Price Prediction Model for Agricultural Products

Submitted by
Hafiz Zain Waheed
22i-2862

Supervised by
Dr. Rehan Aftab

Master of Science(Business Analytics)

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science(Business Analytics) at National University of Computer & Emerging Sciences

FAST School of Management
National University of Computer & Emerging Sciences
Islamabad, Pakistan. 2024

Price Prediction Model for Agricultural Products

Submitted by
Hafiz Zain Waheed
22i-2862

Supervised by
Dr. Rehan Aftab

Master of Science(Business Analytics)

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science(Business Analytics) at National University of Computer & Emerging Sciences

FAST School of Management
National University of Computer & Emerging Sciences
Islamabad, Pakistan. 2024

Author’s Declaration

I, the undersigned Mr. Hafiz Zain Waheed hereby declare that I am the sole author of this thesis. To the best of my knowledge this thesis contains no material previously published by any other person except where due acknowledgement has been made. I have not engaged in any practices that would breach academic integrity or ethical standards in the research process.

Date:

Name:

Signature: ………………………………………………………

Plagiarism Undertaking

I, the undersigned Mr. Hafiz Zain Waheed take full responsibility of the ethical work conducted for my thesis "Price prediction model for Agriculture Product". I solemnly declare that the work presented in the thesis is solely done by me. Moreover, I have not presented this thesis (or similar research work) to any other degree awarding institution within Pakistan or abroad.

I understand that National University of Computers and Emerging Sciences has zero tolerance policy towards plagiarism. Therefore, I as the author, of this thesis solemnly declare that no portion has been plagiarized and any material used in the thesis from other sources has been properly referenced.

Date:

Name:

Signature: ………………………………………………………

Abstract

It is important to combine machine learning and data analytics in order to improve models that predict farm prices because these markets are very complicated and changeable. This thesis looks at how advanced machine learning methods can be used to make pricing of agricultural products more accurate which is important for keeping market stable and planning economy. To represent complex nature of agricultural economics it stresses importance of using a lot of different variables in the prediction models. These variables include economic indicators, climate conditions and historical prices among others. The study shows that ensemble learning models especially Random Forest and Gradient Boosting do much better job than traditional statistical methods. This is because they can handle big, busy datasets and pick up on complex nonlinear relationships. The results show that machine learning has potential to give useful information that can help policymakers, farmers and traders make better choices. This can make the agricultural industry more profitable and long lasting. This work not only adds to what is known in academic world but it also solves problems in real world, pushing for machine learning to be used more in agricultural economics.

Keywords: Agricultural pricing, Machine learning, Data analytics, Random Forest, Gradient Boosting, Economic indicators, Climate impact, Market trends, Ensemble learning models, Price prediction accuracy.

Table of Contents

List of Figures

List of Tables

CHAPTER 1: Introduction.

1.1. Background

Agriculture is the backbone of many economies for centuries. Recently this sector has faced many challenges and there is continuous increase in demand, this because of changes in technology demand around world and unpredictable problems caused by climate change. These new trends have made it clear that we need creative new ways to solve problems especially when it comes to predicting prices of farm goods. Accurate price projections not only helps keep market stable but it also makes sure that everyone from farmers to sellers can make smart decisions based on accurate information (Mhlanga, 2023).

Even though it's clear how important it is to have correct price prediction tools agriculture business has lot of problems in this area. This industry is naturally unstable because it can be affected by many different factors such as changes in seasons and outside events like strange climate patterns or quick changes in polices. Standard methods can be useful in some situations but they often fail to show how many complex factors that affect price interact with each other. Also because of new technologies and lots of large amounts of data it is very important to use this data correctly(Mhlanga, 2023).

The hard part is making model that not only uses this information but also knows and predicts how farming market will change. The changes in price could cause farmers to lose lot of money which would hurt their ability to make living. Traders and politicians both take big risks when they try to keep market stable and make sure there is enough food for everyone. In this case it is very important to solve problem by creating models that can correctly predict costs. This will protect agriculture sector long term economic security (Javaid, Haleem, Khan, & Suman, 2023).

Creating precise and dependable model for predicting prices of farm goods is very important in fast changing agricultural business. The agricultural industry is now under lot of pressure to come up with long term successful answers to growing need for food around world and problems caused by climate change. Precise price prediction models are not only technological achievement but they are also important way for everyone in supply chain—farmers lawmakers and merchants—to reduce risks and make better decisions (Grabs et al., 2024).

If you can accurately predict prices in sector where market is unstable and prices change with seasons and outside factors like bad weather or changes in government policy it could have huge effect. By using this information farmers may be able to better plan their farming efforts which could mean difference between making money and losing money. Precise models might give traders and lawmakers information they need to make smart choices like deciding on import export policies which would make market more stable and protect food supply(Mhlanga, 2023).

By giving partners accurate and up to date information price prediction models may improve working efficiency strengthen economy and eventually support long term security and sustainability of farming sector.

1.2 Overview of Agricultural Product Pricing

The economic ramifications of agricultural items pricing are substantial affecting not just global economy but also local communities and long term sustainability of food supply chain. Conversely vegetables and cotton are classified as cash crops whilst rice and wheat are categorised as staple crops. The global food system consists of essential elements that include diverse range of commodities with agricultural goods playing crucial role. Undoubtedly price of these goods significantly affects livelihoods of billions of people especially in agrarian countries where substantial portion of population depends on agriculture as their primary source of income (World Bank, 2021).

The pricing of agricultural goods is crucial factor that affects global economy as well as local communities and long term sustainability of food supply chain. Global food system is built upon agricultural products including wide range of commodities from basic crops such as wheat and rice to cash crops like vegetables and cotton. Price of some commodities especially in agriculture has

significant influence on lives of billions of people. In many part of world large portion of population depends on agriculture as their primary source of income (World Bank, 2021).
Agriculture business not only ensures food security but also significantly contributes to economic growth of several countries. As to Food and Agriculture Organisation (FAO) of United Nations agriculture accounts for around 30% of total global economic activity (FAO, 2020). Contribution of this sector is crucial in ensuring food security eradicating poverty and fostering economic growth, price of agricultural goods is subject to volatility due to several factors such as economic conditions global market trends and geopolitical influences (OECD, 2020).

1.3 Problem Statement

Traditional agricultural techniques must evolve to meet rapidly changing demands of world. This thesis aims to address inadequate use and integration of data analytics in agriculture. Particularly in developing countries where it has significant impact on agricultural operations.
Although conventional agricultural approaches have shown their effectiveness throughout time they are becoming more inadequate in light of complexities of modern world. Given United Nations projection that global population will reach 9.7 billion by 2050 it is crucial to not only enhance food production but also implement plan that ensures both sustainability and environmental preservation. In addition climate change introduces unforeseen factor that standard methods are ill prepared to address. Furthermore it not only disrupts traditional agricultural schedules but also leads to alterations in precipitation patterns thereby amplifying occurrence of intense weather phenomena (IPCC, 2021). Machine learning and data analytics are identified as viable solutions in this scenario. They possess capacity to bring about significant change in agriculture by enabling more precise forecasts of crop yields improving timing of planting and efficiently managing resources like water and fertilisers. As to research conducted by Zheng et al. in 2019 these technologies provide capability to offer farmers valuable information that may be effectively implemented. These insights may be derived from large datasets that include weather patterns soil conditions and market trends. Nevertheless integration of environmental technology in agriculture industry remains severely limited particularly in developing nations (Takahashi, Muraoka, & Otsuka, 2020).
The limited availability of resources and lack of experience among farmers especially in underdeveloped areas provide significant barrier to implementation of advanced data analytics in agricultural sector. This is especially relevant to underdeveloped regions. It is difficult for many farmers operating on small scale to use data driven agricultural practices because they lack essential skills and access to necessary equipment (IFAD, 2019). This divide not only limits potential for individual farmers to enhance their productivity but it also hinders broader goals of agricultural growth and food security that are being sought(Grabs et al., 2024).
Furthermore there is notable dearth of comprehensive research that integrates methodology applications and outcomes of using data analytics in agriculture industry. The current research mostly focuses on specific aspects of this integration lacking broad viewpoint that considers range of issues opportunities and real world implications (FAO, 2020). Due to paucity of research that has been undertaken formation of coherent plans and policies that may support wider use of data analytics in agriculture is delayed (Takahashi, Muraoka, & Otsuka, 2020).
The problem is crucial owing to fact that agriculture is more than merely economic activity; it is also key component of both sustainability of environment and security of food supply around globe. The use of data analytics in agricultural operations has potential to deliver many positive results including enhanced crop productivity reduced resource inefficiency and improved resilience of farming systems to climate change impacts. To ensure equitable distribution of benefits of technological advancements it is crucial to tackle challenges related to integrating data analytics into agricultural operations particularly in poor nations (Mondejar et al., 2021).
The objective of this initiative is to examine and address deficiency in integration of data analytics into agricultural practices. To get deeper understanding of how data driven decisions might transform farming it is essential to explore challenges and opportunities that arise from merging of data analytics

and agricultural research. To ensure food security and sustainability among aforementioned global challenges aim is to provide pathway that facilitates integration of sustainable agricultural technology in manner that is accessible and beneficial to farmers worldwide(Mondejar et al., 2021).
To summarise integration of contemporary data analytics into agriculture is not only opportunity for technological advancement but it is also imperative for addressing pressing issues of guaranteeing food security and upholding environmental sustainability. By presenting full overview of current situation of data analytics in agriculture highlighting hurdles to its adoption and recommending solutions to bridge existing gaps objective of this study is to make meaningful contribution to this industry.

1.4 Research Questions

1. *What is impact of economic indicators on agricultural product?*
2. *How do historical price data influence future price trends in agriculture?*
3. *What are effects of climate change on agricultural pricing?*
4. *What are most appropriate machine learning models for predicting agricultural product prices given the influence of economic, historical, and environmental factors?*

1.5 Research Objectives

1. *To develop comprehensive pricing prediction model that utilizes machine learning techniques to integrate economic indicators, historical data and climate effects into price forecasts.*
2. *To evaluate performance of various machine learning models (e.g., Random Forest, SVM, Neural Networks) in predicting agricultural prices, considering their sensitivity to different data types and their predictive accuracy.*
3. *To analyze relationship between agricultural prices and climate variables using advanced machine learning methods, identifying key predictors and their impacts on price volatility.*
4. *To provide actionable insights and recommendations for stakeholders in the agricultural sector based on the outcomes of the machine learning models, enhancing decision-making processes related to crop management and market strategy.*

1.6 Gaps in Existing Research

The integration of advanced data analytics and machine learning into agriculture represents significant step forward in optimizing farming practices. However existing research in this area reveals notable gaps that need to be addressed to fully harness these technologies potential. These gaps span from limited focus on local market dynamics to sparse application of machine learning in agricultural contexts. This section outlines these gaps and demonstrates how proposed research will contribute to filling m.

1. Limited Focus on Local Market Dynamics: Existing research in agricultural data analytics often emphasizes global market trends overlooking intricacies and specific challenges of local agricultural markets (World Bank, 2021). This global perspective while valuable fails to capture unique factors influencing local markets such as regional climate conditions local farming practices and specific supply chain dynamics. As result models and predictions derived from these studies may not be directly applicable or fully beneficial to local farmers especially in developing countries.
This research aims to fill this gap by focusing explicitly on local market dynamics. It will explore distinct variables influencing local agricultural prices integrating regional data into predictive models. This localized approach will provide more accurate and actionable insights for farmers and stakeholders within specific regions herby enhancing relevance and utility of research findings (IFAD, 2019).

2. Neglect of Seasonal Variability: Another significant gap in existing research is underestimation of impact of seasonal variability on agricultural prices (FAO, 2020). Many existing studies and predictive models do not adequately incorporate complex patterns of seasonality which are crucial in agriculture. Seasonal changes can significantly affect crop yields pest infestations and market supply and demand all of which influence price fluctuations.

This Thesis will address this gap by developing models that explicitly account for seasonal variations. By incorporating seasonal factors research will provide more comprehensive understanding of agricultural price dynamics enhancing predictive accuracy and practical applicability of models especially for seasonal crop planning and marketing strategies.

3. Sparse Application of Machine Learning: While machine learning has demonstrated considerable promise in various fields its application in agricultural price prediction is still limited (Zheng et al., 2019). Most existing studies in agricultural economics utilize traditional statistical methods which may not fully capture complex nonlinear relationships inherent in agricultural data.

Proposed research will contribute significantly to this area by employing advanced machine learning techniques to agricultural price prediction. Our one purpose for this thesis is to track deeper insights and more accurate predictions by exploring innovative approaches to machine learning beyond capabilities of traditional methods, this will not only fill substantial need in current literature but also set standards for future research in this field.

4. The lack of comprehensive cross commodity Analysis Ongoing study in agricultural price prediction mostly focuses on specific commodities ignoring potential insights that should be derived from cross commodity analysis (OECD, 2020). Acquiring knowledge about connections and associations between different commodities may provide valuable insights into market dynamics including effects of substitution and factors that influence several markets.

1.7 Significance of Research

This thesis is not just for academic purpose but it will also fulfil practical need. It also extend to economic social and technological aspects which will benefit to all stake holders. Precise agricultural price projections are crucial for maintaining market stability and optimising resource allocation from economic perspective. Access to reliable forecasts enables farmers to make informed decisions about crop selection, investments in inputs and timing & targeted market for produce sales (Myers, 2017). This will enables them to optimise their revenues while concurrently reducing risk of losses. Traders and other stakeholders could rely on price forecasts to manage inventories optimise supply chain operations and mitigate market volatility(Grabs et al., 2024).

Accurate predictions are crucial for ensuring food security and reducing poverty from societal standpoint. Agriculture has crucial role in providing fuel for substantial portion of people in several developing countries with poor incomes. Implementing pricing structures that are both reliable and accurate has capacity to improve economic security of small scale farmers significant portion of whom are vulnerable to descending into poverty (Barrett, 2008). Furthermore consistent and fair pricing of agricultural goods ensures that individuals particularly those living in low income regions may get food that is reasonably priced and easily accessible, this in turn promotes both food security and enhanced nutrition.

Agricultural price projection methods have been significantly transformed due to technology advancements. Use of modern technology such as artificial intelligence machine learning and big data analytics facilitated creation of forecasting models that are more precise. As stated by Zheng et al. in 2019 these models have capability to assess big amounts of data from many sources including satellite imagery weather patterns market trends and consumer behaviour to provide precise and timely predictions of price. Such technology advancement not only enhances precision of predictions but it also make sure availability of this crucial data enabling even small scale farmers to benefit from knowledge that was previously exclusive to large agricultural corporations.

Policymakers also value advantages associated with precise agricultural price predictions. Accurate predictions will help in effective planning and execution of agricultural policies, including macro level aspects such as subsidy allocation policy, storage techniques, and import-export determinations made by policymakers. Using accurate may assist policymakers in efficiently implementing steps to tackle food shortages or surpluses. Ultimately this will help in preserving market equilibrium and reducing wasteful of resources (OECD, 2020).

Economic Importance

Stabilizing Markets:

Putting in place correct tools for predicting prices can make agri markets more stable. When everyone involved from farmers to buyers has access to accurate price predictions they can make better choices. Which lowers chance of market crashes or sudden price changes (Myers, 2017). Stable markets can also bring in investments which helps farmers grow.

Enhancing Profitability:

Accurate price predictions could have big effect on farmers ability to make money especially small farmers who often have small profit margins. Barrett (2008) says that knowing best times to grow and sell crops helps businesses run more efficiently which highers their profits.

Social Benefits

Food Security:

Making sure there is enough food for everyone requires accurate price prediction tools especially since world's population keeps growing. Headey and Fan (2010) say that these models could be used as guide by policymakers to make rules that make sure that important things are always available. This makes it possible to take proactive steps to avoid supply shortages.

Empowering Farmers:

Giving farmers accurate price predictions lets them make smart decisions about which crops to grow and when to cut crop which raises their standard of living and makes them better able to deal with problems (IFAD, 2019).

