Analyzing Global Live Pigs Export Market Using Machine Learning Models

Author: Zemelak Goraga

Course: DSC680-T301 Applied Data Science (2251-1)

Project 2 Milestone 3: Final White Paper

Professor Amirfarrokh Iranitalab

Date: 10/19/2024

1. Summary

This project analyzed the global live pigs export market from 1961 to 2013 using machine learning models to uncover patterns, optimize strategies, and address market inefficiencies. The dataset, sourced from FAOSTAT via Kaggle, provided insights into export quantities and values, identifying trends, deviations, and strategic opportunities. Key machine learning techniques, including K-means clustering, association rule mining, and Principal Component Analysis (PCA), were applied to segment countries based on their export behaviors, explore patterns linking export quantities to values, and visualize market clusters.

Results indicated moderate correlation (0.66) between export quantities and values, showing that as export volumes increase, their economic value tends to rise, though other factors may also influence this relationship. The descriptive analysis revealed fluctuations in average export quantities, ranging from 305.4 heads in 2008 to 723 heads in 2012, while export values varied from $189,600 in 2010 to $642,100 in 2007. Clustering analysis identified three distinct groups, with regions like Africa and Australia & New Zealand falling into separate clusters. Association rule mining revealed strong patterns, with a 95% confidence that high export quantities would lead to high values.

These insights provide stakeholders with actionable strategies to optimize export performance, capitalize on strategic advantages, and drive sustainable growth in the global live pigs export market.

2. Business Problem:

The live pigs export market plays a vital role in global food security, economic stability, and agricultural growth. However, challenges such as fluctuating demand, trade barriers, and inefficiencies in export practices persist. Traditional analysis often fails to capture the market's complexities, leading to missed optimization opportunities.

This project uses machine learning models to analyze export data and identify patterns, providing stakeholders with actionable insights to optimize export strategies, improve risk management, and promote long-term sustainability.

3. Background:

Pork demand is growing globally, making the live pigs export market crucial for meeting this demand. The market is influenced by factors such as economic conditions, trade policies, and consumer preferences. Traditional analysis methods fail to capture the complex interactions between these factors, limiting the ability to predict market performance accurately.

This project leverages machine learning models to analyze export quantities and values, revealing trends, deviations, and clusters within the market, equipping exporters, policymakers, and trade organizations with the knowledge needed to optimize their strategies.

4. Data Explanation:

The dataset, sourced from FAOSTAT via Kaggle, covers global live pigs export statistics from 1961 to 2013. Key variables include country, year, export quantities (heads), and export values (USD). This comprehensive dataset allowed for time series analysis, correlation studies, clustering of countries, and extraction of association rules.

The dataset's structure enabled detailed analysis, revealing trends and deviations that provide actionable insights for decision-making in the live pigs export market.

5. Methods:

The analysis followed a structured data science process, beginning with data preprocessing and progressing through advanced machine learning model implementation.

Data Preprocessing: The raw dataset was cleaned by handling missing values through imputation techniques, ensuring no significant gaps in the analysis. Additionally, new features, such as the value-to-quantity ratio, were engineered to provide deeper insights into the relationship between export quantities and their corresponding economic values. Data normalization was also applied to ensure that variables with varying scales were treated consistently in the analysis.

Exploratory Data Analysis (EDA): This phase involved generating various visualizations, such as line plots and heatmaps, to identify key trends in export quantities and values. Correlation analysis was conducted to assess the relationship between export volumes and their economic value, revealing patterns that informed later model development. EDA also helped identify deviations in value-to-quantity ratios, indicating potential inefficiencies or strategic advantages.

Model Development: Several machine learning models were implemented to extract meaningful insights. K-means clustering was applied to group countries based on their export behaviors, uncovering distinct market segments. Association rule mining using the Apriori algorithm was performed to identify patterns between export quantities and values, while Principal Component Analysis (PCA) was used to reduce the dimensionality of the data and visualize cluster separation effectively.

