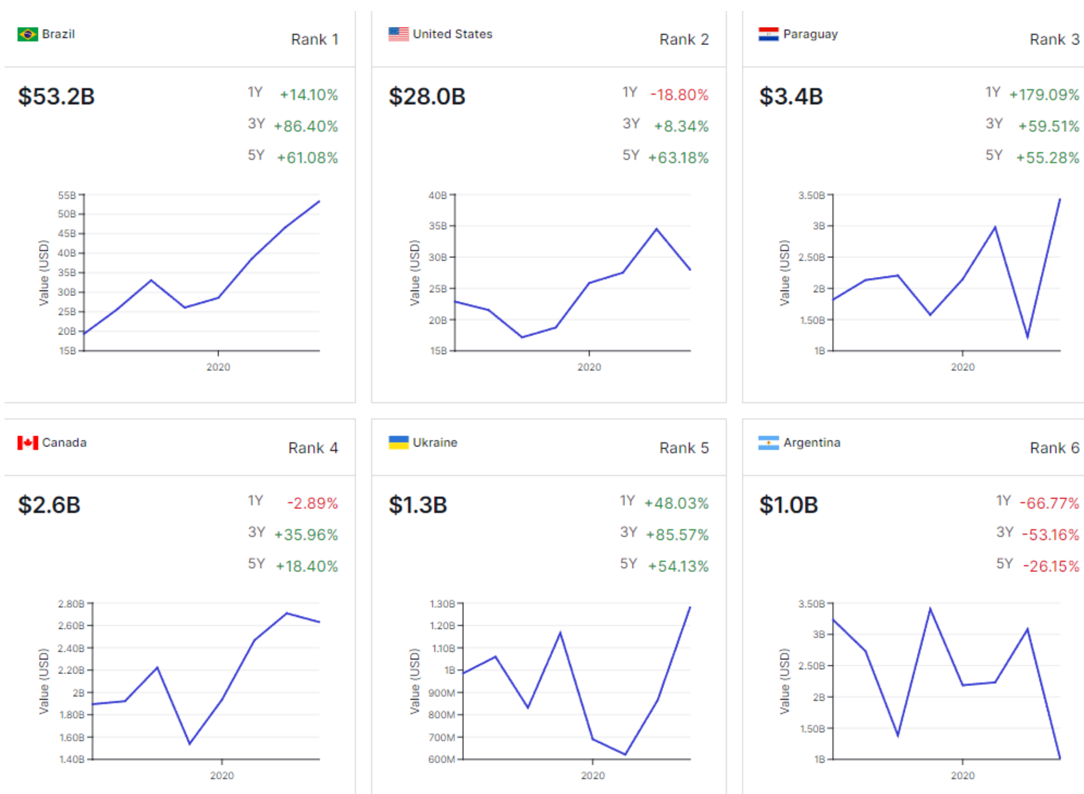
Source: https://www.tridge.com/intelligences/corn/export

Today, corn is very actively used in the production of most food products and is becoming a new alternative source of environmentally friendly energy (biofuel), becoming a product of significant influence on the world economy. Also, corn is one of the main products for feed intended for farm animals.

The growth of trade from leading corn supplier countries highlights the growth of corn consumption in food and energy markets, in turn providing growth opportunities for developing economies such as Brazil, Argentina, Ukraine, Romania and South Africa.

- Soybean production:

Figure 5. Top 6 countries by soybean export values



Source: https://www.tridge.com/intelligences/soybean/export

The global Soybean market is valued at $93 billion in 2023, with production reaching 628.48 million tons in 2022 showing a growth of 63% over the past five years. The leader in soybean production by a wide margin from other countries is Brazil, together with the United States. Ukraine and Paraguay have significantly increased their soybean exports over the past five years. This growth is due to a significant increase in soybean consumption and its popularization in food production.

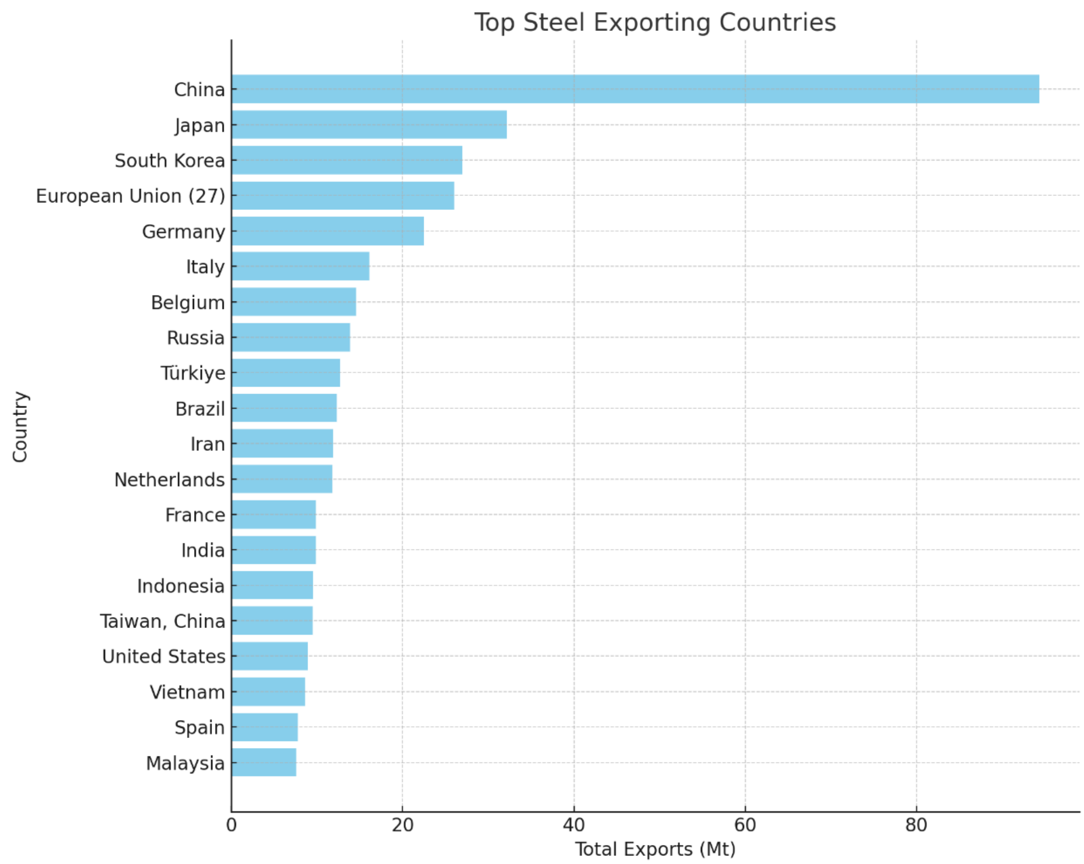
Soybeans are becoming an increasingly important part of the diet of people who want to maintain a healthy diet, consume quality proteins and replace meat. The production of such raw materials is an excellent opportunity for economic growth for developing countries, which are the leaders in the world supply of soybeans.