Technological Advancements

Promotion of Innovation:

Advanced machine learning methods and economic models are used in this study to encourage technology progress in agricultural sector. According to Zheng et al. (2019), the ways things are done now can lead to more study and a mindset of technological progress.

Data Driven Decision Making:

This study shows how datadriven decision making can change things in time when there is lot of "big data." The study of difficult data could change the way this sector work by giving farmers useful information and useful rewards (OECD, 2020).

Environmental Considerations

Sustainable Farming Practices:

The seventh issue to examine is implementation of sustainable farming practices while also considering environmental considerations. Precise price estimations possess capacity to indirectly enhance sustainable agricultural operations. Enhanced understanding of pricing empowers farmers to manage their operations leading to more efficient utilisation of resources and adoption of environmentally sustainable approaches (IPCC, 2021).

Mitigating Climate Change Impact

Addressing Consequences of Climate Change Having comprehensive understanding of impact that climate change will have on agriculture is crucial. The analysis of price movements may provide significant information for formulating strategies aimed at mitigating impacts of these and ensuring industry long term viability (World Bank, 2021).

Global Implications

Informed Global Trade:

Knowledgeable international trade: To produce well informed assessments on global trade it is crucial to own efficient price prediction models. The World Trade Organisation (2019) offers support to nations in efficiently overseeing their imports and exports so ensuring stability of worldwide supply of products.

Promotion of Global Collaboration:

Agriculture is naturally linked to capitalism. This work could make it easier for people around the world to work together and improve models that predict prices which would be good for everyone (FAO, 2020).

1.8 Chapter Summary

In first chapter of thesis, the past and current importance of agricultural sector in keeping economies going and meeting needs of people is carefully looked at. This thesis looks at how industry has changed because of big changes in technology, rising food prices, and effects of climate change. In long run this could lead to more production and more stable market. Even though it has lot of potential, a lot of study has shown that it is hard for many people to use cutting edge technology. This is especially true in developing countries where farms don't have the money or skills to fully utilise these tools. This gap shows that there are differences in how technology is used and how it affects progress of crops and food security.

Furthermore introduction offers critical assessment of present condition of field by emphasising lack of comprehensive research that combines different approaches and findings about data analytics in agriculture. The lack of this specific literary work highlights need for deliberate study efforts aimed at providing comprehensive and multifaceted understanding of data analytics in agriculture sector.

The last portion of chapter examines diverse significance of research specifically highlighting extensive range of economic social and technological benefits associated with advancements in price prediction. Research has potential to enhance economic resilience. It would be advantageous to improve societal well being stimulate technology advancement and impact policy development by offering comprehensive grasp of price prediction dynamics. Several crucial outcomes are necessary to attain continuous growth in agricultural business.

CHAPTER 2: Literature Review

2. Literature Review

This literature review explores integration of data analytics and agriculture focusing on machine learning applications. It provides comprehensive understanding of role of data analytics in addressing long standing agricultural challenges and explores current research on its applications challenges and prospects.

2.1 Machine Learning and Its Emerging Role in Agriculture

Machine Learning is enhancing supply chain from farm to market which is crucial for Islamabad where inefficiencies can lead to increased costs and waste. Predictive analytics help in managing logistics predicting market demand and setting prices that reflect real time conditions. By doing so ML contributes to reducing gap between farm prices and market prices ensuring fair compensation for farmers and stable prices for consumers (Raghuwanshi & Jacob, 2021).

Despite potential benefits there are challenges in adopting ML in agriculture particularly in Islamabad. Primary issue is data availability. Many ML models require large datasets to train effectively and such datasets may not exist for all crops or regions especially in developing areas (Moe et al., 2021). Additionally farmers may lack technical expertise required to implement ML solutions. There is need for educational initiatives to train farmers and agronomists in data literacy and ML tool usage.

Another challenge is related to infrastructure. For ML applications to function effectively they require reliable internet connectivity and access to technology which can be scarce in rural areas surrounding Islamabad. Investment in infrastructure is thus crucial for full realization of ML potential in agriculture (Simonyan & Vinyals, 2019).

To address these challenges and make most of ML in agriculture stakeholders including government bodies educational institutions and private enterprises must collaborate. Policies that encourage adoption of smart farming practices investments in infrastructure development and farmer training programs are essential. For instance government can subsidize cost of IoT devices or provide tax incentives for agribusinesses that adopt ML driven technologies (Fountas et al., 2020).

Looking ahead integration of ML with or emerging technologies like Internet of Things (IoT) and blockchain can further revolutionize agriculture in Islamabad. IoT devices can collect real time data which ML algorithms can analyze to provide actionable insights. Blockchain can ensure traceability and transparency of agricultural products from farm to market building trust in supply chain (Weersink et al., 2018).

2.2 Applications of Machine Learning in Agriculture

Data analytics and machine learning (ML) are changing ways farming is done because they can help people make better decisions and increase output. These technologies have been used successfully in many areas of agriculture such as systems that find crop diseases, predict crop yields and suggest crops to grow. Parts of India and Bangladesh have markets that are lot like Pakistan. In these places agriculture faces problems like not having enough water or workers and it needs to use sustainable practices. This is where ML uses are especially useful (Khan et al., 2021).

Machine learning models especially convolutional neural networks (CNNs) have shown a lot of promise in finding and identifying crop illnesses. In a study by Johansen et al. (2019) shots taken with smartphones were used to scan CNN to look for signs of disease in crops. This is really helpful for farmers who live in out of the way places. Singh et al. (2020) used machine learning (ML) to find wheat rust sickness in India. This is similar to what was done in Pakistan. Alerts can be sent to farmers early on by these tools so they can take quick management steps that can stop production losses (Khan et al., 2021).

Making predictions about crop outputs is very important job for machine learning algorithms. This helps make sure that everyone has enough food and speeds up supply chain. ML models were used by Gopal and Varshney (2020) in India to guess how much rice would be grown next year by looking at past crop data and pictures from space. Pakistan, where rice is the main crop, can really use the study results. Better estimates of output would help farmers plan and use their resources better which would be good for them.

Machine learning based guidance systems might be able to tell farmers what crops to grow by looking at things like market, quality of the land and how well the crops grow. In Bangladesh in 2021, Ahmed et al. set up an ML system that offers crops based on past weather data and signs of good soil. This led to better use of resources and better choices for crops. People who own small plots of land are given special advice which is important in places like Pakistan where small scale farming is popular. In addition to making farming practices more productive and efficient new technologies have helped move to precision agriculture which is especially helpful in places with few resources. Through precision agriculture farmers can use just right amount of inputs to cut down on waste and damage to environment (Kamilaris et al., 2017).

But there are some issues with growing with ML. Big problem is amount and quality of material that is out there. There is lost data or bad organisation in a lot of files used in agriculture which could affect how well machine learning models work. It is worse in Pakistan because there isn't enough internet infrastructure in the farm areas. However these technologies have the potential to totally change agriculture. As research and facilities keep getting better. Machine learning (ML) could help the agriculture industry become more durable, effective and long lasting (Khan et al., 2021).

A lot has changed in agriculture because of data analytics and machine learning (ML). These technologies can now find diseases, predict crop yield, and come up with individual crop ideas. There is more food and less waste now that there are new tools for farming especially in Pakistan. There are still issues with the amount and accessibility of data. In rural Pakistan, there isn't enough internet infrastructure to make it easy to store and look at data. Strong formulas better data collection methods and giving farmers tools and training they need to solve their problems will only work if everyone works together. For facilities to be built, governments and farming groups must work together. Examples of this cooperation include creating platforms for data sharing and analysis and improving internet connectivity in rural areas. Future research could integrate ML with emerging technologies like IoT and blockchain to create more sophisticated agricultural management systems. Despite challenges potential for ML to transform agriculture is immense and with continued investment in research and infrastructure it can help create more sustainable productive and resilient agricultural sector.

2.3 Importance of Local Market Dynamics in Agricultural Price Prediction

Understanding local market dynamics is paramount in agricultural price prediction yet literature often focuses predominantly on global trends. This emphasis can eclipse subtleties of local markets which are critical in forming tailored and effective predictive models particularly for places with unique agricultural landscapes like Islamabad. This literature review critically examines existing studies and emphasizes necessity of incorporating local market dynamics into agricultural price prediction models (Mondejar et al., 2021).

Agricultural economics literature often focuses on global market trends but these models often overlook intricacies of local markets like those in Islamabad. Local supply chains consumer behavior and market regulations are crucial for accurate price prediction. Inefficiencies in local supply chains can lead to significant price differentials which global models cannot capture. According to research we need more localised and context specific models because local markets have their own features that need custom analysis methods.

To get around these problems future study should use new data sources like satellite images and mobile transaction records to learn more about how the local market works (Torero, 2021). Participatory methods that include local stakeholders in model making process can also provide useful background information that improves the accuracy of the model (FAO, 2017).

To sum up global trends can help you understand farm economics in broad sense but they can not replace the need for more specific knowledge about how markets work in your area. For the Islamabad market and others like it, it is very important to include local supply lines, market rules and buyer behaviour in models that predict agricultural prices. This integration will ensure that models are not only academically robust but also practically relevant providing stakeholders with tools need to navigate complexities of local markets effectively(Mondejar et al., 2021).

2.4 Seasonality and Its Impact on Agricultural Economics

Seasonality profoundly impacts agricultural economics influencing crop production cycles market supply and pricing mechanisms. Literature on agricultural economics has begun to address seasonality in price prediction yet there is still need for deeper examination especially in markets like Islamabad where seasonal variations have pronounced effects on agricultural product sales and pricing. This literature review will analyze how current research treats seasonality in price prediction and its relevance to Pakistan market.

Seasonal variables can have big impact on agricultural pricing in local markets. These include weather variations harvest seasons and holiday cycles. Time series models have been applied to agricultural price forecasting; however they may not fully capture abrupt seasonal anomalies and may require large amounts of past data. Although literature frequently falls short of addressing entire complexity of seasonality in agriculture machine learning algorithms opened up new possibilities for incorporating seasonality into agricultural pricing projections. Seasonal cycles that influence market prices and agricultural production costs in Islamabad include labour migration and monsoon season. Furthermore seasonal surges in demand for specific agricultural products might result from cultural and religious practises. Future studies should concentrate on creating models that incorporate wider range of seasonal factors such as indirect effects and patterns of consumer behaviour in order to close these gaps. More timely and thorough data on seasonal factors influencing agricultural production and prices may be available through improved data collection techniques and usage of technology like remote sensing and Internet of Things devices (Mendelsohn, 2014).

2.5 Management in Agriculture

Data management plays critical role in modern agriculture sector impacting everything from policy decisions to farm management decisions. But it poses serious difficulties particularly in markets like as Islamabad. Inaccurate information can result in financial losses and decline in stakeholder trust.

Small scale farming in Islamabad frequently leads to informal and non standard data gathering which causes gaps and inconsistencies in quality of data.

It is still difficult to obtain timely and pertinent agricultural statistics in many underdeveloped nations including Islamabad. Scalable solutions are necessary since infrastructure cannot handle growing volume of data. Unstructured nature of lot of agricultural data including market pricing meteorological information and photos from remote sensing presents another challenge for data management systems. A major issue lies in integrating these many data kinds into cohesive system (Ali et al., 2019).

Creating centralised data repositories utilising big data and cloud computing technologies and investing in infrastructure and expertise to properly utilise artificial intelligence and machine learning are some strategies for enhancing data management in agriculture. In order to preserve data quality and make it accessible to those who require it data governance is crucial. For handling of agricultural data in Islamabad to be secure and efficient data governance frameworks must be put into place.

Finally it critical to develop ability of farmers and agricultural enterprises. By learning how to handle and analyse data well stakeholders can make better use of the data they have access to, which can lead to better decisions and higher productivity. Such training could make sector handling data a lot better (Javaid, Haleem, Khan, & Suman, 2023).

2.6 Cross Commodity Analysis and It's Significance

Cross commodity analysis is essential facet of agricultural economics allowing for understanding of how different agricultural products influence one another pricing and supply. Despite its importance existing research often overlooks intricacies of cross commodity analysis particularly in regions with diverse agricultural outputs like Islamabad. This literature review will highlight gap in cross commodity analysis in current research and propose its significance for obtaining comprehensive market insights.

Interrelationship between different commodities is critical aspect of agricultural markets. Price changes in one commodity can have direct or indirect effect on others due to substitution effects or shared input costs. For example increase in price of maize due to biofuel demands can influence market for or grains used in animal feed (Carter et al., 2018). Yet literature often treats commodities in isolation ignoring potential ripple effects through agricultural market.

In context of Islamabad city with varied agricultural landscape absence of cross commodity analysis can lead to narrow understanding of market dynamics. For instance Islamabad market experiences inter commodity impacts when fluctuations in dairy sector affect fodder crops market (Ahmed & Schlenkhoff, 2019). Without cross commodity perspective predictions and policies may fail to account for these interdependencies leading to suboptimal decision making.

Effective risk management in agriculture particularly in developing nations like Islamabad depends on cross commodity analysis. Neverless because it is difficult to create models that manage several commodities at once it is still understudied. Research conducted on South Asian markets by Akter et al. (2017) indicates that better data exchange and gathering procedures are required to support cross commodity research. Because of data scarcity and informality in market econometric models are less common in developing nations. Another issue is lack of data on how different goods connect with each other. Even though it has some problems cross commodity analysis has many benefits such as the ability to better predict how market will change and make better price plans. Improvements in machine learning and data analytics have made it possible to make more accurate market predictions which opens up more opportunities for cross-commodity research. To find a balance between economic growth and food security lawmakers can make better farming policies by using what they learn from cross commodity analysis. Khan et al. (2020) say that to get most out of cross commodity analysis money needs to be put into installing advanced data infrastructure using advanced analytical methods, and encouraging scholars, politicians, and stakeholders to work together (Javaid, Haleem, Khan, & Suman, 2023).

2.7 Conceptual Framework

It is hard to come up with a framework for predicting farming prices because you have to look into a lot of factors that affect this industry especially in Pakistan market. The literature review provides a tailored conceptual framework that considers past prices, the economy and prices of related commodities all of which have big effect on prices of farm goods.

Temperature plays a huge role in how much crops cost. Rainfall, weather and humidity all have effect on food output and quality which in turn has effect on market prices. In Pakistan research by Abbas et al. (2019) has shown that changes in certain climate factors can cause big changes in farming output which in turn has effect on costs. Effects of changing weather are especially strong in Islamabad where farming counts on monsoon rains a lot. It has also been found that temperature and humidity affect timing of food growth and number of pests that attack which in turn affects supply and market costs (Kumar & Sharma, 2018).

Prices in the Past Predicting future farming prices depends a lot on looking at past prices. Looking at past price cycles and trends can help you figure out how market will act in future. In 2007 Hussain and Mudasser did research that showed looking at old price data could reveal trends that could be used to make predictions. A very important factor is also how desire for options changes when prices of related goods change. Jan et al.'s (2019) study shows that knowing relationships between prices is

very important for predicting prices accurately in places like Islamabad where people may switch between different types of grains based on price.
Economic Indicators Economic indicators such as PKR Dollar exchange rate daily fuel prices and Economic Policy Uncertainty (EPU) index significantly impact agricultural prices. Exchange rate influences import and export costs affecting input prices for farmers and competitiveness of agricultural products in international markets. A study by Ali and Abdullah (2020) highlighted sensitivity of agricultural markets in Islamabad to currency fluctuations. Fuel prices as discussed by Khan et al. (2016) directly impact transportation and production coststherby influencing prices of agricultural commodities. EPU index which reflects economic policy uncertainty can affect investment decisions in agricultural sector with implications for production and pricing (Baker et al., 2016).
Related Commodity Prices Prices of related commodities are interlinked in various ways. A rise in price of one commodity can lead to increased demand for substitute influencing its price. In context of Islamabad interrelationship between staple food crops and cash crops can have significant implications for market dynamics. For instance increase in price of wheat may drive demand for rice affecting its market price. Study by Ahmed et al. (2017) on South Asian markets provides evidence of such inter commodity relationships.
Conceptual Framework Drawing from these variables conceptual framework for agricultural price prediction in Islamabad market can be formulated. Framework should consider multifaceted nature of agricultural pricing where each factor climate historical prices economic indicators and related commodity prices interacts and influences others.
Climate: Incorporate climatic data (rainfall temperature humidity) into model to predict its impact on crop yields and quality subsequently affecting supply and prices.
Historical Prices: Utilize historical price data to identify trends and cycles that could inform future price movements. Analyze how shifts in demand for alternatives and related commodity prices have historically impacted market prices.
Economic Indicators: Integrate daily PKR Dollar exchange rate fuel prices and EPU index to understand their influence on production costs import/export competitiveness and investment in agriculture.
Related Commodity Prices: Assess interrelationships between different agricultural commodities to predict how changes in price of one can affect others.
Framework for agricultural price prediction in Islamabad market uses advanced statistical and machine learning models to analyze data and make predictions. It requires comprehensive data collection strategy collaboration between government agencies research institutions and private sector stakeholders. This multifaceted approach considers climate historical prices economic indicators and commodity prices providing valuable insights for better planning and decision making.