Model Evaluation: The models were evaluated using several metrics. For clustering, silhouette scores assessed the degree of separation between clusters, while association rules were evaluated based on support, confidence, and lift metrics. Visualizations further enhanced the understanding of the patterns uncovered, ensuring the models' effectiveness in addressing the research questions.

6. Results:

Trends in Export Quantities and Values:

Export quantities fluctuated between 305.4 heads (2008) and 723 heads (2012). Export values varied from $189,600 (2010) to $642,100 (2007) (Fig. 1). Variability across countries and periods indicated uneven export performance. A correlation of 0.66 showed a moderate positive relationship between export quantities and values, suggesting that increased export volumes generally lead to higher economic returns, although other factors also play a role (Fig. 2).

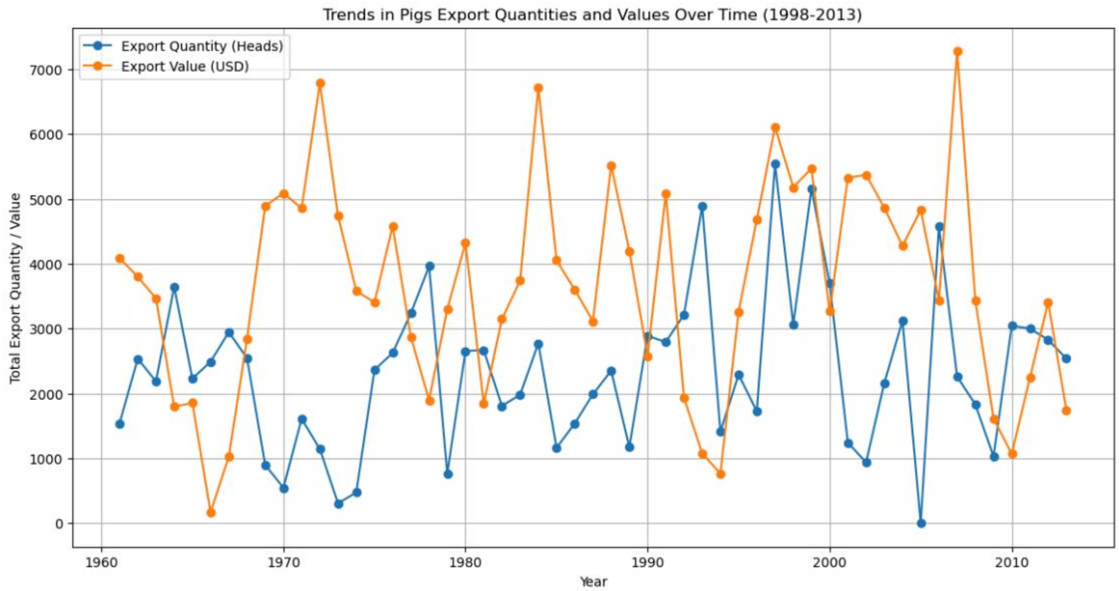

Fig. 1. Trends in Pigs Export Quantities and Values Over Time (1998 – 2013)