- Steel production:

The volume of steel production is about 3489 million metric tons. In recent years, the steel market has shown stable and moderate behavior. Over the past 5 years, the level of production has grown by an average of 1%. The leading manufacturing countries have also not changed over time. It is worth noting that China is the largest manufacturer by a very decent margin from other countries, but most of the steel is consumed within China itself.

Even though steel ranks third in Ukrainian exports, in the world arena, Ukrainian steel ranks only 14th in terms of trade volumes, 6.2 million metric tons. As expected, the leaders in steel exports are Asian countries (China, South Korea and Japan). The countries of the European Union, the USA and Germany buy the most steel on the market.

Figure 6. Steel export in 2023 by country



Source: https://worldsteel.org/data/world-steel-in-figures-2024/

2.2. Related studies

The World Bank's report "Forecasting Industrial Commodity Prices" from the Commodity Markets Outlook series (April 2023) presents a wide range of analysis on the topic of "Forecasting Industrial Commodity Prices". This material examines methods for forecasting prices of key commodities. And the report pays special attention to machine learning methods for time series analysis.

Along with the development of technologies and the growing availability of programming, storing and processing large amounts of data, the topic of developing and popularizing machine learning methods is increasingly appearing in studies and reports of leading financial and analytical institutions and organizations.

This is justified by the significant potential of machine learning in strengthening the security of global commodity markets and creating comfortable conditions for developing economies in the world, especially Ukraine. After all, it is the supply of raw materials that is often the main export product and, accordingly, a source of income for countries with an underdeveloped financial and production field. Accurate price forecasting will help make more effective decisions in the field of finance and management for both the private sector and the public sector.

Machine learning methods are distinguished by their effectiveness due to their ability to remember impressive amounts of data, for example, 20 years of corn futures prices, and at the same time the model can take into account many other factors, such as financial indices and macroeconomic indicators. Due to this advantage, the accuracy of forecasts significantly exceeds the results of standard econometric models such as ARIMA and VAR. It is also worth noting the ease of application of machine learning methods due to ready-made libraries with code, but the need for large amounts of data to train the model is still the main barrier that slows down the development of this technology for time series analysis.

A significant contribution to the application of machine learning (ML) in economics is presented in the review by Storm, Baylis, and Heckelei (2019) titled "Machine learning in agricultural and applied economics", published in the European Review of Agricultural Economics. This particular study covers both early work on the application of machine learning methods dating back to the early 1990s and modern research and practice.

The key subject of the references and literature of my research is the use of neural networks (NN), which have deservedly become the most widely used ML method in economic forecasting.

Recurrent neural networks (RNN) have become the mouthpiece of machine learning for time series forecasting. Among all the varieties of RNNs, the long short-term memory (LSTM) model is one of the most effective tools when dealing with long-term dependencies, such as the subject of my research - commodity price forecasting. The unique feature of LSTM is its ability to process sequential data and “remember” relevant information over long time spans, as well as “forget” irrelevant details that do not contribute to predicting future values. This makes LSTM an excellent tool for economic analysis, where historical trends greatly influence future values of variables.

Unlike the ARIMA model, which requires a clear definition of the lag structure and the assumption of linear relationships, the LSTM model automatically learns temporal patterns in the data. This is achieved using so-called entry, exit, and forget gates that control the flow of information in the network. These gates allow LSTM to maintain a balance between short-term and long-term memory. This ability to quickly remember and forget makes the model highly effective for forecasting in volatile and nonlinear markets.

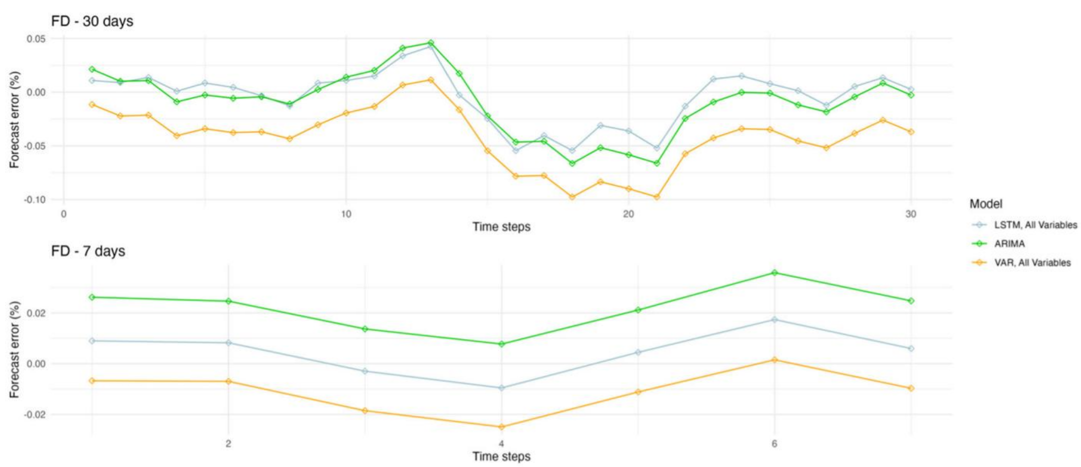
Another important advantage of the LSTM model is its ability to add additional explanatory variables to the model in addition to the main time series. Macroeconomic indicators and financial indices can greatly improve the accuracy of forecasts.

In the context of my specific research, I would consider in the future the integration of weather change data and climate change indices in key regions of food commodity producing countries.

I would also like to highlight one of the recent scientific papers studying the application of LSTM-RNN in the agro-industrial sector. This paper by Brignoli (2024) studies a specific application of LSTM-RNN to forecasting agricultural commodity futures prices. It was this paper that inspired me to study this topic in more detail to popularize machine learning in Ukrainian economic analysis.

Figure 7 shows the result of comparing the performance of classical econometric models Arima and VAR against the LSTM-RNN model.

Figure 7 Performance comparison of LSTM versus classical models for 30 and 7 day forecast horizon.



Source: Brignoli et al. (2024), Machine learning to predict grains futures prices, Agricultural Economics

I would also like to highlight one of the recent scientific papers studying the application of LSTM-RNN in the agro-industrial sector. This work by Brignoli (2024) studies the specific application of LSTM-RNN to forecast prices of agricultural commodity futures. It was this paper that inspired me to study this topic in more detail to popularize machine learning in Ukrainian economic analysis.