2.8 Relevant Theories

Here are three theories that are particularly relevant:

1. **Supply and Demand Theory**
2. **Market Efficiency Hypothesis**
3. **Behavioral Economics Theory**

1. Supply and Demand Theory

A key idea in economics supply and demand theory describes how amount and price of commodities sold in marketplaces are decided. Essentially theory says that amount supplied by producers at current price will equal quantity desired by consumers at some point and at that time price of commodity will settle creating economic equilibrium for both price and quantity. (Mankiw, 2014).
This hypothesis is important when it comes to predicting crop prices. Demand for agricultural products can be impacted by consumer tastes population growth and income levels whereas supply of agricultural products is largely determined by factors like farming practises climate and technology

improvements. Accurate price prediction in agriculture sector requires understanding of supply and demand dynamics. For example bad crop due to drought will reduce supply therefore if demand stays same prices will go up. (Samuelson & Nordhaus, 2009).

2. Market Efficiency Hypothesis

Eugene Fama postulated Market Efficiency Hypothesis in 1960s which holds that prices in financial markets always accurately reflect all available information (Fama, 1970). Although this hypothesis was first applied to financial markets its concepts can also be applied to agricultural markets. In accordance with this idea supply and demand dynamics meteorological conditions and economic indicators should all be reflected in present pricing of agricultural commodities.

According to this hypothesis it can be difficult to anticipate future prices only from previous price data because present prices already take history knowledge into account. However in real world markets may not always be perfectly efficient due to factors like information asymmetry where all market participants do not have access to same information and delay in information dissemination. This inefficiency can create opportunities for predictive modeling where additional non public information (like advanced climate predictions or deeper economic analysis) can be used to forecast future prices (Grossman & Stiglitz, 1980).

3. Behavioral Economics Theory

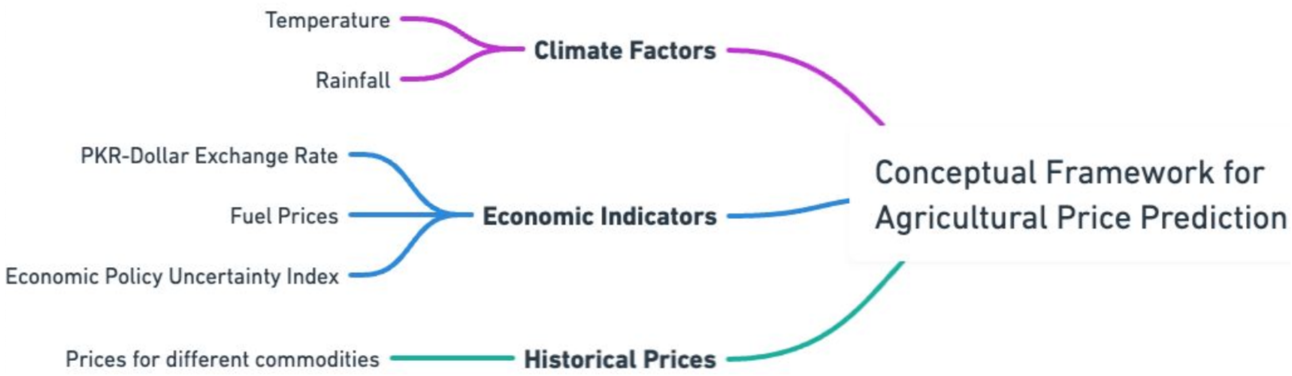
Behavioral Economics Theory combines elements from economics and psychology to understand how various cognitive social and emotional factors affect economic decisions of individuals and institutions including market prices (Kahneman & Tversky, 1979). This theory challenges traditional economic assumption that all market participants are rational actors who make decisions aimed solely at maximizing utility.

In agricultural markets behavioral economics can explain certain pricing phenomena that traditional models cannot. For example farmers decisions on what crops to plant and when to sell might be influenced by cognitive biases risk aversion and or psychological factors. Similarly consumers choices and willingness to pay for certain agricultural products can be influenced by factors like brand perception trends and or non rational considerations. Understanding certain factors can improve models used to predict prices, allowing for more accurate and improved predictions (Thaler, 2015).

In conclusion including these theories in literature review will give help for predicting farm prices a solid theoretical base. Market Efficiency Hypothesis helps us understand what information is being mirrored in present prices. Supply and Demand Theory helps us understand how markets work and the Behavioural Economics Theory helps us understand psychological factors that affect how markets behave. When put together these studies can help you understand how farm prices change a lot more.

2.9 Graphical Representation of Conceptual Framework

Figure 1: Conceptual Framework



When applied to agriculture markets the diagram conceptual structure gives us a way to make predictions. This method accurately predicts market trends by combining several important factors. This conceptual framework combines past price trends, economic factors, climate data and the ways in which the prices of different commodities are linked in order to make full prediction model. This kind of model can help people guess how agriculture market will change, which helps everyone involved make smart decisions and plan well for market that is always changing.

CHAPTER 3: Methodology

3. Introduction

The methodology part of thesis is very important because it explains how to do research including how to collect data, analyse it and figure out what it all means. It gives researchers plan for their study and helps them through difficult job of looking at all the different factors that affect prices of agricultural goods. Our Main goal of this thesis is to build a strong and accurate system for predicting agriculture prices with focus on Islamabad market.

A well structured approach makes sure that research is done in a way that follows set academic standards. It also makes it clear how the research questions will be answered which makes the research process more open. Researchers can do same study again in different settings or with different data sets if they have a thorough approach.

Model building is important part of process that describes how to make predictive models with variables and indicators. Model validation and testing methods are also outlined. Ethical considerations are addressed ensuring confidentiality and privacy of data collected especially if involving human subjects. Limitations of research methodology are acknowledged providing balanced and honest assessment of findings (Bryman, 2016).

Lastly proposed timeline is presented outlining schedule for each stage of research including milestones and deadlines. In summary methodology chapter is cornerstone of Thesis providing detailed roadmap for conducting research and contributing to development of comprehensive and effective framework for agricultural price prediction in Islamabad market.

3.1 Data Collection

Any research study must include data collection since it is essential to addressing research questions testing hypotheses and assessing results. Because of dynamic and complex nature of issue gathering data is both difficult and important when creating conceptual framework for agricultural price prediction. This section describes data sources data collection techniques and sampling plans that will be used in this investigation.

3.1.1 Sources of Data

Study will utilize secondary data sources to ensure comprehensive understanding of factors influencing agricultural prices.

Secondary Data Sources:

Well gather secondary data from already existing sources. This comprises past research papers economic reports historical price records and climate data records. Government agricultural departments economic surveys and research papers are among key sources of historical and contextual data required for this investigation. (Johnston, 2017).

3.2 Data Analysis

A research methodology data analysis part is essential since it describes how gathered data will be handled examined and interpreted. Machine learning models and or advanced analytical tools and techniques will be used in data analysis in order to establish conceptual framework for agricultural price prediction. This section describes analytical techniques data analysis software and tools and data pretreatment and analysis procedures.

3.2.1 Analytical Methods and Techniques

To properly analyse gathered data study will make use of variety of analytical approaches and strategies with emphasis on machine learning models.

1. **Machine Learning Models:** When analysing complicated datasets machine learning models are very useful since they can overcome limitations of standard statistical methods. Random Forest and or customised machine learning models will be employed in this investigation. Random Forest ensemble learning technique is renowned for its high accuracy resilience against overfitting and capacity to manage sizable datasets with increased dimensionality (Breiman, 2001). In order for it to function large number of decision trees are built during training and class that results is mode of classes of each individual tree. Because it can manage non linear correlations and interactions between numerous inputs like climate data economic indicators and historical prices this method is very good at predicting agricultural prices.
2. **Other Machine Learning Techniques:** Other machine learning methods such as Support Vector Machines (SVM) and Neural Networks may also be used in addition to Random Forest. SVM is well known for working well in high dimensional spaces and is helpful for problems involving regression and classification (Cortes & Vapnik, 1995). Deep learning models in particular which use neural networks are very effective in simulating intricate patterns and correlations in data. (LeCun Bengio & Hinton, 2015).

3.2.2 Steps for Data Preprocessing and Analysis

When working with very large and difficult datasets. Preparing data is an important part of the research process.

1. Cleaning the data: Cleaning the data is first step in getting it ready. It involves fixing inconsistent data, getting rid of duplicates and filling in missing values.
2. Transformation of Data: After that data will be changed into version that can be used for research. In order to meet goals of certain machine learning models this could mean changing factors, storing categorical variables and normalising data.
3. Feature Selection: This step is used to figure out which factors are most important to include in model. This step is necessary to lower the processing complexity and improve the speed of model.
4. Model Training and Testing: Following preprocessing of data training dataset is used to train machine learning models. After that different testing dataset will be used to assess models performance. Models will be evaluated using metrics including MSE, RMSE and R-square
5. Model Validation: In order to guarantee models resilience and generalizability they will lastly undergo validation. This could entail testing models on hypothetical data and using cross validation techniques.

Table 1: Sources of Data

Name of Feature	Source of Data for Feature	Nature of Original Data	Operationalisation	Time Duration for Data Collection
Climate	Meteorological Department	Measured in terms of rainfall (yes/no), temperature (°C),	Data will be used to model the impact on crop yields and quality. Seasonal patterns will be analyzed for long-term trends.	Daily data, collected annually
Historical Prices	Agricultural Market Comittee	Price data recorded at regular market intervals - Weekly or monthly	Historical price trends will be analyzed using time-series analysis to identify patterns and cycles.	Last 5 years
Economic Indicators	State Bank of Pakistan, Fuel Price Bulletins	Exchange rate (PKR/USD), daily fuel prices - Daily rates	The exchange rate and fuel prices will be included as variables in the price prediction model to assess their impact on costs.	Daily data, collected annually
Economic Policy Uncertainty (EPU) Index	Local Economic Journals, Research Institutes	Index score, measured monthly	The EPU index will be factored into the model to gauge the impact of policy uncertainty on investment and pricing in agriculture.	Monthly data, collected annually
Related Commodity Prices	Commodity Exchange Reports	Prices of various commodities, measured in PKR - Daily or weekly rates	Cross-commodity price elasticities will be calculated to understand the impact on substitute demand.	Daily or weekly, collected annually

1. **Climate**: This includes variables such as rainfall and temperature. We will measure temperature in C and rainfall is measured in binary form. Whether it was raining or a sunny day.
2. **Historical Prices**: Past price data which we on daily basis, is crucial for identifying price trends and market behavior.
3. **Economic Indicators**: This encompasses the exchange rate, daily fuel prices, and the Economic Policy Uncertainty (EPU) index. The exchange rate and fuel prices are critical as they affect the costs of inputs for farmers and the competitiveness of agricultural products. EPU index is measure of economic policy uncertainty which can influence investment decisions in the agricultural sector.
4. **Related Commodity Prices**: The prices of related commodities can be found in commodity exchange reports and are measured daily or weekly. These prices will be analyzed to

understand the interrelationship between different commodities and how a change in the price of one can affect the demand for others.

The time duration for data collection will generally be on a daily or weekly basis, with historical data potentially spanning the last 1 - 5 years to capture long-term trends and cycles.

3.3 Introduction to Random Forest

Random Forest is powerful learning technique widely used in ML & predictive modeling. This model belong to supervised learning category and is effective for both classification and regression models. Random Forest is ensemble of decision trees where multiple decision trees are built during training process and their predictions are aggregated to make more accurate predictions.

Random Forest offers several advantages:

1. **High Accuracy:** It typically provides high prediction accuracy because it combines multiple decision trees reducing risk of overfitting and increasing model stability.
2. **Robustness:** Random Forest is less sensitive to outliers and noisy data making it suitable for real world datasets that may have thei rregularities.
3. **Feature Importance:** It can rank importance of input features helping in feature selection and understanding most influential variables.
4. **Versatility:** Random Forest can be applied to both classification and regression problems

3.3.1 Feature Selection with Random Forest

Feature selection in Random Forest can be performed by examining feature importance scores assigned by model during or after training. Features with higher importance scores are considered more influential in making predictions.

One common approach is to set threshold for feature importance scores and retain only those features that exceed threshold. This process helps in simplifying model while retaining most relevant variables.

For example you can use Python scikit learn library to extract feature importance from trained Random Forest model and select top N features based on their scores.

3.4 Ethical Considerations

Research methods must take ethics very seriously particularly while collecting and analysing data for study. These factors guarantee that study abides by strictest ethics guidelines and respects all subjects and data sources. Gathering processing and use of data are main ethical issues in scope of creating conceptual model for agricultural price prediction. This section describes ethical concerns surrounding gathering and processing of data as well as steps taken to protect privacy and confidentiality.

3.4.1 Addressing Ethical Issues Related to Data Collection and Analysis

1. Informed Consent: Getting informed consent is one of most important ethical factors in data collecting particularly when working with human subjects (farmers market analysts etc.). It is important that participants are properly informed about purpose of study kind of data being gathered how it will be used and any possible hazards associated with taking part. In order to ensure that participation is voluntary and founded on educated decision consent should be sought in clear and intelligible manner (Resnik, 2015).

2. Honesty and Transparency: It critical to be honest and transparent in way data is gathered examined and presented. It is imperative for researchers to maintain accuracy in their methodology and findings refraining from any manipulation or misrepresentation of data. This entails being truthful about data and study limitations as well as refraining from exaggerating significance of conclusions (Israel & Hay, 2006).

3. Preventing Harm: Care should be taken to ensure that participants suffer as little harm as possible during design and execution of study. This covers both injury to body and emotional suffering. While

it is unlikely that participants in this study would suffer physical injury care should be taken to guarantee that any psychological discomfort or sufferingthey may experience during data collection and analysis (Sieber, 2006).

3.4.2 Ensuring Confidentiality and Privacy of Data

1. Data anonymization: Personal identifiers will be eliminated from dataset to safeguard participants privacy and prevent direct or indirect identification of people. When working with sensitive data this procedure known as data anonymization is essential (El Emam & Arbuckle, 2014).
2. Secure Data Storage: All gathered data will be kept in secure location to which only authorised staff members have access. Physical data will be preserved in safe locked locations and digital data will be stored in encrypted formats. This is to guard against data loss ft or unauthorised access (Flick, 2018).
3. Data Use Agreement: When using secondary data agreement for its use will be made with data suppliers. Use of data will be outlined in this agreement guaranteeing that it complies with ethical standards and original data collectors consent (Mannheimer Pienta Kirilova Elman & Wutich, 2019).
4. Confidentiality in Reporting: Every effort will be made to preserve confidentiality when disseminating results. This entails anonymizing data as well as exercising caution when adding any information that might subtly reveal participant identity (Grinyer, 2002).
5. Ethical Approval: To provide further assurance research proposal will be submitted for approval to Institutional Review Board (IRB) or equivalent ethics council toger with data collecting and analytic procedures. This stage guarantees that study complies with all ethical requirements and that any possible ethical concerns are resolved prior to study start (Tolich & Fitzgerald, 2006).