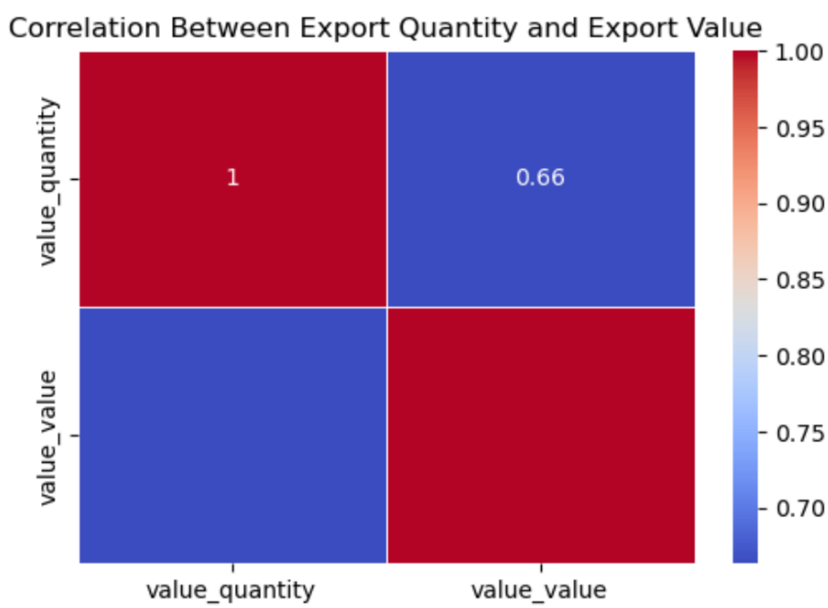

Fig. 2. Correlation between Export Quantity (heads) and Export Value (US$)

Countries Comparison & Deviations in Value-to-Quantity Ratios:

Top 10 Countries by Export Quantity (1998-2013):

Eastern Africa (4,318 heads) and Australia & New Zealand (4,150 heads) led the exports, followed by Oceania (3,534 heads) (Fig. 3).

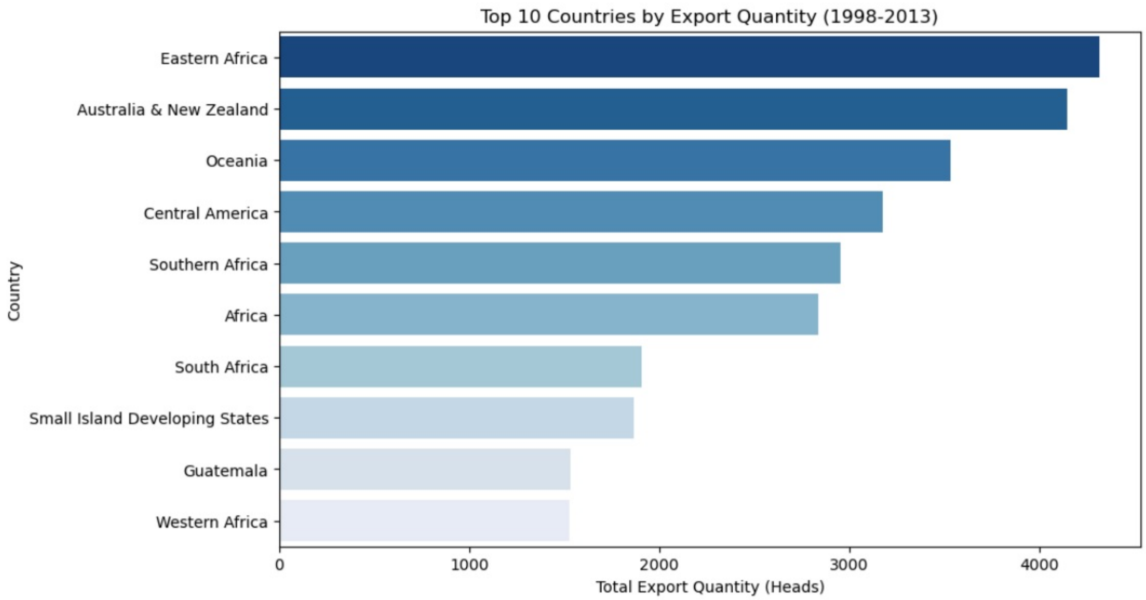

Fig. 3. Top ten countries by Export Quantity (1998 – 2013)

Top 10 Countries by Export Value (1998-2013):

Central America led with $7,652,000, followed by Low Income Food Deficit Countries ($7,085,000) and Oceania ($6,767,000) (Fig. 4).