The authors of the study compared the performance of classical econometric models Arima and VAR against the LSTM-RNN model.

As part of the study, the authors found that LSTM-RNN outperforms traditional econometric models over relatively long forecasting horizons (30 days). The mean absolute error (MAE) is on average 10% lower than that of classical Arima and VAR models.

However, over a short distance of 7 days, LSTM-RNN did not particularly distinguish itself in terms of super-performance. The classical Arima model turned out to be more effective over a short distance.

It was the results of this study that inspired me to study the effectiveness of the model over long time periods, because it is over a long distance that LSTM-RNN shows unique effectiveness that is unattainable for classical econometric analysis models.

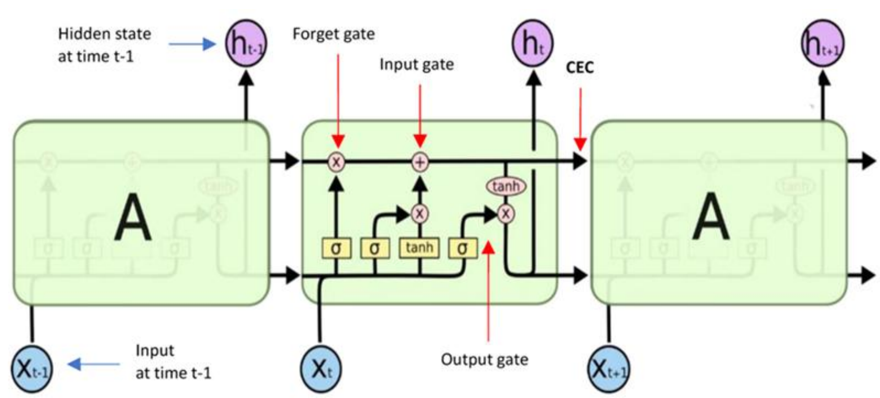
Based on my analysis of the literature and my own vision for the development of machine learning applications in economic analysis, I can draw the following conclusion: machine learning models, in particular LSTM, provide the widest possible range of analysis of variables that can affect the subject of prediction. In combination with the ability to remember historical trends over a long period of time, this creates an unprecedented tool for generating forecasts that can make a significant contribution to ensuring stability in food markets and predicting crisis stages in the market.

CHAPTER 3. METHODOLOGY

Time series analysis is one of the main tools for forecasting a value in the future. In this particular study, futures prices at a daily frequency for 20 years will be used to analyze and forecast prices for Ukrainian export products, which will allow us to form a large enough dataset for training the LSTM model. When working with machine learning, it is extremely important to have data for a long period of time, and unfortunately, most of the Ukrainian export products are not presented on exchanges, or have been presented only recently (sunflower oil, ores, titanium, etc.).

To obtain forecasts for futures of Ukrainian export products, this study uses the LSTM (Long Short-Term Memory) recurrent neural network. This model, which uses machine learning, is excellent for analyzing massive time series due to its feature - the presence of memory cells and the ability to manage it. From Figure 8 it can be observed that each cell has three gates - Input, Output and Forget. Thanks to these cells, the model can forget data when needed and remember data for a long time and more effectively analyze changes in the time series, which contributes to more accurate forecasts.

Figure 8. Machine learning to predict grains futures prices, Brignoli 2024



Source: Brignoli et al. (2024), Machine learning to predict grains futures prices, Agricultural Economics

LSTM (Long Short-Term Memory) is a recurrent neural network (RNN) designed to work with time series.

Advantages of the model for forecasting raw materials:

- LSTM is excellent at remembering important dependencies over long periods of time, which makes it suitable for analyzing time series, as in our case with a 24-year data set.

- Unlike traditional models such as ARIMA, which work better with linear dependencies, LSTM is excellent at identifying and exploiting complex nonlinear dependencies.

- LSTM can easily handle irregular time series and take into account both seasonal and trend components of the data.

Main parameters of model:

- The number of units in LSTM layers is the number of neurons in each LSTM layer that allow the model to capture more complex dependencies.

- Dropout - helps prevent model overfitting by randomly excluding neurons during training.

- Batch size - the number of examples needed to update the model weights at each training step. Smaller values make training more accurate but slower, while larger values can speed up the process but with a loss of accuracy.

- The number of epochs - the number of times the model goes through the entire dataset during training.

Step-by-step plan for applying the model in research with Python software:

1. Applying normalization (MinMaxScaler) to bring the data to the range from 0 to 1

2. Generating input data for LSTM with a sequence of 60 day steps to predict the next value

3. Creating the LSTM model
4. Compiling the model using the mean squared error loss function and the adam optimizer.
5. Training the model on the prepared data.
6. Predicting values for the test data.
7. Applying model evaluation metrics (MAE, RMSE, R^2) for each product item.
8. Comparing real and predicted values on graphs to evaluate the accuracy of the model.

The proposed methodology based on the LSTM model provides high accuracy in forecasting prices for key Ukrainian export goods over long time periods. It is adapted to work with large volumes of data, contains mechanisms for processing seasonal and nonlinear trends, and takes into account structural changes in the data. These advantages make it an ideal tool for analyzing time series in the context of unstable markets.

CHAPTER 4. DATA

To model and forecast prices for leading Ukrainian export commodities, this study uses data on futures prices collected from MacroTrends and YahooFinance. The author collected prices on a daily basis over a period from 2000 to 2024 year.

Table 1. Descriptive statistics

	count	mean	std	min	25%	50%	75%	max
wheat	5908	540.116	183.822	244.75	404.0	515.75	656.312	1425.25
corn	5921	412.481	161.42	183.5	309.5	374.0	515.75	831.25
soybean	5954	1005.52	335.755	418.5	762.312	975.625	1303.62	1771.0
steel	5962	247.49	133.008	47.04	167.642	228.34	293.265	718.72

Source: https://www.macrotrends.net/, https://finance.yahoo.com

As Table 1 shows, the average price of soybean futures is $1005.52, which is the highest among all the selected products. Considering the large difference in the average price of soybeans with wheat ($540.12), corn ($412.48), and steel ($247.49), we can conclude that there is significant demand and significant costs for the production of this commodities.