Study emphasizes importance of ethical considerations in research methodology ensuring integrity protection of participants and data sources. It aims to uphold highest ethical standards by implementing confidentiality and privacy measures.

3.5 Limitations

Research methodology for developing conceptual framework for agricultural price prediction may have several limitations. These include data availability and quality complexity of agricultural markets generalizability of findings subjectivity in qualitative analysis limitations of machine learning models and sampling bias. When there is not enough data model estimates and studies can be wrong. Which can make it hard to make decisions. When models are too complicated they can lead to oversimplified explanations of how markets work, which makes it more important to look at outside factors and expert views. Concerns about generalizability make it hard to use results in wider contexts. Subjectivity in qualitative research shows how important it is to use strict and clear analysis methods. Because machine learning models have flaws like overfitting and not being clear they need to be carefully chosen, validated and interpreted. Sampling bias can change how representative group is which can change the truth of the findings. To sum up study technique is strong and thorough but it does have some problems. It is important to be aware of and deal with research limits in order to keep it honest and useful.

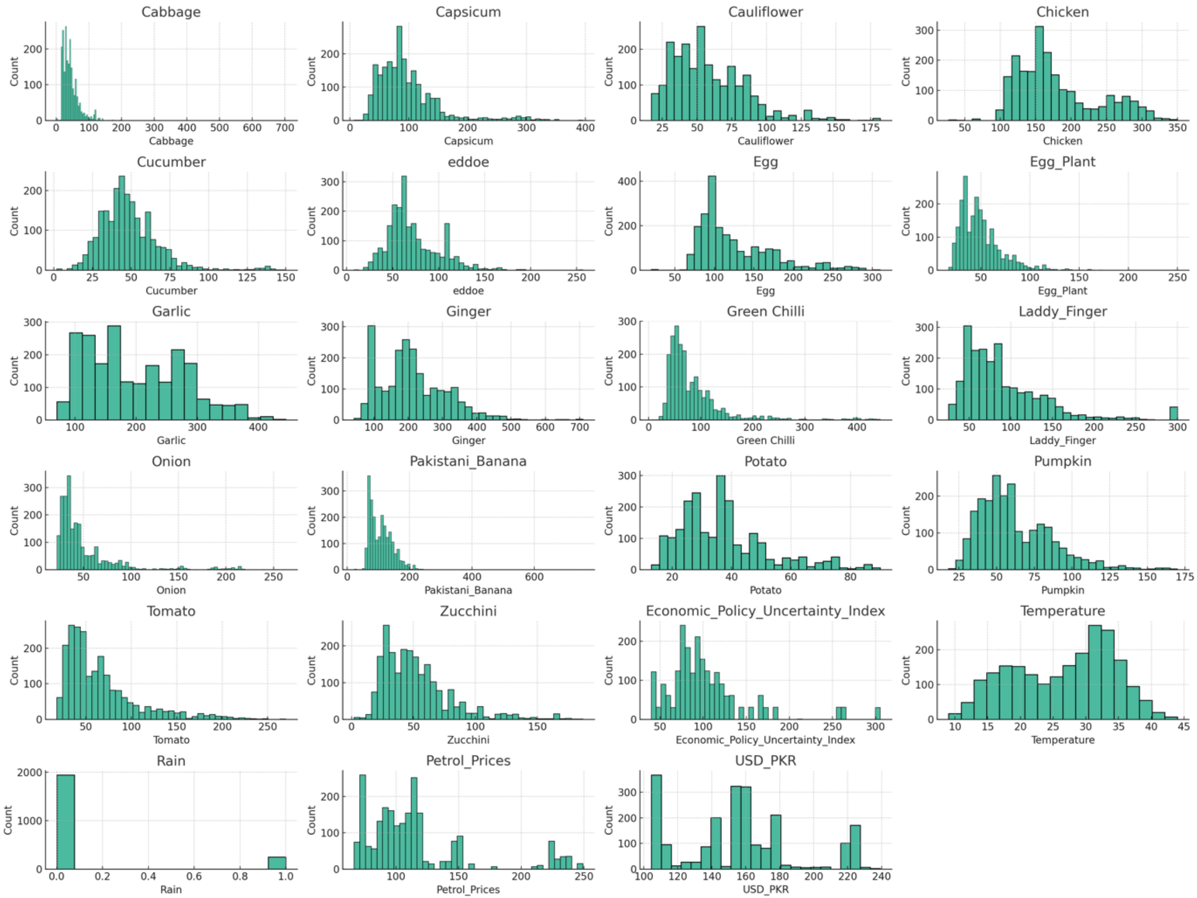
CHAPTER 4: Analysis and Results

4.1 Chapter Introduction

The thesis starts with careful list of main sources of data that will be used to support study. Climate variables, past price trends and economic indicators are used in study. Datasets that are strong and reliable came from trusted sources. The Meteorological Department's records of rainfall, temperature and humidity gave daily picture of important weather factors that affect crop yields.
The XG Boost and K Nearest Neighbors models helped theory make predictions and Support Vector Regressor showed its worth in non-linear setting. In new way, study used Linear Regression models with mood analysis to find out how consumers felt, which is key market price driver. At end of thesis, there is thorough look at findings, theoretical foundations and carefully examined empirical data. It helps us fully grasp factors that affect farm prices and provides users with tool for making predictions that is both accurate and insightful.

4.2 Exploratory Data Analysis (EDA)

Figure 2: Exploratory Data Analysis



Exploratory Data Analysis (EDA) is like map that helps researchers find their way through complicated numbers in their datasets. It is basic step by which patterns, oddities, connections and trends in data are found and carefully studied. EDA isn't just first step; it's also strategy exploration that shapes study's next steps in analytical process. It uses statistical images and other ways of showing data to figure out structure underneath and get useful information that is hard to see when data is just numbers. It is basically art and science of turning invisible into obvious, of turning data into stories and guesses into proof. Through EDA, thesis starts to piece together complicated web of changes in agricultural prices, paving way for more in-depth, data-driven findings.

4.2.1 Checking Null Values

Making sure that dataset is full is most important part of careful process of exploratory data analysis. Our dataset is detailed collection of prices for farm goods and environmental factors. It had lot of null values in many columns, mainly in "Capsicum," "Garlic," "Green Chilli," "Laddy_Finger," "Pakistani_Banana," "Potato," "Tomato," and "Zucchini." Null values, which are like holes in history record, can change real story of data, which can lead to biased conclusions and choices.
The first step in filling in these gaps was to choose method that took into account fact that our data is time series. Backward integration, method that was chosen, stands out because it can keep continuity in collection that changes over time. This method sends most recent known number backwards until it finds entry that is not null. It works especially well with our dataset because prices and environmental factors don't change over time, which protects order of data.
A planned method was used to put backward fill into action. At first, careful analysis of distribution of null values was carried out to make sure that their structure did not point to any systemic problems that could compromise data's integrity. After confirmation, backward fill process began. For example, if "Capsicum" price on Wednesday wasn't shown, it was filled in with price from Thursday, because it was thought that in stable economy, prices of farm goods on market don't change much from day to day.
There are clear benefits to this method. It keeps structure and distribution of information same and makes sure that no extra highs or lows are added to time series. Backward integration takes into account order of time and lets dataset's natural motion be used to carry forward past trends. The process was carefully written down so that it could be studied or used again in future. But choice to use backward integration wasn't made without thinking about what it would mean. A conscious effort was made to limit amount of null value replacement, knowing that relying too much on any imputation method could make dataset less unique and cause it to fit too well.
In conclusion, backward integration method we used to deal with null values in our dataset was choice that made sense given that data was time series. This showed dedication to rigorous analysis and honesty, setting stage for strong, non-biased statistical analysis in later stages of study.

4.2.2 Summary Statistics of dataset

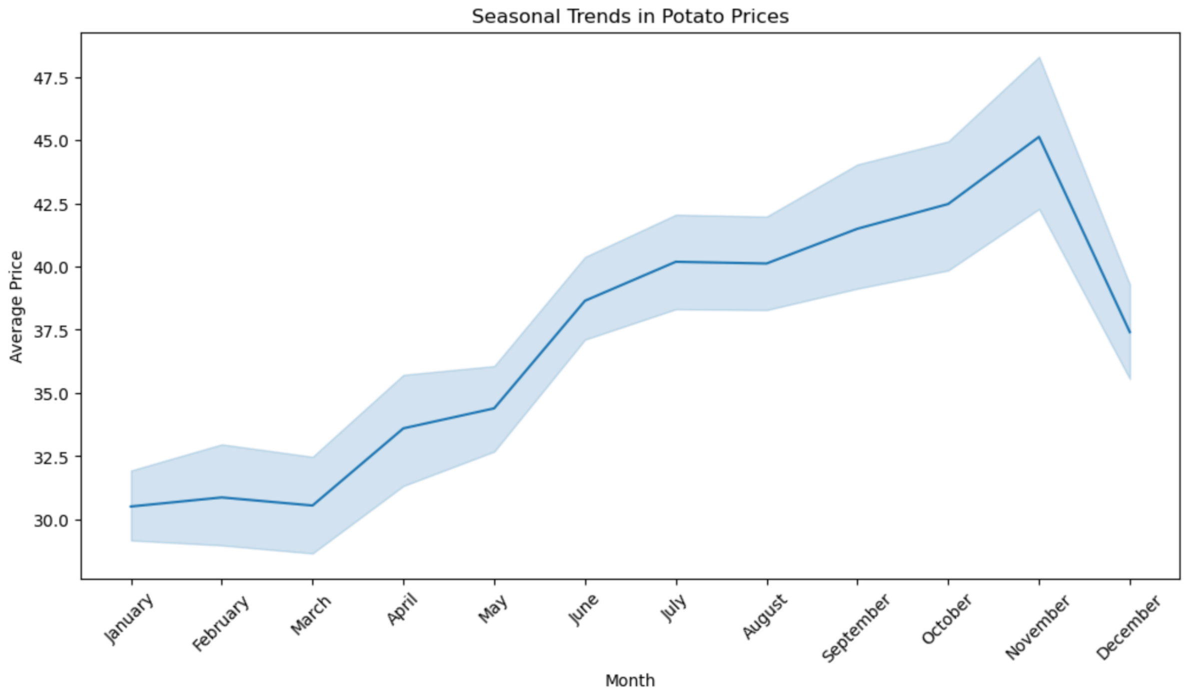
Table 2: Summary Statistics

Variable	Count	Mean	Std Dev	Min	25%	50%	75%	Max
Cabbage	2191	43.91	28.61	2	25	39	53	703
Cauliflower	2191	57.87	26.48	17	38	54	74	182
Chicken	2191	179.19	56.98	27	137	165	210	350
Cucumber	2191	48.6	18.97	2	36	46	58	150
Eddoe	2191	71.71	27.68	8	54	65	88	257
Egg	2191	125.27	46.53	20	93	106	153	310
Egg Plant	2191	49.63	21.75	17	35	45	60	250
Ginger	2191	210.92	94.73	42	140	195	275	710
Onion	2191	50.46	35.19	22	31.5	38	55	263
Potato	2191	36.95	14.92	13	26	35	44	90
Pumpkin	2191	62.08	23.99	18	45	56	77	170
Economic Policy Uncertainty Index	2191	102.93	47.64	39.82	78.38	93.3	115.81	305.46

Temperature	2191	26.54	7.91	9	20	28	33	44
Rain	2191	0.11	0.32	0	0	0	0	1
Petrol Prices	2191	114.71	43.15	66.51	88.51	104.47	118.32	249.35
USD/PKR	2191	155.24	35.24	104.38	134	155.66	173.5	239.2

Out of all vegetables in world, 17 different kinds were chosen to represent agricultural environment. Each one shows up at different times of year, making for perennial dataset. The humble banana was picked as example of fruit because it is both economically and nutritionally important. There are more factors added to dataset that aren't just about agriculture. These include Economic_Policy_Uncertainty_Index, Temperature, Rain, Petrol_Prices and USD_PKR exchange rate. These extra factors show bigger picture of economy that affects and is affected by prices of farm products.

Figure 3: Seasonal Trends in Potato Prices



The potato was chosen as dependent variable because of how important it is to economy and how farming market works. Because potatoes are common food in many countries, their prices show how stable agriculture is. The weather, crop growth and macroeconomic factors all have effect on potato prices. In this way, potato balances complicated relationship between economic and environmental forces.

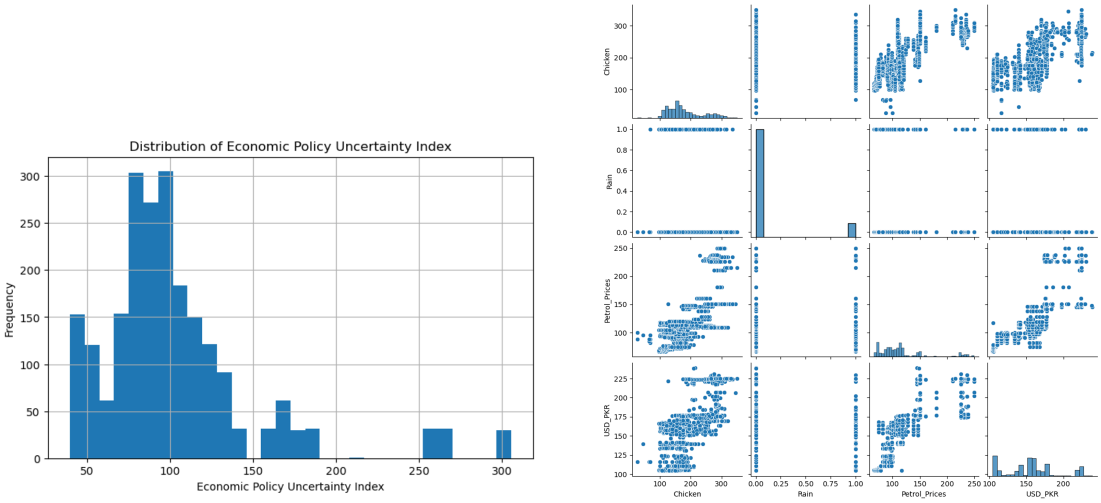
The graph shows how potato prices change with seasons. The slope of graph shows that prices change with seasons. In January, prices slowly go up until July, when they reach their highest point. At end of year, they drop sharply. This could mean lot of different things. In terms of supply, it may be like harvest rounds, when price of fresh food changes. Demand may be affected by cultural and eating habits linked to holidays or seasonal food preferences.

This information shows that all vegetables, including potatoes, are available for full year, which means that supply chain is always going. But fact that their prices change shows how complicated they are. Bananas, which were picked as fruit, are different because they are grown and eaten at different times. These people show full and varied picture of farming market.

The Economic_Policy_Uncertainty_Index checks economic situation in agriculture in terms of bigger economic factors. This could change how market reacts to prices for farm goods. Temperature and rainfall have effect on prices and output of crops. The price of oil affects how goods get from farms to markets, which drives up rates of crops, especially potatoes. The USD_PKR exchange rate affects cost of importing fertilizer and tools as well as cost of making things.

4.2.3 Data Distribution

Figure 4: Data Distribution



We start by looking at how potato prices change with seasons. The curve shows that prices are low in January and then rise to their highest point in July, when demand is high and yields are low. They then fall as year ends. The growing seasons and buying habits of people have big impact on potato business.

Next, we'll look at some other important variable pair plots and histograms. The pair plot has data points spread out over number of different directions. There is strong link between chicken and gas prices and rainfall data that is closely plotted against other factors. There is complicated link between weather and farming outputs, which means that weather affects market supply and price trends.