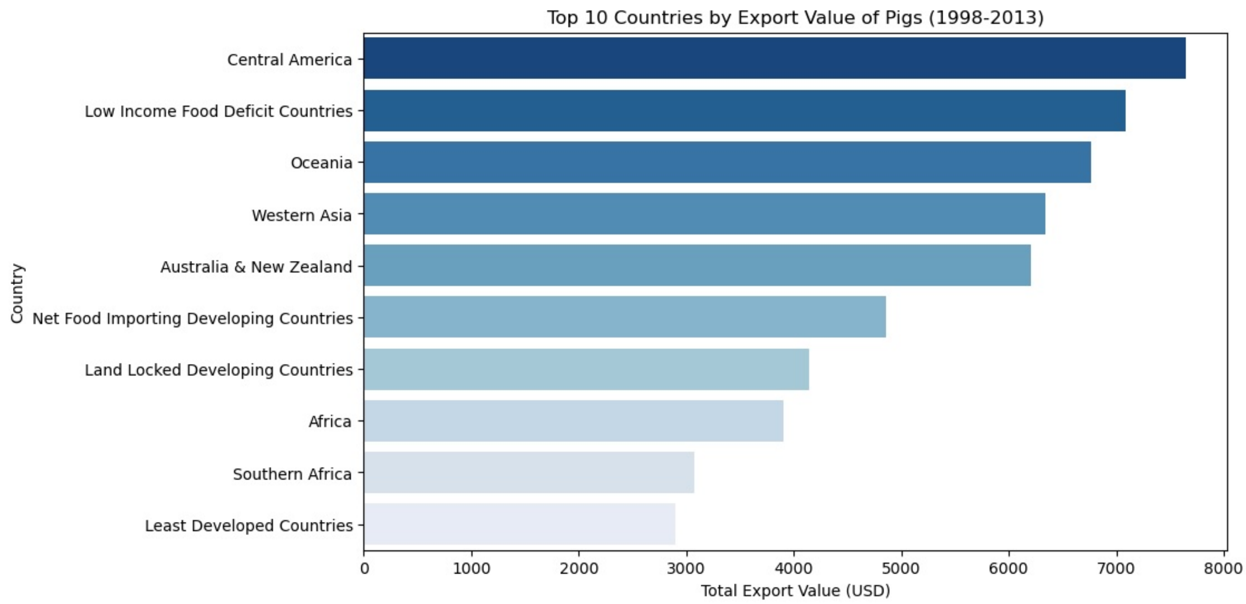

Fig. 4. Top ten countries by Export Value (1998 – 2013)

Value-to-Quantity Ratio Deviations:

Significant deviations in value-to-quantity ratios were observed, highlighting inefficiencies or strategic advantages. For example, the Caribbean in 1984 had a high value-to-quantity ratio (1.87), suggesting an unusually high export value relative to the quantity of live pigs exported. Countries like Australia & New Zealand in 2007 had a balanced ratio, reflecting efficient trade practices.

Clustering of Exporting Countries:

K-means clustering identified three distinct clusters:

Cluster 0 included countries like Africa and Southern Africa, characterized by higher export values but lower quantities.

Cluster 2 included high-volume exporters like Australia & New Zealand and Oceania, reflecting their large-scale production and competitive pricing (Fig. 5).

The silhouette score of 0.41 indicated moderate separation between clusters, although some overlap exists.