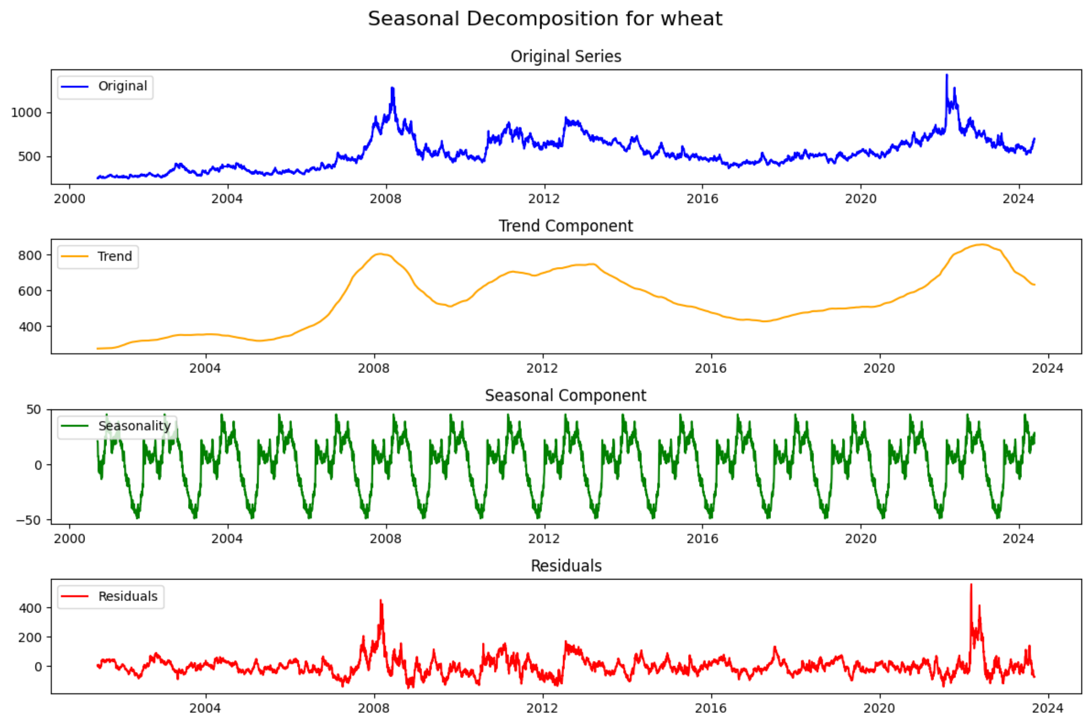
In addition to the high average price, soybeans are distinguished among other raw materials by their high volatility. The standard deviation of soybean futures prices is $335.76. Wheat and corn also have quite large standard deviations, which can be due to the seasonality of production of these products and the economic and political situation in key supplier countries, such as Ukraine. In general, this market can be described as quite volatile, and the need for accurate price forecasting for these products is high. Steel, in turn, has a standard deviation of $133.01, which is the smallest value among all the data sets in this study. This may be due to a fairly stable supplier, which supplies more than half of all steel on the market - China.

It is also worth noting that soybean futures have the largest range of Min and Max values - from $418.5 to $1771. This fact confirms that the soybean price is prone to volatility. The range of Min and Max values for steel futures is from $47.04 to $718.72, which is also quite large. Compared to soybeans and steel, corn and wheat have narrower ranges - $183.5 - $831.25 for corn and $244.75 - $1425.25 for wheat), but still quite large.

The median values (50%) for all export products are closer to the average, indicating a symmetrical price distribution.

To better understand the price behavior of key Ukrainian export products, it is necessary to analyze their trends and seasonality:

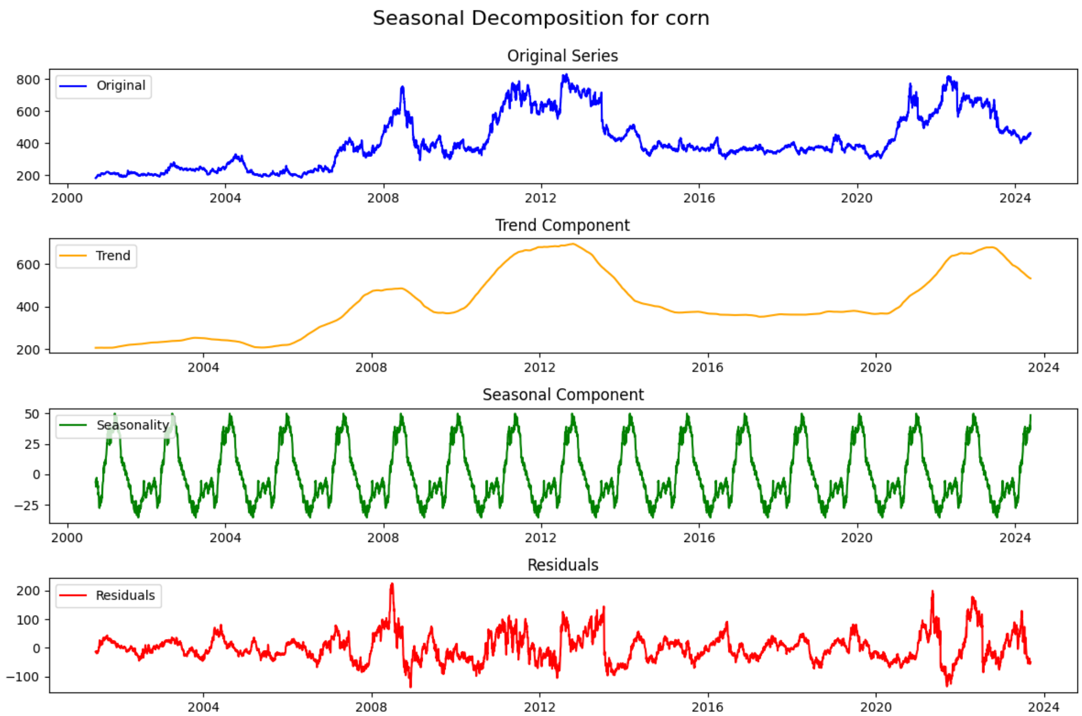
Figure 9. Seasonal decomposition for wheat



Source: Made by the author using Python software based on the data used in this study

The seasonality analysis confirms the presence of distinct annual cycles in wheat prices and highlights their sensitivity to crises, such as those observed in 2008 and 2020–2022.