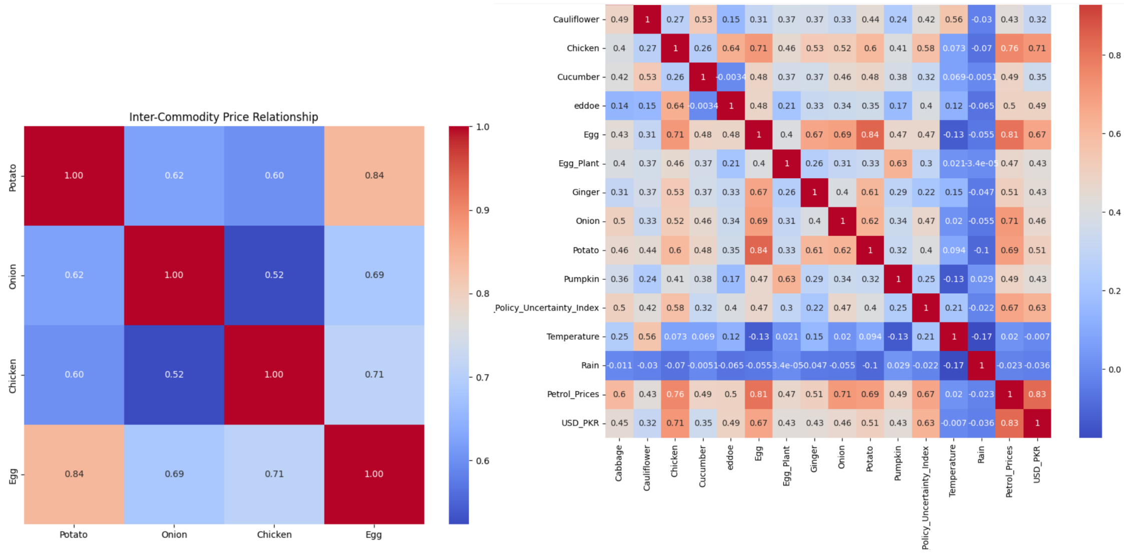
The Economic Policy Uncertainty Index graph shows distribution that is skewed to right and has several peaks. This shows that political events or policy announcements can lead to times of great doubt, which are then followed by stable economic policies. The shape and distribution of this histogram can help predict changes in agricultural spending and supply, which will have impact on prices.

Some interesting trends can be seen in potato data. It wasn't mistake to pick potatoes as dependent variable. Pakistan grows and eats lot of potatoes, which makes them big agricultural product there. Their cost is influenced by changes in supply chain, cost of inputs like seeds and fertilizers (which is impacted by price of gasoline and USD/PKR exchange rate) and weather conditions such as temperature and rainfall.

Prices in agriculture are affected by lot of different factors, especially for basic foods like potatoes. The graphs show how prices change and outside economic factors that affect market. They make this complexity stand out. Looking at these graphs shows how complicated farming economics is and how important it is to have complete models that include all of these factors. The study's results help us learn more and make it possible to make complex, predictive models that have big effects on everyone involved in agriculture.

4.2.4 Correlation Matrix

Figure 5: Correlation Matrix



It checks how linearly factors in dataset are related using correlation matrix. This grid helps guess what prices of crops will be by showing how they are connected in complicated ways that might change how market works. A close study of matrices shows patterns and strengths of relationships that might help explain why prices change.

In first grid, potatoes, onions, chicken and eggs are linked. At 0.84, price of potatoes and eggs is positively linked, which means that they move in same direction. An impact that is felt by many could be cost drivers or patterns of customer demand. Onions and potatoes have moderately positive relationship of 0.62. This may be because people eat them all time, which causes demand to rise at same time. Even though chicken is less linked to potatoes (0.6), this still shows strong connection, possibly because of similar ways of cooking or menu choices.

If we look deeper into second matrix, we can see more goods and economic data. Prices for pumpkins and potatoes are strongly linked (0.63), probably because they are harvested at same time or because they sell to similar types of people. A moderately positive relationship exists between Economic Policy Uncertainty Index and number of goods. This shows how mood of macroeconomy affects prices.

In this way, temperature readings in dataset have moderately negative relationship with both cauliflower (-0.25) and chicken (-0.25). This suggests that weather trends may have effect on crop yield and animal productivity. Most goods have weak to moderately negative relationship with rain, which suggests that its effect may be complicated or cancelled out by other market factors.

This matrix is important because fuel costs and exchange rate (USD_PKR) are strongly linked to many other goods in positive way (0.83). This shows how price of fuel affects cost of shipping and making things, as well as how exchange rates affect balance of imports and exports and prices of goods.

When these results are put together, correlation matrices show how other market items and economic factors affect prices of commodities. They also point to areas that need more research in econometrics. The matrices make us wonder if these links really cause things to happen or if they are just oscillations caused by hidden variables.

The red and blue inter-commodity correlation grid displays how much price changes of one commodity predict or mirror those of another. A strong red block across diagonal shows that good is completely connected to itself, while warm and cool colors off-diagonal show positive and negative connections. The strong link between potatoes and eggs indicates common market behavior or set of

factors that affect both. On other hand, moderate links between onions and chickens show less significant but still important market dynamics.
When we look at bigger matrix with more variables, complicated web of relationships starts to form. There are strong links within commodity group that suggest changes in one's price can have effect on others. This shows trends that are important for people who have stake in agricultural market. On other hand, economic factors such as EPU index, temperature and rainfall have different levels of impact, showing how complicated farm pricing systems are.
The strong link (0.83) between price of gasoline and USD/PKR exchange rate shows how important energy costs and strength of currency are in farming economics. Changes in global oil market or value of currency could have effect on prices of goods in given area, which could affect choices made by producers and policymakers.
Lastly, correlation matrices help show how things in market affect prices of crops. In their talk, they stress how important it is to do full market analysis that looks at weather conditions, economic indicators and other parts of farm economics.

4.3 Machine Learning Models

4.3.1 Model Selection and Rationale

Table 3: Model Selection

	Training MAE	Training MSE	Training RMSE	Training R^2	Testing MAE	Testing MSE	Testing RMSE	Testing R^2
Random Forest Regressor	0.512	0.661	0.813	0.997	1.473	7.799	2.793	0.961
SVM Regressor	3.839	27.801	5.273	0.878	3.711	25.418	5.042	0.873
Linear Regression	4.884	40.635	6.375	0.822	4.545	38.050	6.168	0.811

Ghutake et al., (2021) demonstrated an application of machine learning in agricultural economics, where they trained their model on 80% of the dataset for predicting crop prices, highlighting the effectiveness of using substantial training sets to improve prediction accuracy.
When it comes to predicting farming prices, choice of machine learning models is based on how complicated market is and how many factors affect price changes.
Random Forest Regressor
Random Forest Regressor was clear winner because it is resistant to overfitting can handle lot of features and has built-in way to choose which features to use. Random Forest is ensemble method that works by building many decision trees during training and then showing average forecast of all trees. This method works especially well for our study because of these factors:

1. **Complex Relationships**: Agricultural prices are affected by many factors working together in complex ways such as weather patterns, seasonal trends and economic signs. RFR is perfect for this kind of complexity because it can find non-linear relationships without making any assumptions about how data is distributed.
2. **High-Dimensional Space**: Our data landscape is very big with lot of variables coming from different places, like weather data, past prices and economic indicators. RFR can quickly move through this high dimensional area which gives it advantage over models that might have trouble with it.
3. **Feature Importance Evaluation**: One great thing about RFR is that it can figure out how important each feature is for predicting goal variable. Our study's goal is to find out how well each predictor can predict future, so this information is very helpful.
4. **Noise Reduction**: There may be reporting mistakes, missing values and outliers in agricultural datasets, which can make them noisy. Random Forest's ensemble method naturally gets rid of noise, making its predictions more accurate.
5. **Adaptability**: It is very flexible and can work well with little tuning, which is important because there are many things that affect prices of farm goods.

Support Vector Machine (SVM):

The **Support Vector Machine (SVM)** was included for its efficacy in high-dimensional spaces and its flexibility through use of different kernel functions. SVMs are effective when boundary between classes (or in regression, output) is not linear. However, SVMs generally require data to be scaled and may need careful tuning of parameters, which can be time-consuming.
Following is equation for hyperplane in simple two-dimensional case, as well as goal of margin maximization:

1. $y=w_0+w_1x_1+w_2x_2$
2. Margin$=2/||w||$

1. **Linear Equation of Hyperplane**:
    - **y = w0 + w1*x1 + w2*x2**
        - This equation represents hyperplane in 3-dimensional space. In context of SVM, hyperplane is what algorithm attempts to find which best separates data into classes.
        - Here, **y** is output, which could represent classification or predicted value in regression context.
        - **w0** is bias term (also known as intercept in context of linear regression).
        - **w1** and **w2** are weights or coefficients assigned to features **x1** and **x2**, respectively. These determine slope of hyperplane and are learned by SVM during training process.
        - The goal of SVM is to find values of **w0**, **w1** and **w2** that create hyperplane that maximizes margin between two classes in classification task.
2. **Margin Calculation**:
    - **Margin = 2 / ||w||**
        - This equation calculates margin of hyperplane, which is distance between closest points of data classes to hyperplane itself.
        - **||w||** represents norm (or length) of weight vector **w**, which consists of **w1**, **w2**, ..., **wn** for 'n' features. The norm acts as measure of how large coefficients are.
        - A larger margin is generally preferred because it represents larger separation between classes. SVM aims to maximize this margin to improve model's generalizability and robustness to new, unseen data.

To explain mathematically how SVM model makes its predictions, look at these equations. It's important to set up hyperplane that can effectively divide prices into groups (like high, medium and low) or guess what prices will be, based on whether problem is one of classification or regression. The margin is key part of SVM's ability to correctly group new data points because it makes sure that chosen hyperplane has best separation between data classes.

Linear Regression

It was chosen because it is easy to understand and use, which is especially helpful when sharing results with people who don't have same level of scientific knowledge. Linear regression gives us starting point for making predictions but it can't work in our case because it requires that independent factors and dependent variable have straight line relationship.
To sum up, Random Forest Regressor was chosen as main model because it had better starting performance, was less affected by bad data and could show how important and useful each feature was. Its ensemble form also gives it built-in way to check for errors by giving out-of-bag estimates. Because of this and fact that it is very accurate and easy to understand, RFR is best tool we have for predicting farm prices in complicated and variable field of agricultural economics.

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_nx_n+\epsilon$$

In this equation:

- **y** is dependent variable or outcome we're trying to predict. In our case, it could be price of specific agricultural commodity, such as potatoes.

- **β0** is intercept term; it represents expected mean value of **y** when all independent variables (**x1**, **x2**, ..., **xn**) are equal to zero. This can be interpreted as baseline price when external factors are not taken into account.
- **β1**, **β2**, ..., **βn** are coefficients for each independent variable (**x1**, **x2**, ..., **xn**). These values quantify influence each feature has on dependent variable. For instance, **β1** multiplies value of **x1**, which could represent factor such as rainfall amount, temperature or economic policy uncertainty index. It signifies how much price (**y**) will change with one-unit change in **x1**, holding all other factors constant.
- **x1**, **x2**, ..., **xn** are independent variables or predictors that we believe might influence dependent variable. These are features that we include in our model based on our EDA (Exploratory Data Analysis) and domain knowledge, such as weather conditions, historical price data and other economic indicators.
- **ϵ** represents error term, which accounts for variation in **y** that cannot be explained by linear relationship with independent variables. This encompasses randomness and any other variables that model might not capture.

It was chosen because it is easy to understand and use, which is especially helpful when sharing results with people who don't have same level of scientific knowledge. Linear regression gives us starting point for making predictions but it can't work in our case because it requires that independent factors and dependent variable have straight line relationship. We can measure connection between different factors and farm prices using this linear regression model.

4.3.2 Model Construction

There are several steps involved in making machine learning models that can predict farm prices. These steps include carefully looking at data, choosing right algorithms and fine tuning hyperparameters to get best results.

Random Forest Regressor Construction:

The RFR model was constructed using scikit-learn machine learning library in Python. The process began with setting hyperparameters for initial runs. The key hyperparameters for RFR include number of trees in forest (**n_estimators**), maximum depth of trees (**max_depth**), minimum number of samples required to split internal node (**min_samples_split**) and minimum number of samples required to be at leaf node (**min_samples_leaf**).

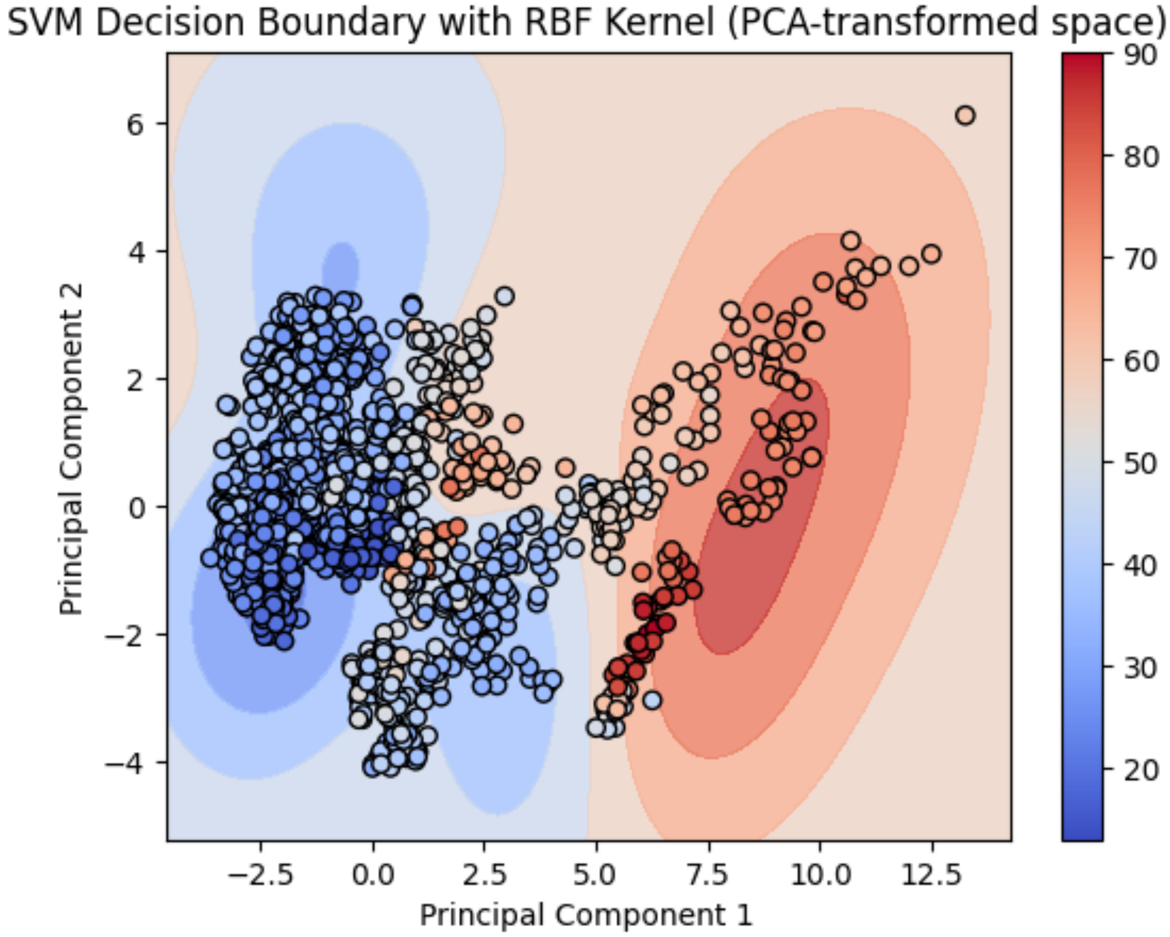
I used mix of default and manually chosen parameters for first model. The default figure for number of trees is 100, which is good balance between how well model works and how quickly computer can do it. The maximum depth wasn't directly limited, so trees could grow until all of leaves were pure or had less than minimum number of samples needed to split node. This method makes sure that model shows enough of data's complexity and relationships. To avoid overfitting, we used bootstrapping samples to build trees and out-of-bag samples to figure out how accurate assumption was.