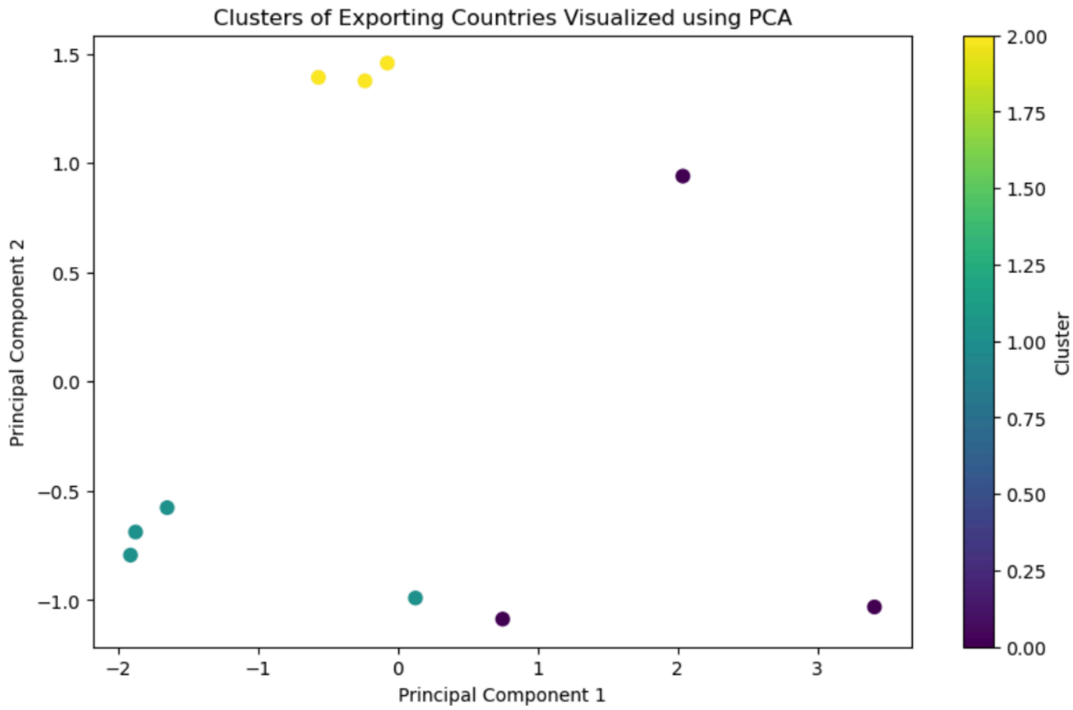

Fig. 5. Clusters of Exporting Countries Visualized using PCA

Association Rule Mining:

Strong rules like (Export Value) → (Export Quantity) had a confidence level of 95% and a lift of 1.98, showing that higher export values are frequently associated with higher export quantities. Other rules, such as (Export Value, Import Quantity) → (Export Quantity), indicated the interconnected nature of export and import activities.

PCA:

PCA effectively reduced data complexity, helping visualize clustering results. Countries like Africa and Australia & New Zealand were clearly separated along PCA1 and PCA2, reflecting differences in export volumes and values (Fig. 6).