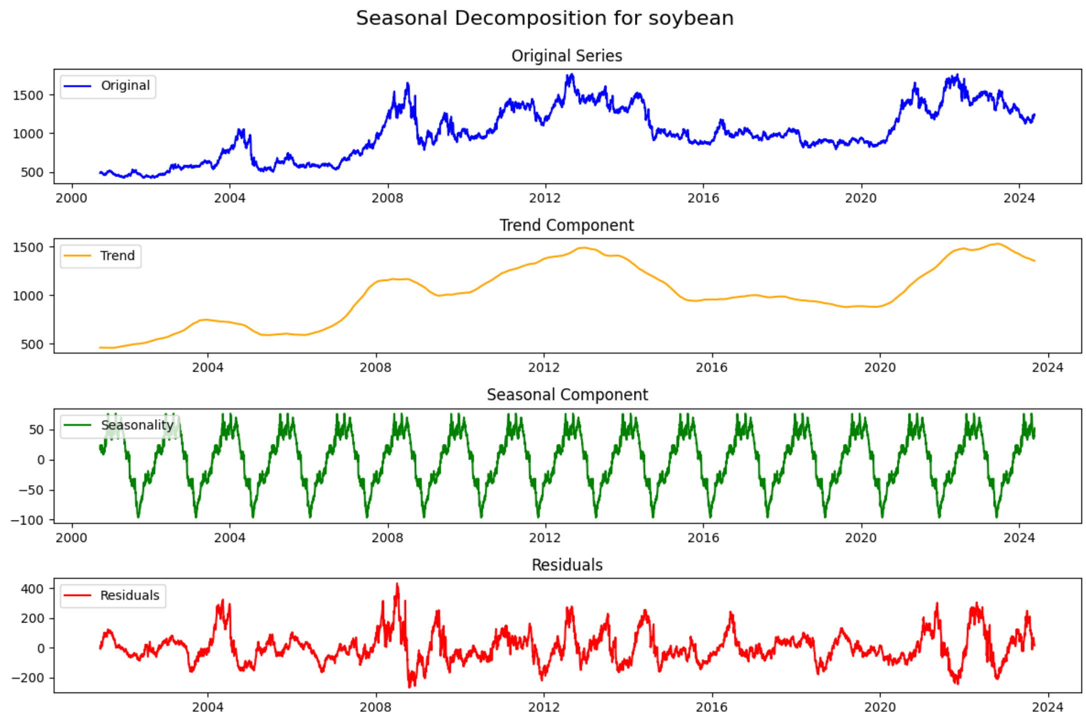
Figure 10. Seasonal decomposition for corn



Source: Made by the author using Python software based on the data used in this study

Seasonal analysis of corn prices shows consistent annual cycles with notable sensitivity to market volatility during the 2008 and 2020–22 crises. Corn and wheat have similar seasonal patterns.

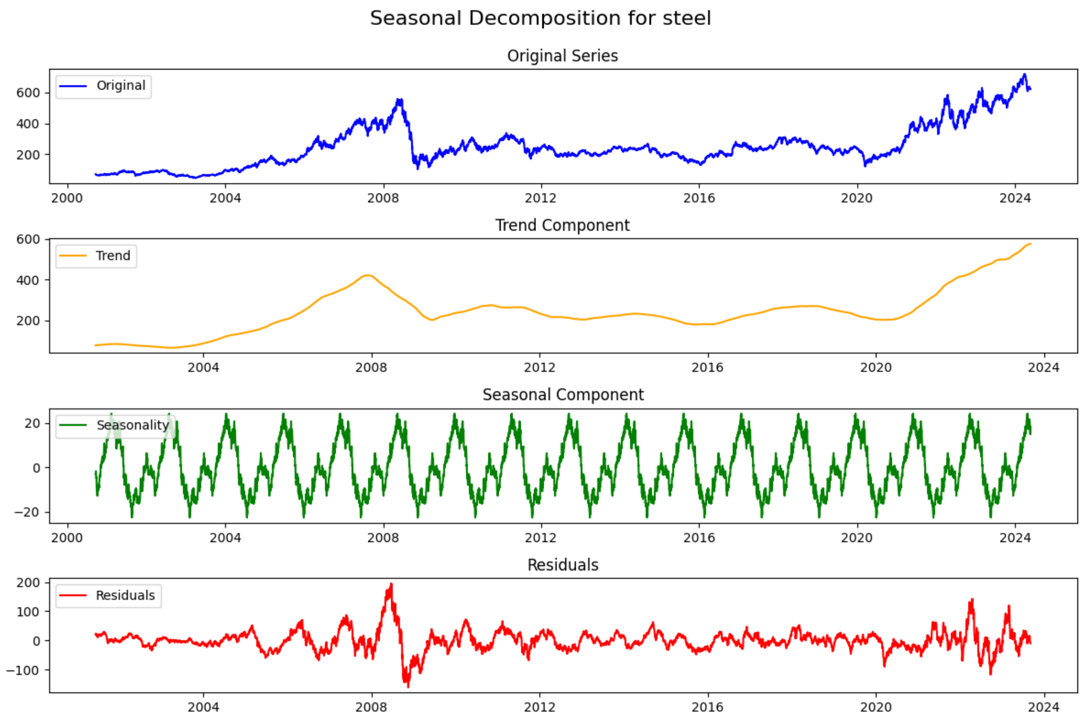
Figure 11. Seasonal decomposition for soybean



Source: Made by the author using Python software based on the data used in this study

Seasonal analysis of soybean prices also shows significant annual cycles. Compared to wheat and corn, soybean prices show a similar seasonal pattern, but with higher volatility. This suggests that soybeans are more sensitive to economic shocks.

Figure 12. Seasonal decomposition for steel



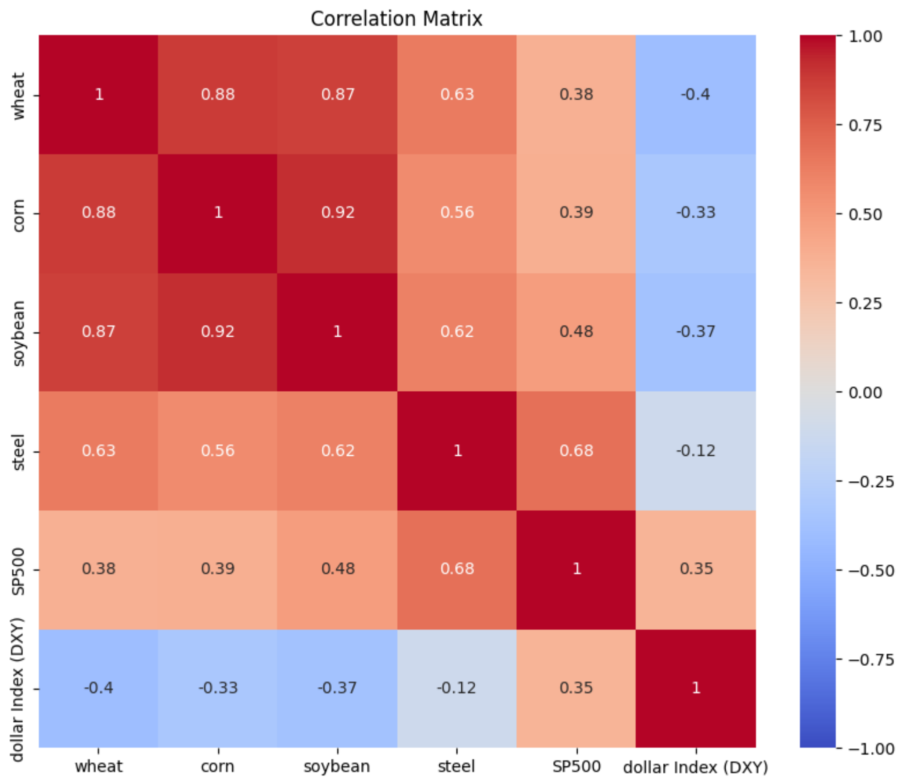
Source: Made by the author using Python software based on the data used in this study

Seasonal decomposition of steel prices shows fairly stable behavior in the long term. Unlike agricultural crops such as wheat, corn and soybeans, steel does not depend on the aspect of yield and weather conditions.