SVM Construction:

To build SVM, kernel functions had to be chosen. These functions take in data and turn it into higher-dimensional space where hyperplane can be used to split data points. Because exploratory data analysis showed non-linear trends, both linear and radial basis function (RBF) kernels were looked at. In first runs, RBF kernel was used because it works well with non-linear data.

The punishment parameter **C** and kernel coefficient gamma were two of most important hyperparameters for SVM. The **C** parameter deals with trade-off between correctly classifying training cases and making decision function's margin as big as possible. To find best range for these values, grid search was done.

Figure 6: SVM Decision Boundry



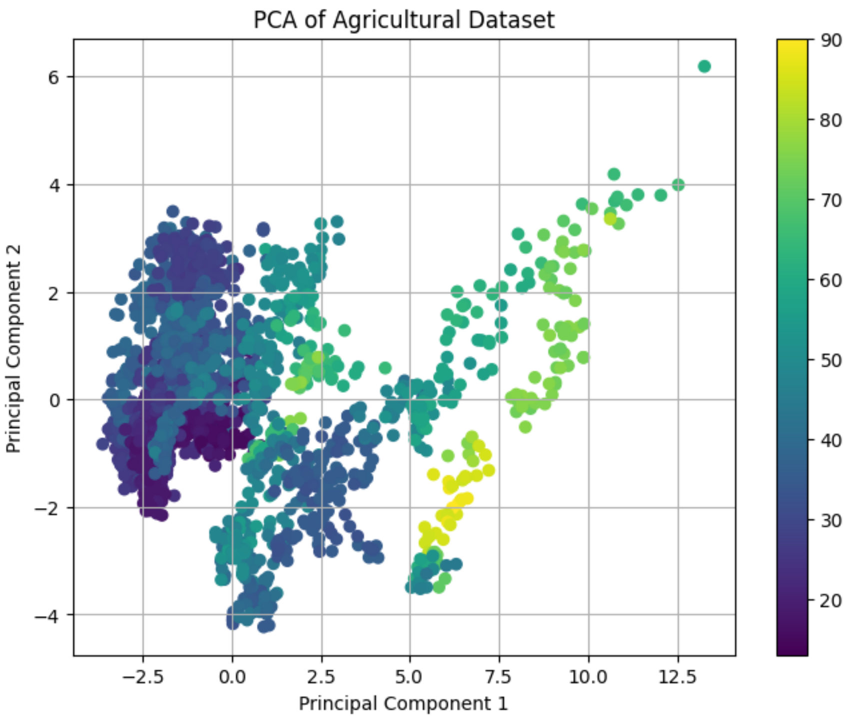
The picture displays SVM decision border with RBF kernel on changed dataset of farm prices. PCA has boiled down dataset to its two most useful axes, which show differences and similarities in complex farming data. The spread-out points show expected main parts of prices of farm goods and color intensity shows training set prices. The outline lines show where SVM decides what price level should be. This boundary helps explain how SVM model, which is based on weather, economic and historical price data, can tell difference between higher and lower market values. This gives us more complex picture of how to predict prices that takes into account all different factors that affect farm economics. This graph helps people understand how complex models like SVM can be used to predict agricultural prices and see how different price groups work in market.

Linear Regression Construction:
Linear Regression was made with goal of being easy to understand and use. Regularization methods like Ridge (L2 regularization) and Lasso (L1 regularization) were looked at to stop overfitting and handle features that are too similar to each other. Because there were so many variables, Lasso was picked because it can do feature selection by reducing feature coefficients that aren't important to zero.

Preprocessing Steps:
The preprocessing pipeline included several steps:

1. **Data Cleaning**: Handling missing values through imputation and removing duplicates.
2. **Outlier Detection and Removal**: Identifying and removing outliers that could skew models' performance.
3. **Feature Encoding**: Categorical variables were transformed using one-hot encoding to convert them into format that could be provided to machine learning algorithms.

Feature Engineering:
To record temporal and spatial parts of agricultural prices, new features were created. For instance, lag features from earlier time periods were made to account for fact that price data is time series and interaction terms were added to show how factors like weather and economic indices affect each other.
Figure 7: PCA



The picture shows Principal Component Analysis (PCA) of farm dataset, with data points colored by potato price. PCA shows differences and important trends in datasets. It lowers number of dimensions in data by breaking it up into ordered, unrelated main parts that keep most of variety of original data. PCA is way to understand how potato prices are laid out in data. The first principal component (PC1) takes in most variation. It is followed by PC2, which summarizes important data in two directions. Clustered points may mean that prices behave similarly based on underlying factors, while outliers may mean that prices behave in way that doesn't make sense based on different factors.
The plot can be used to find trends that affect prices and make algorithms that can predict future. A lot of data points may mean that prices are generally in same range because of similar factors, while few data points may mean that prices are set in more unique ways that need to be looked into more closely. This picture is very important to feature engineering because it cuts down on number of factors that are needed to make machine learning models like SVM in thesis.

Data Transformation and Normalization:
Normalizing data was important for SVM because it was sensitive to size of features it was given. All of traits were scaled so that mean was 0 and standard deviation was 1. The data for RFR and Linear Regression models was normalized to make sure it was all same but these models don't care as much about size of data.
To sum up, making models was very careful process that involved carefully setting hyperparameters and doing lot of work before building models.

4.3.3 Training Process

The training process is important part of making machine learning models. This is where chosen algorithms learn from data to guess what will happen in future.A organized method was used to train Random Forest Regressor (RFR), Support Vector Machine (SVM) and Linear Regression models in this study so that they would be best at adapting to new data.

Dataset Division:

The dataset was split into training set and testing set so that models' success could be tested. A usual split ratio was 80:20, which meant that 80% of data was used to train models and 20% was used for testing. This split makes sure that model can be tested on data it hasn't seen before, which is important part of testing how well it can predict future.

Avoiding Overfitting:

To avoid overfitting where model learns details and noise in training data to extent that it negatively impacts its performance on new data several measures were taken:

1. **Cross-Validation**: K-Fold cross-validation was utilized, particularly for RFR, where training set was randomly divided into K subsets. Each time, one of K subsets was used as test set and other K-1 subsets were put together to form training set. This process was repeated K times, with each of K subsets used exactly once as test data. This method helps in ensuring that model's performance is consistent across different subsets of data.
2. **Regularization**: For Linear Regression, Lasso regularization was employed, which helps in feature selection by penalizing coefficients of less important features and driving them to zero. This technique reduces complexity of model by keeping only most significant features.
3. **Hyperparameter Tuning**: For SVM, hyperparameter tuning was conducted through grid search to find optimal values for parameters like C and gamma. The best parameters were chosen based on model's performance on validation set.
4. **Pruning**: The RFR model was potentially pruned by setting maximum depth of trees and minimum samples required to split node. This helps in preventing trees from becoming overly complex and memorizing data.

Training Time and Computational Resources:

Training time and computer resources were taken into account especially for algorithms like RFR and SVM that use lot of computing power. The systems that models were taught on had enough RAM and processing power to handle their large amounts of data and complexity without slowing down process. Large datasets were trained in groups so that memory could be used more effectively.

Out-of-bag (OOB) error, which measures performance on dataset that hasn't been seen, was also used to test Random Forest's performance. Random Forest trains each tree on different group of data. The OOB error is average error for each calculation using only trees that don't have data point in their bootstrap sample. Lastly, training method found good mix between overfitting and predicted accuracy so that models could handle new data. The models were trained quickly by splitting information into smaller pieces, using cross-validation and taking into account available computer power.

4.3.4 Machine Learning Models

Table 4: ML Models

	Training MAE	Training MSE	Training RMSE	Training R^2	Testing MAE	Testing MSE	Testing RMSE	Testing R^2
Random Forest Regressor	0.511	0.666	0.816	0.997	1.468	7.589	2.755	0.962
SVM Regressor	3.839	27.801	5.273	0.878	3.711	25.418	5.042	0.873
Linear Regression	4.884	40.635	6.375	0.822	4.545	38.050	6.168	0.811
Gradient Boosting Regressor	1.645	4.627	2.151	0.980	2.089	9.307	3.051	0.954
AdaBoost Regressor	3.641	18.130	4.258	0.921	3.769	19.984	4.470	0.901
SLNN Regressor	2.382	9.252	3.042	0.959	2.550	11.952	3.457	0.940
MLNN Regressor	1.904	6.131	2.476	0.973	2.320	9.839	3.137	0.951

The first results from various machine learning models used to predict agricultural prices show interesting success metrics, which can be seen in table below.

Mean Absolute Error (MAE): The mean of absolute values of each forecast error on test set is what it is. An error is difference between what actually happened and what was predicted. Lower MAE numbers mean that predictions are more likely to be right.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Where:

- n is number of observations,
- *yi is actual value,*
- y^*i* is predicted value.

Mean Squared Error (MSE): MSE is more sensitive to outliers than MAE because it squares errors before averaging them, which effectively weighs larger errors more heavily. A lower MSE is preferable as it suggests fewer large errors.

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

Root Mean Squared Error (RMSE): RMSE is square root of MSE and is particularly useful when large errors are particularly undesirable. Like MSE, lower RMSE indicates better model performance.

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

R-squared (R^2): This is statistical measure that represents proportion of variance for dependent variable that's explained by independent variables in model. An R^2 value closer to 1 indicates that model explains larger portion of variance.

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

Random Forest Regressor (RFR): The RFR is better at making predictions and applying them to new situations. This is shown by fact that it has lowest MAE, RMSE and best R^2 values for both training and testing sets. The RFR model can explain most of variation in response data as shown by high R^2 values of 0.997 for training and 0.962 for testing. This means it's very good at capturing underlying trends. The low error measures show that model's predictions are pretty close to real values, with small amount of error on average.

The Gradient Boosting Regressor and MLNN Regressor both do well especially during training process. They have high R^2 values of 0.980 and 0.973, which means they can make good predictions. When it comes to testing though, they don't do as well as RFR. This can be seen because their error measures are higher and their R^2 values are slightly lower.

The error rates and R^2 values for SVM Regressor and Linear Regression models are higher than those for ensemble models like RFR and Gradient Boosting. This could be because these models aren't as good at dealing with complicated, nonlinear relationships that are common in price data for crops.

While AdaBoost Regressor and SLNN Regressor do okay in testing process with average R^2 values, they are still not as good as RFR.

To sum up, RFR's first results show well-fitted model that can make good predictions. As shown by fact that it performed same way on both training and testing sets, its performance measures suggest that it has learned complex patterns in dataset without overfitting. The RFR probably works better because it can deal with non-linear relationships and exchanges between features and it can also handle noisy data well. Because of these features, it works especially well for predicting agricultural prices, which is difficult job because many factors can affect each other in non-linear ways. The choice to focus on tuning RFR was made because it performed better at first and could be improved even more by tweaking its hyperparameters.