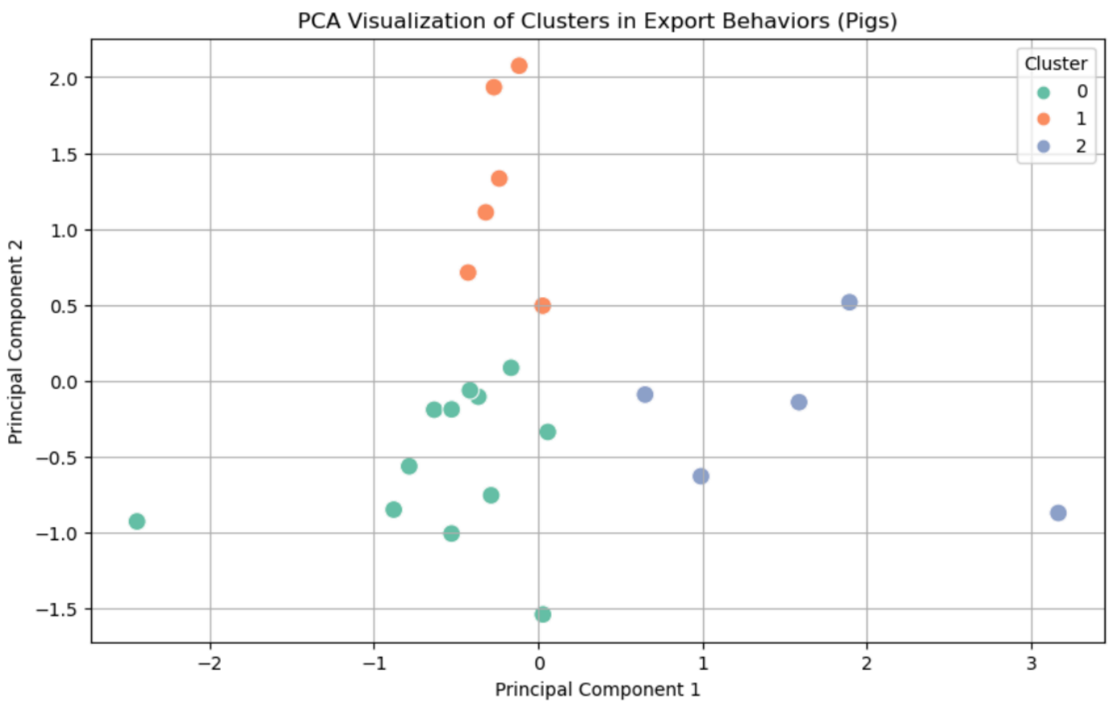

Fig. 6. PCA Visualization of Clusters in Export Behaviors (Pigs)

7. Discussion of Analysis Results:

The results of this project provide valuable insights into the global live pigs export market and its dynamics over time. By using machine learning techniques, I was able to explore five key research questions, each revealing critical trends, inefficiencies, and strategic opportunities within the export market. This section will discuss the findings of each research question in detail, providing context for how the results can benefit stakeholders in optimizing strategies and improving market performance.

1. What are the trends in export quantities and values, and how do they correlate over time?

The analysis of export quantities and values from 1998 to 2013 revealed significant fluctuations in both metrics, indicating that the live pigs export market is subject to various influencing factors such as economic conditions, trade policies, and consumer demand. For example, the mean export quantity ranged from 305.4 heads in 2008 to 723 heads in 2012, reflecting how global events during these years, such as economic crises or trade disruptions, likely influenced export performance.

Similarly, export values showed substantial variability, with mean values ranging from $189,600 in 2010 to $642,100 in 2007. These fluctuations highlight the sensitivity of the market to external factors, as well as the interdependency between export volumes and their economic value. The correlation between export quantities and values was moderate at 0.66, indicating that while higher quantities tend to correspond with higher values, other factors (such as market pricing, product quality, or regional trade agreements) may also play a role in influencing value independently of quantity.

This correlation suggests that stakeholders can generally expect increased economic returns when export volumes rise, but they should also consider external factors that may disrupt this relationship. The variability in both export quantities and values highlights the importance of adaptability and market awareness in maintaining export performance.

2. Which countries or periods show significant deviations in value-to-quantity ratios, highlighting inefficiencies or strategic advantages?

The analysis of value-to-quantity ratios revealed significant deviations in certain countries and time periods, offering insights into potential inefficiencies or strategic advantages. A key example is the Caribbean in 1984, which exhibited a notably high value-to-quantity ratio of 1.87, coupled with a z-score of 3.89. This suggests that the Caribbean achieved a disproportionately high export value relative to the quantity of live pigs exported. Such deviations could be due to a variety of factors, such as higher market prices, premium product quality, or advantageous trade agreements that allowed the region to export at a higher economic value than other countries with similar export volumes.

On the other hand, countries like Australia & New Zealand in 2007 had a value-to-quantity ratio close to 1.0, indicating a more balanced and efficient export strategy where the quantity and value were closely aligned. This suggests that these countries operated with strategic pricing or efficient trade practices, avoiding inefficiencies that might arise from over- or under-pricing their exports relative to volume.