It is also worth noting that key steel exporters are fairly stable geopolitical players, which also affects the sustainable growth and stability of the steel market.

To check how exactly these products correlate with each other, It is necessary to place the data on a correlation matrix and supplement it with key financial and economic indicators in order to check how Ukrainian major export goods react to changes in these indicators.

Figure 13. Correlation matrix



Source: Made by the author using Python software based on the data used in this study

The correlation matrix shows the relationships between the studied commodities in the dataset.

As expected, grain crops show a high correlation with each other. Corn and soybeans (correlation 0.92), Wheat and corn (correlation 0.88), Wheat and soybeans (correlation 0.87). The price of these raw materials in most cases moves in the same vector. S&P 500 and steel, in turn, also show a significant correlation with a value of 0.68. The price of steel follows the growth of the economy and production, which is reflected by the S&P 500 index.

It is also worth noting that the Dollar Index (DXY) and grain crops wheat, corn, and soybeans have negative values -0.4, -0.33, and -0.37. Based on this fact, we can conclude that the strengthening of the dollar reduces the price of raw materials and demand accordingly.

CHAPTER 5. RESULTS

To assess the accuracy of the model's forecasts for prices of key Ukrainian export products, this study uses the following indicators:

- Mean Absolute Error (MAE) is a metric for measuring the average absolute difference between the actual value and the one predicted by the model.

y - true value for the i observation

x - predicted value for the i observation

n - number of observations

$$\left| \frac{\sum_{i=1}^{n} y_i - x_i}{n} \right|$$

- Root Mean Squared Error (RMSE) is a metric for measuring the square root of the average squared deviations between predicted and actual values.

y_i - true value for the i observation

x_i - predicted value for the i observation

n - number of observations

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}$$

- Coefficient of Determination (R^2) is a metric for measuring the proportion of the variance in the dependent variable is explained by the model.

y_i - true value for the i observation

$\hat{y}_i$ - predicted value for i observation

$\overline{y_i}$ - mean (average) of the actual values

n - number of observations

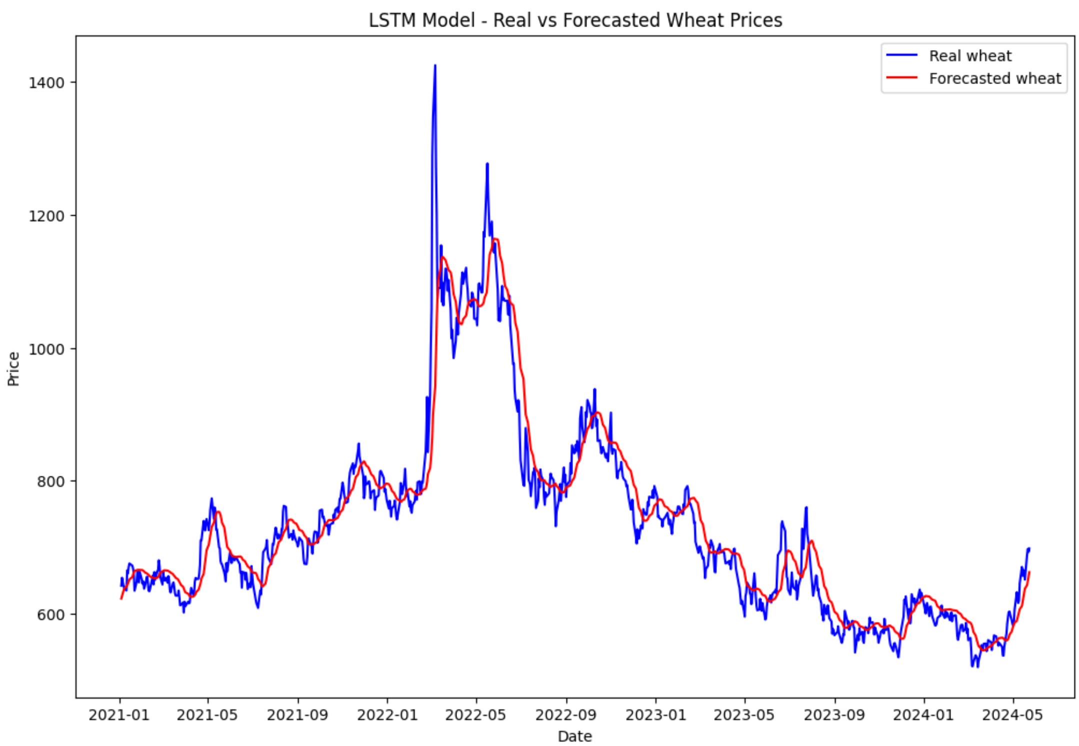
$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Table 2. Predicted Wheat Value Evaluation