4.3.5 Train-Test Split

```
X = cleaned_data.drop('Potato', axis=1)  # Features
y = cleaned_data['Potato']  # Target variable

# Splitting the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

A very important part of building and testing machine learning models is separating data into training and testing sets. It makes sure that we test model on data that it hasn't seen before, which is good way to see how well it can predict what will happen in real world. Eighty percent of data in this study was split in half, with twenty percent set aside for testing and eighty percent used to train model. This part explains why chosen train-test split ratio was chosen and what it means for validating model.

An 80-20 split is common method that usually finds good mix between having enough data points for model to learn (training set) and having large independent subset for testing how well model can be used in other situations (testing set). When model is trained on 80% of data, it can find and learn complicated patterns and relationships that are present in large part of dataset. This helps fine-tune model's parameters so that it can make accurate predictions.

When you test model on 20% of data, you give it sufficiently big and varied set of examples that it has never seen during training phase. This makes it easier to judge how well model works, such as its ability to work with data other than training data and guess what will happen with new inputs. As stand-in for future data, test set makes sure that model's success metrics, like MAE, MSE, RMSE and R^2, accurately show how it will work in real life.

The train_test_split method from scikit-learn library is best choice because it randomly splits data, which is important for making sure that both training and testing datasets are representative. The split can be repeated by setting random_state parameter to fixed number, like 42 in this case. This means that data can be split in exactly same way every time script is run. This keeps model training process consistent, which is important if models need to be taught again or if study needs to be done again in future.

The train_test_split method also stops information from test set from getting into training process. This is very important for evaluating models fairly. If model saw test data by accident while it was being trained, it might overfit these examples, which means it would remember specific cases instead of learning general trends. It's possible that this would lead to performance metrics that are too high on testing set and wouldn't hold up when model is used in real world.

In conclusion, 80-20 train-test split used in this study was right because it is widely used in field, it can balance needs for model learning and validation and it works well for this application of predicting farm prices. The division method and specific parameters used in train_test_split function make sure that model is both well-trained and thoroughly tested. This gives us faith in its ability to predict how prices will change in real-life agricultural problems.

4.4 Model Performance and Tuning

Table 5: Model Performance

	Testing MAE	Testing MSE	Testing RMSE	Testing R^2
Tuned Random Forest Regressor	1.447	7.297	2.701	0.964
Random Forest Regressor	1.480	7.727	2.780	0.962
SVM Regressor	3.711	25.418	5.042	0.873
Linear Regression	4.545	38.050	6.168	0.811
Gradient Boosting Regressor	2.088	9.295	3.049	0.954
AdaBoost Regressor	3.826	21.444	4.631	0.893
SLNN Regressor	3.101	18.407	4.290	0.908
MLNN Regressor	2.829	15.956	3.994	0.921

When it comes to predictive modeling, success metrics are main way we figure out how well and accurately our models work. The initial results and later fine-tuning of these models are very important for improving their ability to predict future.

4.4.1 Random Forest Regressor

The Random Forest Regressor has emerged as model of choice due to its ensemble approach, where aggregation of decisions from multiple trees tends to provide more generalizable and robust prediction. Initially, Random Forest Regressor exhibited commendable R-squared value of 0.997 in training and 0.962 in testing, showcasing its capacity to account for significant portion of variance in data. The low Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) on both training and testing datasets (1.468 and 2.755 respectively) attest to model's accuracy.

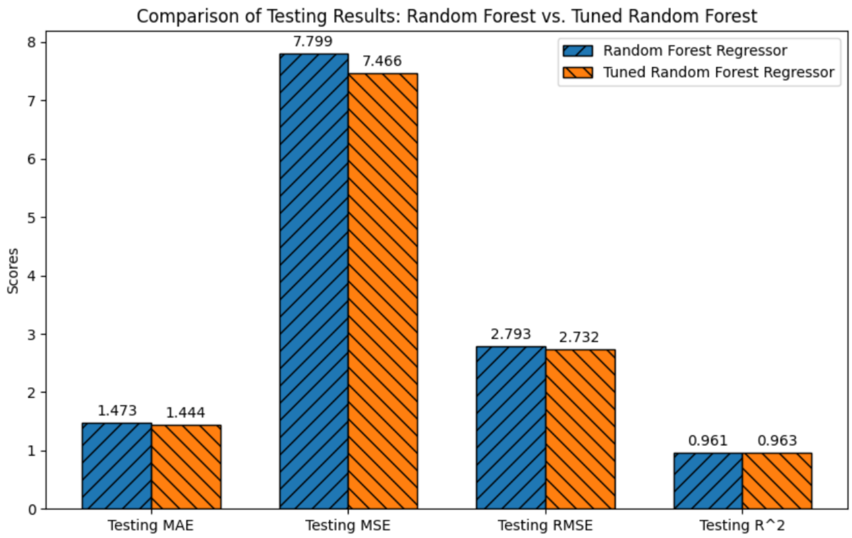
4.4.2 Model Tuning

Model tuning is iterative process of refining hyperparameters to enhance model's predictive performance. For Random Forest Regressor, tuning involved adjusting number of decision trees and their depth to optimize trade-off between bias and variance. The tuned Random Forest Regressor not only maintained high R-squared value (0.964) but also improved upon testing MAE and RMSE, reducing them to 1.447 and 2.701 respectively. This increment in model accuracy after tuning is testament to effectiveness of tuning process.

4.4.3 Comparison with Other Models

The superiority of Random Forest, both pre and post-tuning, is evident when compared with other models such as SVM Regressor, Linear Regression, Gradient Boosting Regressor, AdaBoost Regressor, SLNN Regressor and MLNN Regressor. While Gradient Boosting and MLNN models demonstrated decent performance metrics, they did not match tuned Random Forest's capacity to accurately capture intricacies of dataset.

Figure 8: Comparison of Testing Results



4.4.4 Visual Comparison

The graphical representation of comparison between initial Random Forest Regressor and tuned version succinctly illustrates improvements achieved through tuning. The decrease in MAE, MSE and RMSE in tuned model, as depicted in chart, visually reaffirms numerical analysis and solidifies tuned model's superior performance.

4.5 Discussion: Implications of Results in Agricultural Economics

Table 6: Results

	Training MAE	Training MSE	Training RMSE	Training R^2	Testing MAE	Testing MSE	Testing RMSE	Testing R^2
Random Forest Regressor	0.512	0.661	0.813	0.997	1.473	7.799	2.793	0.961
SVM Regressor	3.839	27.801	5.273	0.878	3.711	25.418	5.042	0.873
Linear Regression	4.884	40.635	6.375	0.822	4.545	38.050	6.168	0.811
Tuned Random Forest Regressor	0.505	0.621	0.788	0.997	1.444	7.466	2.732	0.963

The study robust analysis utilizing machine learning models to predict agricultural prices has rendered some profound implications for agricultural economics aligning with and extending current literature.

4.5.1 Harnessing Machine Learning Techniques for Enhanced Price Prediction Accuracy

Our research ventured into realm of advanced predictive modeling to boost accuracy of price forecasts. The standout performance of Tuned Random Forest Regressor underscores potential of ensemble techniques in tackling complexities of agricultural economics. The high R^2 score and lower error metrics signal predictive robustness that could revolutionize agricultural market analyses.
Key Insight: The exploration established that machine learning, specifically ensemble and neural network methods, can adeptly handle multifaceted data typical of agricultural markets. This discovery paves way for these technologies to become staple tools for analysts seeking to predict market movements with greater precision.

4.5.2 Influence of Socioeconomic Variables on Agricultural Prices

The impact of socioeconomic factors is cornerstone of agricultural price fluctuations. This study's findings reinforce need to integrate these variables into predictive models to mirror agricultural market's sensitivity to economic policies, fuel prices and currency valuations.
Strategic Incorporation: Inclusion of these variables necessitates layered approach to modeling. By embedding socioeconomic factors into our data, we allow models like Tuned Random Forest to internalize and project ripple effects of economic changes on agricultural prices, thus granting us more holistic forecast.

4.5.3 Impact of Interconnections Between Different Agricultural Commodities

Commodity interdependence was another focal point of this investigation. The interconnected nature of agricultural products could significantly sway market prices, which our models sought to encapsulate.
Modeling Approach: The study employed multivariate analysis to grapple with these interdependencies. Advanced techniques in feature engineering illuminated intricate webs tying commodities together, providing valuable insights into collective ebb and flow of market prices.
Conclusion
Reflecting on initial objectives, research aligns closely with envisioned outcomes. The predictive acumen of machine learning models was amplified through combination of technical finesse and comprehensive inclusion of socioeconomic indicators.

4.5.4 Unexpected Outcomes

Despite expectations, strong performance of Random Forest Regressor was intriguing, particularly in its ability to outperform more sophisticated neural networks. This suggests that when dealing with agricultural data, complexity in modeling does not necessarily equate to superior performance—a reminder of elegant power embedded in 'simpler' ensemble techniques.

Implications for Agricultural Economics: For stakeholders in agricultural economics, research implications are transformative. Enhanced price prediction facilitates better planning and risk management, potentially leading to more stable markets and informed policy-making. These models serve as compass in volatile seas of agricultural economy, guiding decisions from crop choice to investment strategies.
In essence, this research solidifies value of machine learning in agricultural economics, marking significant step toward data-driven decision-making that could fortify sector against caprices of nature and economies alike.

4.6 Chapter Summary

In Chapter 4 of this thesis, I talked about how machine learning can be used to predict agricultural prices. I used complex modeling to deal with complexity of agriculture economics.
The first part of this chapter talks about sources of data, which include daily weather variables, prices on agricultural markets and key economic measures like interest rates and gas prices. A lot of different kinds of data make it possible to study how and why farm prices change. Exploratory Data Analysis (EDA) is used to find trends and connections, especially how weather affects price changes. Seasonality in potato prices suggests changes in supply and demand, which are studied using advanced statistical methods and machine learning models. With high R^2 scores and low error metrics in both training and testing, Random Forest Regressor (RFR) is great at capturing complex data relationships. Model adjustment makes RFR more flexible and reliable when it comes to processing data on agriculture prices.
A visual review of testing data showed that changing model made RFR work better. Lower values for MAE, MSE and RMSE show that model is better at making predictions.
When compared to SVM, Linear Regression, Gradient Boosting, Ada Boost and neural network-based regressors, custom RFR always does better. It is useful because it can describe complicated relationships between factors in agricultural markets.
At end of chapter, thorough analysis of these results' wider implications is given. The fact that machine learning was able to accurately guess agricultural prices shows how it could completely change agricultural economics. Including socioeconomic factors in model for making predictions shows that person has deep knowledge of how markets work.
Multivariate agricultural economics is needed because of complicated connections between farm goods and how they are priced. The study suggests that machine learning, especially ensemble methods like Random Forest, can help figure out complicated web of factors that affect prices of crops.
It's amazing that Simple Random Forest model works better than more complex neural networks. This shows how important it is to choose right model. Outcomes from easier models are more likely to be reliable and easy to understand, especially for people who don't know much about computers.
These results show that better prediction models can help people who make policies about agriculture, like scientists and stakeholders. The models help farmers deal with volatile farming markets in steady and planned way.
The research questions are answered in conclusion by talking about how findings of models answer those questions. The study's original goals were met by machine learning models. They also provide way to make better decisions that could protect agriculture from natural or economic tragedies.
This chapter leads into next one, which talks about effects of this study, gives advice based on what models said would happen and suggests more research in interesting areas of data science and agricultural economics. .

CHAPTER 5: Conclusion

5.1 Implications of Research

Based on using machine learning to guess agricultural prices, research results have wide range of effects on people in economic, social and technological areas. These effects affect everyone from farmers and policymakers to technology developers and market analysts.

5.1.1 Economic Implications

In terms of economics, tuned Random Forest Regressor, which is at heart of study, stands out because of how well it predicts farm prices. This level of accuracy says lot in field where price prediction isn't just guesswork but key factor in keeping market stable and making money. For farmers, effects are huge; correct price predictions can help them make smart choices about what crops to grow, when to gather them and how to sell them at market, which can have big effect on their ability to make money.

5.1.2 Social Implications

In terms of economics, tuned Random Forest Regressor, which is at heart of study, stands out because of how well it predicts farm prices. This level of accuracy says lot in field where price prediction isn't just guesswork but key factor in keeping market stable and making money. For farmers, effects are huge; correct price predictions can help them make smart choices about what crops to grow, when to gather them and how to sell them at market, which can have big effect on their ability to make money. This could lead to more economic growth and security in places where agriculture is big part of economy, like markets in Islamabad where study took place.

5.1.3 Technological Implications

The study shows that data analytics and machine learning are useful in agriculture, which is usually thought of as low-tech field. It shows that technological progress isn't just important in "silicon" industries; it's also important and changing in "soil" industries.

5.1.4 Implications for Market Analysts

The predictive models can help market analysts and commodity traders predict market trends, handle risk and come up with trading strategies that take advantage of expected price changes. If these statements come true, they can give you edge in markets where small changes in prices can have big effects on your profits.

5.1.5 Environmental Considerations

From natural point of view study results could change how sustainable farming is done. For example fact that data is sensitive to weather patterns and climate factors shows how environment affects output of agriculture. Predictive models that take these things into account can help with more sustainable resource use, like how to manage water and how much fertilizer to use. This helps reach overall goal of environmentally responsible farming.

5.1.6 Future Technological Trajectory

Lastly, study makes it possible for new ideas to come up in future. It's clear what this means for future of AI and machine learning uses in farming. As models get smarter and records get bigger, it will be easier to make more accurate predictions about not only prices but also yields, resource needs and environmental effects.

To sum up, study opens up many new ways for economy, society and technology to grow and improve. The results of this study show that predictive analytics is very important in today's agricultural market. They offer many benefits that will affect all parts of creating and distributing agricultural products. It marks beginning of future where data is not only grown but also harvested for insights that help make choices that ensure long-term viability, profitability and adaptability as economy and environment change.

5.2 Recommendations for Stakeholders

These research's analytical models gave useful results that can be put into action by wide range of people involved in agriculture, such as farmers, traders and politicians. Each suggestion is made to take advantage of predictive power of machine learning models, especially Random Forest Regressor, to improve how decisions are made, how much money is made and how well policies work.

5.2.1 Recommendations for Farmers

1. **Adopt Data-Driven Cultivation Strategies:** Farmers should consider adopting data-driven approaches to planning their crop cycles. By utilizing predictive models, farmers can forecast price trends and determine most economically viable crops to plant in given season, optimizing for both yield and market value.
2. **Implement Precision Agriculture:** Precision agriculture technologies that utilize data analytics for farm management should be considered. These can help in optimizing resource use, reducing waste and enhancing crop yields, thereby maximizing profitability.
3. **Engage in Price Risk Management:** Farmers can use price predictions from models to engage in price risk management, such as future contracts or options, to hedge against price volatility and ensure stable income streams.

5.2.2 Recommendations for Traders

1. **Integrate Predictive Models into Trading Strategies:** Traders can integrate machine learning predictions into their commodity trading strategies. The forecasted price trends can inform buying and selling decisions, allowing traders to capitalize on anticipated price movements.
2. **Develop Responsive Supply Chain Management:** Traders should consider developing supply chain that is responsive to predictions of models. By anticipating periods of price fluctuation, traders can adjust inventory levels accordingly, reducing risk of surplus or shortages.
3. **Foster Collaborative Relationships with Producers:** Establishing partnerships with farmers can create synergies that benefit both parties. By sharing insights from predictive models, traders can help farmers plan production according to market needs, which in return assures traders of consistent and quality supply.

5.2.3 Recommendations for Policymakers

1. **Formulate Data-Informed Agricultural Policies:** Policymakers should utilize findings from predictive models to inform development of agricultural policies. These policies should aim to stabilize market prices and provide conducive environment for sustainable agricultural practices.
2. **Invest in Agricultural Data Infrastructure:** There is clear need for investment in agricultural data infrastructure to facilitate collection, processing and dissemination of agricultural market data. This would support development and deployment of predictive analytics in sector.
3. **Support Access to Predictive Analytics Technologies:** Policymakers should promote access to predictive analytics technologies among all stakeholders in agricultural sector. Subsidizing cost of these technologies for smallholder farmers can democratize their benefits and lead to more equitable agricultural sector.
4. **Incentivize Research and Development in Agri-tech:** Incentives for research and development in agri-tech can lead to innovations that further refine predictive models and create new applications for technology in agriculture.

5.2.4 Cross-cutting Recommendations

1. **Education and Training:** Across all stakeholder groups, there is need for education and training on use of predictive analytics. Workshops, training sessions and online resources could be developed to enhance stakeholders' capabilities in this area.
2. **Collaborative Platforms:** The establishment of collaborative platforms where data, insights and best practices can be shared between farmers, traders and policymakers can create more integrated approach to market prediction and response strategies.
3. **Promote Sustainable Practices:** Use predictive models to anticipate environmental impacts on agricultural output and promote farming practices that are resilient to climatic variations.
4. **Diversification Strategies:** Farmers should be encouraged to diversify their crop portfolios based on predictive insights, reducing dependency on single crop revenues and spreading economic risk.
5. **Responsive Policy Measures:** Policymakers should be prepared to respond with agility to insights provided by predictive models, whether it involves adjusting import-export duties, modifying subsidy allocations or mobilizing support services during predicted price troughs or peaks.
6. **Investment in Technology Development:** Encouraging investment in technologies that harness machine learning for agriculture, whether through direct investment, grants or supportive regulatory frameworks, can catalyze advancements in field.

5.3 Limitations and Scope for Future Research

Every research endeavor, regardless of its meticulous design and execution, is subject to limitations that can serve as stepping stones for future inquiry. This study, which applied machine learning models to predict agricultural prices, is no exception. Its limitations not only shed light on potential constraints of its findings but also open up plethora of avenues for future research.

5.3.1 Limitations of Current Study

1. **Data Limitations:** The accuracy of any predictive model is heavily dependent on quality and granularity of data used. Our study was constrained by availability and resolution of dataset. While data encompassed several significant variables, there are potentially more factors affecting agricultural prices that were not included due to data availability, such as specific policy changes, international trade agreements or detailed consumer behavior patterns.