Understanding these deviations is critical for stakeholders seeking to optimize export performance. By identifying inefficiencies, exporters can adjust pricing strategies or negotiate better trade agreements to maximize value. Similarly, countries with strategic advantages can capitalize on these strengths to maintain or improve their market positioning.

3. Can machine learning models cluster exporting countries based on similar export behaviors?

The K-means clustering model provided deeper insights into the export behaviors of different countries by grouping them based on their export quantities and values. The model segmented the countries into three distinct clusters, each characterized by unique export patterns:

Cluster 0: This cluster included countries like Africa, Southern Africa, and Western Asia. These countries were characterized by higher export values but lower export quantities, suggesting that they may focus on premium markets or engage in more value-driven export strategies. This could be due to strategic positioning, specialized products, or stronger pricing power in global markets.

Cluster 1: This cluster contained countries like Cambodia and Melanesia, which exhibited lower export quantities and values. These countries may have limited market presence or face challenges such as lower demand, higher competition, or less efficient export systems.

Cluster 2: Australia & New Zealand and Oceania were part of this cluster, which was characterized by higher export quantities, suggesting that these regions were high-volume exporters. This could be indicative of efficient production systems, strong trade partnerships, or greater access to global markets.

The silhouette score for clustering was 0.41, indicating moderate separation between the clusters. While the clusters are reasonably well-defined, there is some overlap, suggesting that while countries within each cluster share similar behaviors, other factors outside of export quantity and value may influence export patterns.

For stakeholders, this clustering provides a framework for market segmentation. Exporters can tailor their strategies based on their cluster characteristics, whether focusing on high-volume or high-value markets.

4. What patterns do association rules reveal about frequently occurring high export quantities and values?

The association rule mining process uncovered strong patterns between export quantities and values, providing insights into frequently co-occurring export characteristics. A key rule identified was (Export Value) → (Export Quantity), which had a confidence level of 95% and a lift of 1.98. This suggests that higher export values are strongly associated with higher export quantities, reinforcing the predictable relationship between these two metrics.

Another important rule was (Export Value, Import Quantity) → (Export Quantity), with a confidence of 98% and a lift of 2.05, indicating a strong association among export values, import quantities, and export volumes. This rule highlights the interconnectivity between import and export activities, suggesting that countries with higher import volumes may also see a corresponding increase in export quantities.

These association rules provide practical insights for stakeholders looking to optimize their export strategies. Exporters focusing on increasing volumes can expect higher economic returns if market conditions (such as pricing and demand) remain favorable. Additionally, the strong relationship between import and export quantities suggests that exporters should also consider the role of imports in shaping export performance, as high import volumes may create opportunities for re-exporting or increasing domestic supply chains.

5. How do export behaviors cluster based on multiple features, and can PCA effectively visualize these patterns?

Principal Component Analysis (PCA) was applied to reduce the complexity of the dataset while retaining meaningful distinctions between export behaviors. The first two principal components

(PCA1 and PCA2) successfully visualized the clustering results, providing a clear view of how countries' export behaviors varied.

For example, countries like Africa and Australia & New Zealand were clearly separated along the PCA1 and PCA2 axes. Africa had negative PCA1 and PCA2 values, reflecting lower export quantities and values, while Australia & New Zealand had slightly higher PCA1 values, indicating stronger export volumes. Similarly, Central Asia and India were separated along PCA1, reflecting different dynamics in export behavior.

This visualization provides stakeholders with a clear and intuitive way to understand export patterns across different regions. By analyzing how countries are positioned on the PCA plot, exporters and policymakers can identify market segments, evaluate their relative market position, and develop strategies to optimize their export performance based on these insights.

8. Conclusions:

This project demonstrates the effectiveness of applying machine learning techniques to analyze the global live pigs export market, uncovering key patterns and insights that traditional methods might overlook. The results showed moderate correlation (0.66) between export quantities and values, indicating that while higher export volumes generally lead to higher economic value, other factors also play a role in determining export success. Significant deviations in value-to-quantity ratios, such as the Caribbean in 1984, highlighted areas where inefficiencies or strategic advantages may exist.