Commodity	MAE	RMSE	R^2
Wheat	28.791490	46.772557	0.902379

Source: Model evaluation output

Figure 14. Real vs Forecasted Wheat Prices



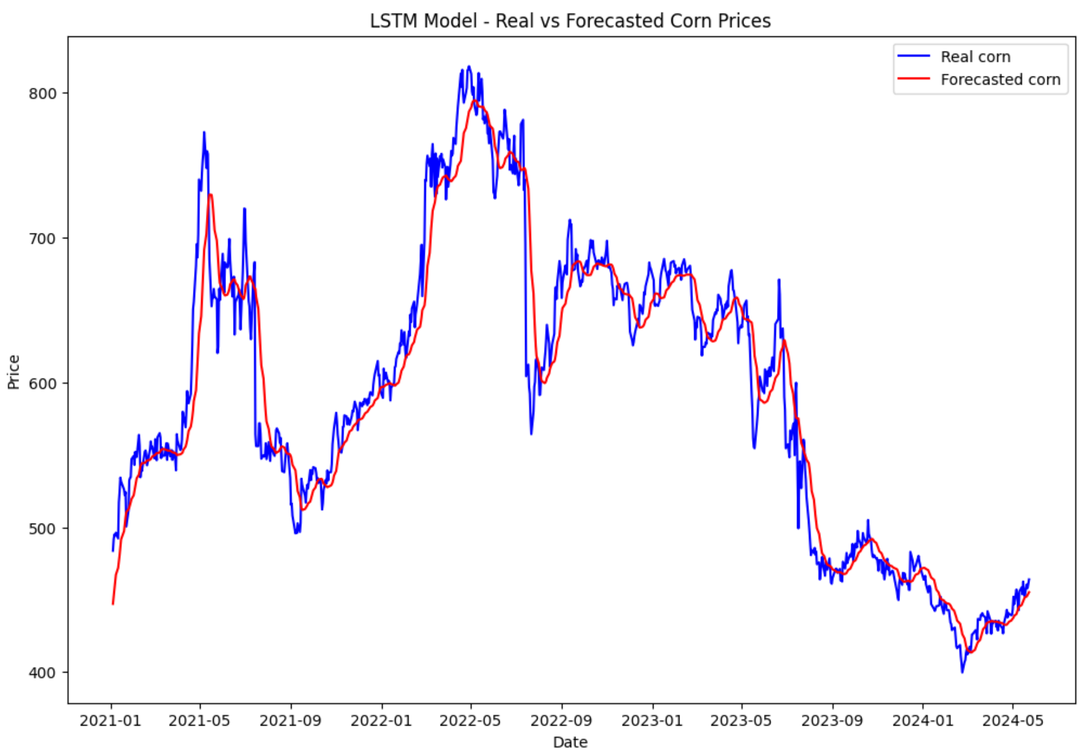
Source: Made by the author using Python software based on the forecast used in this study

Table 3. Predicted Corn Value Evaluation

Commodity	MAE	RMSE	R^2
Corn	17.820165	26.564439	0.932935

Source: Model evaluation output

Figure 15. Real vs Forecasted Corn Prices



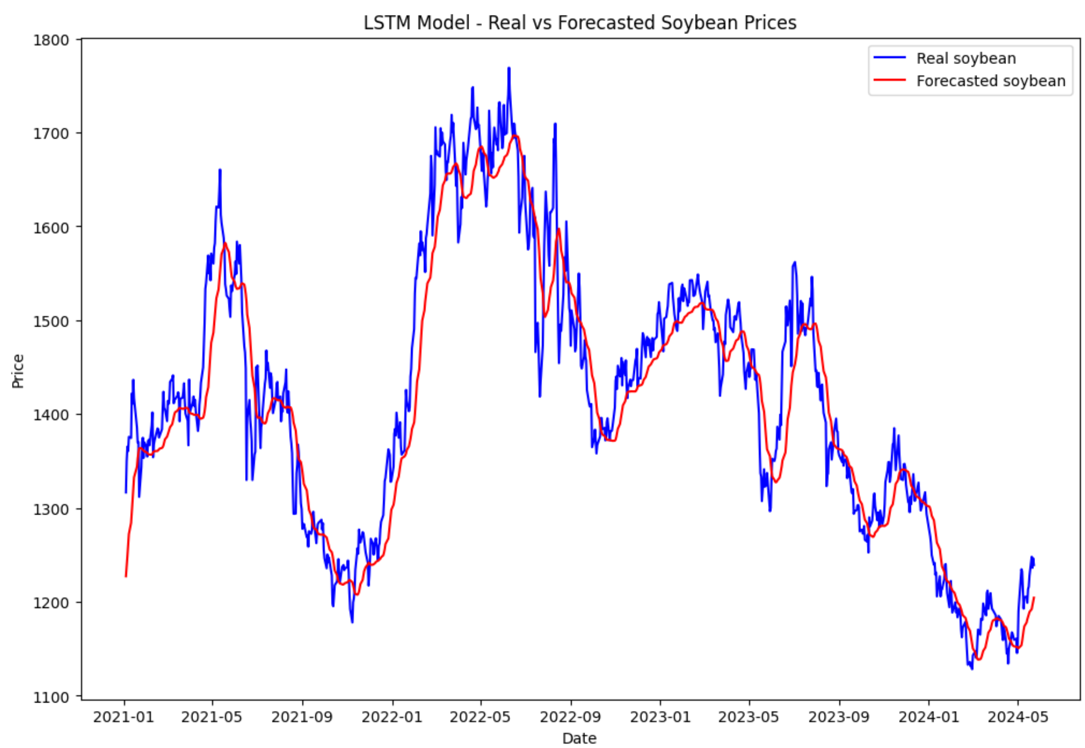
Source: Made by the author using Python software based on the forecast used in this study

Table 4. Predicted Soybean Value Evaluation

Commodity	MAE	RMSE	R^2
Soybean	36.270230	47.817555	0.895197

Source: Model evaluation output

Figure 16. Real vs Forecasted Soybean Price



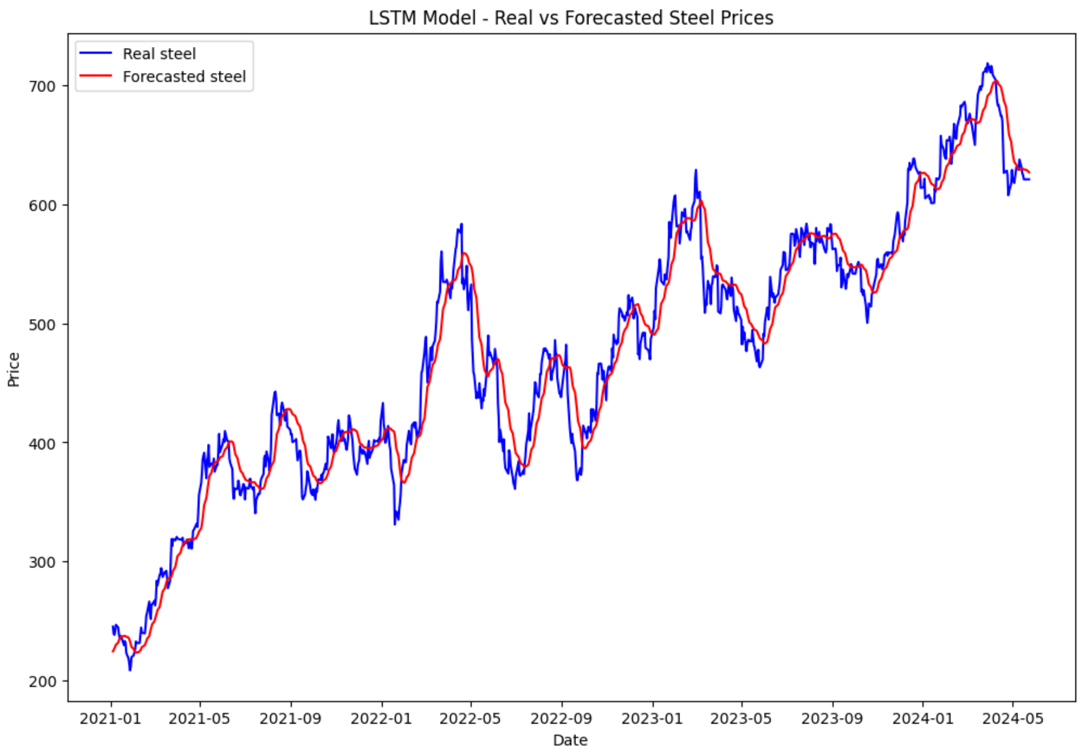
Source: Made by the author using Python software based on the forecast used in this study