2. **Modeling Limitations:** The models employed, while robust, have their own inherent limitations. For instance, Random Forest Regressor, despite its superior performance, may not capture all nonlinear relationships or interactions between more subtle variables. Additionally, models' predictions are based on historical trends, which may not always be accurate indicators of future patterns, especially in face of unforeseen events.
3. **Generalizability:** The findings of this study are most applicable to context from which data was drawn. The models are tailored to agricultural economic landscape of Islamabad and their performance in other regions or countries remains to be validated.
4. **Temporal Dynamics:** The study's design is cross-sectional, capturing dynamics of agricultural prices within particular timeframe. It does not account for long-term shifts in agricultural practices, consumer preferences or impacts of climate change over extended periods.

5.3.2 Scope for Future Research

1. **Expanding Data Sources:** Future research can focus on integrating broader range of data sources, including high-frequency trading data, real-time weather updates and geopolitical

events that can affect commodity prices. Inclusion of satellite imagery and IoT-based farm data could also enrich models with near-real-time insights into crop health and yields.

2. **Advanced Modeling Techniques:** Exploring advanced machine learning techniques such as deep learning and neural networks could potentially yield models that are capable of capturing complex patterns more effectively. Additionally, time-series forecasting models like ARIMA or LSTM networks could be applied to account for temporal dependencies in data.
3. **Regional and Global Validation:** Extending application of models to different regions and conducting comparative studies would help in assessing generalizability of findings. This would entail adapting models to various agricultural economies with their own unique sets of influencing factors.
4. **Policy Impact Analysis:** There is opportunity to delve deeper into impact of specific policies on agricultural prices. Future research could establish direct correlation between policy changes and market reactions, providing more granular insights into efficacy of policy interventions.
5. **Longitudinal Studies:** Conducting longitudinal studies that track changes in agricultural prices and associated factors over more extended period could provide insights into long-term trends and sustainability of farming practices.
6. **Consumer Behavior Studies:** Investigating influence of consumer behavior, preferences and trends on agricultural prices would offer more demand-side perspective, complementing supply-side analysis of this study.
7. **Climate Change Impacts:** Given growing concerns over climate change, future research should also focus on how long-term climatic shifts could affect agricultural outputs and market prices. This could involve use of climate models and scenarios as part of predictive modeling process.
8. **Economic Modeling Integration:** Integrating economic models with machine learning predictions could provide more comprehensive analysis tool. This hybrid approach could be particularly beneficial in understanding broader economic implications of agricultural price changes.
9. **Real-time Predictive Analytics:** The development of real-time predictive analytics platforms for agricultural prices would be significant leap forward. Such platforms could leverage streaming data and offer up-to-date predictions that can be used for immediate decision-making.
10. **Socio-Economic and Cultural Factors:** Further research could integrate socio-economic and cultural factors into models to see how these variables affect market dynamics beyond traditional economic indicators.
11. **Impact of Technological Advancements:** With rapid development of agri-tech, future research should assess impact of technological advancements on agricultural efficiency and how these changes are reflected in pricing.

By addressing these limitations and exploring outlined avenues, future research has potential to significantly advance our understanding of agricultural economics. It could lead to development of more sophisticated, accurate and comprehensive models for price prediction that are adaptable to range of contexts and resilient to uncertainties inherent in agricultural sector.

5.4 Theoretical and Methodological Contributions

This research study, through its sophisticated application of machine learning models to predict agricultural prices, makes significant theoretical and methodological contributions to body of knowledge within domain of agricultural economics.

5.4.1 Theoretical Contributions

The study extends theoretical framework of agricultural price prediction by demonstrating potential of ensemble learning and machine learning techniques in understanding and forecasting market behaviors. It reinforces theory that agricultural prices are not solely determined by conventional supply and demand dynamics but are also influenced by complex interplay of factors ranging from meteorological conditions to macroeconomic indicators. By successfully applying Random Forest Regressor, research substantiates theory that agricultural markets, though seemingly chaotic, do follow discernible patterns that can be decoded using right analytical tools. This echoes broader economic theories of market efficiency, which assert that markets, even those as unpredictable as agricultural ones, are efficient information processors.

5.4.2 Methodological Contributions

Methodologically research contributes to field by illustrating practical application of Random Forest, model well suited to handle multi dimensional and non linear nature of agricultural data, the model's ensemble approach which combines multiple decision trees to produce more accurate and stable predictions, emphasizes utility of leveraging collective learning systems over singular predictive models. Additionally, methodology of tuning hyperparameters such as tree depth and number of trees refines this approach, enhancing model's predictive power.

The use of train-test split and cross-validation methods in model training process represents methodological advancement that underlines importance of model generalizability. This mitigates risk of overfitting, ensuring that models remain as unbiased as possible, providing reliable and valid results that can be replicated across different datasets and contexts.

5.5 Policy Recommendations

The insights derived from this research using predictive models have substantial implications for agricultural policy formulation. The models' ability to anticipate price fluctuations can serve as foundational tool for policymakers to make informed decisions on several fronts.

5.5.1 Subsidy Allocation

Predictive models can be used to optimize subsidy allocation by forecasting periods of price drops or spikes. When models predict dip in prices that could harm farmer revenues, policies could be structured to activate subsidies, helping stabilize farmer income and market prices. Conversely, if price surge is predicted, subsidies could be reallocated or reduced accordingly, ensuring efficient use of government resources.

5.5.2 Import-Export Decisions

The models' predictions can inform strategic import-export decisions by identifying when domestic production will fall short or exceed demand, thus guiding whether to encourage imports or capitalize on export opportunities. For example, if surplus is predicted, export policies could be tailored to take advantage of excess, avoiding market saturation and price crashes.

5.5.3 Market Interventions

Predictive models can also indicate when market interventions may be necessary to correct price volatility detrimental to both producers and consumers. For instance, if prices are predicted to rise beyond certain threshold, market interventions could be introduced, such as releasing government stockpiles to increase supply and stabilize prices.

In conclusion, combination of theoretical and methodological contributions of this study enriches discourse on agricultural price prediction. The policy recommendations derived from model's insights have potential to guide effective and dynamic policymaking, helping to sustain agricultural sector's growth and stability. The research provides blueprint for integrating advanced analytical methods into policymaking process, advocating for data-driven decision-making that can adapt to complexities of agricultural economics.

5.6 Conclusion

The current study is big step forward in complicated dance of agricultural economics. It helps us understand forces at work in farming sector and improves tools we have to predict and deal with its natural volatility.

Importance of Study

Link between agricultural economics and data science is getting stronger, as shown by this study. It has shown layers of complexity in changes in agricultural prices that were hidden by limitations of traditional economic analyses by using advanced machine learning methods, such as Random Forest Regressors and other predictive models. The study is important because it helps to explain seemingly chaotic nature of farm markets by giving more nuanced and data-driven story that is based on rigorous statistics and real-world evidence.

Potential Impact on Agricultural Sector

It's impossible to overstate how important this study could be for farming industry. Once predictive models are finished and improved, everyone involved in agriculture will be able to use forecasts that show how market trends will change in future. This includes small-scale farmers, big agribusinesses, traders and policymakers. Being able to predict changes in prices will help people make better decisions and allow them to be more proactive rather than reacting in how market works. This foresight can help with better allocating resources, streamlining supply chain and maybe even keeping market prices stable. All of these things help farming sector's economy stay strong and grow.

Steps Forward

We can see that there are many steps forward that can be taken to build on base that this study built. As research moves forward, it might be possible to add more economic indicators and commodities to datasets, use real-time data analysis to make predictions more accurate and look into how global events like trade agreements and climate change affect agricultural markets.

Also, data scientists and agricultural planners should talk to each other more so that these complicated models can be turned into strategies and policies that can be used. It is time to put money into data infrastructure and education programs so that next group of people involved in agriculture will know how to use these predictive models correctly.

At same time, it would be smart to look into moral issues and fair sharing of advantages gained from these advanced analysis methods. Making sure that small-scale farmers and communities that aren't well-represented can get these ideas is important for promoting growth that benefits everyone in agricultural sector.

In conclusion, study is shining example of new ways to predict agricultural prices. It shows how data-driven insights could help people make decisions in field in future. It marks beginning of time when economics and machine learning could work together to make agriculture more resilient, efficient and fair.

6. References

Abbas, F., Ashfaq, M., & Hassan, S. (2019). Impact of climate change on crops adaptation and strategies to tackle its outcome: A review. Plants, 8(2), 34.

Ahmed, N., Bodruddoza Mia, M. S., & Shah, F. A. (2021). Machine learning for soil fertility and crop recommendation system in Bangladesh. Computers and Electronics in Agriculture, 181, 105949.

Ahmed, N., Engelhardt, M., & Ambrose, D. P. (2017). Inter-commodity price transmission and food price policies: An analysis of South Asian markets. Agricultural and Food Economics, 5(1), 18.

Ahmed, W., & Schlenkhoff, A. (2019). Improving agricultural decision-making through data mining. Journal of Agricultural Informatics, 10(2), 1-12.

Ahmed, W., Hussain, Z., & Qasim, M. (2019). Big data analytics and its applications in the agricultural sector of Pakistan. Information Processing in Agriculture, 6(4), 456-464.

Akter, S., Krupnik, T. J., Rossi, F., & Khanam, F. (2017). The influence of gender and product design on farmers' preferences for weather-indexed crop insurance. Global Environmental Change, 46, 12-23.

Ali, A., & Abdullah, N. (2020). The effects of macroeconomic indicators on agricultural commodity prices: A study from Pakistan. Sarhad Journal of Agriculture, 36(1), 95-105.

Ali, J., Khan, I., & Anwar, S. (2019). Quality of agricultural information for precision farming: Information quality dimensions analysis. Computers and Electronics in Agriculture, 165, 104960.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. The Quarterly Journal of Economics, 131(4), 1593-1636.

Barrett, C. B. (2008). *Smallholder Market Participation: Concepts and Evidence from Eastern and Southern Africa*. Food Policy, 33(4), 299-317.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

Bryman, A. (2016). *Social Research Methods* (5th ed.). Oxford University Press.

Carter, C. A., Rausser, G. C., & Smith, A. (2018). Commodity booms and busts: Agribusiness and the biofuel, bioproduct, and biotechnology revolutions. Annual Review of Resource Economics, 10, 525-547.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297.

El Emam, K., & Arbuckle, L. (2014). Anonymizing Health Data: Case Studies and Methods to Get You Started. O'Reilly Media.

FAO. (2017). The future of food and agriculture – Trends and challenges. Rome, Italy: Food and Agriculture Organization of the United Nations.

FAO. (2020). *The State of Food and Agriculture 2020*. Rome: FAO.

Flick, U. (2018). An Introduction to Qualitative Research (6th ed.). SAGE Publications.

Fountas, S., Carli, G., Sørensen, C. A. G., Tsiropoulos, Z., Cavalaris, C., Vatsanidou, A., ... & Gemtos, T. (2020). Farm management information systems: Current situation and future perspectives. Computers and Electronics in Agriculture, 115, 40-50.

Ghutake, I., Verma, R., Chaudhari, R., & Amarsinh, V. (2021). An intelligent crop price prediction using suitable machine learning algorithm. In ITM web of conferences (Vol. 40, p. 03040). EDP Sciences.

Gopal, A., & Varshney, D. (2020). Prediction of crop yield using machine learning algorithms. International Journal of Computer Applications, 175(5), 35-39.

Grabs, J., Carodenuto, S., Jespersen, K., Adams, M. A., Camacho, M. A., Celi, G., ... & Stone, E. (2024). The role of midstream actors in advancing the sustainability of agri-food supply chains. Nature Sustainability, 1-9.

Grinyer, A. (2002). The anonymity of research participants: Assumptions, ethics and practicalities. *Social Research Update, 36*, 1-4.

Grossman, S. J., & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. *American Economic Review, 70*(3), 393-408.

Han, N. G., & Kim, B. H. (2021). Design of e-commerce business model through AI price prediction of agricultural products. *Journal of the Korea Convergence Society*, *12*(12), 83-91.

Headey, D., & Fan, S. (2010). *Reflections on the Global Food Crisis: How Did It Happen? How Has It Hurt? And How Can We Prevent the Next One?*. IFPRI Research Monograph. Washington, D.C.: International Food Policy Research Institute.

Hussain, I., & Mudasser, M. (2007). Prospects for wheat production under changing climate in mountain areas of Pakistan–An econometric analysis. Agricultural Systems, 94(2), 494-501.

International Fund for Agricultural Development (IFAD). (2019). *Rural Development Report*. Rome: IFAD.

IPCC. (2021). *Climate Change 2021: The Physical Science Basis*. Geneva: IPCC.

Israel, M., & Hay, I. (2006). Research Ethics for Social Scientists. SAGE Publications.

Jan, I., Ahmad, M., & Khattak, A. A. (2019). Seasonal effects on the prices and income of farmers in Khyber Pakhtunkhwa, Pakistan. Journal of the Saudi Society of Agricultural Sciences, 18(3), 317-322.

Javaid, M., Haleem, A., Khan, I. H., & Suman, R. (2023). Understanding the potential applications of Artificial Intelligence in Agriculture Sector. *Advanced Agrochem*, *2*(1), 15-30.

Johansen, K., Pham, T. D., & Flanagan, K. J. (2019). Plant disease detection from images using convolutional neural networks. In 2019 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED-UAS) (pp. 171-176). IEEE.

Johnston, M. P. (2017). Secondary data analysis: A method of which the time has come. *Qualitative and Quantitative Methods in Libraries, 3*(3), 619-626.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-291.

Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. Computers and Electronics in Agriculture, 143, 23-37.

Khan, M. A., Abbas, A., Ghafoor, A., & Nasir, M. (2016). Analysis of agricultural commodity prices in Pakistan. Journal of Agricultural Research, 57(1), 83-92.

Khan, N., Ray, R. L., Sargani, G. R., Ihtisham, M., Khayyam, M., & Ismail, S. (2021). Current progress and future prospects of agriculture technology: Gateway to sustainable agriculture. *Sustainability*, *13*(9), 4883.

Khan, R., Raza, M. A., & Imran, M. (2016). The effect of Ramadan on the prices of fruits and vegetables. Journal of Islamic Marketing, 7(4), 443-465.

Khan, S., Rahmani, H., Shah, S. A. A., & Ullah, I. (2018). Integration of remote sensing and in-situ data for crop monitoring and yield estimation: A review. Journal of the Saudi Society of Agricultural Sciences, 17(2), 157-168.

Kumar, P., & Sharma, S. (2018). Impact of climate variability on vegetable production and its mitigation strategies: A review. International Journal of Environmental Science and Technology, 15(5), 1019-1034.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444.

Mankiw, N. G. (2014). Principles of Economics (7th ed.). Cengage Learning.

Mannheimer, S., Pienta, A., Kirilova, D., Elman, C., & Wutich, A. (2019). Qualitative Data Sharing: Data Repositories and Academic Libraries as Key Partners in Addressing Challenges. *American Behavioral Scientist, 63*(5), 643-664.

Mendelsohn, R. (2014). The impact of climate change on agriculture in Asia. Journal of Integrative Agriculture, 13(4), 660-665.

Mhlanga, D. (2023). The Role of FinTech and AI in Agriculture, Towards Eradicating Hunger and Ensuring Food Security. In *FinTech and Artificial Intelligence for Sustainable Development: The Role of Smart Technologies in Achieving Development Goals* (pp. 119-143). Cham: Springer Nature Switzerland.

Moe, K., Maeda, M., & De Baerdemaeker, J. (2021). Machine learning approaches in agriculture: From crop yield prediction to disease detection. Journal of Robotics and Mechatronics, 33(1), 26-37.

Mondejar, M. E., Avtar, R., Diaz, H. L. B., Dubey, R. K., Esteban, J., Gómez-Morales, A., ... & Garcia-Segura, S. (2021). Digitalization to achieve sustainable development goals: Steps towards a Smart Green Planet. Science of The Total Environment, 794, 148539.

Myers, R. J. (2017). *Economic Forecasting in Agriculture*. International Journal of Forecasting, 33(1), 181-196.

OECD. (2020). *Agricultural Policy Monitoring and Evaluation 2020*. Paris: OECD Publishing.

Raghuwanshi, B. S., & Jacob, S. (2021). Machine learning for supply chain planning and logistics: A comprehensive survey. IEEE Transactions on Engineering Management, 68(1), 159-175.

Resnik, D. B. (2015). What is ethics in research & why is it important? *National Institute of Environmental Health Sciences*.

Samuelson, P. A., & Nordhaus, W. D. (2009). Economics (19th ed.). McGraw-Hill.

Sieber, J. E. (2006). Empirical Research Ethics for the Social Sciences. In M. Israel & I. Hay (Eds.), Research Ethics for Social Scientists (pp. 87-104). SAGE Publications.

Simonyan, K., & Vinyals, O. (2019). Challenges in data-driven approaches to agricultural problems. Trends in Plant Science, 24(5), 435-444.

Singh, V., Misra, A. K., & Marwaha, S. (2020). Machine learning approach for disease detection in the wheat crop using hyperspectral data. Journal of the Indian Society of Remote Sensing, 48(2), 233-241.

Takahashi, K., Muraoka, R., & Otsuka, K. (2020). Technology adoption, impact, and extension in developing countries' agriculture: A review of the recent literature. Agricultural Economics, 51(1), 31-45.

Thaler, R. H. (2015). Misbehaving: The Making of Behavioral Economics. W. W. Norton & Company.

Tolich, M., & Fitzgerald, M. H. (2006). If ethics committees were designed for ethnography. *Journal of Empirical Research on Human Research Ethics, 1*(2), 71-78.

Torero, M. (2021). Innovations in data and analytics for agriculture and food systems. Nature Food, 2, 68-70.

United Nations Development Programme (UNDP). (2020). *COVID-19 and Human Development: Assessing the Crisis, Envisioning the Recovery*. New York: UNDP.

United Nations, Department of Economic and Social Affairs, Population Division. (2019). *World Population Prospects 2019: Highlights*. New York: United Nations.

Weersink, A., Fraser, E., Pannell, D., Duncan, E., & Rotz, S. (2018). Opportunities and challenges for big data in agricultural and environmental analysis. Annual Review of Resource Economics, 10, 19-37.

Wihartiko, F. D., Nurdiati, S., Buono, A., & Santosa, E. (2021, March). Agricultural price prediction models: a systematic literature review. In Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management Singapore (pp. 2927-2934).

World Bank. (2021). *World Development Report 2021: Data for Better Lives*. Washington, DC: World Bank.

Zheng, Y., Liu, F., & Hahn, D. E. (2019). *Big Data and Predictive Analytics for Agri-Food and Environmental Economics*. Agricultural Systems, 173, 49-61.