Clustering analysis revealed distinct export behaviors, grouping countries based on similarities in export quantities and values. Countries like Africa and Australia & New Zealand showed different strategies, with Africa focusing on value-driven exports and Australia & New Zealand being high-volume exporters. Association rule mining identified strong relationships between high export quantities and values, offering practical insights for improving export strategies.

The project highlights the value of advanced analytics in optimizing export performance, enabling stakeholders to make data-driven decisions that enhance market efficiency and profitability. By leveraging these insights, exporters, policymakers, and trade organizations can better position themselves in the global market, improve decision-making, and drive sustainable growth in the live pigs export sector.

9. Assumptions:

The project assumed the FAOSTAT dataset accurately reflects global export trends. Relationships between export quantities and values were assumed to remain consistent, and selected machine learning models were presumed to capture underlying patterns. External factors, such as economic shocks, were not fully considered.

10. Limitations:

The project relied on historical data from 1961-2013, potentially missing recent market shifts. Models didn’t fully account for unpredictable geopolitical or economic factors. Data quality issues, including missing values, may have impacted accuracy, and the analysis focused on economic indicators, excluding consumer behavior or environmental influences.

11. Challenges:

Challenges faced during the project include ensuring data quality and handling large, complex datasets. The selection and optimization of machine learning models required extensive experimentation, and the market's sensitivity to economic volatility posed difficulties in predicting future trends. Addressing potential biases in the models and ensuring interpretability of results were also significant challenges.

12. Future Uses/Additional Applications:

The methodologies used in this project can be extended to other agricultural products, enabling stakeholders to optimize trade strategies across various sectors. Integrating real-time data sources can enhance market monitoring, and predictive models can support supply chain optimization, policy impact analysis, and risk management. The insights generated can inform strategic decision-making in multiple industries, including education, sustainability, and international trade.

13. Recommendations:

To optimize export performance, stakeholders should leverage advanced analytics and adapt strategies based on export trends and deviations. Clustering analysis helps segment markets, while association rules provide insights for decision-making. Machine learning models offer actionable insights to enhance trade strategies, improve market efficiency, and drive sustainable growth in the live pigs export sector.

14. Implementation Plan:

The implementation plan includes a phased approach:

Phase 1: Preparation and Capacity Building (Weeks 1-4) – Engage stakeholders, enhance data infrastructure, and provide training.

Phase 2: Model Integration and Testing (Weeks 5-8) – Customize and validate models, followed by pilot testing.

Phase 3: Full-Scale Implementation (Weeks 9-16) – Deploy models, monitor performance, and adjust as needed.

Phase 4: Evaluation and Continuous Improvement (Weeks 17-20) – Assess model performance, gather feedback, and refine strategies.

15. Ethical Assessment:

Ethical considerations included ensuring data privacy, mitigating biases in machine learning models, and maintaining transparency. Promoting data-driven insights responsibly ensures fairness and sustainability in global trade practices.

16. References

Abdelbaki, W., Zreikat, A. I., Cina, E., Shdefat, A., & Saker, L. (2023). Crop prediction model using machine learning algorithms. Applied Sciences, 13(16), 9288. https://doi.org/10.3390/app13169288

Cambridge Core. (2020). iCROPM 2020: Crop modeling for the future. The Journal of Agricultural Science. https://www.cambridge.org/core/journals/journal-of-agricultural-science/article/icropm-2020-crop-modeling-for-the-future/94520DEBD9EA9785EB3F2B75BC3AC5F4

Food and Agriculture Organization of the United Nations (FAO). (n.d.). FAOSTAT data. http://www.fao.org/faostat/en/#data

Kaggle. (n.d.). FAOSTAT: Food and agriculture data. https://www.kaggle.com/datasets/faoorg/faostat-food-and-agriculture-data