Table 5. Predicted Steel value evaluation

Commodity	MAE	RMSE	R^2
Steel	18.626640	24.055938	0.954031

Source: Model evaluation output

Figure 17. Real vs Forecasted Steel Prices



Source: Made by the author using Python software based on the forecast used in this study

CHAPTER 6. CONCLUSIONS AND RECOMMENDATIONS

LSTM (Long Short-Term Memory) showed excellent results despite the large data array and the forecast duration of several years. Given the speed, ease of implementation and accuracy of forecasts, we can conclude that the use of machine learning to forecast prices for Ukrainian export products is an excellent tool that leading economists and analysts of Ukraine should use and scale in their work.

As the author of the study, I would like to draw attention to the lack of indices of Ukrainian products. Indices of the cost of the main products of Ukrainian exports will allow for a narrow-scale analysis that will help the Ukrainian government and exporting enterprises make more effective management decisions.

Investments in the field of data collection and the creation of appropriate online resources where financial indicators will be published will allow taking the analysis of Ukraine's export activities to another level.

Modern standards and technologies for data analysis and visualization should be an integral part of the priority for the development of Ukrainian businesses and financial institutions. It is also worth noting the need to develop an academic level of teaching business intelligence subjects.

REFERENCES

Arroyo-Marioli, F., Khadan, J., Ohnsorge, F., and Yamazaki, T. (2023). *Forecasting industrial commodity prices: Literature review and a model suite.* Policy Research Working Paper No. 10611. World Bank Group, Development Economics.

Brignoli, P. L., Varacca, A., Gardebroek, C., and Sckokai, P. (2024*). Machine learning to predict grains futures prices. Agricultural Economics,* The Journal of the International Association of Agricultural Economists.

Storm, H., Baylis, K., and Heckelei, T. (2020). *Machine learning in agricultural and applied economics. European Review of Agricultural Economics*, 47(3), 849–892.

World Bank. (2023). *Commodity Markets Outlook: Special Focus - Forecasting Industrial Commodity Prices*, April 2023. World Bank Open Knowledge Repository.

APPENDIX

Code used in the research

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn.preprocessing import MinMaxScaler

from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

from keras.models import Sequential

from keras.layers import LSTM, Dense

file_path = 'sample_data/Master_dataset.xlsx'

df = pd.read_excel(file_path)

columns = ['Date', 'wheat', 'corn', 'soybean', 'steel']

df = df[columns]

df['Date'] = pd.to_datetime(df['Date'])

df = df.sort_values(by='Date')
```

```
# Remove missing values and duplicates

df = df.dropna().drop_duplicates()

# Define the split date for training and testing sets

split_date = '2020-12-31'

# Data up to the end of 2020

train_data = df[df['Date'] <= split_date]

# Data from the beginning of 2021

test_data = df[df['Date'] > split_date]

# List of columns to forecast

model_columns = ['wheat', 'corn', 'soybean', 'steel']

# Initialize a dictionary to store predictions and metrics

predictions_dict = {}

metrics_dict = {}

for col in model_columns:

    print(f"Processing {col}...")

    # Normalizing data to a range from 0 to 1
```

```
scaler = MinMaxScaler(feature_range=(0, 1))

scaled_data = scaler.fit_transform(df[[col]])

# Create training datasets

train_col_data = scaled_data[:len(train_data)]

test_col_data = scaled_data[len(train_data):]

x_train = []

y_train = []

# For each element, a "window" of 60 previous values (x_train) is formed to predict the current value (y_train)

for i in range(60, len(train_col_data)):

    x_train.append(train_col_data[i-60:i, 0])

    y_train.append(train_col_data[i, 0])

x_train, y_train = np.array(x_train), np.array(y_train)

# The x_train data is converted into a three-dimensional array (samples, time steps, features) for LSTM compatibility. This adds a "features" dimension to work with a univariate time series.

x_train = np.reshape(x_train, (x_train.shape[0], x_train.shape[1], 1))

# Creating and setting up an LSTM model

model = Sequential()
```

```
model.add(LSTM(units=50, return_sequences=True, input_shape=(x_train.shape[1], 1)))

model.add(LSTM(units=50, return_sequences=False))

model.add(Dense(units=25))

model.add(Dense(units=1))

model.compile(optimizer='adam', loss='mean_squared_error')

# Train the model

model.fit(x_train, y_train, batch_size=32, epochs=1)

# Test the model

test_inputs = df[col][len(df) - len(test_col_data) - 60:].values

test_inputs = test_inputs.reshape(-1, 1)

test_inputs = scaler.transform(test_inputs)

x_test = []

for i in range(60, len(test_inputs)):

    x_test.append(test_inputs[i-60:i, 0])

x_test = np.array(x_test)

x_test = np.reshape(x_test, (x_test.shape[0], x_test.shape[1], 1))
```

```
# Predicting the test set results

predictions = model.predict(x_test)

predictions = scaler.inverse_transform(predictions)

# Store the predictions in the dictionary

predictions_dict[col] = predictions

# Calculate metrics

mae = mean_absolute_error(test_data[col], predictions)

rmse = np.sqrt(mean_squared_error(test_data[col], predictions))

r2 = r2_score(test_data[col], predictions)

metrics_dict[col] = {'MAE': mae, 'RMSE': rmse, 'R2': r2}

print(f"{col} - MAE: {mae}, RMSE: {rmse}, R2: {r2}")

# Plot the results for each commodity

plt.figure(figsize=(12, 8))

plt.plot(df['Date'][len(train_data):], df[col][len(train_data):], label=f'Real {col}', color='blue')

plt.plot(df['Date'][len(train_data):], predictions, label=f'Forecasted {col}', color='red')

plt.title(f'LSTM Model - Real vs Forecasted {col.capitalize()} Prices')

plt.xlabel('Date')
```

```
    plt.ylabel('Price')

    plt.legend()

    plt.show()

metrics_df = pd.DataFrame(metrics_dict).transpose()

print("\nModel Evaluation Metrics:")

print(metrics_df)
```